Interpretability

Studies in Granular Computing

University of Sheffield

Muhammad Zaiyad bin Muda

Department of Automatic Control and Systems Engineering

The University of Sheffield

A thesis submitted to The University of Sheffield for the degree of

Doctor of Philosophy

December 2022

ABSTRACT

The first stage in creating data-driven soft computing models is to derive information from data, towards developing the structure of computational models. An effective way to extract information from data is through granular computing (GrC), which is inspired by the way human naturally groups similar objects together. Using GrC and Fuzzy Sets in modelling engineering systems, one can describe systems in a very transparent and interpretable way. GrC has been applied in the literature, however, there are still issues with data uncertainty and its impact on the interpretability of Fuzzy Logic systems. The consideration for data uncertainty is very important for real applications, where data usually comes from measurements/sensing, with inherent noise and uncertainty. Hence, this thesis aims to investigate methods to address uncertainty in GrC, methods to address variable importance in GrC, methods to address granular overlapping in GrC, and methods to evaluate the resulting impact on the interpretability.

A new data-driven modelling framework based on GrC, Fuzzy Logic, and uncertainty measure is proposed, in which the uncertainty (in this thesis captured via conflict) between information granules is modelled using Shannon entropy. The issue of interpretability due to overlapping is addressed with a new iterative data granulation mechanism that controls the amount of granule overlapping using R-value, a metric that represents the ratio of overlapping areas among categories in a data cluster. In order to characterise the importance of data features, Weighted GrC (W-GrC), a new iterative data granulation technique with evolving feature weighting is proposed. The feature weights are determined based on the current information granules (within-granule variances) and adaptively change in each iteration. Since W-GrC is studied in both Type-1 and Type-2 systems, a new interpretability index for Type-2 Fuzzy Logic systems is proposed based on Nauck's index, taking into account both upper and lower membership functions.

A thorough set of simulations based on UCI datasets are conducted to demonstrate the effectiveness of each of the frameworks proposed in this thesis. The simulation results demonstrate the potential of all proposed frameworks in improving the predictive accuracy while maintaining good level of interpretability.

ACKNOWLEDGEMENTS

I would like to express my greatest gratitude to my supervisor, Professor George Panoutsos for giving me the opportunity to do PhD under his supervision. I really appreciate his guidance and endless supports throughout my PhD journey.

My great appreciation goes to Adrian, Kai and Dr. Iñaki Esnaola who helped me in the technical aspects of this research. Many thanks also go to my second supervisor, Dr. John Oyekan for his comments and advice.

I would like to thank the Malaysian Government and Universiti Teknologi MARA (UiTM) for the financial supports and the opportunity to study in The University of Sheffield.

I am indebted to my parents, brothers and sisters for supporting and encouraging me to pursue this degree. Finally, I would like to thank my greatest support system; my beloved wife, Noor Khafifah Khalid and my sons Zaim, Zamil, and Zarif who supported and encouraged me throughout the studies, and who made this journey possible.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
BEP	Back Error Propagation
FIS	Fuzzy Inference System
FL	Fuzzy Logic
FOU	Footprint of Uncertainty
GrC	Granular Computing
GrC-FLS	Granular Computing – Fuzzy Logic System
GT2-FLS	General Type-2 Fuzzy Logic System
GT2-FS	General Type-2 Fuzzy Set
GT2-RBFNN	General Type-2 Radial Basis Function Neural Networks
IT	Information Theory
IT2-FLS	Interval Type-2 Fuzzy Logic System
LMF	Lower Membership Function
MF	Membership Function
NI	Nauck's index
NF	Neural Fuzzy
RBFNN	Radial Basis Function Neural Network
RMSE	Root Mean Square Error
SD	Standard Deviation
SVM	Support Vector Machine
T1-FS	Type-1 Fuzzy Set
T1-FLS	Type-1 Fuzzy Logic system
T2-FS	Type-2 Fuzzy Set
T2-FLS	Type-2 Fuzzy Logic system
UMF	Upper Membership Function
W-GrC	Weighted Granular Computing

CHAPTER 1

INTRODUCTION

Discovering connections between data sets and organising data together can be accomplished with the help of data clustering methods. Fuzzy C-means and the Mountain approach are examples of clustering strategies that have been applied to help in the design of fuzzy systems [1]. The major drawback of such techniques is that the quality of the partitions depends on the priori parameters such as the number and location of the initial cluster centres [2]. Granular Computing (GrC), which is motivated by the human perception in grouping similar features, provides a straightforward method of extracting information from data sets.

GrC allows for the grouping of data based on similarity features and furthermore keep in the data space information and data properties. In GrC, the most important concept is the definition of the compatibility measure, which can be purely geometrical (size of granules, volume of granules), density driven (ratio of cardinality versus granule volume), similarity driven (data overlap), proximity or function based [1]. Instead of being generated by an algorithm, the information granules develop from the original data. GrC's transparency and the additional information gathered during the grouping process make this methodology perfect for integrating with Fuzzy Logic systems. The scope in GrC is the development of information granules – objects that provide a way of organising information about the data and the existing relationships [3].

Fuzzy Logic (FL) rule-based models can be created in one of two ways: by employing expert knowledge or by applying data mining techniques [4]. However, the first approach is limited due to the availability and consistency among the experts. Therefore, data mining techniques can be utilised to discover the knowledge (relationships and structure) from data. This information can be used to establish the

Fuzzy Logic parameters in data-driven FL modelling that is effective in identifying a wide range of complex nonlinear systems [5].

In this thesis, FL models are developed based on a granular computing algorithm suggested by Pedrycz and Bargiela [6], known as granular clustering or iterative data granulation. This algorithm is comparable to the idea of agglomerative hierarchical clustering that starts with treating each data point as an individual cluster, or granule, and merges the most similar clusters up until a predetermined termination condition is met [7]. The most significant difference between these two algorithms is that in iterative data granulation, the compatibility measure is used as the merging criteria. The compatibility measure is a function that comprises the distance between information granules and the information density of the obtained granule [8].

1.1 PROBLEM STATEMENT

High quality information granules may lead to the formation of representative FL rule-bases that capture the behaviour of real systems. However, in reality, data or information is always uncertain as the result of information deficiency, i.e. incomplete, imprecise, vague or contradictory. Imprecise measurement of data also may contribute to the problem of data uncertainty. In the presence of outliers for example, information granules cannot capture accurately the underlying data. This will inhibit the interpretability of an FL rule base, and causes inaccurate representation of the system. As a consequence, the information granules formed using the iterative data granulation may be affected, hence leading to the formation of inaccurate and inconsistent FL rule-bases. Therefore, it is necessary to take into account the data uncertainty during the data granulation process. This can be achieved by quantifying the uncertainty using the concept of information theory, and utilising this information in a new compatibility measure to guide the data granulation process.

In the conventional iterative data granulation, all features or input variables are treated as equally important [9]; the weight for each feature is the same without considering their degree of relevance or importance. However, this approach is not necessarily ideal, since there are always variables that are more significant or relevant to a particular task than the others. Ideally, the more important features are assigned with higher weighting than the less important features. In the literature, most of the

works regarding iterative data granulation [10], [11] use equal weights for all features, while some other works attempted to assign fixed predetermined weights for the granulation process [12]. Due to its iterative-hierarchical nature, this approach might not be the best since the information provided by the granules in each iteration is changing according to the current parameters of granules (i.e. cardinality, variance, distance).

Another issue in the iterative data granulation is the class overlapping which can be a source of uncertainty. Most conventional clustering algorithms aim at forming a number of disjoint clusters, or granules where an object belongs to only one granule [13]. Traditional clustering algorithms such as k-means may function effectively especially when the group boundaries are distinct and the data is free of outliers. However, most real-world data sets have overlapping information, therefore some data objects or patterns might be a part of multiple granules [14]. Due to numerous issues that need that the granules be inclusive, for example overlapping information that exists in the real world data sets [14] and incorrect representation of the underlying data structure [15], there is a need to work with overlapping granules. So far, the compatibility criteria of the iterative data granulation do not incorporate any parameter that is able to control the amount of the overlapping among the granules.

The motivation to enhance the development of information granules is to form high quality FL rule-bases. One of the reasons for using Fuzzy Logic systems (FLSs) is the application of linguistic variables and rules that are interpretable by humans [16]. Interpretability represents the systems' ability to describe the real system in a comprehensible way [17]. Unlike accuracy, whose calculation is straightforward, the computation of interpretability is highly subjective and challenging. This is due to different knowledge and experience of each individual in interpreting a Fuzzy Logic system [18], hindering the interpretability to be assessed in a standard numerical way. In type-1 Fuzzy Logic systems (T1-FLSs), most of the works related to interpretability use various types of interpretability indices, such as number of rules, total rule length, average rule length, Nauck's index and fuzzy index [18], [19]. However, interpretability studies in type-2 Fuzzy Logic systems (T2-FLSs) are scarce. This is due to higher complexity in the T2-FLSs, especially because it involves two separated membership functions in the Fuzzy Logic sets, namely upper and lower membership functions;

hence there is a need to formulate appropriate numerical measures to capture interpretability in T2-FLS.

The scope of this thesis is to research human-centric intelligent computational frameworks and systems, such as information granulation, fuzzy sets theory and Fuzzy Logic systems, artificial neural networks and interpretability measures for Fuzzy Logic systems.

1.2 RESEARCH AIMS AND OBJECTIVES

Since iterative data granulation is relatively new computing technique, there are still research gaps that can be explored on the development of new computational frameworks in particular in dealing with uncertainties and class overlapping during the granulation process and its possible combination with FL. Moreover, up to date this algorithm has not been reported yet to incorporate evolving feature weighting technique, an embedded method that allows the more important features to have higher influence in the data granulation than the less important features, for a given iteration. This research is not limited to the Type-1 Fuzzy Logic systems, therefore the enhanced information granules are utilised towards building higher order FL systems – e.g. Type-2). Since the interpretability is the motivation of using FL rule base system, this thesis explores the quantification of interpretability for Type-2 Fuzzy Logic systems.

The aims of this research work are to investigate methods to address uncertainty in GrC, methods to address variable importance in GrC, methods to address granular overlapping in GrC, and methods to evaluate the resulting impact on the interpretability. In order to accomplish these aims, this PhD research work focuses on the following objectives:

1. To identify and develop mathematical methods for quantifying uncertainty accumulation in the granulation process and use this information to enhance the iterative data granulation towards the building the Fuzzy Logic rule-based systems.

2. To investigate and introduce a new computational framework for capturing the importance of features during iterative data granulation and analyse the effectiveness of the new framework in the development of Type-1 Fuzzy Logic

systems (T1-FLSs) models while also assessing the impact on rule base interpretability.

3. To develop the new compatibility measure in the iterative data granulation that includes an additional parameter to control the amount of overlapping among the information granules.

4. To extend the feature weighted (Objective 2) computational framework to Radial Basis Function Neural Network (RBFNN) and General Type-2 Radial Basis Function Neural Network (GT2-RBFNN) systems.

5. To investigate and develop new interpretability measures for Type-2 FLSs and use this to study the impact of feature weighting (as in Objectives 2 and 4) in terms of the systems' interpretability.

1.3 CONTRIBUTIONS

The main contributions of this thesis are:

1. A method to quantify uncertainty during the iterative data granulation process is introduced. Shannon entropy, an important parameter within the information theory framework is used to numerically quantify the uncertainty, and this information is utilised in guiding the merging of the information granules with the aim to minimise the uncertainty during data granulation. A new compatibility measure incorporating the amount of uncertainty is developed, and the enhanced information granules are used to build Fuzzy Logic rule-based systems. This contribution is based on Chapter 3 of this thesis and the outcome of this work has been presented and subsequently published in the 7th International Conference on Soft Computing & Machine Intelligence (ISCMI).

2. A new GrC algorithm called weighted GrC (W-GrC) is introduced to assign and update the feature weight based on the feature importance. The main idea of W-GrC is to compute and update the feature weights according to the within-granule variances in each iteration, meaning that the feature-weighting

technique is embedded in the data granulation process. The new and enhanced information granules are used to build Type-1 Fuzzy Logic systems (T1-FLSs) rule-bases. The impact of W-GrC is investigated in terms of the system's predictive accuracy and interpretability, in which the interpretability is measured using Nauck's index. This contribution is based on Chapter 4 of this thesis and the outcome of this work has been presented and published in 20th UK Workshop on Computational Intelligence, and has been awarded with the Best Student's paper (2021).

3. The issue of class overlapping in data granulation is addressed. Specifically, a new iterative data granulation algorithm is introduced, by integrating a parameter called R-value to model the overlapping among the information granules in the compatibility measure. The aim is to allow an instance to belong to more than one granule, rather than only one granule as in the conventional disjoint types of granulation. The effectiveness of overlapped GrC is studied in the classification tasks within the T1-FLSs framework.
4. A new measure for interpretability in type-2 Fuzzy Logic systems (T2-FLSs) is introduced based on the Nauck's index. Since the study of Fuzzy Logic systems' interpretability in the literature is limited to mainly T1-FLSs, a new index is developed, taking into account both membership functions in T2-FLSs (upper and lower membership functions). To validate, the W-GrC developed in (2) is extended to Radial Basis Function Neural Network (RBFNN) and General Type-2 Radial Basis Function Neural Network (GT2-RBFNN) models. The impact on the interpretability is quantified using the proposed interpretability measure. This contribution is shown in Chapter 6 of this thesis and the outcome of this work has been published in Expert Systems journal.

1.3.1 PUBLICATIONS

Journal Papers

- M. Z. Muda, A. R. Solis, and G. Panoutsos, "An Evolving Feature Weighting Framework for Radial Basis Function Neural Network Models," Expert Systems, e13201, pp. 1-14, 2022, doi: https://doi.org/10.1111/exsy.13201.

Conferences

- M. Z. Muda and G. Panoutsos, “An Evolving Feature Weighting Framework for Granular Fuzzy Logic Models,” In: Jansen, T., Jensen, R., Mac Parthaláin, N., Lin, CM. (eds) Advances in Computational Intelligence Systems. UKCI 2021. Advances in Intelligent Systems and Computing, vol. 1409. Springer, Cham, pp. 3-14, 2021, https://doi.org/10.1007/978-3-030-87094-2_1.

- M. Z. Muda and G. Panoutsos, "An Entropy-Based Uncertainty Measure for Developing Granular Models," 2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI), pp. 73-77, 2020, doi: 10.1109/ISCMI51676.2020.9311589.

1.4 THESIS OUTLINE

This thesis is organised as follows:

Chapter 2 provides a literature review on the current research in granular Fuzzy Logic systems and human-centric computational intelligence systems. It starts with covering the review of Fuzzy Logic systems and neuro-fuzzy modelling, for both Type-1 and Type-2 systems. Next, highlights are given to the current research trend in the interpretability of Fuzzy Logic systems, covering various types of interpretability measures. Finally, the concept of granular computing and the development of information granules is described.

In chapter 3, the uncertainty occurring by the iterative data granulation process is quantified via the use of information theory. Here, the iterative data granulation algorithm, which is the focus in this thesis is described in detail, where the computation of the compatibility measure is detailed. As a comparison, other clustering algorithms are also presented in this chapter. Highlight is also given to the Granular computing based Fuzzy Logic modelling framework that is applicable in other chapters (Chapter 4-6). This refers to the translation of information granules obtained from GrC to Fuzzy Logic rule-bases. Finally, the effectiveness of integrating the uncertainty measure in the iterative data granulation is demonstrated in Type-1 Fuzzy Logic systems.

Chapter 4 presents a new GrC algorithm called weighted GrC (W-GrC) that assigns and updates the feature weight based on the feature importance. The difference between feature selection and feature weighting is described, as well as the computation of feature weights based on within-granule variances (in each iteration). The feature weights are allowed to evolve as the granulation process progresses according to the inbuilt feature-weighting mechanism. The information granules obtained from W-GrC are translated into Type-1 Fuzzy Logic systems (T1-FLS) rule-bases and the new framework is validated using classification problems (data from UCI machine learning repository).

Chapter 5 tackles the issue of class overlapping in data granulation. A parameter known as R-value is used to model the overlapping between granules, the calculation of this parameter is shown. R-value is integrated into the compatibility function and provides a way to control the amount of overlapping among the granules. This method permits an object to belong to one or more granules rather than just one. The new GrC with overlapping measure is demonstrated by using datasets from UCI Machine Learning Repository.

Chapter 6 focuses on the interpretability measure in Type-2 Fuzzy Logic systems. The weighted GrC in Chapter 4 is demonstrated in Radial Basis Function Neural Network (RBFNN) and General Type-2 Radial Basis Function Neural Network (GT2-RBFNN). A new approach in measuring the interpretability based on Nauck's index is proposed, and the computation is exemplified in detail. The NI is assessed for both W-GrC and conventional GrC in type-1 and type-2 models, and the impact on the interpretability is investigated.

Chapter 7 - Conclusions and Future Works conclude all contributions from all chapters and provide recommendations for future works.

CHAPTER 2

A BACKGROUND TO GRANULAR COMPUTING AND FUZZY LOGIC MODELLING

This chapter provides a literature review on the current research in granular computing and Fuzzy Logic systems within the context of human-centric computational intelligence systems. The focus of this chapter is on Fuzzy Logic systems, Neuro-fuzzy modelling, granular computing and on the interpretability of Fuzzy Logic systems.

2.1 FUZZY LOGIC SYSTEMS

Identifying a precise model representation of a system is not always practical, or feasible, when dealing with real-world problems, especially for complex systems. Due to the large number of interacting parameters, which humans might not be able to interpret at once, it can be difficult to gain a thorough description or understanding of the system's behaviour [20]. Fuzzy Logic (FL) rule-based systems based on Fuzzy Logic sets and Fuzzy Logic theory provide a potent yet succinct approach for representing complex systems [21].

Fuzzy Logic systems (FLS) are a very useful tool in building human-centric models [21]. FLS's strong capacity to model/represent complex systems and provide system transparency is largely attributable to its use of IF-THEN statements, which are straightforward language principles that can be understood in obtaining an output value [22], [23]. Considering FLS as models has other advantages too, for example for their ability in handling any non-linear problems [24] as well as model's interpretability due to the linguistic nature of FLS [21].

2.2.1 FUZZY LOGIC SETS

A Fuzzy Logic set is a mathematical object in which membership is continuous [25]. Despite being crisp (as in classical set theory), it has a membership function which ascribes a component to a membership grade. In classical set theory, the membership of one element is either 1 or 0, or simply means that it belongs to that set, or does not. For example, given a set A, the membership function for each $x \in X$ is either:

$$\mu_A(x) = 1 \; or \; \mu_A(x) = 0 \tag{2.1}$$

where X is a collection of objects denoted generically by x [25].

In contrast, fuzzy set theory allows for a gradual evaluation of an element's membership in a set. This is stated using a membership function with a value in the real unit range [0, 1] [26]. According to Zadeh [27], a Fuzzy Logic set A in X is a set of ordered pairs:

$$A = \{\big(x, \mu_A(x)\big) : x \in X\} \tag{2.2}$$

where $\mu_A(x)$ is the membership function that maps X to a real number in the interval [0,1].

The purpose of the membership function (MF) is to establish the relationship between each input and the degree of membership used to assess usability [28]. In addition to mapping the MF, it is also necessary to determine the shape of the MF. Commonly used shapes include triangular, trapezoidal, Gaussian [26] and bell shape [22] as shown in Figure 2.1. The membership functions for each shape can be expressed as [22], [29] ,[30], [31]:

- Gaussian MF

$$f(x; c, \sigma) = e^{\frac{-(x-c)^2}{2\sigma^2}} \tag{2.3}$$

- Triangular MF

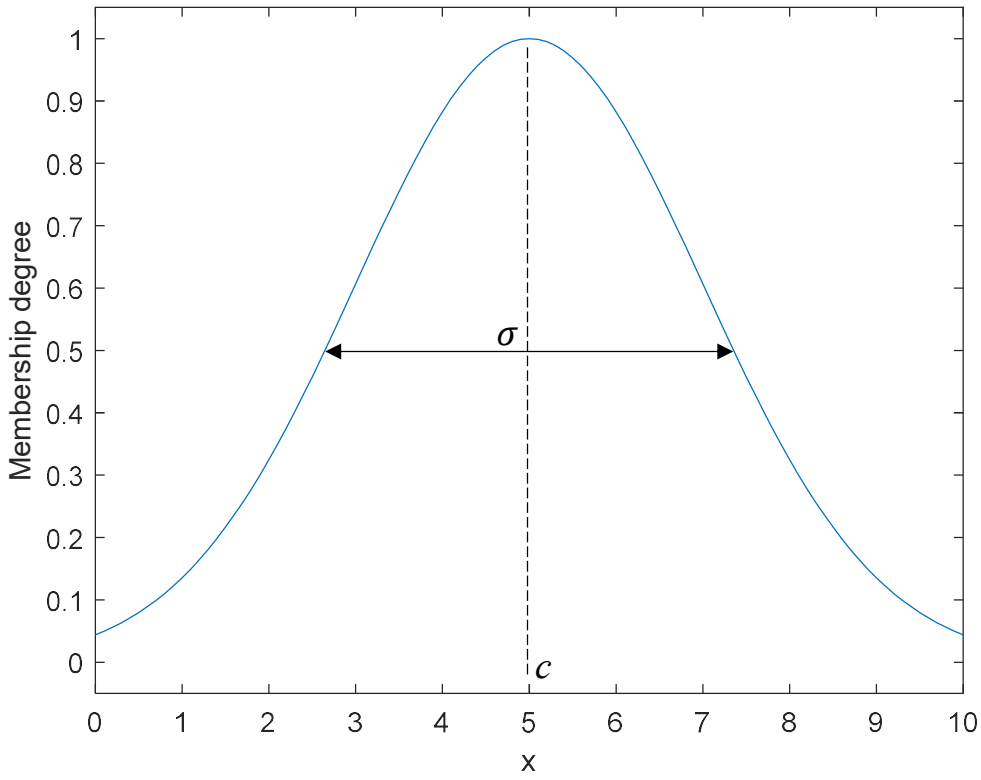
$$f(x; a, b, c) = \begin{cases} 0, \; x \leq a \\ \frac{x-a}{b-a}, a \leq x \leq b \\ \frac{c-x}{c-b}, b \leq x \leq c \\ 0, \; x \geq c \end{cases} \tag{2.4}$$

- Trapezoidal MF

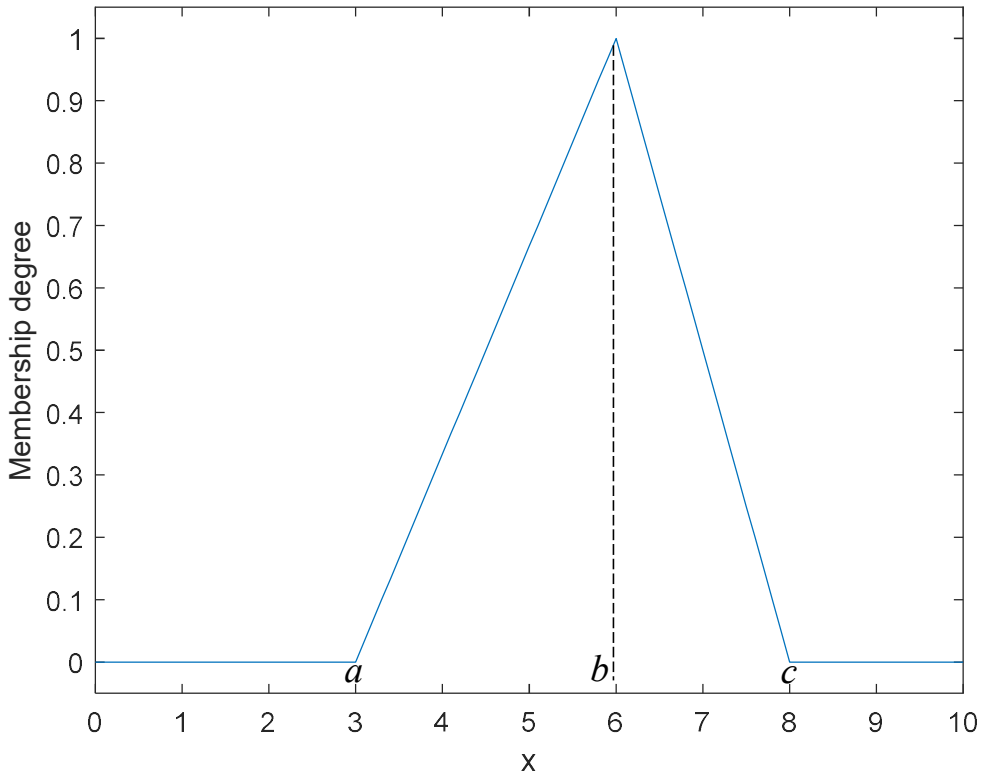
$$f(x;a,b,c,d) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ \frac{d-x}{d-c}, & c \le x \le d \\ 0, & x \ge d \end{cases} \tag{2.5}$$

- Generalised Bell MF

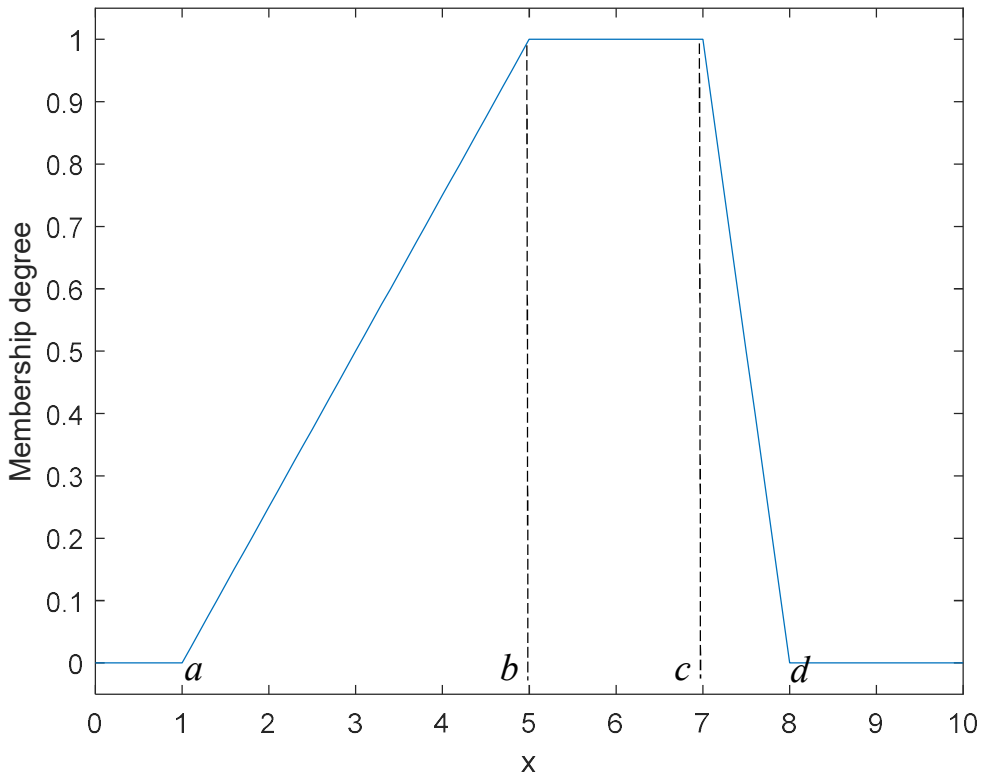
$$f(x;a,b,c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \tag{2.6}$$



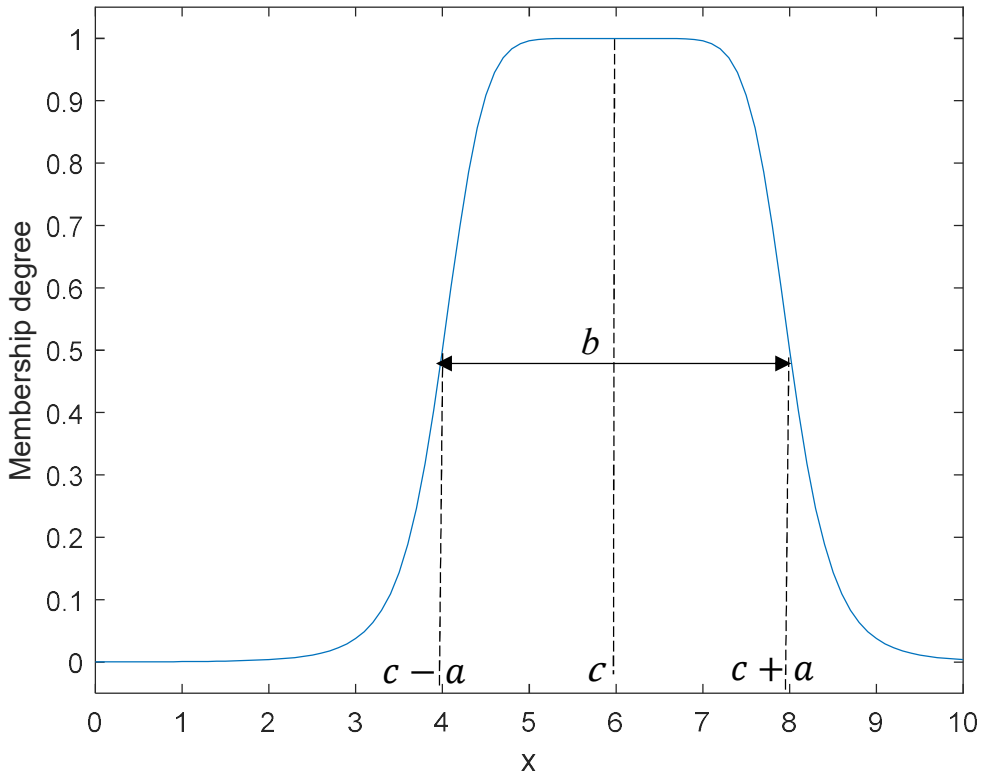
(a) Gaussian MF



(b) Triangular MF



(c) Trapezoidal MF



(d) Bell MF

Figure 2.1: Shapes for Membership Functions (MFs)

2.2.2 FUZZY LOGIC SYSTEMS

The general configuration of FL model consists of five main components, which are Fuzzy Logic rule base, fuzzy inference engine, database, fuzzifier and defuzzifier [4] as shown in Figure 2.2. The fuzzifier converts the numerical inputs into degrees of membership to the Fuzzy Logic sets of the input variables [32]. In this stage, membership functions are used to represent Fuzzy Logic sets graphically [33]. This process of converting crisp input into linguistic variable uses the MFs stored in the database and is called fuzzification.

The inference engine then utilises the fuzzy measurements to evaluate the control rules kept in the fuzzy rule base [34]. The rule base represents a set of IF-THEN rules that have been identified during the information granulation level. The database contains all of the values for the parameters of the rule-based model [4]. Inference engine is where the combination of rules occurs [35] while the final step,

defuzzification occurs in the defuzzifier. Here, the fuzzy output is transformed back to crisp value.

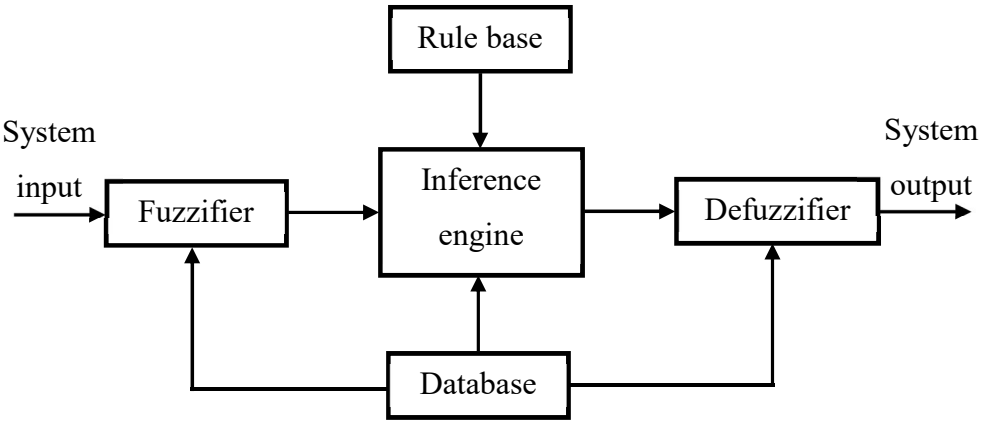


Figure 2.2: The general configuration of Fuzzy Logic model

A Mamdani-type Fuzzy Logic system rule has the following form [36], [22]:

$$R^i : \text{IF } x_1 \text{ is } F_1^i \text{ and } \ldots \text{IF } x_m \text{ is } F_m^i \text{ THEN } y \text{ is } G^i \tag{2.7}$$

where R^i denotes the i-th rule, $i = 1,2, \ldots, N$, and N is the total number of rules. F_m^i and $G^i$are Fuzzy Logic sets in the input space, $x_1, \ldots, x_m$ is the FL system's input and y is the output. In contrast, the consequent propositions of Takagi-Sugeno FL models [37], are functions of the antecedent propositions rather than FL propositions [4]:

$$R^i : \text{IF } x_1 \text{ is } F_1^i \text{ and } \ldots \text{IF } x_m \text{ is } F_m^i \text{ THEN } y \text{ is } f_i(x) \tag{2.8}$$

2.2.3 NEURO-FUZZY SYSTEMS

Combining two or more artificial intelligence techniques is known as hybridisation [24]. A number of hybridisation techniques have been applied to combine with FL systems, for example with neural network (known as neuro-fuzzy) [38], Hybrid GT2-Support Vector Machine (GT2-SVM) [39] and Hybrid Fuzzy Genetic Algorithm [40].

Researchers are increasingly turning to neuro-fuzzy systems in an effort to address challenges such as the elicitation of FLS from data, or the lack of transparency of neural networks. Even though a neural network can be a very powerful tool for forecasting [41], however, it is often treated as a black box, lacking transparency [42]. This limitations led to the formation of neuro-fuzzy systems [42]. Moreover, these two algorithms are similar in terms of their uncomplicated algorithmic procedure (instead

of a complex mathematical analysis) and tunable parameters, that allows the hybridisation process [43].

In neuro-fuzzy systems, the models are developed in two steps: identification of FL rules and tuning of rules [44]. First, the FL rules are extracted from numerical data, and then the neural network adjusts the membership functions through the learning process [41]. Therefore, the resulting neuro-fuzzy system has the advantages of the two; learn like neural networks, and express knowledge similar to that of humans using FL systems' capacity for deriving linguistic rules, hence improving the transparency of the whole system [45].

2.3 TYPE-2 FUZZY LOGIC SETS AND SYSTEMS

Fuzzy Logic provides a framework to model the uncertainty prevalent in the majority of real world problems [46]. In reality, information could be imprecise, contradictory, or consist of ambiguity (information that can be interpreted in many ways), therefore leads to uncertainty. These various degrees of uncertainty have a significant impact on the decision-making process. More details about the types of uncertainty are described in Chapter 3.

Considering the current study of fuzzy sets (FSs), the idea of determining the true meaning of uncertainty in linguistic variables has led to the developed three primary FS representations - type-1 fuzzy sets (T1-FSs), interval type-2 fuzzy sets (IT2-FSs), and general type-2 fuzzy sets (GT2-FSs) [46] , as well as other higher order (Type-n) [47]. As the most basic type of linguistic variable representation in this situation, T1-FSs can only specify a limited level of imprecision or ambiguity [46].

However, in more complicated circumstances, it can be very challenging to determine the exact numerical value of an entity's membership or to provide a specific membership value for any ambiguous entity. As a result, membership functions in T1-FSs may also experience uncertainty due to improperly formed fuzzy rules [48].

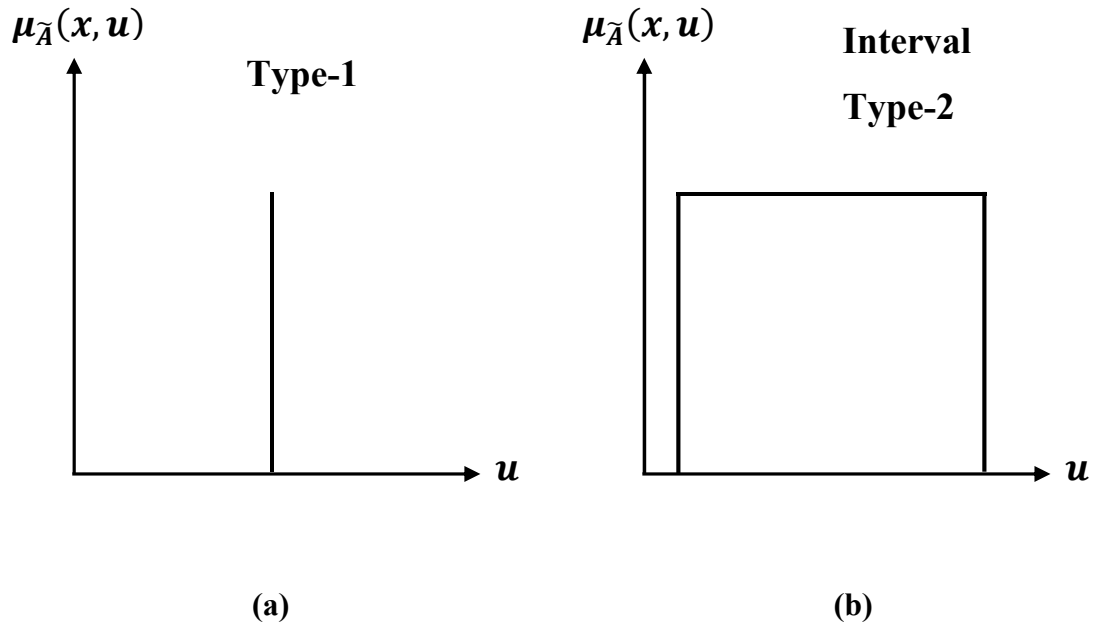
Type-2 fuzzy sets (T2-FSs) offer extra design degrees of freedom, which can be very helpful in the presence of lots of uncertainties [49]**.** Unlike T1-FSs that have a crisp degree of membership, the degree of membership in T2-FSs is fuzzy, which is very helpful when the exact value of MF is hard to obtain [50]. Thus, the resulting type-2 Fuzzy Logic Systems (T2-FLS) have the potential to achieve higher performance than

their type-1 counterpart [50], especially when dealing with noisy data and different word meaning [51]**.** This trait of FS has attracted many researchers to use T2-FLS, for example the works presented in [52], [53], [54] and [10].

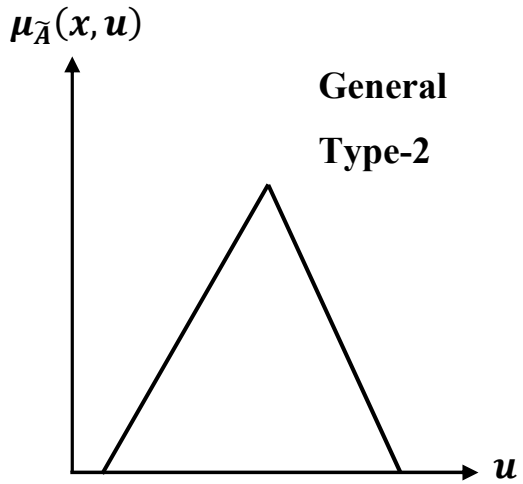
2.3.1 GENERAL TYPE-2 FUZZY LOGIC SETS

Due to their ability to outperform their T1 counterparts in terms of system performance, Type-2 FLS have received the majority of attention, however, the focus within the last two decades is mainly on interval type-2 (IT2) FLS [55], [56]. General type-2 fuzzy sets (GT2-FSs) on the other hand, give a way to describe larger levels of uncertainty due to the additional degrees of freedom that its third dimension offers [46].

Instead of using interval FSs (as in IT2-FLSs) [55], GT2-FLSs employ T1-FSs as their secondary MFs**.** In other words, the secondary membership function of a GT2-FS is itself a type-1 fuzzy set [56]**.** Therefore, one may expect that the structure and design of GT2-FLSs is more complex compared to their T1 and IT2 counterparts with more parameters to be tuned. However, the output of a GT2-FLS can be easily determined with the presence of new representations [56] namely z-slices [57], alpha-planes [58] and alpha-cuts [59]. The secondary membership functions for Type-1, Interval Type-2 and General Type-2 (in 2-D) are shown in Figure 2.3.



(a) **(b)**



(c)

Figure 2.3: Secondary membership functions for Type-1, Interval Type-2 and General Type-2 (reproduced from [57])

In Figure 2.3 (a), there is only one value in the secondary MF domain for T1-FS. Therefore, a is certainly the membership degree for the particular x value, indicating that there is not uncertainty related to the primary MF [57]. Whereas in IT2-FS (in Figure 2.3 (b)), the uncertainty in determining the secondary MF is at the peak; with a secondary membership of 1 being assigned to each point in the primary MF interval of [a,b]. The uncertainty in GT2-FSs illustrated in Figure 2.3(c), can be modelled to any degree between T1 and IT2-FSs, for instance, by a triangular secondary MF. As a result, GT2-FSs can accurately represent the uncertainty in the third dimension, from almost minimal uncertainty to maximum [60].

The bivariate membership function $\mu_{\tilde{A}}(x,u) \subseteq [0, 1]$ defines the GT2-FS (denoted as $\tilde{A}$) , with the primary variable is $x \in X$ .Thus $\tilde{A}$ is defined as:

$$\tilde{A} = \{(x,u),\ \mu_{\tilde{A}}\ (x,u) | \forall x \in X, \forall u \in J_x \subseteq [0,1]\ \} \tag{2.9}$$

in which $0 \leq \mu_{\tilde{A}}\ (x,u)| \leq 1$.

An α-plane is characterised by $\tilde{A}_\alpha$, is the union of the primary MFs of $\tilde{A}$ with α $(0 \leq \alpha \leq 1)$:

$$\tilde{A}_\alpha = \{(x,u),\ \mu_{\tilde{A}}\ (x,u) \geq \alpha | x \in X, u \in [0,1]\ \} \tag{2.10}$$

Figure 2.4 shows the illustration of GT2-FS with trapezoid T1-FSs and respective α-plane. Note that by raising an α-cut to level α one obtains an α-plane, and footprint of uncertainty $FOU\left(\tilde{A}\right) = \tilde{A}_0$ [61]. Footprint of uncertainty (FOU) is the 2-D support of $\mu_{\tilde{A}}\ (x,u)$, which is bounded by lower and upper MFs (or $LMF\ (\tilde{A})$ and $UMF\ (\tilde{A})$, denoted as $\underline{\mu}_{\tilde{A}}(x)$ and $\bar{\mu}_{\tilde{A}}(x)$, respectively.

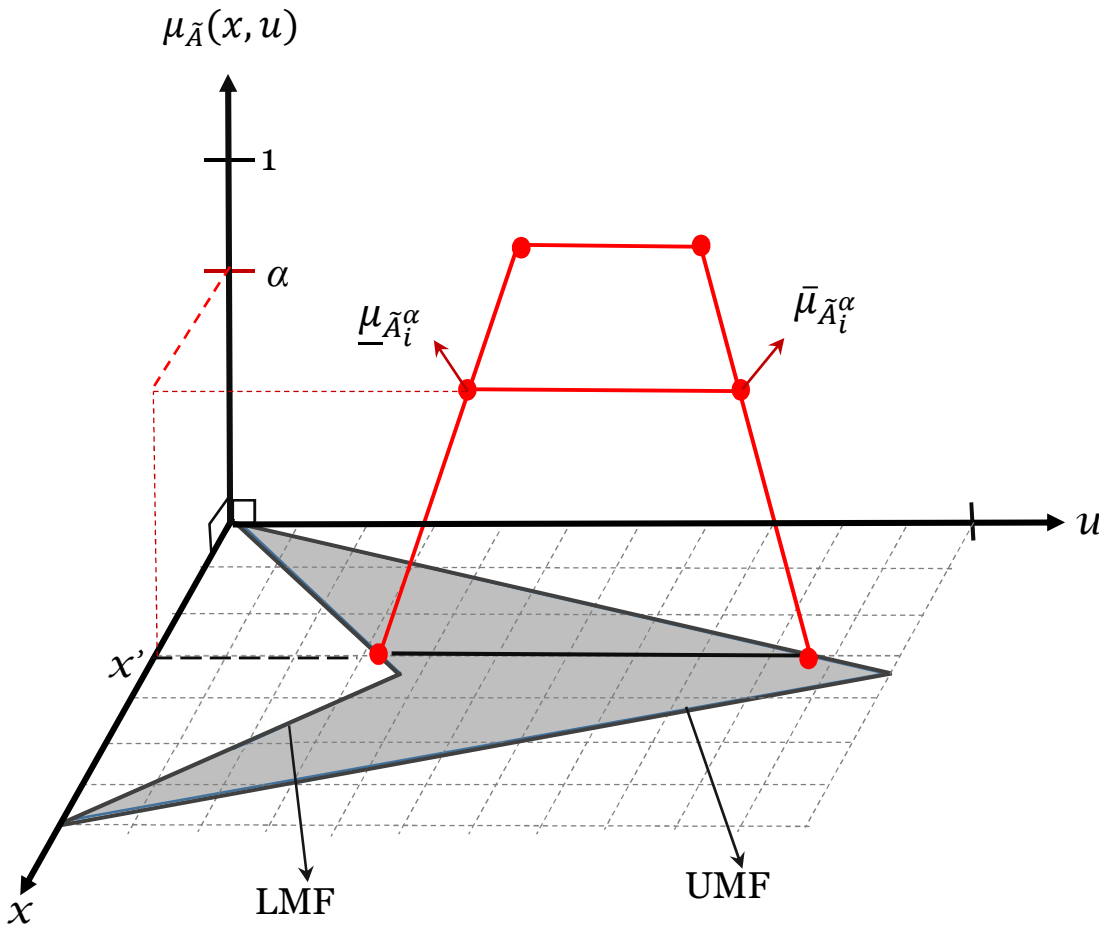


Figure 2.4: Illustration of a GT2-FS with α-plane. (Reproduced from [55])

2.3.2 GENERAL TYPE-2 FUZZY LOGIC SYSTEM

The main components for a type-2 FLS are fuzzifier, rule base, inference engine, type reducer, and defuzzifier [57] as shown in Figure 2.5. In a type-2 FLS, the inputs and outputs of the FLS are represented with type-2 Fuzzy Logic sets. It has similar a structure as the type-1 FLS, where the knowledge is characterised by IF-THEN linguistic rules [22].

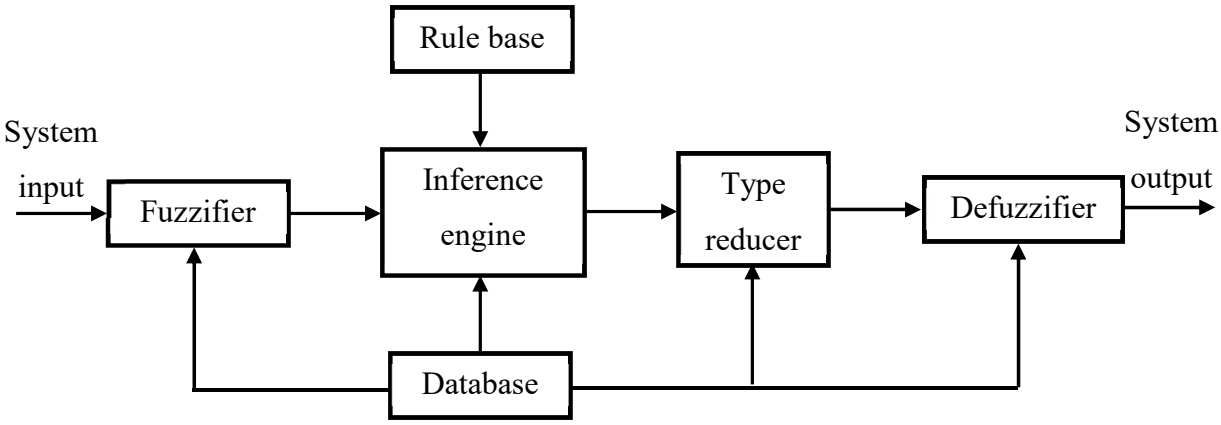


Figure 2.5: The configuration of T2-FLS

The description of each type-2 FLS' components are as follows [57]:

- Fuzzifier: To process within the FLS, the fuzzifier converts crisp inputs into general type-2 Fuzzy Logic sets
- Rule-base: The rules are similar as the IF-THEN form (as in T1-FLS), but the antecedents and consequents are of type-2 Fuzzy Logic sets.
- Database: The database contains all of the values for the parameters of the rule-based model.
- Inference engine: The computational technique used to determine the rules' firing strength for a specific fuzzified input pattern.
- Type reducer: converts the output type-reduced sets (T1-FSs) from the inference engine, which produces T2-FSs as output.
- Defuzzifier: produces a defuzzified crisp number from the type reducer's output.

2.4 INTERPRETABILITY OF FUZZY LOGIC SYSTEMS

The concept of interpretability has become more visible and has begun to gain prominence in light of recent advancements in explainable artificial intelligence (XAI) [62]. FL rule-based systems are useful for extracting knowledge from data [63] and formalise the behaviour of a real system in a human understandable way [64]. In reality, an FLS's interpretability depends on how well its fundamental parts—namely, its rule base and its antecedent and consequent Fuzzy Logic sets—are understood [65]. The interpretability feature of FLS is the key reason to use FLSs in decision making [66], which provides insight into why and how certain results are produced [65].

Even though there is no standard measure to assess interpretability, most researchers would agree on defining the interpretable system as having minimum number of rules and input variables and easily understood rule premises [67]. Unlike accuracy, where the definition is straightforward, the definition of interpretability is rather challenging. Therefore, Gacto et al. [67] proposed a taxonomy that is based on the complexity and semantic interpretability, being measured at both rule base and fuzzy partition level. Complexity-based interpretability measures aim to make the model less complex, and the common indices include number of rules, variables and labels per rule. Semantics-based interpretability methods focus on maintaining the semantics connected to the MFs. They strive to maintain semantic integrity by placing restrictions on the MFs or methods that take into account factors like distinguishability and coverage.

The combination of complexity – semantic measures and rule base – fuzzy partition components prompts analysis of four distinct quadrants:

1. Complexity at the rule base (RB) level (Q1)
2. Complexity at the fuzzy partition level (Q2)
3. Semantics at the RB level (Q3), and
4. Semantics at the fuzzy partition level (Q4).

Table 2.1 shows the double axis taxonomy to analyse the interpretability of FL rule-based systems.

Table 2.1: Double axis taxonomy to analyse the interpretability of FL rule-based systems.

	Rule base level	**Fuzzy partition level**
Complexity-based interpretability	**Q1** Number of rules Number of conditions	**Q2** Number of membership functions Number of features
Semantic-based interpretability	**Q3** Consistency of rules Rules co-firing Rules relevance	**Q4** Coverage Normalisation Distinguishability

Q1. Complexity at the rule base level

The most commonly used measures in this quadrant are the number of rules and number of conditions. The basic principle is that, the rules should be easier to understand if there are fewer of them and they are shorter in length [68]. The FL rules of a fuzzy inference system have the following syntax: if (input fuzzy condition) then (output fuzzy assignment). The rule's antecedent is the input condition, and the rule's consequent is the assignment of the output [69]. According to [67], the maximum number of circumstances in an antecedent to a rule is 7 + 2 different conditions, which corresponds to the maximum number of conceptual entities that an individual can manage.

Recent works that aims at reducing the complexity at the rule base level are reviewed here. Hao et al. [70] proposed an approach for extracting reduced rules based on bias random forest (BRF) and fuzzy support vector machine (SVM). To solve the issues of similar, repetitious, and ineffective conditions and rules brought on by the autonomous learning of each tree in the ensemble method, they developed a reduction approach based on error rate and coverage rate. In [71], a rule-reduced algorithm was suggested. Rules are streamlined by the fuzzy basis functions' sparse encoding. The least angle regression approach is suggested for choosing the most significant rules.

Q2. Complexity at the fuzzy partition level (Q2)

The most common measures in this quadrant are number of features and number of membership functions. Most of the works in this quadrant are related to feature selection and feature reduction. For example Hein et al. [72] provided fuzzy genetic programming (GP) reinforcement learning (FGPRL), a fuzzy GP method that can choose the pertinent state features, estimate the size of the necessary fuzzy rule set, and instantly change all the controller settings. With regard to a specific degree of performance, this GP method can automatically choose the most significant features as well as the most compact fuzzy rule representations.

In another work, Casteillo et al. [63] introduced a method for generating interpretable fuzzy partitions with the best granularity, known as DC*. The method is implemented in two steps: identifying the multidimensional clusters, and further cluster the one-dimensional projections along each dimension at the same time. In 2019, Hajek [73] developed a unique fuzzy system that combines a feature selection and rule extraction element to produce an interpretable system in terms of rule complexity and granularity. A genetic feature selection is employed to exclude unimportant qualities before conducting a comparison study of cutting-edge FRBS like Fuzzy Unordered Rule Induction Algorithm (FURIA) and evolutionary FL rule-based systems. Another work in this area includes the introduction of Fuzzy Feature Rank, a novel feature reduction approach presented in [66].

Q3. Semantics at the rule base level

The important interpretability measures under Quadrant 3 are listed below:

1. Co-firing rules

Co-firing rules are the consequent of having many Fuzzy Logic sets overlap with each other [74]. Guo et al. [75] demonstrated the high level of interpretability by reducing the co-firing rules that lead to many redundant rules. As a solution, they proposed Takagi-Sukeno-Kang (TSK) FL model that integrates bagging and dropout algorithms. The enhanced dropout technique avoids the issue of rules co-firing and only keeps nodes with high activation, enhancing the quality of created rules.

2. Consistency

Consistency [17] is defined as having each combination of antecedents with only one possible consequent label. Souza et al. [76] considered the interpretability when proposing an evolving fuzzy neural network for regression problems. To measure the interpretability, they examined the consistency in a rule base, where there is not overlapping in their antecedents and consequents, or having similar consequents in the case of overlapping. Here, the consistency is measured by determining the similarity of rule antecedents and consequents.

With the same objective to have consistent rules, Gegov et al. [77] suggested a technique for FL systems to simplify the rule base. The approach is based on aggregation of inconsistent rules, or rules with inconsistent input and output linguistic values, given identical permutations of input linguistic values. By substituting each set of incompatible rules with a single equivalent rule, the simplification eliminates redundancy in the FL rule base. Some other recent works regarding the consistency in FL rules can be seen in [78] and [79].

3. Relevance

Relevance is a criterion for FL rules interpretability, being measured by examining the data covered by the antecedent and consequent, and very useful as a rule reduction technique [80]. Rey et al. [80] addressed the FRBS Accuracy- Interpretability trade-off through a multi-objective evolutionary-based rule selection, choosing the most suitable and important rules based on the metrics of accuracy, interpretability and relevance. This approach has demonstrated that, for both scatter and linguistic FRBSs, rule relevance plays a significant role in the Accuracy-Interpretability trade-off.

In another work, Dutu et al. [81] implemented the Selection-Reduction (SR) method that comprises two phases. First, the most relevant rule is selected from each input space, and second, the irrelevant rule is pruned in the reduction phase. Other than rule reduction purpose, the rule relevance is also used to rank the quality of the rules, as presented in [82].

Q4. Semantics at the fuzzy partition level

In this quadrant, the interpretability is viewed at the fuzzy partition level. The interpretability measures in this quadrant involve the properties of the membership functions (MFs), for example the issue of overlapping (related to distinguishability) and the universe covered by the MFs (known as coverage). Hence the most significant measures defined in this framework are:

1. Normalisation

In this criterion, there should be at least one data point in the universe of discourse having maximum degree of membership of one.

2. Distinguishability

A system with a lot of highly overlapping Fuzzy Logic sets that hinder the system's interpretability is not uncommon [83]. Many works have been conducted to deal with highly overlapping MFs in order to promote the distinguishability among the MFs. For example, Ali et al. [84] introduced a novel method named laser simulator inference system to solve the problem of high overlapping among linguistic variables. In this research, the inference of linguistic variable values within a specific range are used to calculate the high overlapping.

With the same objective to minimise the MFs overlapping, Fuchs et al. [83] combined two methods, namely Jaccard similarity and graph theory within a framework known as Graph-Based Simplification (GRABS) to detect the highly overlapped MFs. This allows the similar Fuzzy Logic sets to be eliminated from the rules. Similarly, another research by Shanmugapriya et al. [85] also attempted to find an optimal combination of Fuzzy Logic sets. In this work, they proposed Similarity Estimator (SimE), in which the area of overlap represents the similarity between Fuzzy Logic sets.

3. Coverage

Coverage, or completeness [67] is achieved when the entire universe of discourse for any input variable is fully covered by the MFs [86]. Among recent works within this quadrant include [87], in which Megherbi et al. incorporated the aspect of coverage in fuzzy partition when proposing a Fuzzy Logic controller based on an overlap encoding strategy. The main idea in this work is to define the MFs as overlapped functions instead of separate functions.

In another research, Lloret-Climent et al. [88] used the important parameters in FL systems such as membership functions, inclusion, union and support to develop a systemic approach to complex systems. As a result, the idea of coverage and invariability between sets of variables will automatically emerge as a result of the direct and indirect interactions between variables, which will serve as the foundation for getting fuzzy and/or non-fuzzy connections.

One of the most popular interpretability index that incorporates the aspect of coverage is Nauck's index (NI). NI is a numerical index introduced by Nauck [89] to measure the interpretability of FL rule-based classifiers [90]. NI close to one indicates high interpretability of the FLS, while NI close to zero indicates low interpretability [90]**.** Nauck's index is computed as the product of three components [19]:

- Complexity of FLSs, which is measured as the number MFs of output variables divided by the number of input variables.
- Coverage degree of fuzzy partitions. For this component, the complete coverage is achieved when membership degrees for each element of the domain adds up to 1 for a small number of Fuzzy Logic sets [91].
- Partition index, which is calculated as the inverse of the number of MFs minus one for each input variable [19]. The purpose of this component is to penalise the partitions with high granularity [89].

With the increase of interest in FLSs interpretability, there is significant work in the literature that utilises NI as an interpretability measure. Examples of work using NI within the type-1 FLS framework are [92], [18] and [91]**.** NI also has been demonstrated to be a good indicator of interpretability for special types of FLS, namely hierarchical fuzzy systems (HFSs) as presented in [16], [90], [93].

In type-2 FLS framework, a number of researchers have been using NI in measuring the system's interpretability. In [94], Shukla et al. assess the interpretability using NI after adjusting the parameters of interval type-2 MFs based on genetic algorithm tuning approach. They demonstrated the capability of interval type-2 FLS in modelling uncertainty and achieving good interpretability. In [52] Chandra et al. presented an experimental analysis to address the interpretability quantification (with NI) and accuracy measurement both type-1 and interval type-2 implementation. In

contrast with [94], they demonstrated the capability of type-2 FLS to improve the accuracy of the prediction, but with sacrifice in terms of its interpretability. Interestingly, both works presented a single numerical value of NI, even though an interval type-2 set is described by its upper membership function (UMF) and lower membership function (LMF).

2.5 GRANULAR COMPUTING

The main focus in Granular Computing (GrC) is the development and processing of information granules – formal objects that provide a way of organising information about the data and underlying relationships [3]. GrC as a concept represents the human capacity to perceive the real world under multiple levels of granularity [95]. As a novel multi-disciplinary paradigm that is a crucial component of artificial intelligence, GrC has gained a lot of attention recently. This is due to its capability in providing different levels of knowledge, thus having a major impact on the development and use of intelligent systems, such as classification tasks [95]. Therefore, in the big data era, GrC plays a critical role due to its nature of manipulation, understanding and analysis of the data.

Information granules have been viewed as the basic building blocks that characterise vast amounts of numerical data in an efficient and abstract manner [96]. According to Merriam–Webster's Dictionary, granule is defined as "a small particle; especially, one of numerous particles forming a larger unit". In GrC, the term "granule" refers to any subsets, classes, objects, clusters, and constituents of a universe, which is very similar to the definition given above.

Information granules are essential concepts in human cognitive and decision-making processes [3] [97]. They are defined as collections of objects, typically originating at the numeric level, that are arranged together due to their similarity, functional adjacency, and indistinguishability or alike [3], [97]. GrC includes the features of representing, developing, processing, and exchanging information granules [3].

Several implementations of information granules have been established, including fuzzy sets, rough sets [98], [99], probabilistic sets [100], interval analysis and data clustering [101]. One prominent way to obtain information granules is through clustering approaches, such as DBSCAN, Fuzzy c-means (FCM) and k-means

clustering. In fact, clustering methods, which technically involve creating a partition of the input data, are one of the most well-known examples of ways to produce information granules [98].

Multiple computational techniques that draw inspiration from nature are combined under the data-driven paradigm in the research area known as computational intelligence (CI) [98]. CI is a science that excerpts system structures from big data and identifies models or patterns [102]. According to The Institute of Electrical and Electronics Engineers (IEEE), computational intelligence is defined as "the theory, design, application, and development of biologically and linguistically motivated computational paradigms emphasising neural networks, connectionist systems, genetic algorithms, evolutionary programming, FL systems, and hybrid intelligent systems in which these paradigms are contained"[103]. It has been recognised that these CI techniques have the capability to process inaccurate information and provide good solutions while ensuring robustness and computational tractability [104].

Today, it is well-established that information granules development and CI systems may be integrated within the same framework [98]. As the name implies, granular models are modelling structures that are developed at the level of information granules [105]. Granular neural networks, for instance, provide an intriguing example [98]. For example, Ghiasi et al**.** [106] applied a hybrid algorithm of GrC and artificial neural network (ANN) named as GRC-ANN to estimate the longitudinal dispersion coefficient (LDC). In [107], an interval type-2 (IT2) fuzzy granular neural network dynamic ensemble approach is proposed to handle the uncertainties included in the definition of granular data streams. Other works related to granular neural networks can be seen in [108], [109], and [110].

Another application of information granules can be seen in fuzzy inference systems. In GrC, a granule can be thought of as each rule that makes up a rule-based system [111]. FL rule-based systems are commonly encountered examples of granular models [105]. For instance, Yeom at al. [112] presented the optimised FL-based granular model on the foundation of a hierarchical structure and the best possible information granule allocation. They also proposed Gustafson Kessel (GK) clustering to form the information granules that maintain the coverage among the granules.

In [113], Shan et al. applied fuzzy c-means clustering to develop an interval granular FL model. The key idea in this research is to build an FL model at the level of information granules around the input-error data. Other than fuzzy c-means clustering [114], [115], other clustering algorithms have been applied to build granular fuzzy models, such as k-means [116], hierarchical clustering [117], [21], DBSCAN [118], [119] and iterative data granulation [120], [121] and [51].

2.6 SUMMARY

This chapter provides literature review on the theoretical concepts of Fuzzy Logic systems, comprising the concept of Fuzzy Logic sets and general configuration of Fuzzy Logic systems, both in Type-1 and Type-2 models. Focus is given to the main feature of Fuzzy Logic systems, which is interpretability. In this chapter, the interpretability of Fuzzy Logic systems is categorised in four different quadrants**,** which is based on the complexity and semantic interpretability, being measured at both rule base and fuzzy partition level. In addition, the fundamentals of granular computing and formation of information granules are presented since they are the focus in this thesis.

In the next chapter, the formation of information granules using iterative data granulation is presented. The main contribution is to rigorously quantify the uncertainty during the granulation process and use the information to enhance the iterative data granulation. The enhanced information granules then are used to build Fuzzy Logic rule-based systems. The important concept in information theory, entropy is leveraged to quantify the uncertainty, hence assists in constructing higher quality information granules.

CHAPTER 3

QUANTIFYING AND UTILISING INFORMATION UNCERTAINTY IN GRANULAR COMPUTING

This chapter considers the quantification of uncertainty during the iterative data granulation process. It is proposed here that uncertainty is measured using Shannon entropy, a key constituent of information theory. The main idea is to promote the merging of information granules that have minimum uncertainty to avoid having granules with high disorder in the data distribution. Therefore, the uncertainty is proposed to be included in the compatibility measure to produce higher quality information granules. The impact of this proposal in the modelling of systems, via Fuzzy Logic rule-bases is considered. The final information granules are used to build Fuzzy Logic rule-based systems, and the proposed framework is tested with datasets from UCI Machine Learning Repository, particularly in classification problems.

This chapter aims to create a new framework for iterative data granulation, using Granular Computing, via the use of information theory. The underlying research aims are to a) rigorously quantify uncertainty accumulation in the granulation process, b) the use of this information to enhance the iterative data granulation and c) to use the uncertainty quantification in building higher order FL rule-based systems. There is strong link between information granules and FL rule base, hence by having low uncertainty granules will improve the interpretability of FL rule base.

3.1 INTRODUCTION

When the physical relationships are difficult to understand, a Fuzzy Logic model is appropriate for addressing a non-linear problem [4]. FL models are models whose architecture is based on Fuzzy Logic set structures (Fuzzy Logic sets, Fuzzy Logic

relations, and Fuzzy Logic set operators) [122]. Identifying relationships between the inputs and outputs of a system can be utilised to create a rule-base for the system [22]. The importance of rule-based techniques can be explained by the fact that human experts frequently prefer to express their knowledge in terms of IF–THEN rules, which connect the values of a set of input variables with output variables [123]. FL rules are a powerful tool for capturing and, perhaps more significantly, interpreting the behaviour of input/output data [124].

There are two advantages of FL. The first one is its capability to define complex systems in compact formats. This enables FL to be universal approximators that can realise nonlinear mappings [21]. The second advantage is intuitive interpretability due to its analogue nature. In comparison to "black-box" models like neural networks, interpretability derives from language interpretation, which is near to human thinking. It appears to be a distinctive trait and major advantage [123]. However, this approach is limited because it requires a lot expertise [4]. An expert may not always be able to articulate his or her knowledge directly, and in certain cases, an expert may not even exist for particular systems and case studies [124].

To overcome this issue, knowledge can be discovered from data to determine the FL parameters [4]. Tasks that create and learn FL rules from numeric data are working with various approaches, and one of them is clustering. Some other examples include heuristic methods, neuro-fuzzy algorithms, data mining, and genetic algorithms [125]. A number of clustering approaches have been applied to build FL rule base, such as fuzzy c-means [126], hierarchical clustering [127], mountain method [128] and recently iterative data granulation [129], [130], [9].

Essentially, the formalism linguistic variables and FL rules can be formed by using data granulation [131]. Data clustering is the process of grouping data points that are comparable in some way into clusters. Projecting data clusters onto the dimension of the inputs yields input membership functions [125]. Hierarchical clustering is one of the methods that is most frequently used. The strategy of hierarchical clustering is to build clusters based on hierarchy, and these strategies fall into another two main categories, which are agglomerative and divisive [132]. The former is more widely used compared to the latter [133].

In 2002, Pedrycz et al. introduced a granular computing (GrC) algorithm [6] that has similarity with the concept of agglomerative hierarchical clustering [134]. However, these algorithms differ from one another significantly. In the GrC algorithm proposed, the original data and the information granules are closely related. This is due to the fact that each granule comprises of sub-granules [9], all linked directly to the underlying data. Besides, the compatibility metric provided in GrC is a crucial tool that can be utilised as direction to end the clustering process [135]. Throughout the iterative granulation process, the decrease of compatibility measure can be visualised, signifying the increasing dissimilarity between granules towards the end of granulation process.

The fundamental concept in GrC is the information granulation [4]. The goal of data granulation is to extract information from raw data [129]. It is accomplished through data organisation [6], and results in compressed information granules [129]. Granulation is the process of arranging comparable elements into granules to create coarse-grained worldviews [136]. Information granule is a collection of objects that are drawn together by some constraints [137], such as similarity, indistinguishability, and functionality [136], [138]. In the granulated universe, elements within a granule are considered as a whole rather than individually.

FL models can be established utilising data mining techniques and specialist knowledge, hence leads to the most significant feature of GrC - to discover and excerpt information. GrC imitates the way of human thinking when grouping similar objects or instances [129]. The linguistic FL rules can be created using the information gathered from the GrC, in the form of information granules, to create a structured modelling framework [100]. GrC can be used to transparently create an FL model [9]. The definition of transparent here refers to the relationship between data and information granules in the process of knowledge discovery and the application of this knowledge to create the linguistic rule-base that is useful in deriving a particular prediction.

However, real-world data is frequently associated with uncertainty [139]. Merriam-Webster dictionary defines uncertainty as the state of being uncertain; and Klir [140] defined uncertainty as the result of lack of information, such as incomplete, vague or inconsistent. Systems can frequently handle flawless or perfect data, but the data encountered are always uncertain [139]. Data is always characterised by uncertainty due to imprecise measurement [129]. This increases interest for emerging

knowledge discovery methods that are capable to deal with this challenge. These uncertainties may affect the formation of information granules. Low quality granules will demolish its distinguishability and hence result in inaccurate FL rule base.

In several domains, such as machine learning, pattern recognition, image processing, medical, and data mining, uncertainty quantification is critical [137]. This is something that several academics have looked into. For example, Liang [141] presented the concepts of information entropy and information granulation-based uncertainty measures to measure the underlying semantic capacity of an incomplete information system. Based on entropy and information granulation, Zhang et al. [142] proposed an improved method for causality inference. In this method, the estimation of entropy is enhanced utilising a new framework that employed information granulation as a crucial step. In an incomplete information system, Qian et al. [143] introduced combination entropy CE(A) and combination granulation CG(A), in which the gain function has intuitionistic knowledge content characteristic, i.e., the total number of pairs of elements that can be distinguished from each other on the universe. In another research, Zhang [144] suggested a new feature selection algorithm based on fuzzy information granules and approximate conditional entropy, and demonstrated the accuracy of the approach utilising the entropy.

The concept of uncertainty can be integrated into GrC by integrating uncertainty management as an additional attribute [145]. From the perspective of information granules, Sanchez et al. [145] defined uncertainty as dispersion in a sample of data. Low data dispersion indicates low uncertainty, because all the data points are close to each other in a compact manner, which will produce small standard deviation. There are several efforts to quantify uncertainty to guide the process of iterative granulation. For example, in [130] the authors used a neutrosophic logic concept to measure the hesitation caused by indistinguishability. The measure of hesitation is calculated by using neutrosophic index and this results in a final Granular Computing-Neural Fuzzy (GrC-NF) inference system with a rule-base. However, this method only tackles the problem of fuzziness and neglects the uncertainty due to the randomness of the data. In another research, Baraka et al. [129] proposed an uncertainty measure by using a combination of Shannon entropy and belief theory during the iterative granulation process space. They tackled the problem of conflict between information granules due to similar process conditions with different outcomes.

The novel aspect of this chapter is that information theory is employed as an aiding metric during the iterative data granulation process in order to create FL rule-bases. Information entropy is used as a proxy to uncertainty during the granulation process, hence constructing higher quality information granules. Comparing the outcome to the original GrC modelling framework, the improvement in terms of the predictive performance of the proposed algorithm is demonstrated.

3.2 GRANULATION OF DATA

Before delving into the specifics of the granulation technique, it is useful to review the fundamental principle, which lies on the current clustering algorithms. Clustering, regarded as the most essential question in unsupervised learning, deals with the split of data structures in unknown areas and serves as the foundation for further learning [146]. Saxena et al. [147] categorised the clustering approaches into hierarchical, partitional, density based, grid and model based.

Hierarchical clustering (HC) aims at creating a hierarchy of clusters [148]. There are two methods in HC, agglomerative and divisive. Divisive, also known as a top-down technique, generates a series of clustering schemes with increasing numbers of clusters at each step [149]. Each measure's clustering is derived from the previous one by breaking a cluster into two. In agglomerative hierarchical clustering, every object is considered a cluster at the beginning of the clustering process. Object, in this case, refers to the row of an array [150]. These objects are then will be merged based on distance to form larger clusters until the termination condition is achieved. In each iteration the distance between pair of objects are calculated, resulting in distance matrix of all pairs of objects. The merging process is normally done on the clusters that have minimum distance. One of the advantages of HC is that it does not require the users to specify the number of clusters at the initial stage [151]. This is what differ hierarchical clustering from the partition-based clustering. Moreover, no input parameters are required. However, it is subtle to outliers [147]. It is also not possible to change the cluster for an instance once it has been assigned to one cluster, in the case of misclassification [151].

In partition-based clustering, data are allocated to k-clusters by optimising some criterion function, where k represents the number of clusters. Examples of algorithm within this category are k-means and k-medoid. The main strategy for k-means is to

initialise k clusters of data and update the centre of cluster by iterative computation [147]. While k-means choose the centre of data point as the centre of cluster, k-medoid includes the data point to represent the cluster [146]. This partition-based clustering is famous due to its simplicity and scalability, but it needs the number of clusters to be pre-set [151]. In many situations it is not easy to predict the right number of clusters.

DBSCAN and OPTICS are among the examples of algorithm under density-based clustering [146]. In this approach, the data that is located in the high-density region is considered belong to the same cluster [146]. Two important parameters that need to be observed are radius and number of points in a neighbourhood [146]. In the situation where the density of data space is not even, it is difficult to get high quality clusters. Another limitation is the effect of curse of dimensionality that dissuades its ability to work with high dimensional dataset [151].

Grid-based clustering converts the original data space into a grid structure with a fixed size of clusters, and then execute clustering on the grid rather than the database directly [151]. However, it has difficulty in locating clusters embedded in a low dimensional subspace of a high dimensional dataset [151]. On the other hand, model-based clustering chooses a specific model for each cluster and finds the best fit for that model. There are two categories of model-based clustering algorithms: those that employ statistical learning techniques and those that employ neural network learning techniques [151]. Even though they are successfully implemented for vector quantisation, the parameters still need to be estimated.

3.2.1 Iterative data granulation

Despite the fact that Zadeh introduced the term 'granule' into the area of FL theory, named as information granulation in [152], the term ‘Granular Computing’ was introduced by Zadeh [153], [154] and Lin [155]. Granular Computing is a framework that mimics human cognitive in grouping objects [156]. It mirrors human’s perception and cultural tendency when they group similar objects together and try to reason with granulated data [9]. Information granules that consist of relevant knowledge representations of the data space are the outcome of information granulation.

The clustering technique presented in [6], which arranges data results in the shape of hyper boxes, serves as the foundation for information granulation in this study. The goal of that methodology is to gather data through a data organisation process in the

granules form, which are then appressed based on certain commonalities [157]. It is aimed at achieving an abstraction level by condensing the original data into granules.

The following iterative method carries out the core concept of the granulation approach suggested in [6]:

1. Searching for the two information granules that are the most compatible (with highest compatibility measure) and merging them. The new information granule now consists of both original information granules, and hence reducing the size of the data set.
2. Repeating step (1) until an acceptable level of granulation has been achieved.

The compatibility requirement between any two information granules can be established based on their similarity, distinguishability, or compactness. Figure 3.1 shows the merging process of two information granules A and B, and describes how the compatibility between these two granules is measured, hence determines which pairs of granules to be merged.

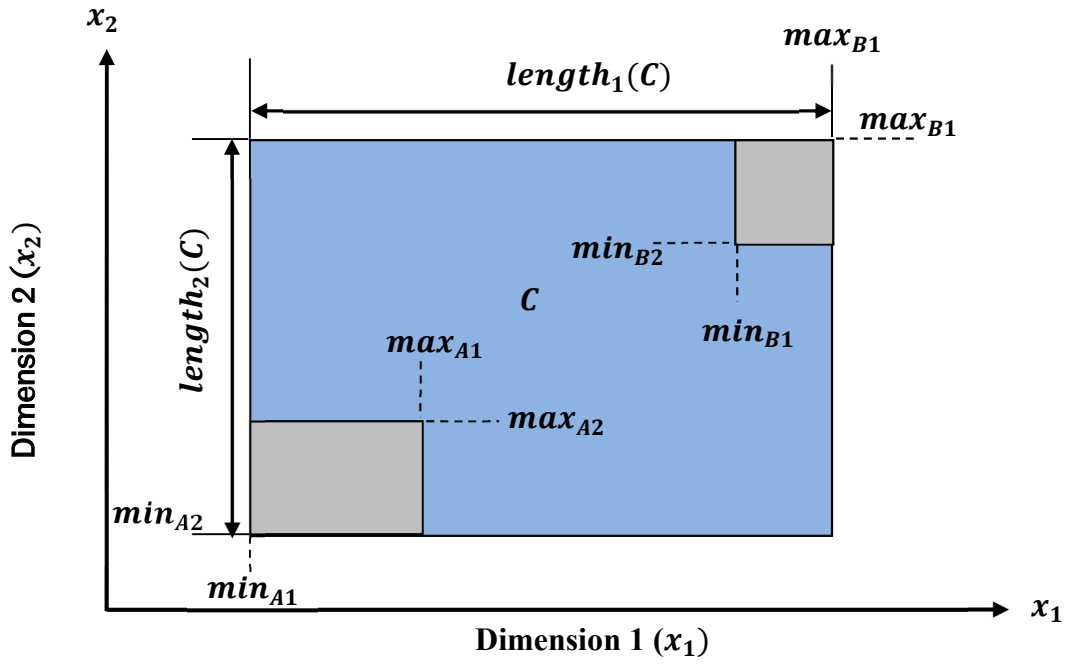


Figure 3.1 The merging of two information granules

In this research the term max_{Ai} and min_{Ai} represent the maximum and minimum value of feature i in granule A, and max_{Bi} and min_{Bi} represent the maximum and minimum value of feature i in granule B. In this 2-dimension example, the x-axis

represents feature 1 and the y-axis represents feature 2. The expression of compatibility $Compat\ (A, B)$ includes two components, which are

i) distance between A and B $(Distance_{A,B})$ and

ii) size of the newly formed granule.

The distance is defined as:

$$Distance\ _{A,B} = (\|min_B - min_A\| + \|max_B - max_A\|)/2 \quad (3.1)$$

which is an average of the two distances. $\|.\|$ denotes the L_p distance, with $p > 1$ ($p = 1$ yields Hamming distance, and $p = 2$ uses Euclidean distance).

The merging of granule A and B produces the new information granule C, whose granularity is represented by the volume $V(C)$:

$$V(C) = \prod_v^d length_v(C) \quad (3.2)$$

Where

$$length_v(C) = \max\ (max_{Bv}, max_{Av}) - \min(min_{Bv}, min_{Av}) \quad (3.3)$$

Therefore, the compatibility is defined in the form:

$$Compat\ (A, B) = 1 - Distance\ _{A,B}.e^{-\alpha\ (C)} \quad (3.4)$$

The following is the reasoning for the compatibility measure in this version. In clustering, two granules are merged when they are closest to the maximum compatibility, $Compat\ (A, B) = 1$. The compatibility measure is not only based on the nearest information granules, but also considers the compactness of the new information granule. Compactness here means that the size of the resulting granule in all dimensions should be the same; hence, the formation of the new granule resembles the shape of a hypercube. The exponential is used for the normalisation purpose, which ensures that all values remain within the unit interval. For example, in the case of a point, the compactness factor is $e^{-0} = 1$, and after several merging process, this value goes down due to the increase of volume. Obviously, the role of the compactness factor here is to ensure the formation of dense granules. The parameter α provides the balance requirement between the distance and compactness.

In another research, Panoutsos and Mahfouf [9] extended the concept of the compatibility measure by replacing the volume of granules with multidimensional density, which is defined as the ratio between cardinality and the multidimensional length. Besides, the feature weight w_v is also introduced, giving more opportunity to assign higher weight for more relevant input feature. Therefore, the compatibility measure used in this chapter is given as:

$$Compat(A,B) = Distance_{MAX} - Distance_{A,B}.\exp(-\alpha \times R) \tag{3.5}$$

in which

$$Density\ R = \frac{C_{A,B}/Cardinality_{MAX}}{L_{A,B}/\ Lengt\ _{MAX}} \tag{3.6}$$

Here, $Distance_{MAX}$ is the maximum possible distance in the dataset, given by the sum of maximum distance in every dimension:

$$Distance_{MAX} = \sum_{n=1}^{d}(distance_v) \tag{3.7}$$

and $Distance_{A,B}$ represents the distance between granule A and B, given by:

$$Distance_{A,B} = \frac{\sum_{v=1}^{d} w_v(D_1 - D_2)}{\mathrm{d}} \tag{3.8}$$

Where

$$D_1 = \max(max_{Av}, max_{Bv}) \tag{3.9}$$

$$D_2 = \min(min_{Av}, min_{Bv}) \tag{3.10}$$

With w_v is the weight for feature v, d is the number of input features; α balances the requirement between distance and density; $Cardinality_{MAX}$ represents the maximum number of objects in the data space; $Length_{MAX}$ is the maximum maximum possible length of a granule in the data set; $C_{A,B}$ is the sum of cardinality of A and B; and $L_{A,B}$is the multi-dimensional length of the new granule, given by:

$$L_{A,B} = \sum_{v=1}^{d}(max_{Xv} - min_{Xv}) \tag{3.11}$$

To exemplify, the calculation of compatibility of two granules to be merged, named as $Compat(A,B)$ is shown here. For simplicity, two-dimensional data is used. All features are treated with same importance, hence the w_v for all features is 1. Figure 3.2

illustrates how two granules *A* and *B* are merged in a normalised data space. It shows some important dimension that need to be accounted. The cardinality of granule *A*, *B, C* and *D* are 6, 7, 6 and 5, respectively. Example of compatibility metrics is shown in Table 3.1. According to Table 3.1, granules 4 and 5 have a compatibility score of 4.7, making them the two most compatible; as a result, these two granules will be combined.

Table 3.1: Compatibility metrics

	G1	**G2**	**G3**	**G4**	**G5**
G1	0	4.41	4.48	4.2	4.15
G2	0	0	4.6	4.57	4.6
G3	0	0	0	4.61	4.53
G4	0	0	0	0	4.7
G5	0	0	0	0	0

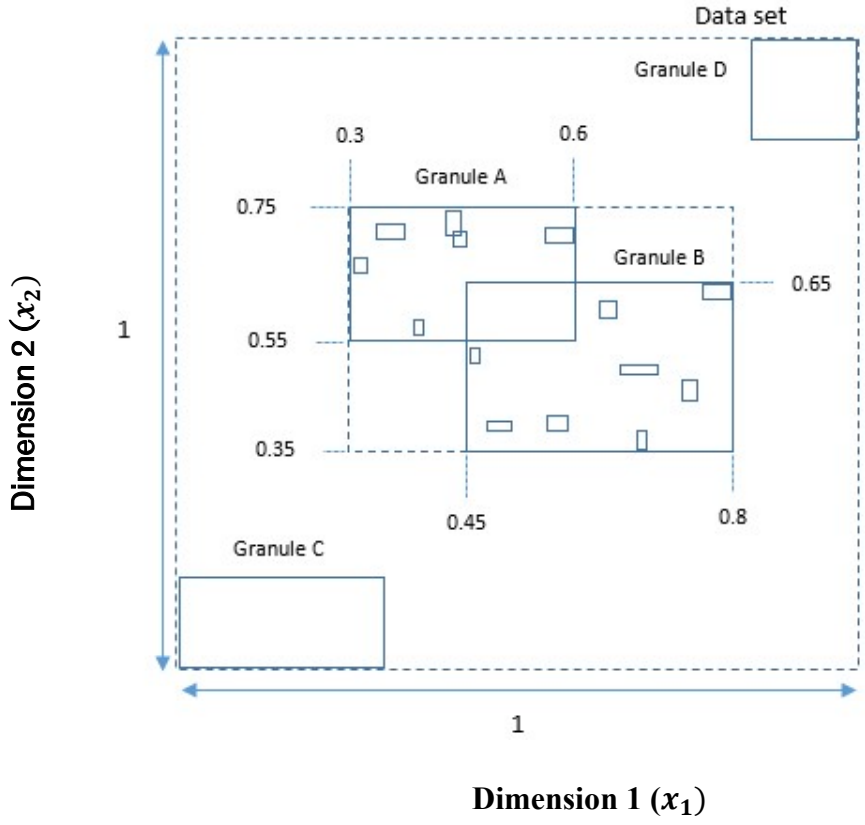


Figure 3.2: Merging granules A and B

Example of calculation of compatibility metrics (based on Figure 3.2):

$$Distance_{MAX} = \sum_{n=1}^{d} (1\ ) = 2$$

$$Distance_{A,B} = \frac{\max(0.6, 0.8) - \min(0.3,0.45)}{2} + \frac{\max(0.75, 0.65) - \min(0.55,0.35)}{2} = 0.45$$

$$C_{A,B} = Cardinality_A + Cardinality_B = 6 + 8 = 14\ granules$$

$$Cardinality_{MAX} = Cardinality_A + Cardinality_B + Cardinality_C + Cardinality_D$$

$$= 6 + 8 + 6 + 5 = 25\ granules$$

$$L_{A,B} = (0.8 - 0.3) + (0.75 - 0.35) = 0.9$$

Given that $\alpha = 0.3$,

$$Compactness\ factor\ R = \frac{14/25}{0.9/2} = 1.244$$

$$C_{A,B} = 2 - 0.45 \times \exp(-0.3 \times 1.244) = 1.69$$

A compatibility metric is constructed after one iteration. Each possible merging's compatibility index is included in this metric. Accordingly, the pair that has the largest compatibility index will be merged. The compatibility index also functions as a monitoring tool in the entire granulation process that can be used as the stopping criterion (hence, determines the number of final granules).

Figure 3.3 shows how the compatibility measure declines throughout the iterative granulation process, based on a synthetic dataset with 400 instances and 2 dimensions. The decrement indicates that the granules merged at the end of granulation process is not as compatible as the pairs at the beginning of the granulation and provides stopping threshold for the process. The maximum compatibility here is 2, which is equal to the number of dimension (normalised).

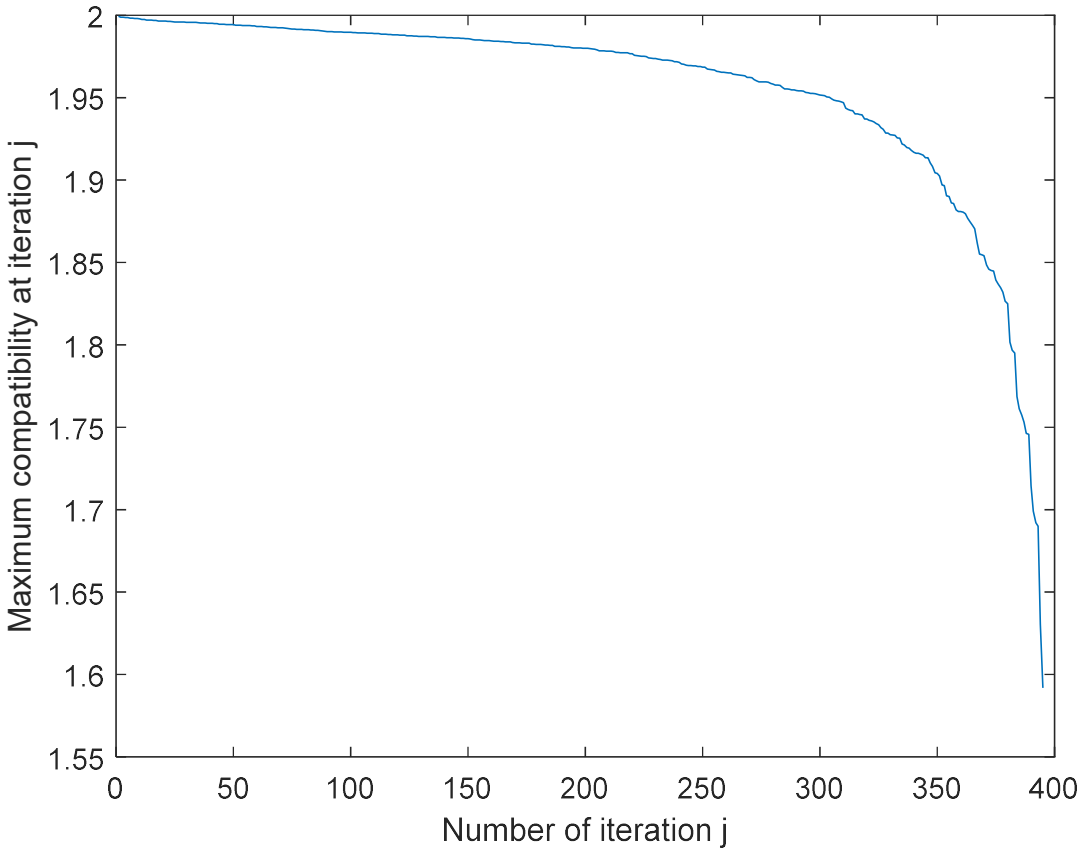
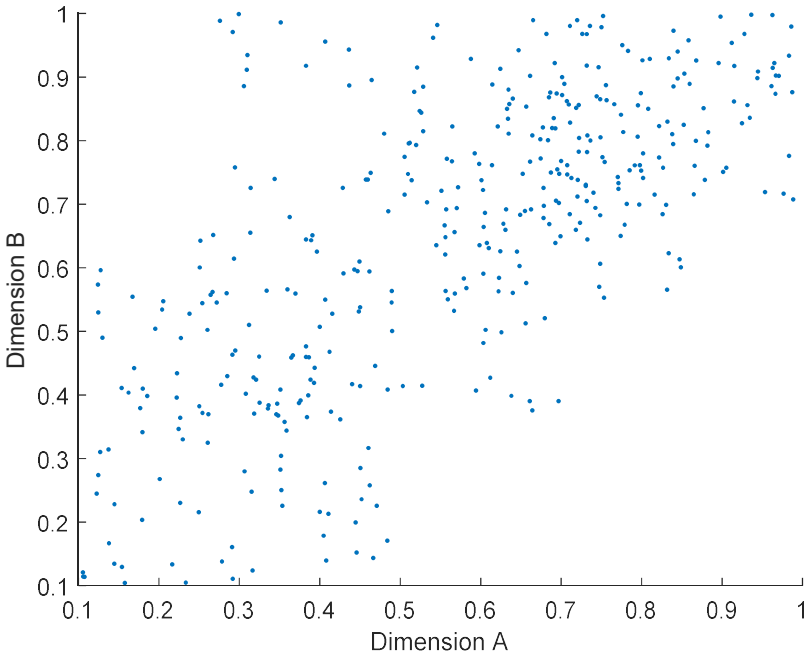


Figure 3.3: Maximum compatibility index

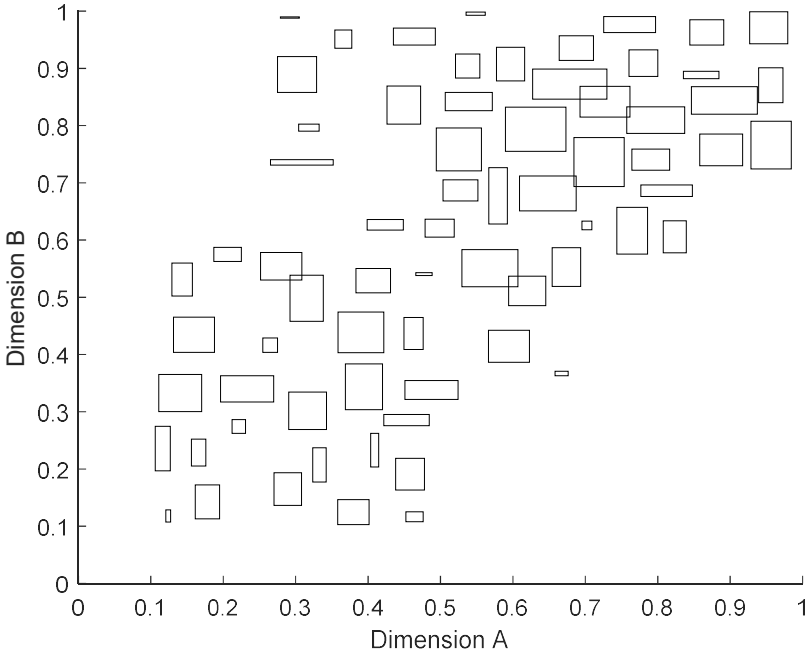
Every object or instance is treated as a granule in the initial iteration, therefore the maximum and minimum limits are identical. The result is the development of hyper box-shaped information granules. Data points are always be included around the edges of the hyper boxes since the compatibility index takes the maximum and minimum points into account when merging two granules (hence promotes the transparency between the original data and the formed information granules).

Figure 3.4 shows a granulation process of a synthetic dataset. The 400 instances are compressed to 70 granules, and finally to five final granules. Since all four inputs are clustered along with the output in order to observe the connection between them, this method can be regarded as supervised learning.

400 data points



70 granules



5 granules

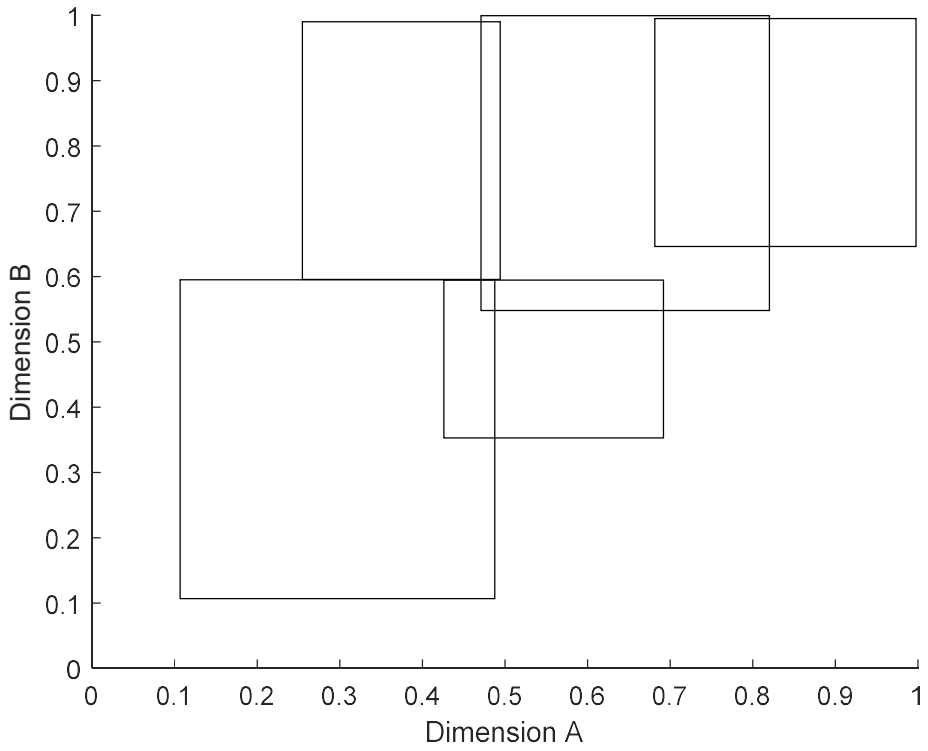


Figure 3.4: Data granulation process

3.3 USING INFORMATION GRANULES TO BUILD A FUZZY LOGIC MODEL

This section describes how the initial Fuzzy Logic rule base is constructed by using the information granules. The example shown here is based on Iris data, which consists of 4 input variables (sepal length, sepal width, petal length and petal width), 1 output variable and 150 instances. The output variable is Iris class: Setosa, Versicolor and Virginica. Five granules are used in this experiment and being translated to FL rule based in the following form:

$$R_i: IF\ Variable_j\ is\ A_j and\ Variable_j\ is\ B_j\ THEN\ \ldots$$

$$\ldots IMPLICATION \tag{3.12}$$

Figure 3.5 shows how five sets of membership functions (MFs) are derived from five information granules. All four dimensions are shown in Figure 3.5 (a) and (b). The centre of Gaussian MFs (c_{ij}) and sigma (σ_{ij}) are represented by the median and the standard deviation of the data in a granule, respectively. An information granule and

an FL linguistic rule are related to one another on a one-to-one basis, means that one granule represents one FL rule. The standard deviation S is defined as:

$$S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}|A_i - \mu|^2} \tag{3.13}$$

where N is the number of data in a granule, A is the random variable vector and μ is the mean of A, defined as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} A_i \tag{3.14}$$

Equation 3.15 shows the created linguistic FL rules. This is example on how one can interpret the rules by examining the relationship between inputs and output. In more complex application, for example manufacturing, a non-expert can easily understand how changes in variables or inputs affect the output. This exercise proves the interpretability of the elicited FL rule base. Moreover, it also proves the capability of GrC in extracting knowledge in a transparent manner. The Granular computing based FL modelling framework is shown in Figure 3.6.

Example of FL rule base is as follow:

Rule 1: IF Sepal length is LOW

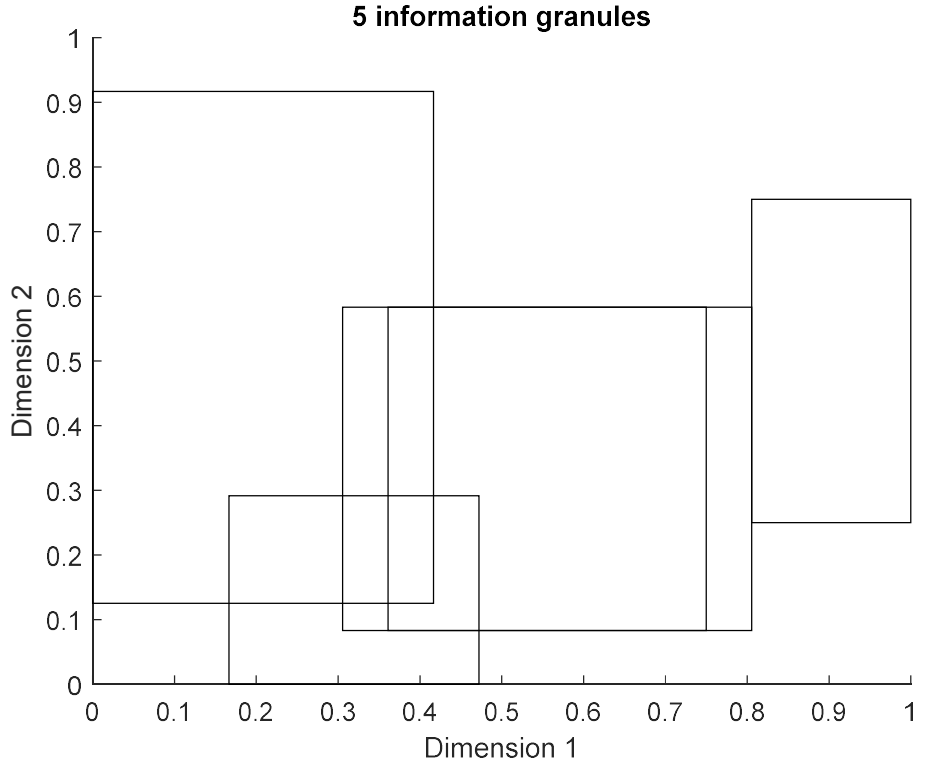
. . . and Sepal width is HIGH

. . . and Petal length is LOW

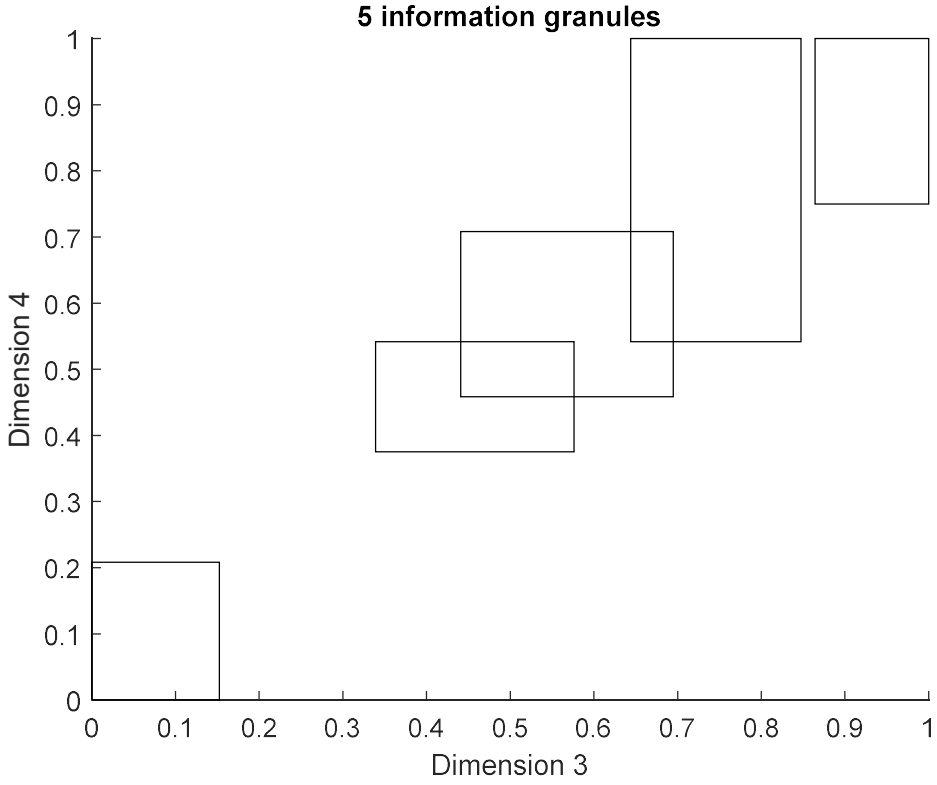
. . . and Petal width is LOW

THEN the species is SETOSA

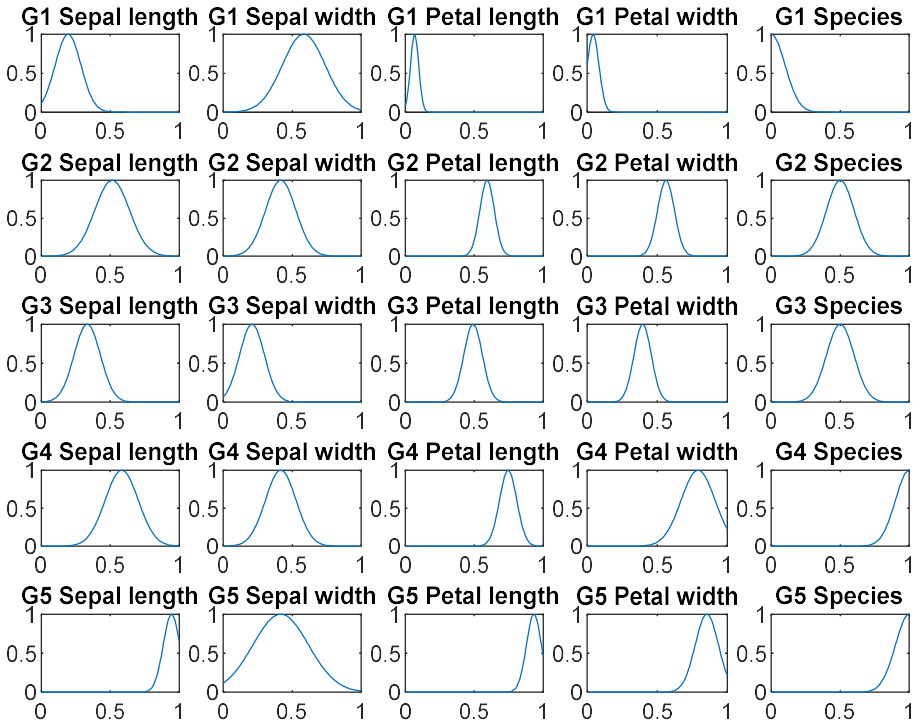
. . . (3.15)


5 information granules
Dimension 2
Dimension 1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

(a)


5 information granules
Dimension 4
Dimension 3
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

(b)



(c)

Figure 3.5: (a) Five information granules observed from Dimension 1 and 2 (Sepal length and Sepal width) (b) Five information granules observed from Dimension 3 and 4 (Petal length and Petal width) (c) Tranformation from information granules to Fuzzy Logic membership functions

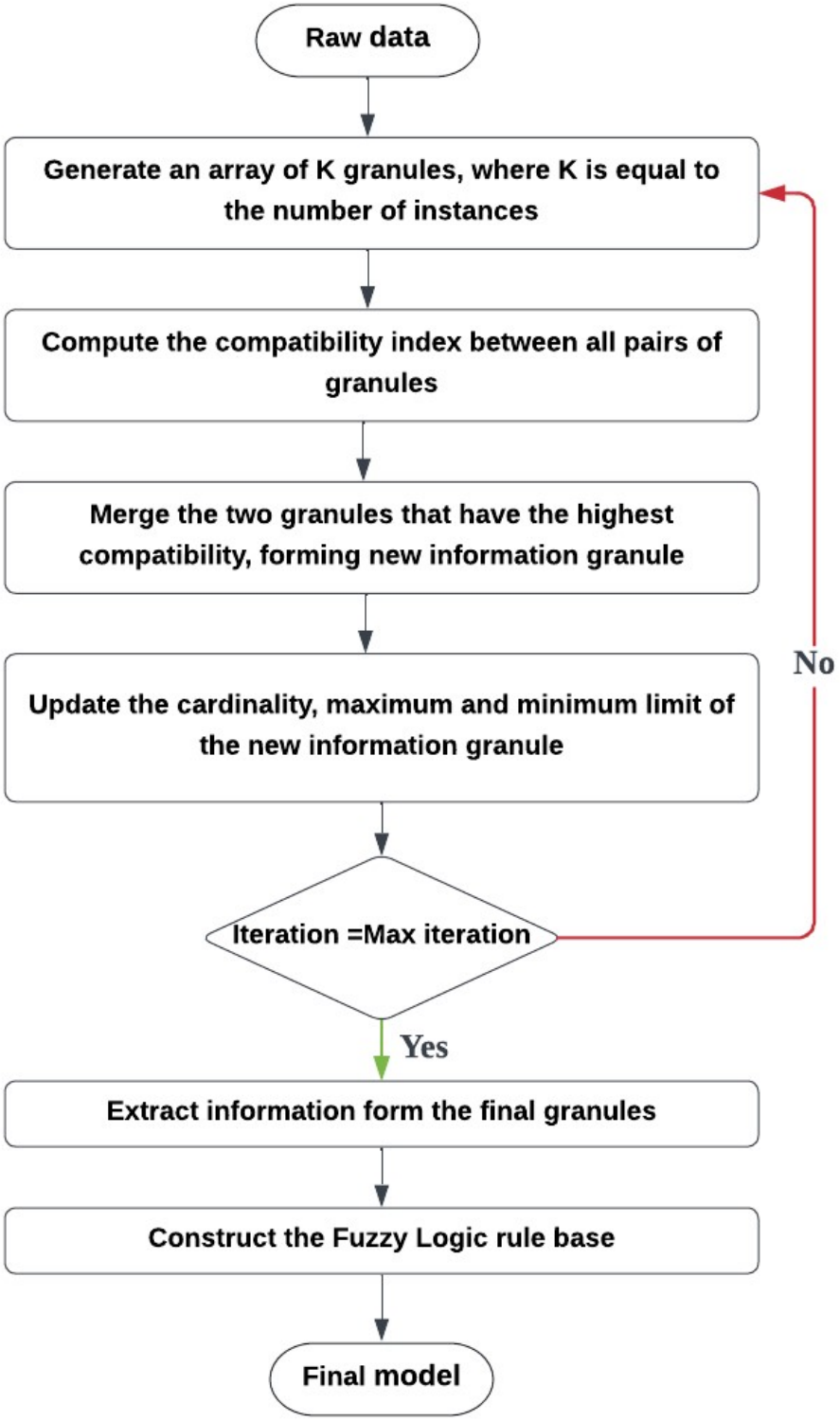


Figure 3.6: Granular computing based Fuzzy Logic modelling framework

3.4 UNCERTAINTY MEASURE IN ITERATIVE DATA GRANULATION

The iterative data granulation presented before is used as a mechanism to execute data mining and to extract meaningful information. It is the very critical step in building data driven soft-computing models, since the initial structure of the models is determined by the final information granules formed at the end of the granulation process. Even though information is the subject of interest in GrC, it is closely related with the concept of uncertainty [158]. In the iterative data granulation, uncertainty could happen due to various reasons, for example, when one of the information granules is an outlier, or when pairs having similar compatibility index during the merging process. For example, there is a considerable likelihood of the fusion of two different granules when there are outliers present. This is so because the geometrical distance is typically utilised as a measure of similarity in most clustering algorithms [159].

Low-quality information granules will make the granular framework harder to recognise and make the FL rule base more difficult to interpret, which will result in an incorrect representation of the system under study. As a consequence, the FL rules that are being created will be inconsistent and conflicting [160]. Therefore, it is critical to tackle the problem of uncertainty during iterative data granulation, as it may affect the quality of the knowledge gained in the form of information granules, and results in low quality FL rule base.

Klir and Yuan [158] identified three types of uncertainty established in five different theories - classical set theory, fuzzy set theory, probability theory, possibility theory, and evidence theory. The first uncertainty type, fuzziness is defined as the lack of sharpness of relevant distinctions. Fuzziness, also known as fuzzy entropy, are used to assess the degree of imperfection (or disorder) in a given case. In order to establish global measurements of the indefiniteness described by Fuzzy Logic sets, such measures characterise the sharpness of the membership functions [161]. The second one is discord, also known as randomness, conflict, confusion or strife [162], [163], which refers to disagreement in selecting between numerous alternatives. Non-specificity, on the other hand occurs when two or more alternatives are left unspecified. It is related to cases where the information is concentrated on sets with cardinality larger than one [162]. Figure 3.7 shows the three basic types of uncertainty.

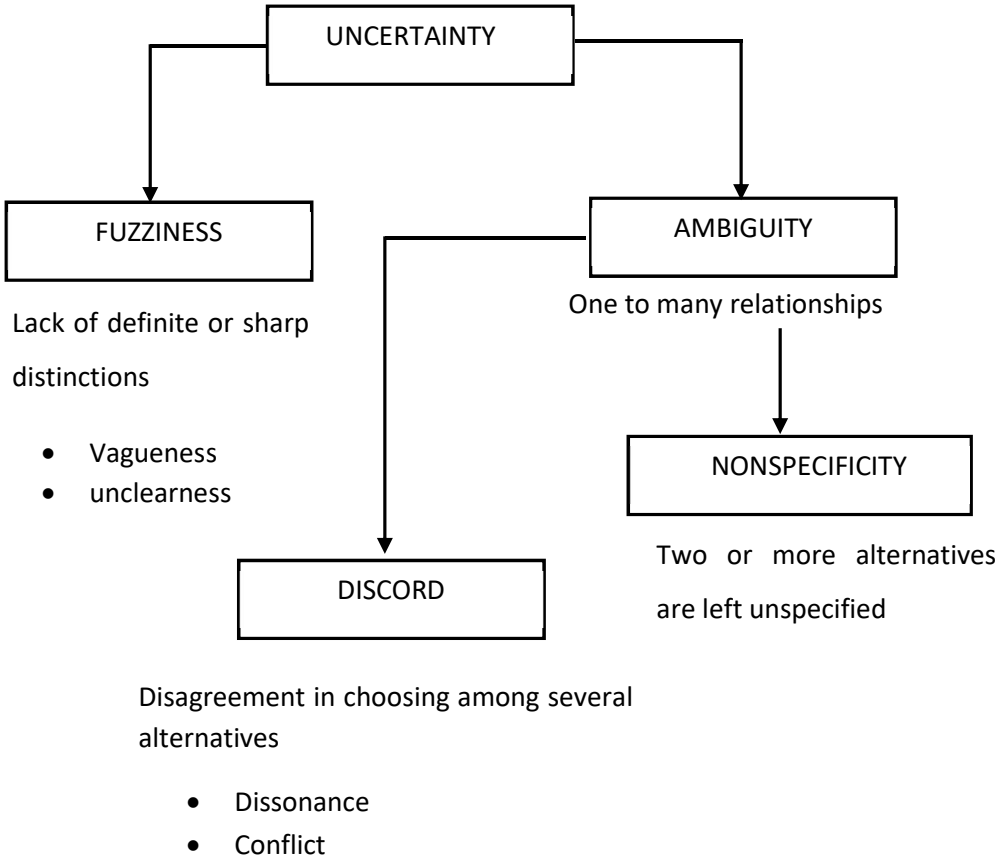


Figure 3.7: Three basic types of uncertainty (reproduced from [34])

Probability theory has always been considered as a fundamental tool to handle uncertainty in machine learning, including clustering [164]. Heavily using probability theory, is information theory (IT) as one of the disciplines that concerns quantifying uncertainty [165]. One of the most important concepts in IT is entropy.

3.4.1 Information theory as a mathematical tool to measure uncertainty

Information theory (IT) is one of the theories that concerns quantifying uncertainty. Shannon entropy is expressed as:

$$H(X) = -\sum_{x\epsilon X} p(x) log_2 p(x) \tag{3.16}$$

where $p(x)$ is the probabilities of occurrence of an event x and X is a random variable. In IT, the dimension or feature is referred to as a random variable.

Conditional entropy, on the other hand, describes amount of uncertainty of one random variable, when given information about another random variable:

$$H(X|Y) = -\sum_{x\in X, y\in Y} p(x,y) log_2 p(x|y) \quad (3.17)$$

where $p(x \mid y)$ is defined as the probability of x given the condition of y.

Another important element in IT is mutual information, or information gain. It quantifies the reduction in uncertainty, when given knowledge of another random variable:

$$I(X:Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = H(X) + H(Y) - H(X,Y) \quad (3.18)$$

Information theory has been widely applied in many areas such as clustering, uncertainty measure, fault diagnosis and condition monitoring. For example, Yin et al. [166] applied Shannon entropy to quantity the uncertainty of the weighted networks and proposed three basic heuristics for game model namely kin selection, resource selection and reciprocity selection in the networks. Jiang et al. [167] proposed a model based on Shannon entropy and probabilistic neural network to quantify the characteristics of vibration signal of rotating machinery. In another research, Liansheng Liu et al. [168] proposed a framework for condition monitoring of an aeroplane engine using the concept of entropy, in which the sensor selection approach applied Shannon entropy in measuring the uncertainty of information in a data series.

Information theory also has been applied in several types of clustering techniques to tackle the issue of uncertainty. Bobek et al. [169] modified the information gain split criterion in Decision Tree to consider the uncertainty of data and proposed to redefine the probability in the entropy equation to search for optimal split criterion. Gullo et al. [170] proposed U-AHC, a new approach for agglomerative hierarchical clustering by introducing new linkage criterion for cluster merging. This research focused on data consists of uncertainty in the attribute level and uses pdfs to model the uncertainty. Before that, Aghagolzadeh et al. [171] also applied IT in hierarchical clustering. They used mutual information as the measure to discover the boundaries between clusters. Using K-medoid algorithm as the platform, Jiang et al. [159] used Kullback-Leibler divergence (KL-divergence), or relative entropy, to capture the difference between two distributions, where uncertain objects were treated as random variables with certain distributions. Chau et al. [172] introduced the popular UK-means clustering algorithm, that incorporates the pdfs of uncertain data in the k-means clustering.

3.4.2 A new framework for iterative data granulation to reduce the uncertainty

As shown before, the selection of pairs of granules to be merged are done using the compatibility measure. Nevertheless, there is a risk of uncertainty occurring throughout the granulation process [129]. The newly created granule, for instance, cannot accurately represent the actual data space in the presence of an outlier. Additionally, there is a tendency to have excessive overlapping. This will generate highly unpredictable information granules and, as a result, producing indistinguishable FL rule-bases.

In this chapter, iterative data granulation is improved to address this problem while taking the uncertainty measure into account. Shannon entropy, which is expressed in equation (3.11), describes the amount of uncertainty in a random variable. Entropy is always used to describe the disorder or unpredictable nature of a dataset. [173]. Therefore, entropy in the proposed GrC compatibility metric, thereby characterises the granules' resistance to merging. Therefore, the goal of the suggested approach is to arrange the instances in a way that minimises uncertainty.

Entropy is a term used to describe the degree of uncertainty in a random variable. However, this does not apply to a random vector. Let X_i be the random variable, and $X = [X_1, X_2, X_3, \dots X_d]$ represents the random vector. Based on the chain rule described in [174], the entropy of a random vector $H(X)$ is expressed as:

$$H(X) = H(X_1, X_2, \dots X_d) = \sum_{v=1}^{d} H(X_v | X_{v-1}, \dots X_1) \tag{3.19}$$

Let say we have two random variables X_1 and X_2, therefore the entropy of the random vector $X = [X_1, X_2]$ is given by:

$$H(X) = H(X_1, X_2) = H(X_1) + H(X_2 | X_1) \tag{3.20}$$

Geometrical distance used in equation (3.5) is vulnerable to outliers [175]. The entropy offers extra information in order to distinguish between uncertain entities with various distributions. The scenario in Figure 3.8 demonstrates the entropy of Granules 1 and 2 in the case of a merging of objects (b1, c2). In this scenario, a two-dimensional data set is employed. It is possible that the instance (b1, c2) will merge with Granule 1 or Granule 2. Despite the fact it shares components (b1 and c2) with both Granules, it is more likely to combine with Granule 2 since its entropy is less than that of Granule 1. The granular computing – entropy based FL modelling framework shown in Figure 3.9.

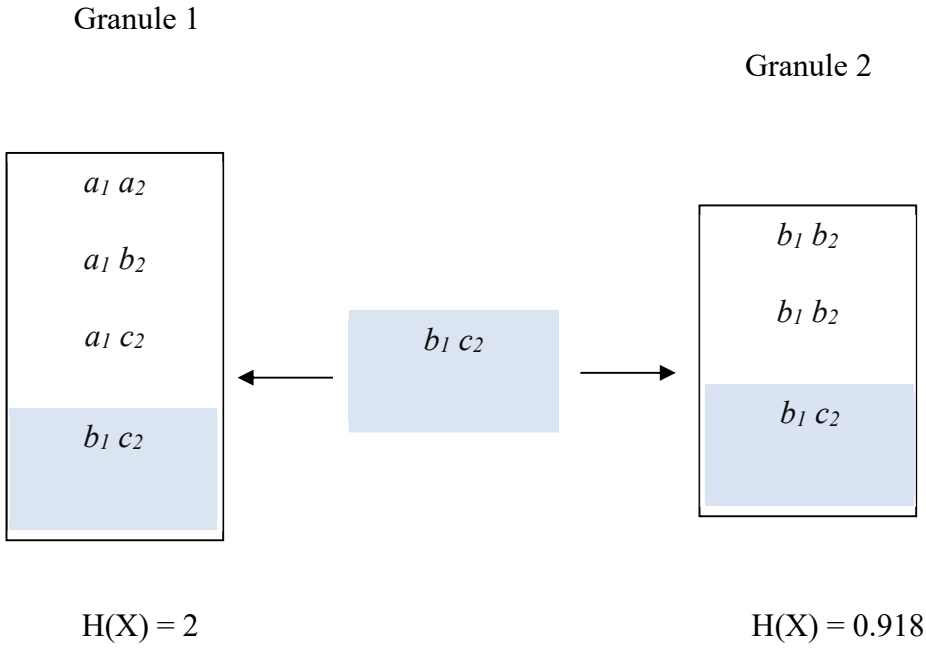


Figure 3.8: Entropy in granules

The entropy H, and the weight w are therefore included in the equation for the proposed compatibility measure:

$$C(A,B) = Distance_{MAX} - (wH + Distance_{A,B}.\exp(-\alpha \times R)) \qquad (3.21)$$

The GrC with entropy measure is then tested with Iris data. The performance between GrC with entropy is compared with GrC without entropy measure. The maximum compatibility of GrC with the entropy measure varies in each iteration, as shown in Figure 3.10. Due to the more competitive merging (since entropy prevents the uncertain objects to be merged), the compatibility measure with uncertainty is lower than the original compatibility. In order to generate granule with the least amount of uncertainty, the iterative granulation process starts to be careful in selecting the best pair to be merged. Figure 3.11 shows the amount of maximum entropy involved in each iteration. It illustrates the low uncertainty at the beginning of granulation, but the amount of uncertainty towards the end is increasing significantly.

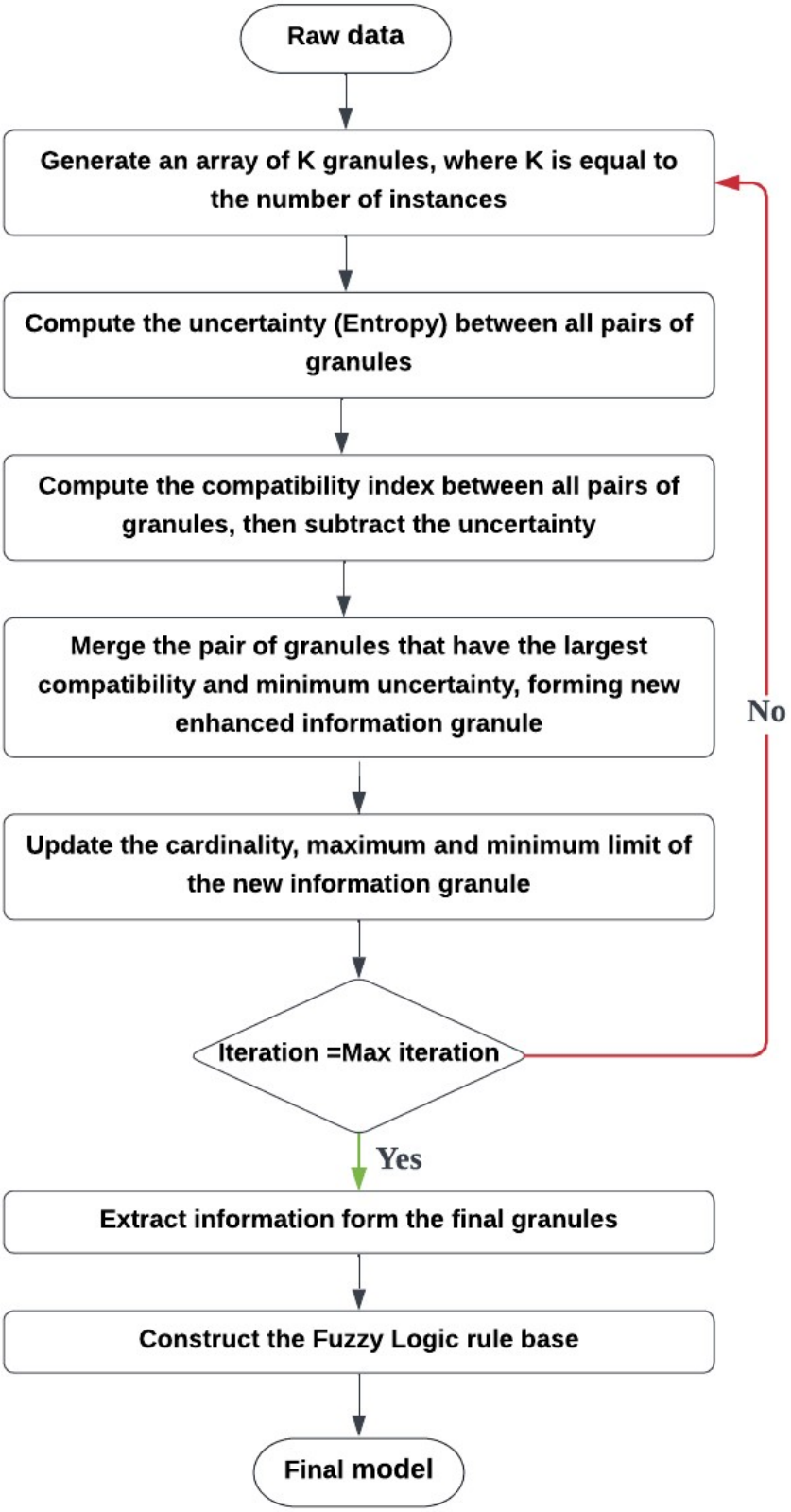


Figure 3.9: Granular computing – entropy based Fuzzy Logic modelling framework

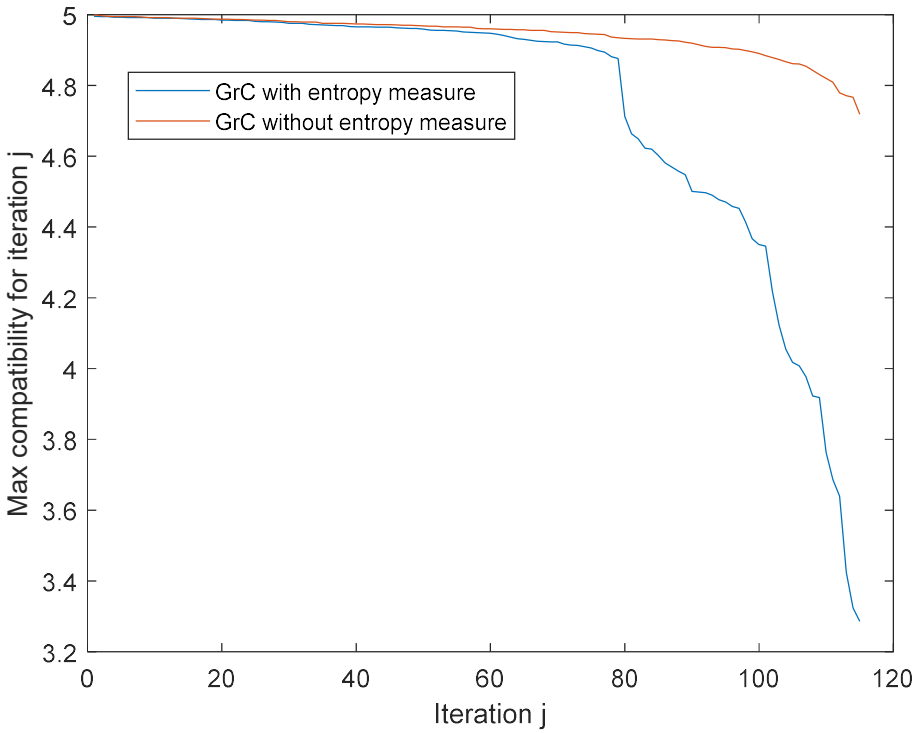


Figure 3.10: Comparison between compatibility index of GrC and GrC with uncertainty measure

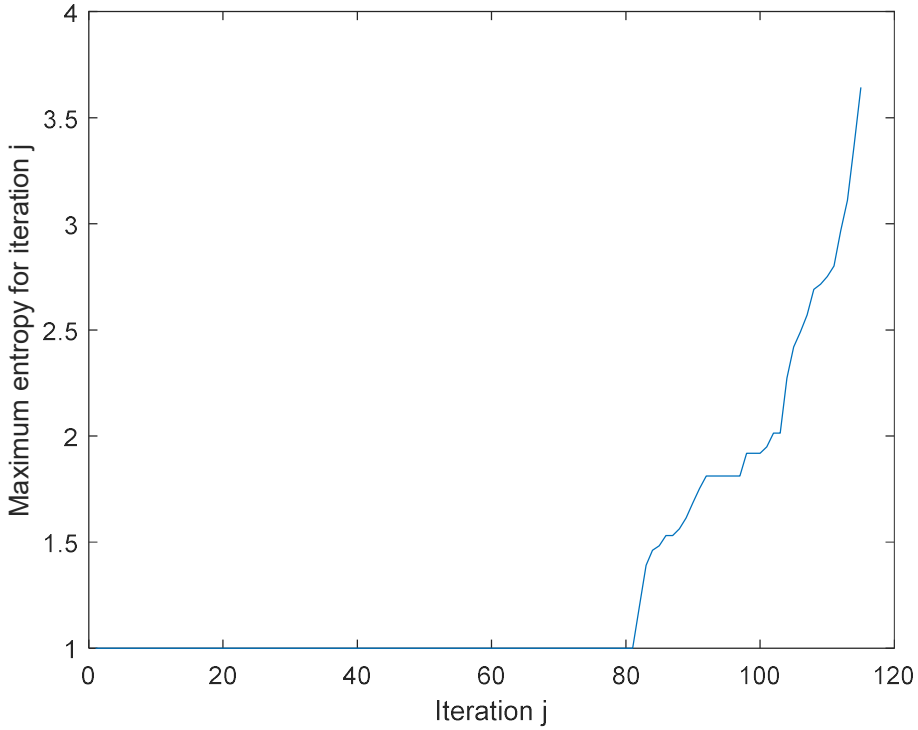


Figure 3.11: Maximum entropy

As shown in Figure 3.11, the entropy increases dramatically after certain number of iterations. This is due to the high 'disorder' of the data in the newly formed granules, indicating the urge to stop the granulation process. The entropy at the beginning of granulation is low due to the merging of similar objects (low disorder), but became larger towards the end of granulation due to higher disorder in the data that are projected to be merged.

3.5 CASE STUDY AND SIMULATION RESULTS

In this section, the comparative study between conventional GrC-FLS and the enhanced GrC-FLS with entropy measure are presented. The dataset used for the presentation here is Iris dataset, a well-established dataset obtained from UCI Machine Learning Repository. It consists of 150 instances, with 4 input variables (sepal length, sepal width, petal length and petal width) and 1 output (Iris class) – Setosa, Versicolour and Virginica. Next, the experiment is extended to another two datasets, which are Wine and Glass. Table 3.2 describes the characteristics of these three datasets.

Table 3.2: Characteristics of data sets: Iris, Wine, and Glass.

Dataset	Number of instances	Number of attributes	Number of classes
Iris	150	4	3
Wine	178	13	3
Glass	214	9	6

The proportion used to partition the datasets into training and testing data is 80:20. To ensure that all variables have the same domain, all variables are normalised to [0,1]. Each method is tested in ten trials, and its predictive performance is assessed by calculating the classification accuracy and the root mean square error (RMSE). The accuracy is measured using the proportion of correctly classified observations over all predictions, whereas the $RMSE$ stands for the model error, expressed as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {e_i}^2} \tag{3.22}$$

where e_i represents the ith observation error and n is the number of observations.

In GrC with entropy, probability mass functions are employed to express the probability of a random variable. A method called Histograms is used, which are used widely as non-parametric density estimators to visualise data and to gain summary quantities, such as the entropy, of the underlying density [176]. The number of bins selected for all input variables is 5, as suggested by [176]. Iris, wine, and glass all have weights of 0.2, 0.1, and 0.3, respectively. Due to the disproportion between the six classes in the glass dataset, the data has to be pre-processed using a bootstrapping technique.

Table 3.3 shows that, on average, the GrC with entropy measure performs better than the traditional GrC, and other clustering algorithms, k-means, and Fuzzy c-means. RMSE represents the root mean square error, SD stands for standard deviation and acc. stands for accuracy. In the case of Iris data, GrC with entropy scores the highest accuracy of 96.33%, with the closest is Fuzzy c-means with 96%. It also achieves the lowest RMSE of 0.0957. The same pattern occurred in Wine dataset, where GrC with entropy obtains the highest accuracy of 95.33%, while Fuzzy c-means scored 95%. However, Fuzzy c-means performs better in terms of RMSE, with 0.088 as compared to 0.0964 in GrC with entropy.

For Glass dataset, the best accuracy is also achieved by the GrC with entropy with 72.09%, followed by the conventional GrC with 69.77%. K-means and Fuzzy c-means score 67.44% and 65.17%, respectively. RMSE for GrC with entropy is also acceptable, at 0.2065. All the results are benchmarked with other related works, for example 97.33% (Iris) and 95% (Wine) in [177], [178] and 75.45% (Glass) in [178].

The GrC with entropy is demonstrated to perform better than the standard GrC and other clustering algorithms. This is due to the fact that penalising the uncertainty in the merging process avoid having uncertain information granules (hence promotes higher quality information granules). As the result, FL rule-bases formed are more distinguishable and provide better predictions as demonstrated in the experiments.

Table 3.3: Predictive performance of GrC, GrC with entropy, k-means and Fuzzy c-means

Dataset	**Iris**		**Wine**		**Glass**	
	RMSE	**Accuracy (%)**	**RMSE**	**Accuracy (%)**	**RMSE**	**Accuracy (%)**
GrC	0.1354	91	0.1233	91.33	0.1968	69.77
GrC with entropy measure	0.0957	96.33	0.0964	95.33	0.2065	72.09
K-means	0.1032	94.33	0.1055	92.67	0.2293	67.44
Fuzzy c-means	0.1257	96	0.088	95	0.2173	65.17

3.6 SUMMARY

This chapter presents a methodology for data capture and modelling that is based on information theory and granular computing. To quantify the uncertainty throughout data granulation, the element of information theory is used. Entropy is demonstrated to be a promising tool to detect uncertainties when there are anomalies present. In order to generate high-quality information granules, the entropy expresses the reluctance of two granules that are about to be combined. It may therefore direct the granulation process to combine the granules with the least amount of uncertainty.

The major goal of this approach is to prevent the formation of highly disordered granules. This keeps the granules from having outliers that would influence their capacity to be distinguished from one another, hence improving the interpretability of FL rule-bases. The framework is evaluated using three datasets from the UCI Machine Learning Repository within the classification problems: Iris, Wine, and Glass, where substantial improvements in prediction accuracy and error reduction are evident.

The presented approach yields the minimum RMSE for Iris and the maximum accuracy across all datasets. This is because low uncertainty granules are produced as a result of penalising the uncertainty during the iterative data granulation. The results

are also compared with other relevant works in the literature, and the benchmark confirms its competitiveness.

Main results and methodology from this work is included in paper titled “An Entropy-Based Uncertainty Measure for Developing Granular Models”. It was presented and published in 2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI) in Stockholm, Sweden. In addition, the outcomes of this chapter were presented in 2021 ACSE PGR Research Symposium in The University of Sheffield.

In this chapter, all input variables are treated as equally important, i.e. $w_v = 1$ and the modelling structure is based on the Type-1 Fuzzy Logic system. In other words, the feature weight of each input variable is neglected throughout the granulation process, even though in reality they have different importance, or weightage. Therefore, in the next chapter, a new iterative data granulation is proposed, where the input variable weight w_v be assigned based on their importance. Moreover, unlike other feature weighting works that predetermine the feature weight (having constant weight throughout the granulation process), the feature weighting in the next chapter allows the weight to be evolved according to the current formation of information granules. This is achieved by embedding the feature-weighting framework in the iterative data granulation.

CHAPTER 4

AN EVOLVING FEATURE WEIGHTING FRAMEWORK FOR GRANULAR MODELS

This chapter presents a new GrC algorithm called weighted GrC (W-GrC) that assigns and updates the feature weight based on the feature importance, iteratively and not a priori. Instead of predetermining the feature weight at the beginning of the granulation process, W-GrC updates the feature weights in each iteration according to the information obtained from current information granules, more specifically the within-granule variances. The embedded feature-weighting algorithm allows the feature weights to evolve throughout the granulation process. Rather than predetermining the feature weights in advance, W-GrC is designed such that it utilises the new information obtained from information granules in each iteration. It is hypothesised here that features may have different importance (as far as grouping goes) throughout the iterative granulation process. The new and enhanced information granules are used to build Type-1 Fuzzy Logic systems (T1-FLS) rule-bases and the new framework is validated using popular datasets: Iris, Wine, and Glass from UCI machine learning repository. Nauck's index is used to measure any impact on the interpretability of FLS in both W-GrC and conventional GrC cases; no significant deterioration of interpretability is observed. Results demonstrate the potential of W-GrC to increase the system's accuracy while maintaining good interpretability.

4.1 INTRODUCTION

The framework of iterative data granulation can be combined with Fuzzy Logic information granulation theory, by means of clustering algorithms [179]. Data is grouped or clustered using such computational frameworks to make it simpler to

comprehend and less complex. The requirement for information simplification, summarisation, and information clarity serve as the driving forces behind this effort [180].

Iterative data granulation aims to intuitively (human-like behaviour) integrate information, for example, the two most compatible granules, until the information has been sufficiently compacted. For the possible newly produced granules, compatibility in this thesis is assessed using distance and density measurements.

The majority of studies on iterative data granulation assign identical weight to each input attribute. However, this approach is not ideal, especially when dealing with data with a high number of attributes, or features [181]; such features may have different weight (importance) as the granulation progresses. In an attempt to organise data collectively, one would naturally exclude features that are less important to the task. It is hypothesised here that features may have different importance (as far as grouping goes) throughout the iterative granulation process. Hence a framework that assigns no weights (importance) to the features, or assigns constant weights a prior may not perform well.

Feature weighting in clustering algorithms is not a new concept. In clustering, Huang et al. [181] proposed the weighted k-means (W-k-means) that shows better performance than the conventional k-means. The feature weighting concept is also applied in hierarchical clustering [182], where Amorim introduced improved version of Ward, called Ward_p that allows a feature to have different weight throughout the clusters.

Although the idea of feature weighting in iterative data granulation has been discussed elsewhere [9], the majority of studies on this technique, including [130] and [129] employ constant weight for each feature. There are not many studies on feature weighting for GrC; one such study is a Fast Correlation-Based Filter in [12] that used symmetrical uncertainty to identify the most important features in welding process. However, in this pre-processing stage (which serves as a filter mechanism), the feature weights are predetermined and their values remain unchanged over the course of the granulation process.

A new GrC algorithm is presented in this chapter that allocates and updates the feature weights according to the significance of the input variables. Using this method,

for a specific iteration, the more significant features are allowed to have a larger influence on the data granulation compared to the less significant ones. Additionally, the feature weighting in this work is included or embedded into the data granulation process itself rather than assigning the weight during the pre-processing phase. This makes it possible to adjust the feature weights in accordance with the information granules that have been created.

Hence, the main objective of this chapter is to study and analyse the use of W-GrC to Type-1 Fuzzy Logic systems (T1-FLS) as well as to examine for the first time how it may affect the interpretability of the rule-bases. UCI datasets such as Iris, Wine, and Glass are used to evaluate the new mechanism. Results show that the new computational framework can improve the classifier's predictive performance results by systematically choosing the feature weight parameter. Additionally, the impact on the interpretability of the developed models is examined in this study (using Nauck's index), and it is demonstrated that W-GrC maintains a good compromise between interpretability and accuracy. The problem here is that a dataset with many variables could become too challenging for a particular algorithm to handle. The hypothesis space is smaller and, hence, easier for a particular algorithm to narrow down to the optimal hypothesis, in the case of dataset with fewer features [183].

4.2 FEATURE SELECTION AND WEIGHTING

One of the most popular techniques for reducing noisy (i.e. redundant) features is dimensionality reduction. There are two most common dimensionality reduction strategies, which are feature extraction and feature selection. Both approaches are capable of increasing learning performance by reducing computing complexity, improving generalisability, and reducing storage requirements [184]. The curse of dimensionality is a term commonly associated with the problems in analysing high-dimensional data [183].

By integrating features, feature extraction methods seek to minimise the dimensionality of data sets, where the new built features are the blend of the original ones [185], as being applied in various approaches such as Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Canonical Correlation Analysis (CCA) [186]. These new features are related to the original features, but not the same; they are converted versions [183] and the task to alter and combine the

original features is rather complicated. Therefore, the transformed features obtained by feature extraction techniques have no physical meaning, which poses a problem for further analysis of new features [184].

In contrast, the feature selection method does not change the feature itself and retains its original meaning to the user [185], and hence provides better readability and interpretability as compared to feature extraction [186]. Feature selection is implemented by an empirical search algorithm that attempts to attain the most suitable feature subset [184]. Next, the difference between feature selection and another approach known as feature weighting is described; a method that is perceived of as a generalisation of feature selection.

4.2.1 FEATURE SELECTION

Filter and wrapper methods are the two main categories of feature selection methods [187]. The filter approach isolates feature selection from classifier learning, ensuring that there is no interaction between the two algorithms' bias. They choose the characteristics that are most significant in terms of a well-known metric from the whole features set derived from the training samples. These attributes, along with the training samples, are then fed into the learning process to create the learning model [184]. The most representative algorithms of the filter model are Relief, Fisher score, and Information Gain based approaches [186]. Filters are scalable; therefore, they are best for problems with many variables. Their main drawback is that they choose important features without considering their impact on the learning algorithm in use [184].

Wrapper approach is a prominent feature selection technique for obtaining the optimal feature subset by optimising a measure corresponding to the feature subset. However, the measure to be optimised in filter-based feature selection approaches is not directly linked to the classifier learning algorithm. Wrappers may get better results, but they take substantially longer to compute than filters [188]. To determine the quality of selected features, the wrapper model uses the predictive accuracy of a predetermined learning algorithm [186]. In the case of clustering, a wrapper method would embed the feature selection algorithm within the clustering algorithm, unlike wrapper that evaluates each feature subset relevancy until the most relevant subset is obtained [183].

4.2.2 FEATURE WEIGHTING

Feature weighting is regarded as a generalisation of feature selection [186], where each feature is multiplied by a weight value according to its ability to identify pattern classes [189]. Unlike feature selection, that assigns a binary weight of 0 or 1 (1 indicates that the feature is selected for the classification, vice versa), feature weighting assigns a value in the interval [0, 1] to determine the importance of a particular feature. Larger value of feature weight w_v indicates that the feature is more relevant and influential in identifying the pattern classes. Filter feature weighting techniques for example, compute feature weights with reference to the label classes (in supervised case) or intrinsic qualities of the data (unsupervised) depending on the learning methodology used [190]. Wrapper approaches, on the other hand, use feedback from a particular machine learning algorithm to approximate feature weights in a black-box iterative style [190].

When calculating the output, it has always been considered that all of the features are equally essential. However, if some features have a greater scale than others, the results can be skewed, compromising the entire algorithm's performance and accuracy [190]. There is no reason why all the features should be treated as equally important. Therefore, in feature weighting, features are no longer selected, but weighted.

Recent literature indicates that a variety of feature weighting methods have been presented with the goal of determining the degree of importance of each feature in deriving the output pattern [190]. Xiaoli et al. [191] proposed LASSO-Based Feature weighting selection method for Microarray Data, with application in biomedical data. They calculated the data's feature-feature inter-correlation and feature-label correlation using Pearson's linear correlation. Zhou et al. [192] introduced a new feature selection approach known as Feature Selection Based on Weighted Conditional Mutual Information (WCFR) that employed standard deviation to adapt the significance between relevancy and redundancy. They demonstrated the effectiveness of WCFR in enhancing the quality of feature subset.

In another research, W-k-means, a modified version of k-means algorithm that can compute the weights of input variables automatically, was presented by Huang et al.[181], in which a new weight for each variable is assigned based on the variance of the within cluster distances of the current partition. Similar approach was implemented

in hierarchical clustering by [182], who introduced a new hierarchical clustering algorithm called $Ward_p$ that has the ability to produce feature weights with the use of the L_p norm.

Using the literature, it is found that Information Theory and Statistical approaches are the two main methods for developing feature weighting algorithms. However, it is important to note that the probability distribution of input variables is required to be known in most of the information-based feature weighting. Since determining the probability distribution can be challenging, discretisation techniques are usually applied, with the disadvantage of information loss. Via statistical-based feature weighting, on the other hand, one can calculate the relationship between the features and the label y via means of statistical measures [190]. As a result, the statistical tests used in these approaches must be carefully chosen.

4.3 EVOLVING FEATURE WEIGHTING GRC

The focus of this chapter is to overcome the difficulty that GrC may have in grouping data with irrelevant or noisy attributes. The overall idea is to integrate feature weighting in the GrC's compatibility measure, to build a clustering framework in which there is homogeneity within clusters, but heterogeneity between clusters. In this work, the feature weighting algorithm is embedded in the iterative data granulation, in which the feature weights evolve (rather than constant) throughout the granulation process.

The Weighted K-Means algorithm (referred to as WK-Means) introduced in [181], which is a slight modification of the K-Means criteria algorithm that applies weights to the variables, served as the inspiration for the research in this chapter. The proposed approach is comparable to the feature selection wrapper approach, in which the feature weights are assigned to a collection of patterns during the clustering stage. The WK-Means algorithm aims at minimising the following object function:

$$W(S, C, w) = \sum_{k=1}^{K} \sum_{i \in S_k} \sum_{v \in V} w_v^{\beta} d(y_{iv}, c_{kv}) \tag{4.1}$$

The equation above is minimised by an iterative method, optimising (4.1) for S, C, and w, where $S = \{S_1, S_2, \dots, S_k, \dots, S_K\}$, $c_k \in C$ is the centroid for each granule k, y_i is an object in dataset Y, v is the input feature (or variable), d is the distance or dissimilarity measure between instance i and centroid of cluster k on the vth feature,

and β is the feature weighting parameter that balances the degree of effect between the weight w and its contribution to the distance.

The parameter β in the criterion (4.1) defines the degree at which weights will have an influence on their contribution to the distance. The WK-Means algorithm uses an iterative optimisation approach aligns with the original K-Means, and as a result, it is influenced by some of its advantages, for example its ability to reach convergence in a finite number of iterations [183].

The feature weight w_v adaptively changes during the clustering. Subject to $\beta > 1$, there are two options for the update of w_v, with S and C fixed:

$$w_v = \begin{cases} 0, \ if \ D_v = 0 \\ \frac{1}{\sum_{j=1}^{h} \left[\frac{D_v}{D_j}\right]^{\frac{1}{\beta-1}}}, if \ D_v \neq 0 \end{cases} \tag{4.2}$$

where h is the number of features where $D_v \neq 0$.

In research regarding GrC, the parameter w_v (feature weight) in equation (3.8) has a fixed value and usually predetermined. This means that, the weights used for all features are constant from the beginning until the end of granulation. However, this is undesirable especially when dealing with noisy features. It is believed that the current information granules do provide some information regarding the relevance of the features during the granulation process. Therefore, in this chapter, equation (4.2) is used to define and iteratively update the weight for each feature v.

Based on equation (4.2), nonzero weight is only assigned to a feature where $D_v \neq 0$. $D_v = 0$ indicates that the vth feature consists of a single value in each granule and therefore, zero weight is assigned to that feature. For example, in the case of supervised learning (where the granulation is done together with the output class), it is very likely that the output class in a particular granule is the same, causing the class variance to be zero. Hence, zero weight for the output class signifies that it does not influence the merging and the granulation itself. In this research, D_v is set as the sum of within granule variance:

$$D_v = \sum_{k=1}^{K} \frac{1}{N-1} \sum_{i=1}^{N} |y_{iv} - c_{kv}|^2 \tag{4.3}$$

where N is the cardinality in the granule k.

As shown in equation (4.2), this equation is not applicable in two undesirable situations:

(i) When $D_j = 0$ for a $j \in V$

In this situation, there is an undesirable division by zero. To mitigate this problem, [181] suggested replacing the zero value with another constant, for example the average dispersion of all features.

(ii) When $\beta = 1$

This would again cause the division by zero. In this research, $\beta > 1$ is consistently used. Note that when $\beta > 1$, the larger D_v, the smaller w_v. Hence, the impact of the variable x_v with large D_v is reduced. This selection aligns with the objective of the algorithm to assign less weight for the variables with higher within-granule variance.

Equation (3.8) is repeated here to show how the distance between two granules A and B will now be influenced by the feature weight w_v. This weighted distance will then be used in the computation of the compatibility measure to determine the next merging process. It is important to note that the weights are not constant, instead they evolve throughout the iterative granulation process, depending on the current granules variance as per in (4.3). This is what distinguish W-GrC from GrC; in conventional GrC the value of w_v is 1 for all input variables.

$$Distance_{A,B} = \frac{\sum_{v=1}^{d} w_v (D_1 - D_2)}{\mathrm{d}} \tag{3.8}$$

Algorithm 4.1 shows how the feature weighting algorithm is embedded in the GrC. In the first iteration, all input features are assumed to have equal weight; i.e. $w_v = 1$ for $v \in d$. However, it is important to note that this is only applicable in the first iteration. As stated before, the feature weight parameter β is set to be larger than 1 i.e. $\beta > 1$. The other conventional GrC setting, for example distance and density weight α and the number of iterations (or the number of final granules) remains as it is.

All features will only be treated equally during the first merging phase, after which the algorithm will adjust the feature weights when calculating distance (and hence the compatibility measure). The granules are updated and all objects in the information granules are tracked after each merging procedure. This is a crucial step in the W-GrC,

since we have to track the within-granule variance during each iteration. The feature weight for each feature are then modified and used in the subsequent iterations.

1. *Initial setting*

   For the first iteration, all features are assigned with feature weight of 1. Feature weight parameter, β is selected (i.e. $\beta > 1$). The rest follows the conventional GrC settings.

2. *Merging of granules*

   Only the first merging process is based on the conventional GrC (granules distances with equal weights), while the rest of iterations will adapt the new evolving feature weights w_v in computing the compatibility measure and selection of pair of granules to be merged.

3. *Granule update*

   Update the cardinality, maximum and minimum limit of the new information granule, and keep track the instances in each granule.

4. *Weight update*

   Update the feature weights based on (4.2). Repeat step (2-4).

Algorithm 4.1: Weighted GrC (W-GrC)

The underlying idea is that input variables with lower within-granule variance should be given higher weights; conversely, input variables with large granule variance should be penalised in the compatibility index. Large sigma (width) MFs would result from high variance. Therefore, using this evolving weight, features that promote the formation of low variance granules are promoted in any specified iteration, and such features are viewed here as being more significant for the evolvement of the granulation process towards the establishment of FL rule-bases for classification tasks. This is one way to distinguish good variables from noisy variables. Figure 4.1 shows the computing framework for weighted GrC models.

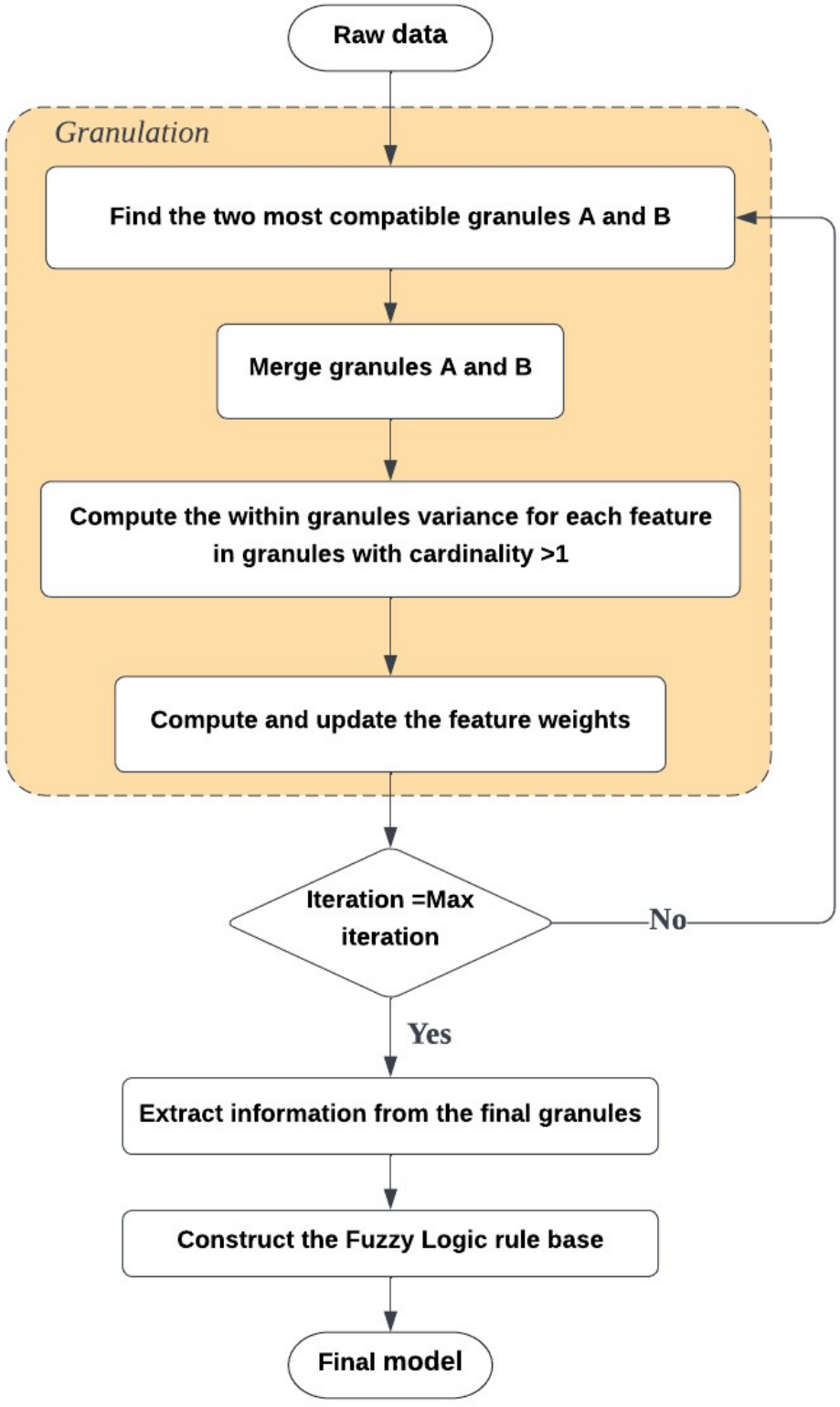


Figure 4.1: Computing framework for weighted GrC models

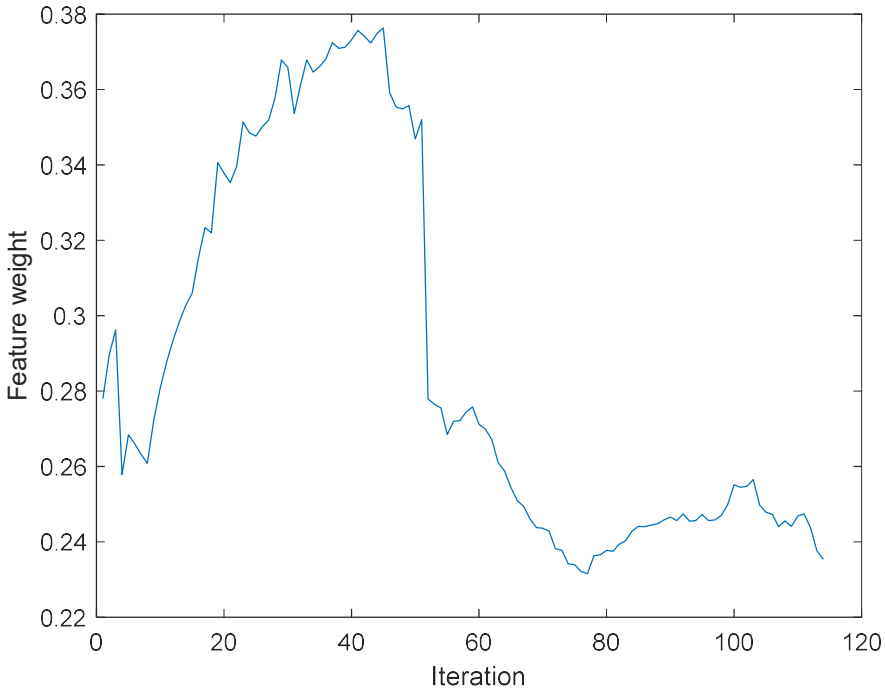
4.4 SIMULATIONS AND EMPIRICAL RESULTS

Focusing on the classification problems, simulations are run on the following three datasets: Iris, Wine, and Glass (UCI Machine Learning Repository). Every feature is resized to the range [0,1]. The data involved are split: 80% for the training and 20% for the testing. The feature weighting parameter's range is set to 2 to 10 (since $\beta = 1$ is considered undesirable situation). The average of 10 trials is used to obtain the root mean square error (RMSE) and the percentage of prediction accuracy.

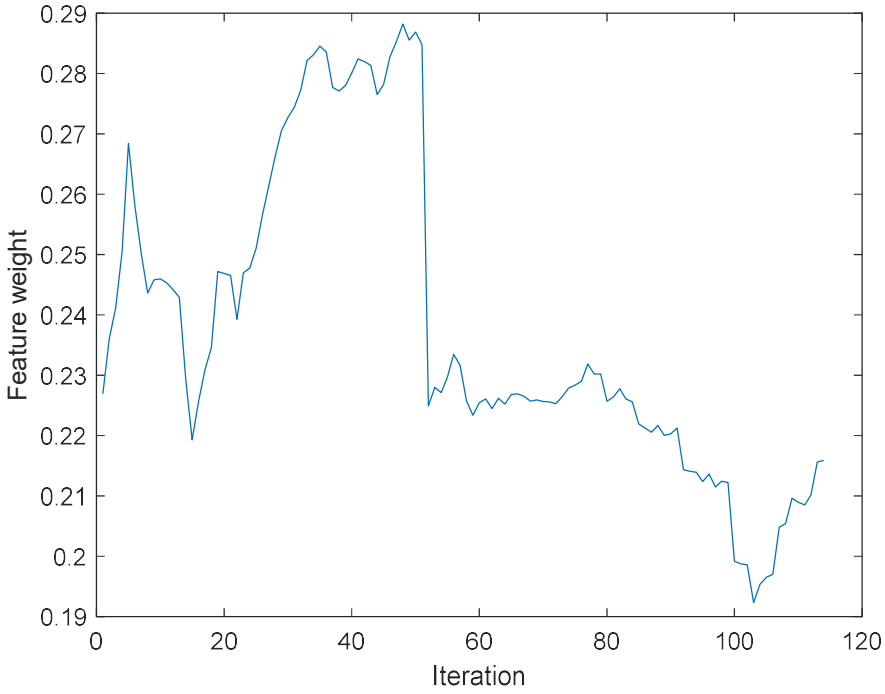
The Iris data comprises 4 input variables and 150 instances. The investigation is then expanded to datasets with larger dimensionality and scalability, such as Wine (13 input variables, 178 instances) and Glass (9 input variables, 214 instances). To regulate the number of objects for each class, the bootstrapping approach is used with Glass data. As a result, there are 371 instances instead of just 214. For reasons of comparison, five granules were chosen for Iris and Wine, whereas 30 granules were chosen for Glass based on prior study [120].

4.4.1 EVOLVING FEATURE WEIGHTING SIMULATIONS

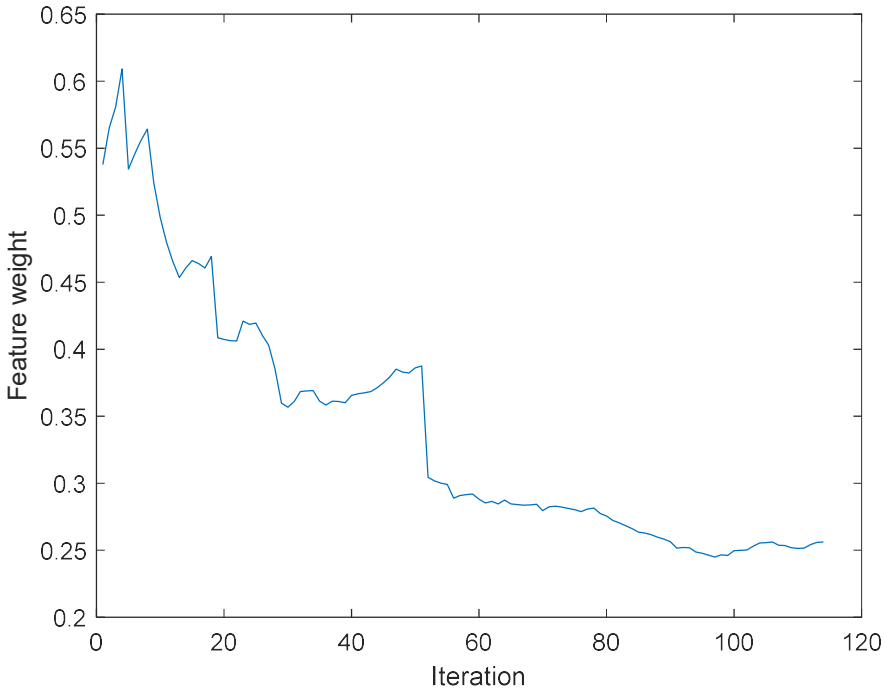
Figure 4.2 provides an illustration of how the feature weights adaptively change during the course of the iterative granulation process based on Iris dataset. All Iris' input variables are shown here: sepal length, sepal width, petal length and petal width. The feature weights are seen to be steady after the second iteration (out of 115 iterations); this is the reason why the weights are plotted starting from this point. This is because only singleton granules (i.e., those with D_v = 0) are involved in the merging process from the start, whereas the feature weights are determined based on the within-granule variances.


Feature weight
Iteration
0.38
0.36
0.34
0.32
0.3
0.28
0.26
0.24
0.22
0
20
40
60
80
100
120

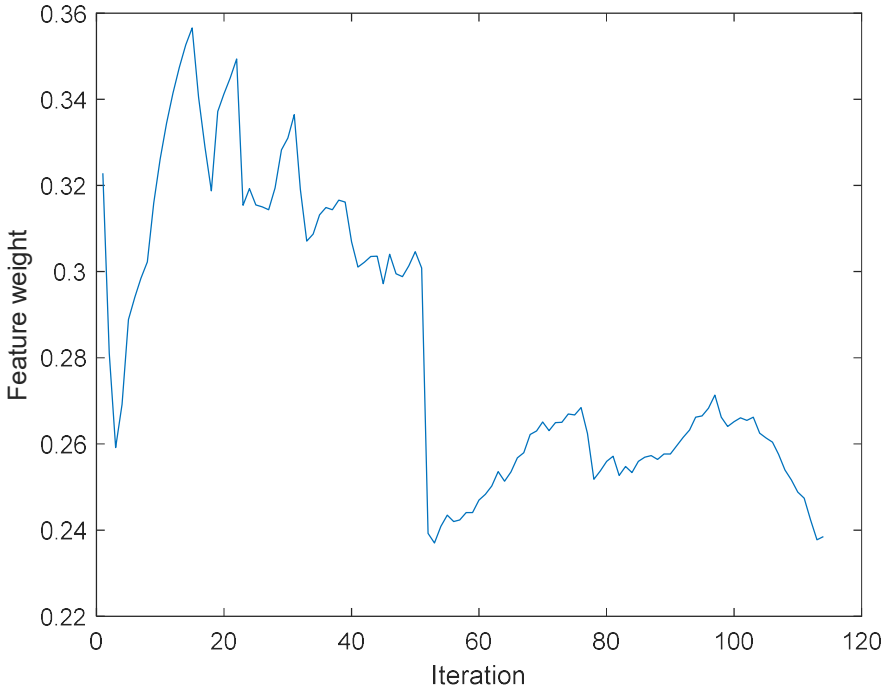
(a)


Feature weight
Iteration
0.29
0.28
0.27
0.26
0.25
0.24
0.23
0.22
0.21
0.2
0.19
0
20
40
60
80
100
120

(b)



(c)



(d)

Figure 4.2: Feature weights throughout the granulation for (a) sepal length (b) sepal width (c) petal length (d) petal width

As indicated in Table 4.1, the feature weight average is calculated and analysed with other metrics like mutual information and feature importance score. Mutual information yields details about the relationship between two random variables and are usually approximated with reference to each variable and the specified class labels [193]. It quantifies the decrease of uncertainty in one variable when information about another variable is given, and the mutual information between two variables X and Y is given by:

$$I\ (Y;X) = \sum_y \int_x\ p(y,x) \log \frac{p(y,x)}{P(y)p(x)} dx \tag{4.4}$$

Equation (4.4) also can be expressed in terms of entropy and conditional entropy, as described by:

$I\ (Y;X) = H(Y) - H(Y|X)$ where the $H(Y)$ is the Shannon entropy that measure the uncertainty in a discrete random variable Y, given by:

$$H(Y) = -\sum_y P(y) \log(P(y)) \tag{4.5}$$

And $H(Y|X)$ is the conditional entropy, that represents the remaining uncertainty when X is known, described as:

$$H(Y|X) = -\int p(y,x)\left(\sum_y p(y|x) \log p(y|x)\right) dx \tag{4.6}$$

The feature importance score ranks the features using a chi-square (χ^2) test [194]. The feature importance score is the negative log of chi-square tests' p-value [195]. This means, the lower the p-value, the higher the importance of the selected feature [194]. When the dependent variable is assessed at a nominal level, the chi-square statistic provides a non-parametric tool for analysing group differences [196]. Figure 4.3 shows the predictor or feature importance score for all input variables in Iris data.

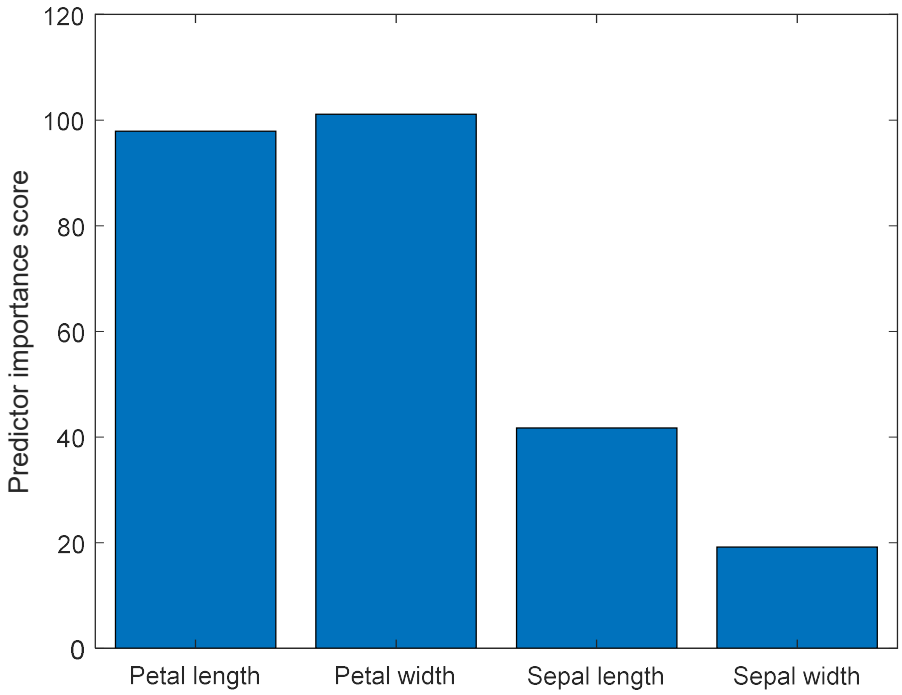


Figure 4.3: Predictor importance score for all input variables in Iris data

This finding suggests that the feature weight ranking is consistent with the other two measures (mutual information and feature importance score); supporting the theory that this suggested approach effectively captures the feature relevance. Petal width is the attribute that all three of these measurements regard as being most significant, followed by petal length, sepal length, and sepal width.

Table 4.1: Comparison of average feature weight in W-GrC with the feature importance score and mutual information (Iris data)

	Average weight (W-GrC)	**Feature importance score**	**Mutual information**
Sepal length	0.2721	41.7358	0.6415
Sepal width	0.2062	19.1551	0.3935
Petal length	0.3072	97.8866	1.2663
Petal width	0.3623	101.1028	1.3245

4.4.2 RESULTS

The performance of W-GrC with various values of β is presented in Table 4.4. The results for the standard GrC, sometimes referred to as the GrC without feature weighting, are shown in the row labelled "no feature weighting." It is seen that the suggested W-GrC outperforms the standard GrC in terms of RMSE and predictive accuracy with appropriate choice of β. The parameter β should be regarded as a hypermeter in this context, which is specified in each circumstance.

For the Iris data, improvement in the performance is achieved at $\beta \in \{3,4,5,6,7,8,10\}$. With 96.33% of correct predictions as opposed to 94% in the standard GrC (an improvement of 2.3%), the maximum accuracy is attained when $\beta = 3$ and $\beta = 6$. In terms of RMSE, all trials score better than the conventional GrC, except for when $\beta = 2$. This proves the capability of W-GrC, even though the number of dimensions is small (only 4 input variables in Iris). Next, the robustness of W-GrC is tested with larger datasets (13 and 10 input variables in Wine and Glass, respectively).

It is clear that the training performance of W-GrC (in Iris) with appropriate input variables weight produces better result than the standard GrC in terms of the predictive accuracy. Moreover, the final rule-bases developed with W-GrC and T1-FLS show good distinguishability as in Figure 4.5 (as compared with the final rule-bases developed with GrC and T1-FLS in Figure 4.4), particularly for the input variables petal

length and petal width. Interestingly, these two input variables are shown to be the most important variables with the highest average weight. Note that there are some fuzzy sets that can be merged, however in this case, the merging process is set to stop at 5 granules (hence 5 fuzzy sets).

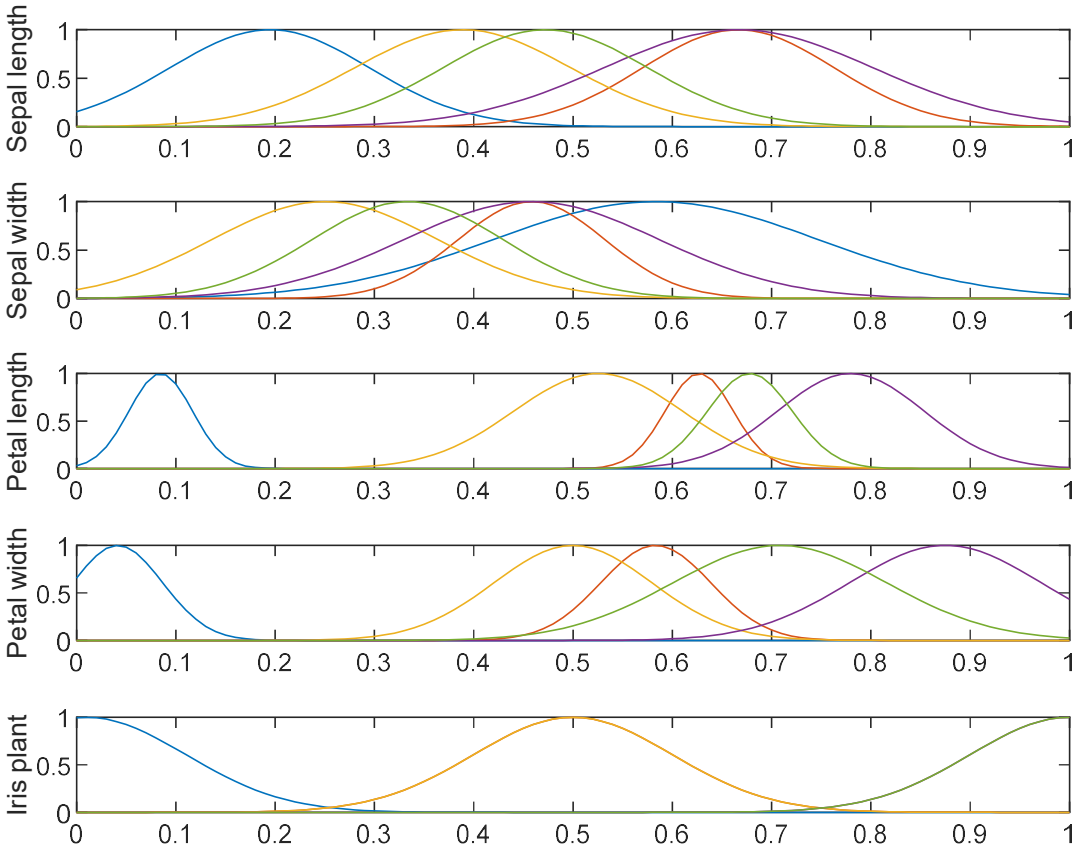


Figure 4.4: The final rule-base of T1-FLS model constructed with GrC

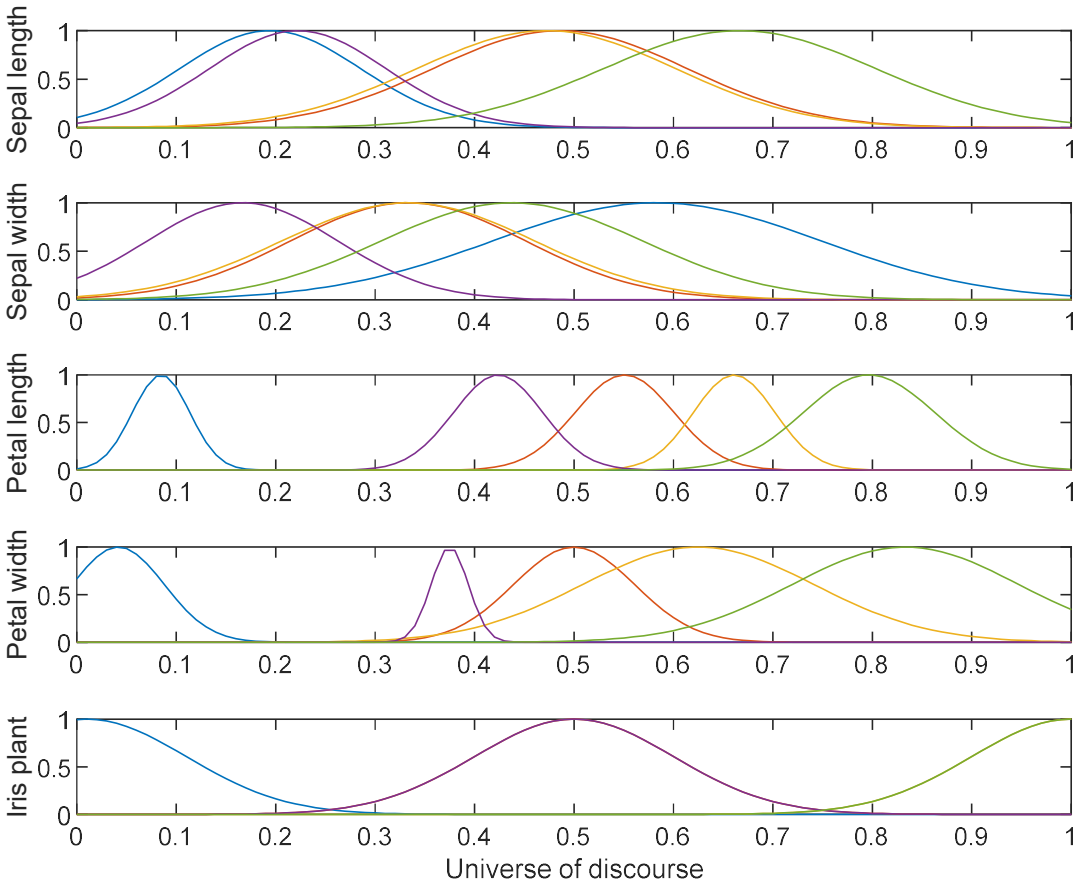


Figure 4.5: The final rule-base of T1-FLS model constructed with W-GrC

Improvement is recorded in the experiment using Wine at a range of β from 3 to 6. Accuracy levels above 90% are observed in every experiment. The best accuracy is obtained when $\beta = 5$ with 95.67% while the conventional GrC scored 92.3%. The improvement in RMSE also is accomplished in the same range of β (from 3 to 6). The trend in Wine is different with the previous dataset (Iris), where the RMSE is improved in most trials. However, it is interesting to observe that most of the good performance started from the value of $\beta = 3$. Table 4.2 shows the average feature weight for Wine dataset.

Table 4.2: Average feature weight for Wine dataset

Input variables	**Average feature weight**
Flavanoids	0.0893
Color intensity	0.0824
OD280/OD315 of diluted wines	0.0823
Hue	0.0821
Proanthocyanins	0.0802
Alcalinity of ash	0.0798
Proline	0.0792
Ash	0.0786
Alcohol	0.0786
Total phenols	0.0781
Magnesium	0.0762
Nonflavanoid phenols	0.0676
Malic acid	0.0643

We can see that greater values of β ($\beta \geq 3$) are preferred to provide superior results in the experiment with the Glass dataset. The highest accuracy of 71.86%, compared to standard GrC's 62.79%, is obtained at $\beta = 5$. Only 32.33% accuracy is achieved when using low value of β (i.e. $\beta = 2$), before the algorithm becomes more stable starting from $\beta = 3$. However, in general, the result for RMSE in the conventional GrC is still better than W-GrC, except for when $\beta = 7$ (0.2020 in the conventional GrC and 0.1980 in W-GrC). Table 4.3 shows the average feature weight for Glass dataset.

Table 4.3: Average feature weight for Glass dataset

Input variables	Average feature weight
Barium	0.1696
Potassium	0.1377
Calcium	0.1372
Refractive index	0.1188
Iron	0.1001
Magnesium	0.0936
Sodium	0.0930
Silicon	0.0922
Aluminium	0.0831

Table 4.4 shows the average root mean square error (RMSE) and percentage of prediction accuracy performance of W-GrC with various β values for all datasets under study. The measures were obtained using ten-fold cross-validation. Overall, W-GrC performs better than the standard GrC, according to Table 4.4. When an adequate value of β is chosen, it provides the best accuracy for all datasets. This is due to the fact that during the iterative granulation process, variables that are more crucial for a particular instance are given higher weights when generating the information granules. However, it should be highlighted that choosing β carefully is essential for achieving superior classification accuracy. As a result, it is recommended that ($\beta \geq 3$) be chosen as the proper value of for this specific case studies (based on the three datasets used).

Results were compared to other studies such as [197] with 96.67% in Iris, [198] with 97.14% in Wine and [199] with 71.66% in Glass. This shows that the performance of W-GrC is comparable with other research works in the literature. Note that there is not any parameters optimisation involved in this experiment.

Table 4.4: Average RMSE and % accuracy performance of W-GrC with various β values, testing (unseen) data, 10 runs per β value.

	Iris		**Wine**		**Glass**	
	RMSE	**Accuracy (%)**	**RMSE**	**Accuracy (%)**	**RMSE**	**Accuracy (%)**
No feature weighting	0.1415	94	0.1173	92.3	0.2020	62.79
$\beta = 2.0$	0.1551	90.67	0.1238	91	0.3365	32.33
$\beta = 3.0$	0.1205	**96.33**	0.1082	94	0.2235	63.02
$\beta = 4.0$	0.1302	94.67	0.1123	92.67	0.2164	69.30
$\beta = 5.0$	0.1253	94.33	0.1033	**95.67**	0.2144	**71.86**
$\beta = 6.0$	0.1251	**96.33**	0.1067	93	0.2165	66.98
$\beta = 7.0$	0.1285	95.67	0.1230	92	0.1980	66.51
$\beta = 8.0$	0.1189	96	0.1342	90.33	0.2219	66.05
$\beta = 9.0$	0.1346	93.67	0.1186	91.67	0.2105	68.14
$\beta = 10.0$	0.1273	95	0.1212	91.33	0.2224	65.81

4.5 INTERPRETABILITY INDEX

Interpretability and accuracy are frequently incompatible goals when building Fuzzy Logic systems (FLS); one might be improved by forgoing the other, a scenario known as an interpretability-accuracy trade-off. For instance, Mamdani-based FLS offers improved interpretability, whereas TSK-based FLS provides improved prediction accuracy [22]. In the context of FLS, interpretability is the feature of a model that enables humans to comprehend a system's behaviour by scrutinising its rule base. In this work, it is determined whether the improved predictive performance has an impact on the models' interpretability by utilising the models that were created using values of β that performs the best in terms of accuracy (according to the results in Table 4.4).

Utilising Nauck's index (NI), the effect of feature weighting on the interpretability index is examined. Nauck [89] established the Nauck's index, a numerical measure, to evaluate the interpretability of FL rule-based classification systems. It comprises three important elements: complexity of FLS ($comp$), average normalised coverage of fuzzy partition ($\overline{cov}$) and average normalised partition index ($\overline{part}$), and is expressed as:

$$Nauck\ index = comp\ \times \overline{cov} \times \overline{part} \tag{4.7}$$

The interpretability index for the new W-GrC and the standard GrC are compared in Table 4.5. It has been proven that W-GrC can produce greater accuracy without statistically significant degradation in the model's interpretability. Less than 2% for the Iris data, and even less for the Wine and Glass case studies, are the impacts on the interpretability index. Due to the large number of rules, it should be noted that Glass's NI is comparatively small (due to 30 rules as opposed to 5 rules in Iris and Wine).

Table 4.5: Comparison of the interpretability index

	Nauck's index	
	W-GrC	**GrC**
Iris	0.3076	0.3129
Wine	0.0929	0.0928
Glass	7.02×10^{-4}	7.07×10^{-4}

4.6 SUMMARY

In this chapter, a new iterative data granulation technique with evolving feature weighting is introduced in order to define the significance of input variables and utilise such weights to guide the information granulation process. Based on the total within-granule variances from the generated granules at each given iteration, the weight for each feature is calculated. The significance of each feature is assessed in order to determine which features are crucial for the calculation of the granules' compatibility measure.

The relevance of features that resulted from averaging feature weights across the data granulation process is shown to be consistent when compared to other approaches like the chi-square test and mutual information; agreement in the feature ranking is confirmed. This agreement between the measure of weight based on the within-granule variances and the other methods give confidence to embed the evolving feature technique in the iterative data granulation.

The proposed W-GrC algorithm outperformed the conventional GrC algorithm in simulation results (UCI classification tasks) in terms of classification accuracy. The more complicated dataset, such as the Glass case study, showed more

significant improvement. The outcomes of the experiment demonstrated the capability of the proposed GrC-Fuzzy modelling framework to handle data of different complexity (features ranging from 4 to 13). Although the study's findings were encouraging, further research might examine case studies with a wider range of complexity and compare the effectiveness of the W-GrC approach to other clustering approaches.

Additionally, this chapter covers the preliminary research on the Fuzzy Logic Systems' interpretability. NI is used to evaluate the interpretability of the generated models, and despite the increased percentage accuracy in the classification tasks, there is no discernible decline in the interpretability. However, the interpretability studies in this chapter is limited to only Type-1 FLSs. The study of interpretability in Type-2 FLSs is covered in chapter 6. In Chapter 6, the W-GrC is extended to Radial Basis Function Neural Networks (Type-1 and Type-2), with more complex datasets such as Breast Cancer (683 instances, 9 input variables) and Cardiotocography (2126 instances, 21 input variables).

Main results and methodology from this work is included in paper entitled “An evolving feature weighting framework for granular Fuzzy Logic models”. It was presented and published in UKCI 2021: 20th UK Workshop on Computational Intelligence in Aberystwyth, Wales, UK, where the authors received the Best Student Paper Award.

In the next chapter, focus is given to tackle another problem in data granulation, known as class overlapping. Rather than operating in a hard clustering approach, the iterative data granulation is aimed at allowing an instance to belong to one or more granules rather than just one. The challenge is to model the overlapping between granules and to propose a new compatibility equation that takes into account the class overlapping.

CHAPTER 5

ADDRESSING CLASS OVERLAP IN GRANULAR COMPUTING

This chapter addresses another issue in data granulation – class overlapping. Conventionally, clustering algorithms aim at producing a number of disjoint clusters, or granules that meet some required criteria. In disjoint clustering, an element or instance is only allowed to belong to one cluster. However, most of real-world data sets contain overlapping information, meaning that certain data objects or patterns may belong to many clusters. For example, a person may belong to more than one group, or organisation. Therefore, the aim of this chapter is to allow an instance to belong to more than one granule, with the assistance of a new parameter known as R-value. R-value is defined as the ratio of overlapping areas among categories in a data cluster. It is used to model the overlapping between granules, and this information is integrated into the new compatibility criterion. Consequently, the new compatibility measure provides additional parameter control for the amount of overlapping among the granules. The new GrC with overlapping measure is demonstrated by using datasets from UCI Machine Learning Repository.

5.1 INTRODUCTION

Clustering involves organising a group of things into classes, with related objects belonging to the same cluster and dissimilar objects belonging to different clusters [200]. A single cluster is assigned to each item in the majority of clustering algorithms, which use hard clustering approaches. On the other hand, fuzzy clustering algorithms provide a structure where each object contributes to the definition of each cluster [200].

There are various classification schemes for clustering algorithms, and one of them is based on the types of produced clusters [201]:

i) Disjoint, where an element only ever belongs to one cluster.
ii) Fuzzy, when one element is included in all clusters but only to a certain extent.
iii) Overlapping, when a component may be a part of multiple clusters.

Most clustering techniques create exclusive clusters, which means that each sample can only be a member of one cluster [13]. When group boundaries are distinct and the data is devoid of outliers, traditional techniques like the k-means algorithm may be successful in appropriately dividing points into groups in many practical data models [15]. The majority of real-world data sets, however, contain overlapping information, meaning that certain data objects or patterns may belong to many clusters [14].

The necessity to work in overlapping clustering is due to the many problems that require the groups to be non-exclusive [201]. This problem naturally arises because many real-world applications require that each observation be assigned to one or more clusters. This issue has been the subject of numerous studies using the overlapping clustering technique [14]. Recent approaches that have been suggested to address this issue are theoretical rather than heuristic in nature, and thus incorporate overlaps in their optimal criteria [202]. Overlapping clusters can occur for a number of reasons: the data may be noisy, the features may not fully capture the information required to distinguish the clusters, or the overlap may be a natural result of the processes that generate the data [203].

In Granular Computing (GrC), the development of a new information granule results in a lack of distinguishability due to the fact that the overlapping is not taken into account in the compatibility measure [130]. However, the work related to overlapping in GrC is limited. Solis [130] developed the neutrosophic approach to add a dimension that allows the compatibility criterion to evaluate the fuzzy entropy (uncertainty) produced during granulation in order to estimate the overlapping behaviour. With the help of this approach, the compatibility search is persuaded to stop looking for prospective granules that could lead to granular overlapping, which would reduce model transparency and compromise the rules' consistency. This method,

however, only aims to attenuate the overlapping behaviour, even though it is recognised that some overlapping is necessary in building FL models.

Therefore, in this chapter, GrC takes into account the overlapping measure (R-Value) to model the overlapping between the granules, and a novel compatibility equation that model the overlapping during granulation is proposed. In this setting, the proposed approach can be viewed as a compromise between hard and fuzzy-clustering methods. This method permits an object to belong to one or more granules rather than just one, leading to the intersection of final granules. Hence, the objectives of this chapter are:

- To study the feasibility of using R-value in quantifying the overlapping between the information granules during the iterative data granulation.
- To propose a new compatibility measure with overlapping measure to enhance the iterative data granulation.

5.2 OVERLAPPING CLUSTERING

Overlapping clustering technique has been studied in many research works [14]. The overlapping k-means algorithm (OKM), which is an extension of the k-means algorithm to produce overlapping clusters, is one of the most used overlapping clustering algorithms [204]. The K-means objective function is modified by the OKM algorithm to take the potential of overlapping clusters into account [205]. In OKM, the condition of $\pi_i \cap \pi_j = \emptyset$ is removed from the objective function, where $\pi = \{\pi_i, \dots \pi_k\}$ represents the set of k clusters. The OKM method relaxed the k-means objective function to permit overlapping clusters [13]. More specifically, eliminating the constraint $\pi_i \cap \pi_j = \emptyset$ allowed the algorithm to have overlapping clusters. OKM algorithm was extended in various ways, including kernel-overlapping k-means (KOKM ϕ) [206] and parametrised OKM methods [207].

Recent works on k-means with overlapping can be found in [14] where the algorithm called 'Enhanced overlapping clustering algorithm for data analysis (eHMCOKE) applied median absolute deviation (MAD) in identifying outliers in datasets and added parameters such as radius of clusters and distance between clusters. These steps enables it to be more effective regarding the overlapping clustering.

However, all these methods are really just extensions of k-means that are very sensitive to initialisation process, where the outcome is subject to the randomly selected initial cluster centroids [13].

In hierarchical clustering, Gama et al. [208] introduced a novel method to obtain hierarchical and overlapping clusters from network data. The proposed method permitted nodes to be a member of more than one group and produced a tiered collection of groupings of the node set depending on the desired level of resolution or similarity. In another research, Jeantet et al. [209] suggested a new approach based on a density requirement that permits clusters to overlap until a high cluster attraction is attained. A directed acyclic graph was used to depict the resulting hierarchical structure, known as a quasi-dendrogram, which combined the accuracy of a less arbitrary clustering with the benefits of hierarchies.

5.3 OVERLAPPING GRC

5.3.1 R-value

In this chapter, R-value introduced in [210] is used to numerically measure the overlapping among the granules. R-value is a measure that can capture the amount of category overlap in a dataset and can be considered as the ratio of samples in the overlapping area [211]. The potential of this measure has attracted many researchers to use R-value in their research works especially when dealing with overlapping for example in [212], [213], and [203].

R-value between the two categories C_i and C_j is represented by $R(C_i, C_j)$ and it signifies the ratio of instances of C_i and C_j which are positioned in the overlapping area of C_i and C_j. R-value utilises normalised values, meaning that R-value ranges between 0 and 1. The overlapping measure in this chapter is given as:

$$R(C_i, C_j) = \frac{1}{|C_i|+|C_j|}[r(C_i, C_j) + r(C_j, C_i)] \tag{5.1}$$

Where

$$r(C_i, C_j) = \sum_{m=1}^{|C_i|} \lambda(|kNN(p_{im}, C_j)| - \theta) \tag{5.2}$$

Where $\lambda(x) = 1\ if\ x > 0, else\ \lambda(x) = 0$, p_{im} is the m-th instance of category C_i, $kNN(p_{im})$ is the set of K-nearest neighbour instances for an instance p_{im},

$kNN(p_{im}, c_j)$ is the set of instances in $kNN(p_{im})$ that belong to the different category c_j, can be defined as:

$$kNN(p_{im}C_{ij}) = \{x | x \in kNN(p_{im}) \wedge x \in C_j\} \quad (5.3)$$

and k is the number of nearest neighbour instances for a given instance. The parameter θ is the threshold value on the number of different class neighbours for considering an instance as belonging to an overlap region [203] within the range $[0, k/2]$. The kNN function is implemented by using a function called $knnsearch$ in Matlab R2020a, which by default, find k-nearest neighbours using input data by using Euclidean distance [214], such that:

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (5.4)$$

In this section, a synthetic dataset is used to exemplify the calculation of R-value. The synthetic data consists of two input variables, Var 1 and Var 2 and one output Class. There are 20 instances, which are equally divided into 2 categories or classes. Figure 5.1 shows the k-nearest neighbour for two points with $k = 3$ and $\theta = 1$. The synthetic dataset used here is shown in Table 5.1.

Table 5.1: Synthetic dataset

Var 1	Var 2	Class
3.5000	3.9000	1.0000
3.4000	3.7000	1.0000
3.3000	2.9000	1.0000
3.4000	3.0000	1.0000
3.5000	2.9000	1.0000
3.5000	3.0000	1.0000
2.8000	4.0000	1.0000
3.3000	3.9000	1.0000
3.6000	3.9000	1.0000
3.8000	4.1000	1.0000
2.9000	3.6000	2.0000
3.3000	3.4000	2.0000
3.6000	3.8000	2.0000

3.5000	4.1000	2.0000
3.9000	3.7000	2.0000
3.5000	4.5000	2.0000
4.3000	5.1000	2.0000
4.6000	4.4000	2.0000
4.1000	4.4000	2.0000
4.0000	4.8000	2.0000

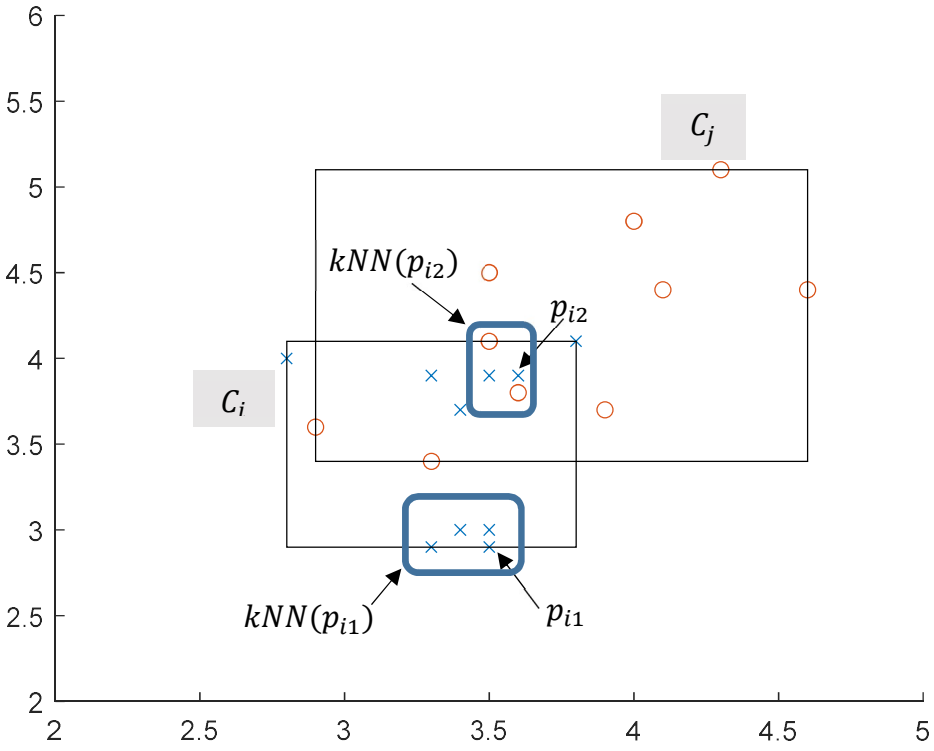


Figure 5.1: K-nearest neighbour for two points

From this synthetic dataset, with $K = 3$ and $\theta = 1$, we found that (from equations 5.2 and 5.3):

$$kNN(p_{i1}, C_j) = 0$$

$$\lambda(|kNN(p_{i1}, C_j)| - \theta) = 0$$

This indicates that the element p_{i1} does not overlap with the category C_j. All three nearest neighbours for p_{i1} are within the same category. However, these values will

change if we increase the value of parameter k. Note that the fourth closest instance from p_{i1} is from a different category.

From Figure 5.1, it is obvious that the instance p_{i2} is located in the overlapping area between C_i and C_j, therefore:

$$kNN(p_{i2}, C_j) = 2$$

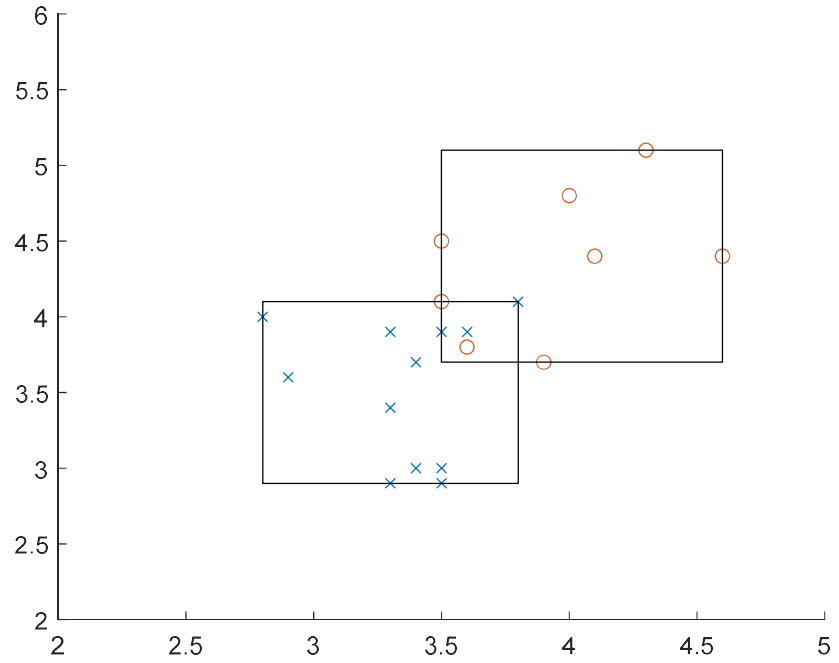
$$\lambda(|kNN(p_{i2}, C_j)| - \theta) = 1$$

Where $\lambda = 1$ shows that the element p_{i2} does overlap with the category C_j. From this figure, 2 out of 3 nearest neighbours for p_{i2} are of different categories. Note that the value of λ will be zero, for example if we change the parameter θ to 2. Repeating these steps for all instances in both categories, the R-value (overlapping measure) for the two granules (computed via Matlab) is:

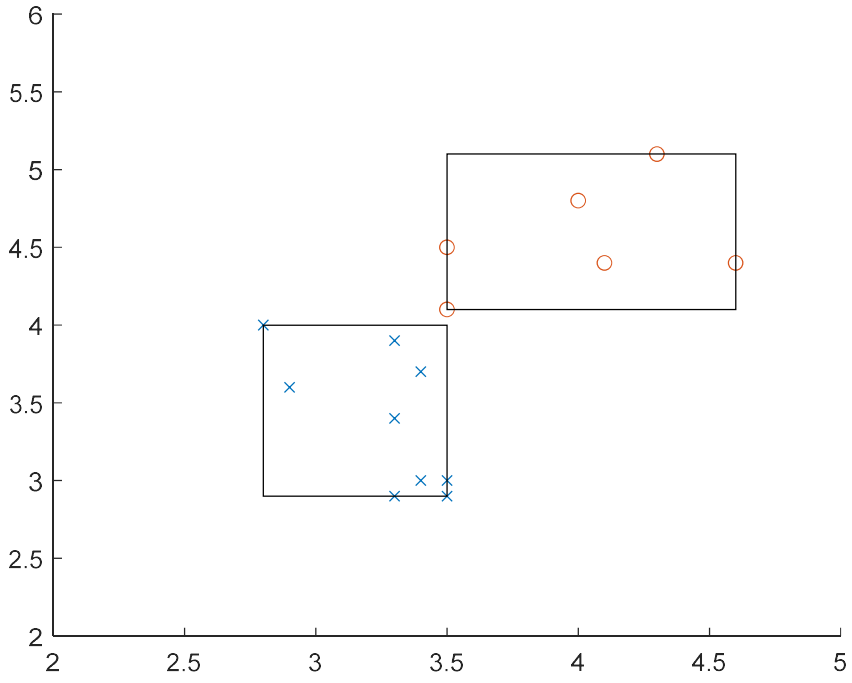
$$R(C_i, C_j) = 0.45$$

A high R-value indicates high overlapping among the granules. To demonstrate this, different datasets are used to see how R-value changes with regards to the overlapping area among the granules. This is shown in Figure 5.2, where datasets with less overlapping (compared to Figure 5.1) is used in the iterative data granulation. The synthetic datasets here are designed to have (a) less overlapping and (b) no overlapping at all.

With R-value, the overlapping in Figure 5.2 (a) and (b) is 0.3 and 0.0667 respectively, which is less than the overlapping measured in Figure 5.1. Note that even though there is no physical overlapping between granules in Figure 5.2 (b), the R-value is not exactly zero. This is due to the closeness of these two granules and the selection of $\theta = 1$, meaning that only one instance is allowed to belong to other category; more than this will be considered as overlapping. $R(C_i, C_j)$ is expected to be zero if the distance between these two granules are increased.



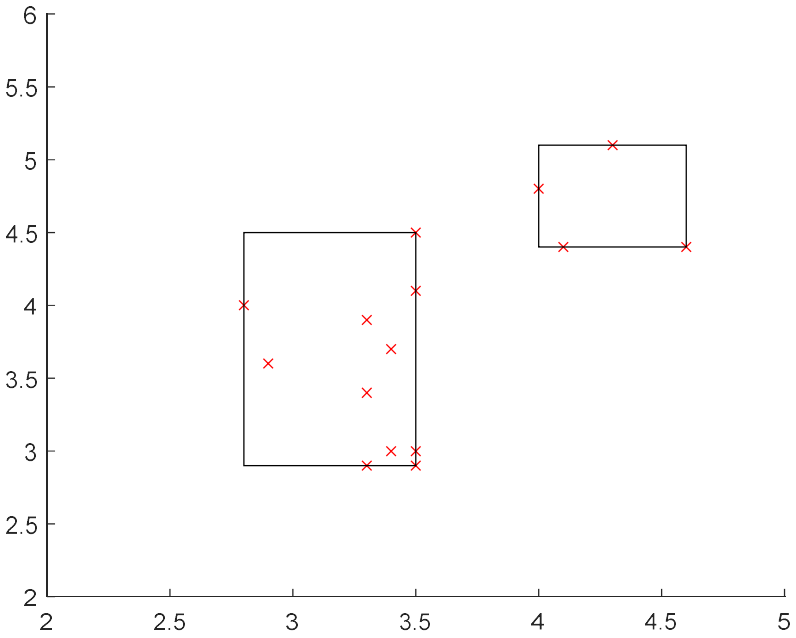
(a)



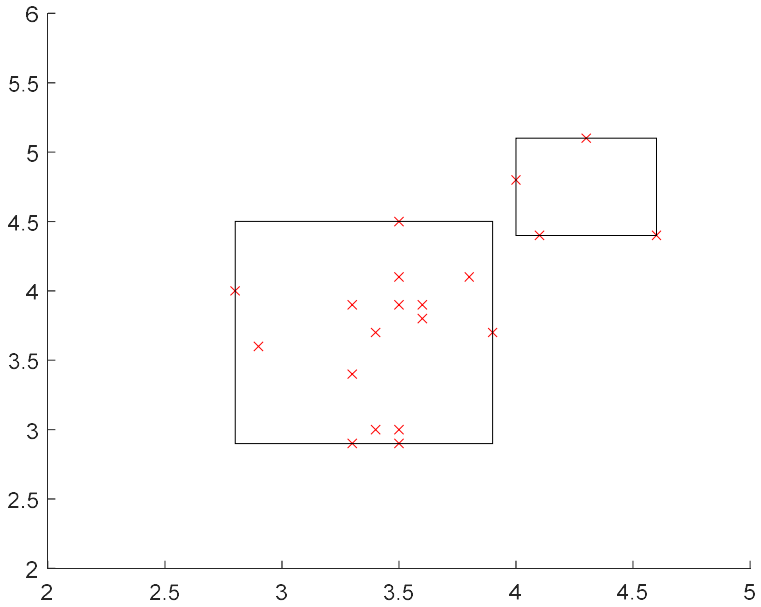
(b)

Figure 5.2: Granules overlapping with (a) $R(C_i, C_j) = 0.3$ and (b) $R(C_i, C_j) = 0.0667$

Next, the R-value is examined with another dataset that consists of only one category. In Figure 5.3, it is observed that the value of $R(C_i, C_j)$ for both dataset is 1. This indicates that the interest to measure the degree of overlapping is for data with different categories. In data granulation perspective, the maximum value of R-value for granules with similar categories means that it will not affect the compatibility criterion, rather it provides additional parameters to model or to 'control' the amount of overlapping among the granules (particularly of different categories).



$(a) \quad R(C_i, C_j) = 1$



$(b) \quad R(C_i, C_j) = 1$

Figure 5.3: Overlapping of two granules with the same categories

5.3.2 A new overlapping measure during the iterative data granulation

In Chapter 3, the use of the compatibility measure in determining the granules to be merged is shown. The compatibility criterion (Equation 3.5) however, does not consider the overlapping measure, and may cause the lack of distinguishability among the granules [130]. Therefore, a novel compatibility criterion that measure the amount of overlapping during granulation is proposed in this chapter.

In this proposed algorithm, the overlapping is not only allowed to occur, but also can be controlled by means of two parameters, k and θ. k is the number of nearest neighbour instances for a given instance, and θ sets the maximum number of instances allowed to belong in class neighbours. For consistency, in this research the values of k and θ are set to be 5 and 1, respectively.

A modification of the compatibility equation is proposed, taking into account the overlapping among the granules. If two granules consist of the same categories, the R-value is always 1, showing that they are favourable to be merged. In the case of two different granules with different categories, the merging of two granules with low

overlapping (R-value) is penalised. Low overlapping indicates that the categories are well separated and well distinguished, hence should not be merged. In an extreme condition where there is a need to merge 2 granules with different categories (as in Figure 5.1 and 5.2(a)), the merging in Figure 5.1 is considered more favorable than in Figure 5.2(a).

The Equation 3.5 is revisited here to show the conventional compatibility measure in GrC:

$$C\ (A,B) = Distance_{MAX} - \ Distance_{A,B}.\exp(-\alpha\ (Density)) \qquad (3.5)$$

where $Distance_{MAX}$ is the maximum possible distance in the dataset, $Distance_{A,B}$ represents the distance between granule A and B and α is the density factor. However, the density is not the topic of research in this chapter, and it remains as it is (as in Equation 3.6). The interest in this chapter is overlapping measures during the data granulation. Hence, the proposed compatibility measure includes the overlapping measure, or R-value, in the compatibly equation as follows:

$$C(A,B) = Distance_{MAX} - \ Distance_{A,B}.\exp(-\alpha\ (Density\ \times\ R-value)) \quad (5.5)$$

The inclusion of R-value in Equation (5.4) promotes the degree of overlapping during the iterative data granulation. With this, GrC is allowed to have overlapping among the granules, and the amount of overlapping can be controlled with parameters K and θ. However, the impact of R-value in this equation is still constrained by the density factor α. Therefore, in the experiment, five different values of α are used ranging from 0.2 to 1. Figure 5.4 shows the framework for overlapping GrC based FL modelling. The third and fourth steps (in shaded boxes) are what differ the overlapping GrC with the conventional GrC, where the overlapping between granules is computed and hence influences the merging during iterative data granulation.

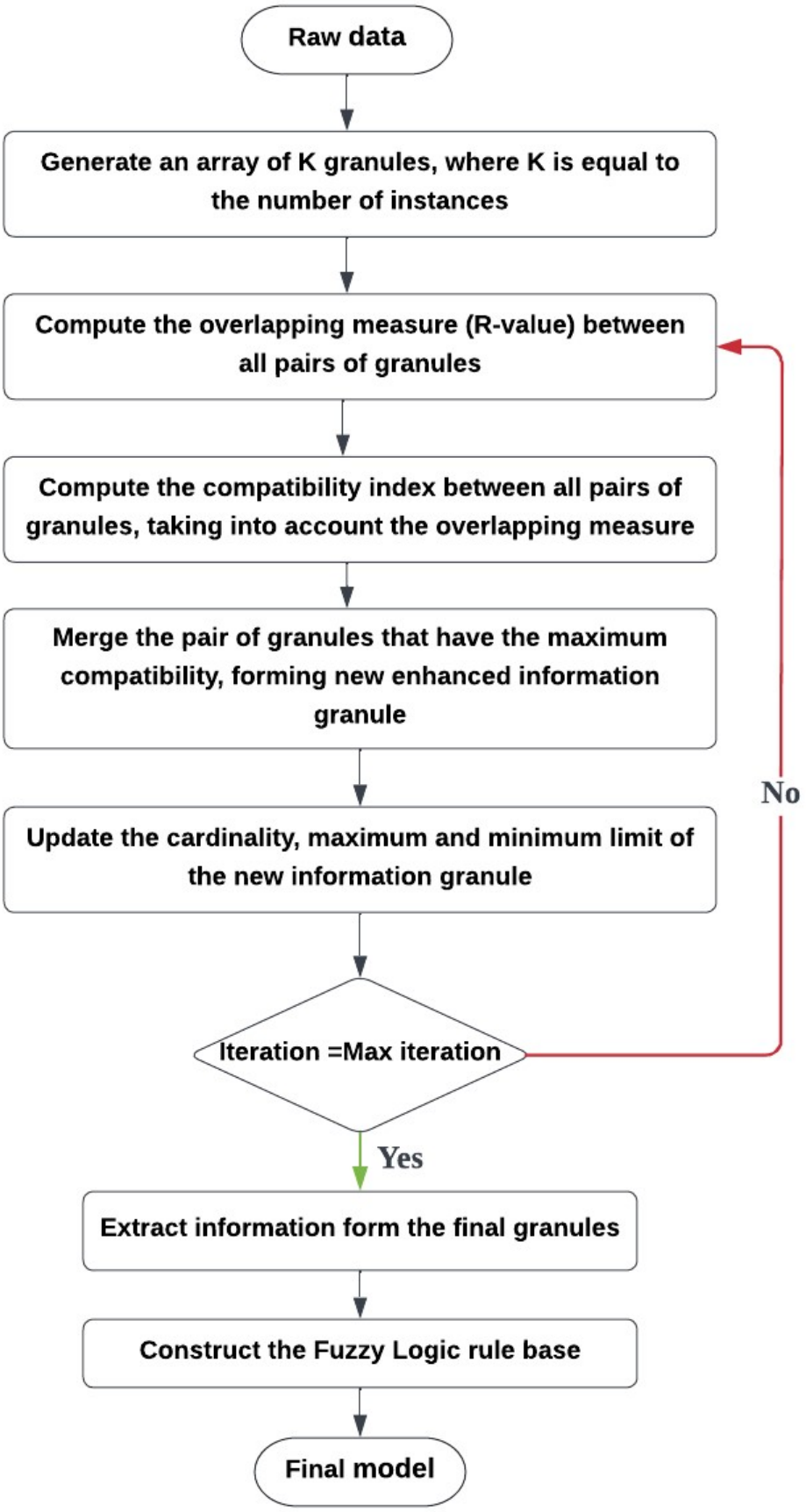


Figure 5.4: Overlapping GrC based Fuzzy Logic modelling framework

5.4 CASE STUDY AND SIMULATION RESULTS

The improvement demonstrated by the GrC with overlapping measures is presented in this section. The datasets - Iris, Wine and Glass from UCI Machine Learning Repository are used to validate the performance of the proposed system. The data is split: 80% for the training and 20% for the testing, and the results are taken from 10 trials in each experiment.

Using Matlab R2020a, the function $knnsearch$ finds the k-nearest neighbour for instance. The values of k and θ are 5 and 1, respectively and are fixed throughout the experiment for consistency. In the experiment, the value of density factor α ranged from 0.2 to 1.0, with an increment of 0.2. The results for Iris, Wine and Glass are presented in Table 5.2 to Table 5.4.

Table 5.2 shows the experiment result for Iris (with 4 input variables, 150 instances) in terms of accuracy and RMSE. From out of five simulations (with different values of α), three simulations of GrC with overlapping measure show better performance in terms of the classification accuracy (i.e. when $\alpha = \{0.2, 0.4, 0.8\}$). The same pattern is also observed in the comparison of RMSE for GrC with overlapping measure and GrC without overlapping measure.

The improvement achieved by the proposed method can be seen more clearly in the Wine dataset (with 13 input variables, 178 instances), where the accuracy of the GrC with overlapping measure consistently outperforms the conventional GrC, as shown in Table 5.3. The highest accuracy is accomplished when $\alpha = 0.6$, where the GrC with overlapping measure scores 96.67%. It is also observed that the conventional GrC can only sustain low values of α ($\alpha \leq 6$) to have good result (i.e. accuracy $\geq$ 90%). In the case of higher α (0.8 and 1.0), the performance of conventional GrC drops to 76.67% and 52.67%, respectively, while overlapping GrC scores 96% and 95.33% . This also demonstrates the robustness of the overlapping GrC in building FL models.

Table 5.4 presents the performance of overlapping GrC in Glass dataset (with 9 input variables, 214 instances). GrC with overlapping measures consistently outperforms the conventional GrC in the Glass dataset. The highest performance is achieved at $\alpha = 0.4$ with 71.16% accuracy. The conventional GrC achieves satisfactory results (above 60%) for all values of α except when $\alpha = 1.0$. From the results, we can

see that even for more complicated datasets (e.g. Glass) GrC with overlapping measure still can produce good result even though high values of α were used.

All the results are benchmarked with other related works, for example 97.33% (Iris) and 95% (Wine) in [177], [178], and 75.45% (Glass) in [178].

Table 5.2: Experiment result for Iris in terms of accuracy and RMSE.

	With overlapping measure			**Without overlapping measure**		
	RMSE	**Accuracy (%)**	**Standard deviation (%)**	**RMSE**	**Accuracy (%)**	**Standard deviation (%)**
$\alpha = 0.2$	0.1164	97.14	2.3	0.1261	96.67	3.00
$\alpha = 0.4$	0.1170	97.62	2.52	0.1298	95.71	2.52
$\alpha = 0.6$	0.1227	96.67	3.33	0.1224	97.14	2.30
$\alpha = 0.8$	0.1178	97.62	3.17	0.1325	96.19	1.26
$\alpha = 1.0$	0.1421	93.33	3.85	0.1249	97.62	3.17

Table 5.3: Experiment result for Wine in terms of accuracy and RMSE.

	With overlapping measure			**Without overlapping measure**		
	RMSE	**Accuracy (%)**	**Standard deviation (%)**	**RMSE**	**Accuracy (%)**	**Standard deviation (%)**
$\alpha = 0.2$	0.1134	94.67	2.98	0.092	93.33	4.08
$\alpha = 0.4$	0.1251	92.67	4.35	0.1386	91.33	4.47
$\alpha = 0.6$	0.0846	96.67	2.36	0.1272	92.00	3.80
$\alpha = 0.8$	0.0945	96.00	4.35	0.2416	76.67	9.13
$\alpha = 1.0$	0.1037	95.33	2.98	0.3082	52.67	29.48

Table 5.4: Experiment result for Glass in terms of accuracy and RMSE.

	With overlapping measure			Without overlapping measure		
	RMSE	**Accuracy (%)**	**Standard deviation (%)**	**RMSE**	**Accuracy (%)**	**Standard deviation (%)**
$\alpha = 0.2$	0.2252	66.05	5.35	0.2285	65.58	4.47
$\alpha = 0.4$	0.1974	71.16	2.08	0.1995	66.05	4.53
$\alpha = 0.6$	0.2142	70.23	5.04	0.2166	65.11	4.65
$\alpha = 0.8$	0.2185	68.37	6.28	0.1630	67.91	5.55
$\alpha = 1.0$	0.2239	63.26	6.24	0.2326	59.53	9.95

5.5 SUMMARY

In this chapter, a new GrC algorithm with the overlapping measure is proposed. A parameter known as R-value is used to model the overlapping between the granules, and being included in the compatibility equation to allow the granules to overlap during the iterative data granulation. This technique results in the intersection of the final granules by allowing an object to belong to one or more granules rather than simply one.

The relevance of using R-value to measure the granules overlapping is also demonstrated in this chapter by using synthetic dataset. It is shown that the interest to measure the degree of overlapping is for data with different categories, in which the proposed compatibility measure provides additional parameters to model or to 'control' the amount of overlapping among the granules.

The proposed overlapping GrC algorithm is validated by using three datasets from UCI Machine Learning Repository - Iris, Wine and Glass. The proposed algorithm shows a better performance in Iris, and the improvement is more obvious in more complicated datasets (Wine and Glass). From the results, the classification accuracy recorded by the GrC with overlapping measure outperforms the conventional GrC even though with higher density factor α ($\alpha \geq 6$). This also indicated that the overlapping GrC is robust and has potential in developing an FL rule base.

So far, the proposed algorithms related to GrC (Chapter 3 to 5) have been demonstrated to work effectively with Type-1 Fuzzy Logic Systems (T1-FLSs). In the next chapter, the weighted GrC (presented in Chapter 4) is extended to Type-2 Radial Basis Function Neural Network (RBFNN) that is functionally equivalent to Type-2 FLSs. The initial study of interpretability in Type-1 FLS has been presented in Chapter 4; hence the interpretability in Type-2 FLS is the focus of the next chapter. Moreover, all works presented in previous chapters (including this chapter) only make use of the initial parameters obtained from the final information granules. In other words, parameters optimisation is not included in training the FL membership functions parameters. Therefore, the optimisation of MFs parameters is included to maximise the predictive performance of the GrC- Neurofuzzy structure.

CHAPTER 6

INTERPRETABILITY INDICES FOR TYPE-2 FUZZY LOGIC SYSTEMS

This chapter presents the extension of weighted granular computing (W-GrC) into Radial Basis Function Neural Network (RBFNN) and General Type-2 Radial Basis Function Neural Network (GT2-RBFNN). The focus in this chapter is the resulting interpretability in Type-2 Fuzzy Logic systems. Nauck's index (NI) is identified as one the main interpretability measures in the literature; however, most of the applications are in the Type-1 Fuzzy Logic systems. Therefore, in this chapter a new interpretability measure based on NI is proposed and described in detail. The NI is assessed for both W-GrC and conventional GrC in type-1 and type-2 Fuzzy Logic models, and the impact on the interpretability of models elicited using GrC is investigated. Results from classification experiments showed that W-GrC has the potential to improve the predictive accuracy performance while maintaining a good rule-base interpretability.

6.1 INTRODUCTION

One of the motivations for using Fuzzy Logic systems (FLSs) is the implementation of linguistic variables and FL rules that can be easily understood by humans [16]. In general, there are two conflicting criteria in building an FL model: predictive accuracy and interpretability [215]. While accuracy indicates the closeness between the real and the model system [215], interpretability refers to the system's capability to describe the behaviour of the real system in an understandable way [17]. In practice, the interpretability measure of an FLS is always dependent on how its rule base, antecedent and consequent are being understood [65].

However, the task of quantifying interpretability is highly subjective and challenging because it depends on many parameters such as the individual experience, knowledge and preference [94]. From the literature, most researchers agree on using some complexity-based indices to represent interpretability, such as number of rules (NOR), total rule length (TRL) and average rule length (ARL) [94]. However, according to Gacto et al. [67] the interpretability measure is not only limited to complexity-based indices, but also includes semantic-based interpretability; both are observed from rule base and partition level. Examples of complexity-based interpretability index are NOR and number of conditions (rule base level) and number of MFs and features (partition level). For the semantic-based interpretability, the popular indices are rules relevance and consistency (rule base level), coverage and distinguishability (partition level).

Nauck's index (NI) [89] is one of the most popular indexes used to measure the interpretability of FLSs [90], [91], [216]. Comprising three main constituents – complexity, partition and coverage, one advantage of using NI is it covers most of the taxonomy of FLSs' interpretability described above (complexity-based interpretability in rule base level and partition level and semantic-based interpretability in partition level). This completeness of NI offers a thorough evaluation on the interpretability of an FLS.

Even though the interpretability has been studied widely in Type-1 FLSs [19], [86], [87] , the same cannot be said for the Type-2 FLSs. Investigations of interpretability in Type-2 FLSs are scarce; some works focused on minimising the number of rule-bases and rule length as appear in [217] [11], [218]. Implementations of NI in Type-2 FLSs can be seen in some research works for example in [94], [219], [220]. However, all these works only presented a single numerical value of NI, even though a type-2 Fuzzy Logic set is described by two separated membership functions; its upper membership function (UMF) and lower membership function (LMF).

The GrC algorithm has been the subject of numerous studies in the construction of data-driven FL models. Recent GrC studies concentrated on the use of GrC to Type-2 FL systems as presented in [121] where the iterative data granulation was used to estimate the FL rule parameters in the GT2-RBFNN models. The GT2-RBFNN model was developed by using the concept of α-planes, and General Type-2 fuzzy sets were

employed in both antecedent and consequent. In another research [221], Baraka et al. presented long-term learning structure that applied the concepts of GrC to excerpt knowledge from original data in the form of interval-valued sets. Other works related to the implementation of GrC into Type-2 Fuzzy Logic systems can be seen in [13] and [14]. The concept of feature weight in GrC has not yet been attempted to be integrated into the Type-2 Fuzzy Logic systems framework.

The main contributions of this chapter are two-fold; the first is to extend the work in Chapter 4, weighted GrC (W-GrC) into Radial Basis Function Neural Network (RBFNN) and General Type-2 RBFNN. Here, the information granules formed from the W-GrC provide the initial parameters for the hidden layer of both Type-1 and Type-2 models. The RBF-NF system's initial structure is optimised using back-error propagation (BEP) to boost the predictive performance. The second contribution is to propose a new interpretability measure in Type-2 Fuzzy Logic systems. Nauck's index is utilised to describe the interpretability of Type-2 FLSs, taking into account - for the first time - both UMF and LMF.

Therefore, the main objectives of this chapter are:

- To implement the weighted GrC into Radial Basis Function Neural Network (RBFNN) and General Type-2 Radial Basis Function Neural Network (GT2-RBFNN).
- To propose a new interpretability measure in Type-2 Fuzzy Logic systems based on Nauck's index.
- To assess the impact of weighted GrC on the FL rule-bases as compared to the conventional GrC, in terms of the interpretability.

6.2 INTERPRETABILITY INDEX FOR FUZZY LOGIC SYSTEMS

Interpretability and accuracy are frequently incompatible goals when building Fuzzy Logic systems (FLS); one might be improved by forgoing the other, a scenario known as an interpretability-accuracy trade-off. In the context of FLS, interpretability is the quality of a model that enables humans to comprehend a system's behaviour by carefully examining its rule base.

One of the most popular interpretability indices is Nauck's index [19]**.** Nauck created the Nauck's index (NI), a numerical index, to evaluate the interpretability of FL rule-based classification systems [19], [89]. It is calculated by multiplying three interpretability constituents: complexity of FLS ($comp$), average normalised partition index ($\overline{part}$) and average normalised coverage of fuzzy partition ($\overline{cov}$) given by:

$$Nauck\ index = comp\ \times \overline{part}\ \times \overline{cov} \tag{6.1}$$

The first constituent, complexity is given as:

$$comp = g\ /\ \sum_{i=1}^{n} v_i \tag{6.2}$$

Where g is the number of classes, n is the number of rules, and v is the number of input variables used in the i-th rule. This component measures the level of complexity of FLSs. The equation implies that the $comp$ will only be 1 if the classifier consists of only one rule per class, and with only one variable used in each rule. For example, two classifiers with 3 and 5 rules may have the same degree of complexity, if the rules are using only one input variable. Note that the number of partitions (or rules) is not being penalised here (in this component).

As an example, in Iris with 4 input variables, 3 output classes, and assuming a classifier consists of 5 rules (granules) (with all 4 input variables are used in each rule), the $comp$ can be computed as:

$$comp = \frac{3}{5\times4} = 0.15 \tag{6.3}$$

However, the equation (6.1) does not imply that the interpretability index is directly proportional to the complexity of FLSs. Therefore, the term $comp$ is suggested to be changed to $simplicity$, since it is inversely proportional to the number of the number of input variables used.

The second component, average normalised partition index ($\overline{part}$) is given as:

$$\overline{part} = \frac{\sum_{j=1}^{m} part_j}{m} \tag{6.4}$$

Where $part_j = \frac{1}{p_j - 1}$, in which p_j is the number of membership functions in the$j$-th input variable. This component ($\overline{part}$) is used to penalise partitions with a high granularity. The maximum $part_j$ (i.e. $part_j = 1$) is achieved when only two

membership functions (or granules) are used. Variables with $p_j < 2$ is not considered in measuring the interpretability index [89].

For example, if a variable x_j consists of 5 membership functions, the value for $part_j$ for that particular variable is given by:

$$part_j = \frac{1}{5-1} = 0.25 \tag{6.5}$$

The value $\overline{part}$ then is taken as the average of $part_j$ for all input variables.

The component cov stands for coverage degree of the fuzzy partition [19]. A complete coverage means that the membership degrees add up to 1 for each element of the domain [89]. If X_j is the domain of j-th input variable partitioned by p_j MFs $\{\mu_j{}^{(1)}, \dots, \mu_j{}^{(p_j)}\}$ then the average normalised coverage of fuzzy partition ($\overline{cov}$) is:

$$\overline{cov} = \frac{\sum_{j=1}^{m} cov_j}{m} \tag{6.6}$$

Where

$$cov_j = \frac{\int_{X_j} \hat{h}_j(x)dx}{N_j} \tag{6.7}$$

with $N_j = \int_{X_j} dx$ for continuous domains.

$$\hat{h}_j(x) = \begin{cases} h_j(x) & if\ \ 0 < h_j(x) < 1 \\ \frac{p_j - h_j(x)}{p_j - 1} & otherwise \end{cases} \tag{6.8}$$

$$h_j(x) = \sum_{k=1}^{p_j} \mu_j{}^{(k)}(x) \tag{6.9}$$

where $h_j(x)$ is the sum of membership degrees for the j-th input variable.

To exemplify the calculation of the coverage, a set of synthetic membership functions for an input variable (x) with $p_j = 3$ is shown in Figure 6.1. It consists of three Gaussian membership functions associated with the input variable labelled as LOW, MEDIUM and HIGH. The standard deviations are set to be constant ($\sigma = 0.2$) and centres are located at $X_1 = \{0,\ 0.5, 1\}$.

Two points of interest are studied here, which are $x_1 = 0.2$ and $x_1 = 0.6$. Based on equation (6.8), the sum of membership degrees at $x_1 = 0.2$ and $x_1 = 0.6$ ($h_1(0.2)$ and

$h_1(0.6)$) are 0.9312 and 1.0178, respectively. For each element of the domain, the sum of membership degrees is expected to be 1 in order to have complete coverage. With reference to equation (6.8), in the first case, the $\hat{h}_j(x)$ will remain the same, i.e. $\hat{h}_j(x) = h_j(x)$. However, since the $h_j(x)$ in the second case is more than one, the value of $\hat{h}_j(x)$ is calculated as:

$$\hat{h}_j(x) = \frac{3 - 1.0178}{3 - 1}$$

$$= 0.9911 \qquad (6.10)$$

Once $\hat{h}_j(x)$ is obtained for all values of X, the integral of $\hat{h}_j(x)\,dx$ can be computed as the area under the graph. In the case where the data is rescaled to the interval of [0,1], the N_i in this chapter is consistently 1 (area of the domain), as illustrated in Figure 6.1.

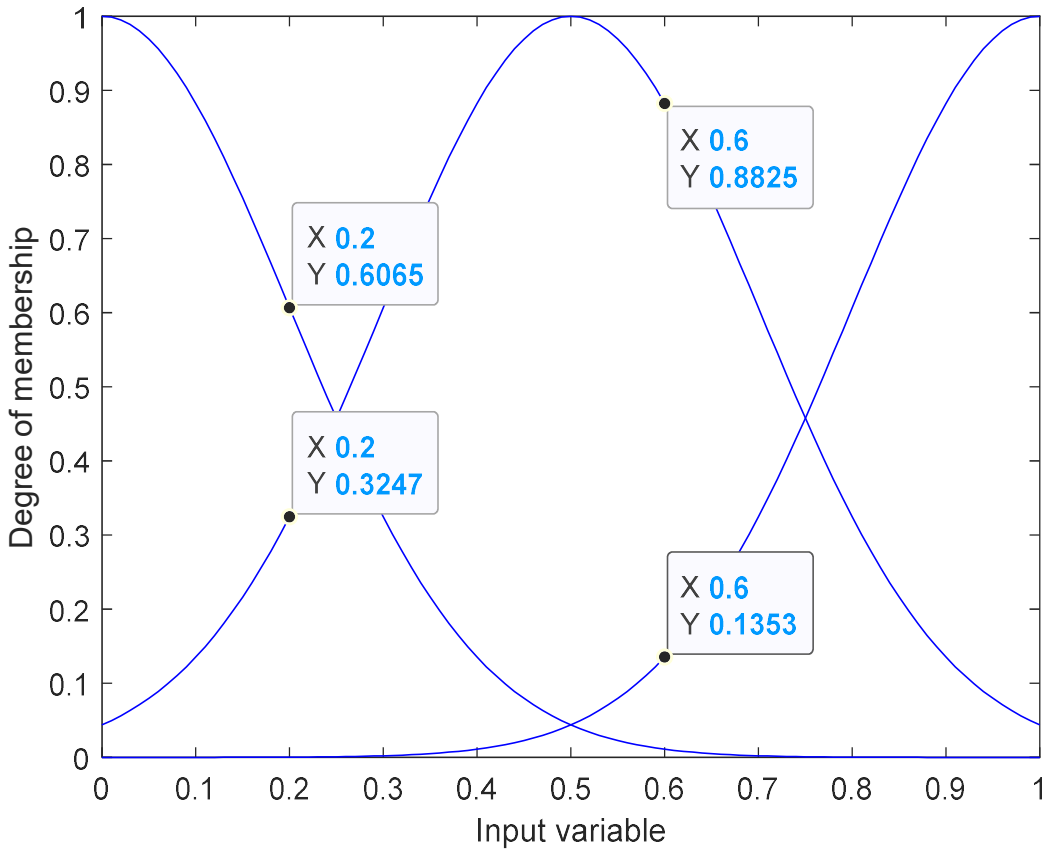


Figure 6.1: Membership functions for an input variable

6.3 RADIAL BASIS FUNCTION NEURAL NETWORK

6.3.1 TYPE-1 RBFNN

Radial basis function neural network (RBFNN), with a good function approximation ability, is a useful tool for modelling nonlinear processes [222]**.** Radial basis function (RBF) networks and Fuzzy Logic inference systems essentially exhibit the same functional behaviour [223], [224]. Figure 6.2 shows the schematic diagram of an RBFNN with n receptive units. The RBF computed by the i-th receptive units is maximum when the input vector is near the centres of that unit [224]**.** Gaussian function is commonly used as the radial basis function in RBF network [225].

RBF network consists of three layers, as shown in Figure 6.2:

1) Input layer - The input layer is the layer to which the inputs apply, and can contain many predictor variables, each of which is linked to a separate independent neuron [226]. The outputs of the input layer are connected directly to the non-linear processing units (neurons) in the hidden layer [227].
2) RBF layer – The RBF layer neurons provide a numerical value by applying their applied inputs to the radial basis function. These sums are transferred to the output layer after being multiplied by linear weights [227]. The number of clusters determines the number of nodes in the RBF layer. In this chapter, the centre of each neuron of RBF layer is obtained from the weighted GrC (W-GrC) algorithm.
3) Output layer - The output layer comprises of a weighted sum of outputs obtained from the RBF layer [228]. Each neuron is associated with weights (z_1, z_2,, . . ., z_p). The RBF layer neuron's output value is multiplied by its weight before being transferred to the summation layer, which adds up the weighted values and displays the sum as the network's output [226].

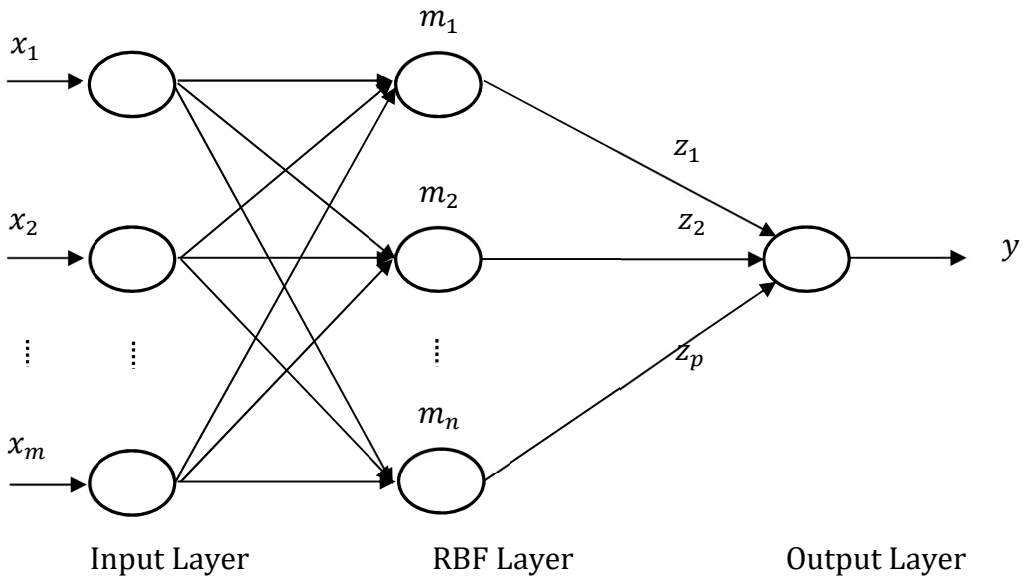


Figure 6.2: RBFNN structure

There are n linguistic rules of the form:

R_i: If $x_1$is $A_{i.1}$ AND x_2 is $A_{i.2}$, AND,…AND x_m is $A_{i.m}$ THEN y is y_i, where $A_{i,j}$ is a Fuzzy Logic set for i-th rule and j-th input variable. The membership value of the input i in j-th rule is defined by Gaussian membership functions:

$$\mu_{ij}(x_j) = \exp\left(\frac{-(x_j - c_{ij})^2}{2\sigma_{ij}{}^2}\right) \tag{6.11}$$

The fuzzification process that takes place in the first layer computes the degree of membership of the inputs for each corresponding fuzzy sets $A_{i,j}$. In the second layer, the output of the k-th is then the firing strength $u_k$of rule k given by:

$$u_k = \mu_{k,1}(x_1)\mu_{k,2}(x_2)\dots\mu_{k,m}(x_m), \quad k = 1, \dots, n \tag{6.12}$$

And can be written as:

$$u_k = \prod_{j=1}^{m} \mu_{k,j}\, x_j \tag{6.13}$$

The rule firing strength is regarded as how closely the current input state complies with each rule in the rule base, with 1 representing perfect compliance [229]. Finally, the output y, which is a crisp value, is calculated as:

$$y = \frac{\sum_{i=1}^{n} u_i z_i}{\sum_{i=1}^{n} u_i} \tag{6.14}$$

The MFs parameters c_{ij} and σ_{ij} (acquired from GrC) are optimised by using the back-error-propagation (BEP) algorithm. The BEP has been proven to be effective in the system optimisation [130].

6.3.2 GENERAL TYPE-2 RBFNN

T2-FLS and T1-FLS have similarity in terms of their linguistic IF…THEN rules. In T2-FLSs, the premise and consequent are of type-2 fuzzy sets (T2-FS). The main difference is that the membership function in T2-FS is a fuzzy set, instead of a crisp number as in T1-FS [51]. When compared to the T1-FS, the Footprint of Uncertainty (FOU) that the T2-FS encompasses gives the T2-FS more degree of freedom, which leads to better modelling of uncertainty [60].

GT2-FS characterised by $\tilde{A}$ is a bivariate membership function $\mu_{\tilde{A}}(x,u) \subseteq [0, 1]$ where the primary variable is $x \in X$. The secondary variable is characterised by $u \in Jx \subseteq [0,1]$ as shown in Figure 6.3. Therefore, $\tilde{A}$ is defined as:

$$\tilde{A} = \{(x,u),\ \mu_{\tilde{A}}\ (x,u) | \forall x \in X, \forall u \in J_x \subseteq [0,1]\ \} \tag{6.15}$$

An α-plane is characterised by $\tilde{A}_\alpha$, is the union of the primary MFs of $\tilde{A}$ with α ($0 \leq \alpha \leq 1$):

$$\tilde{A}_\alpha = \{(x,u),\ \mu_{\tilde{A}}\ (x,u) \geq \alpha | x \in X, u \in [0,1]\ \} \tag{6.16}$$

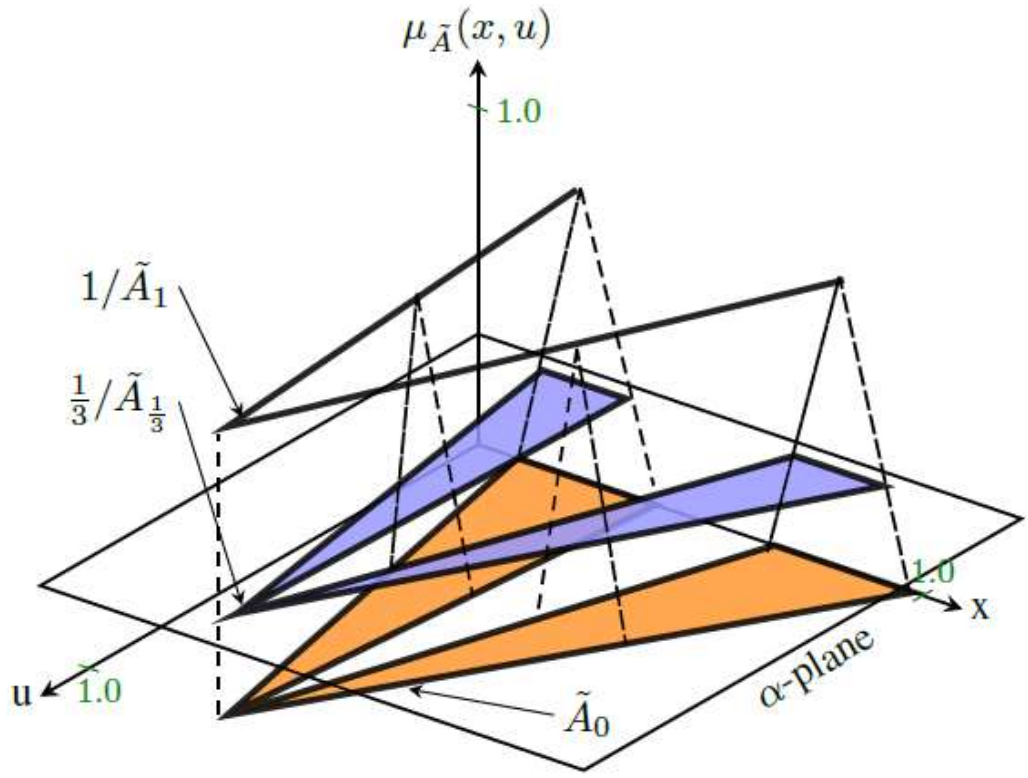


Figure 6.3: α−planes (reproduced from [3]).

Each receptive unit in a Type-1 RBFNN reflects a Fuzzy Logic rule:

$$R^i : \text{IF } x_1 \text{ is } F_1^i \text{ and } \ldots \text{IF } x_v \text{ is } F_v^i \text{ and } \ldots$$

$$\text{IF } x_m \text{ is } F_m^i \text{ THEN } y \text{ is } G^i;\ i = 1, \ldots, N \tag{6.17}$$

The rule is represented by the GT2-FLS and is as follows:

$$\tilde{R}_a^i : \text{IF } x_1 \text{ is } \tilde{F}_1^i \text{ and } \ldots \text{IF } x_v \text{ is } \tilde{F}_v^i \text{ and } \ldots$$

$$\text{IF } x_m \text{ is } \tilde{F}_m^i \text{ THEN } y \text{ is } \tilde{g}^i(\vec{x}_p);\ i = 1, \ldots, N \tag{6.18}$$

where $x_1, \ldots, x_v$ are the input vectors, $\tilde{F}_1^i, \ldots \tilde{F}_v^i$ are the T2-FS, and $i, \ldots, N$ is the number of rules.

Figure 6.4 shows the diagram for General Type-2 RBFNN (GT2-RBFNN). The data are loaded into the GT2-RBFNN as a vector represented by $\vec{x}_p = [x_1, \ldots, x_n]$ with unknown uncertain standard deviation $\sigma_i = [\sigma_i^1, \sigma_i^2]$. Prior to getting the defuzzified

output by averaging the reduced set $[y_l^{\alpha_s}(\vec{x}_p), y_r^{\alpha_s}(\vec{x}_p)]$ the firing interval $F_i^{\alpha_s}$ is utilised to identify the reduced set.

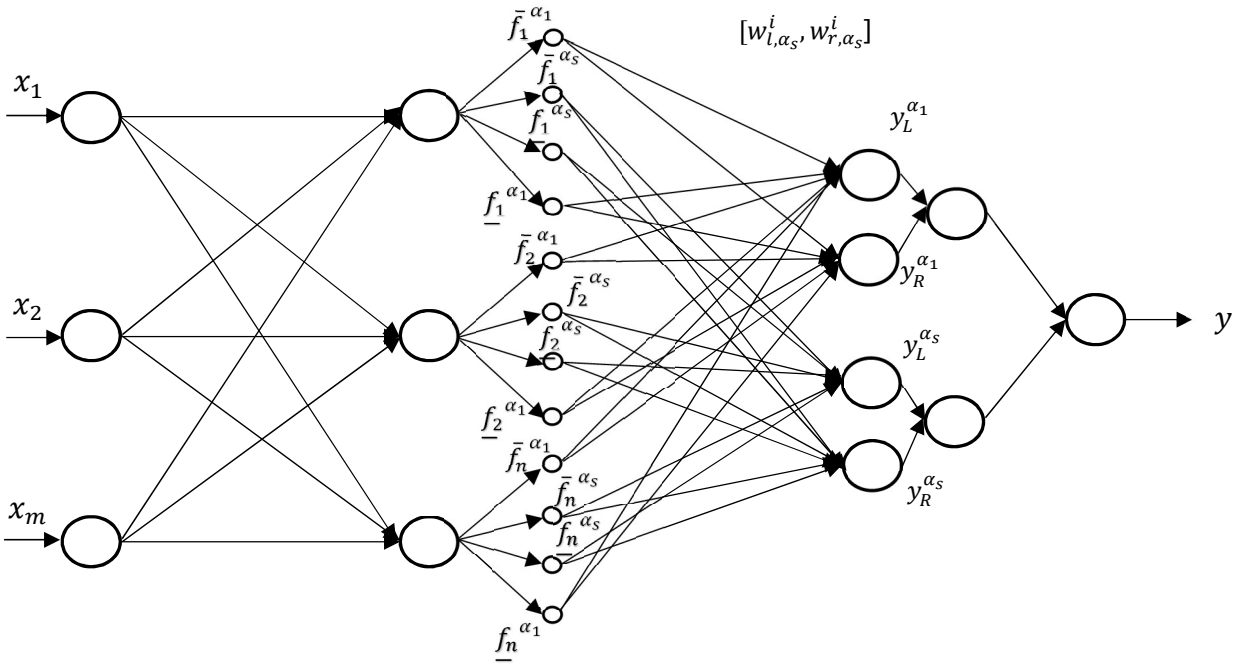


Figure 6.4: General Type-2 RBFNN

The horizontal slice in the General Type-2 RBFNN layer is defined by S firing strengths consisting a set of the lower and upper firing strengths:

$$F_i^{\alpha_s} = \left[\underline{f_i}^{\alpha_s}(\vec{x}_p), \bar{f}_i^{\alpha_s}(\vec{x}_p)\right] \tag{6.19}$$

Where the lower and upper firing strength is given by equation (6.20) and (6.21), respectively:

$$\underline{f_i}^{\alpha_s}(\vec{x}_p) = exp\left[-\sum_{j=1}^{m}\left(\frac{x_j - c_j^i}{\sigma_i^2}\right)^2\right]_{\alpha_s} \tag{6.20}$$

$$\bar{f}_i^{\alpha_s}(\vec{x}_p) = exp\left[-\sum_{j=1}^{m}\left(\frac{x_j - c_j^i}{\sigma_i^1}\right)^2\right]_{\alpha_s} \tag{6.21}$$

Next, the type reduction is implemented to convert the output of the FL inference engine into type-1 fuzzy sets [230]. This step is necessary before obtaining a crisp output through the defuzzification process. In the type reducer layer, the firing strengths are used to calculate the reduced set for α_s level $[y_l^{\alpha_s}(\vec{x}_p), y_r^{\alpha_s}(\vec{x}_p)]$ as in equation (6.22) and (6.23).

$$y_l^{\alpha_s} = \frac{\sum_{i=1}^{L_{\alpha_s}} w_{l,\alpha_s}^i \bar{f}_i^{\alpha_s} + \sum_{L_{\alpha_s}+1}^{N} w_{l,\alpha_s}^i \underline{f}_i^{\alpha_s}}{\sum_{i=1}^{L_{\alpha_s}} \bar{f}_i^{\alpha_s} + \sum_{L_{\alpha_s}+1}^{N} \underline{f}_i^{\alpha_s}} \tag{6.22}$$

$$y_r^{\alpha_s} = \frac{\sum_{i=1}^{R_{\alpha_s}} w_{r,\alpha_s}^i \bar{f}_i^{\alpha_s} + \sum_{R_{\alpha_s}+1}^{N} w_{R,\alpha_s}^i \underline{f}_i^{\alpha_s}}{\sum_{i=1}^{R_{\alpha_s}} \bar{f}_i^{\alpha_s} + \sum_{R_{\alpha_s}+1}^{N} \underline{f}_i^{\alpha_s}} \tag{6.23}$$

where L_{α_s} and R_{α_s} are the switch points [231].

Finally, defuzzification is carried out by the defuzzification layer, which aggregates all horizontal slices. Once $y_l^{\alpha_s}$ and $y_r^{\alpha_s}$ have been determined by using the Enhanced Karnik-Mendel (EKM) [231] type-reduction approach, the type-reduced set can be defuzzified to determine the system's output values (crisp) [11]. The Average of End-Points Defuzzification (AEPD) is employed in this situation [232]:

$$y_p(\vec{x}_p) = \sum_{s=1}^{S} \alpha_s \left[\left(y_l^{\alpha_s}(\vec{x}_p) + y_r^{\alpha_s}(\vec{x}_p)/2 \right) \right] / \sum_{s=1}^{S} \alpha_s \tag{6.24}$$

Numerous techniques based on computational intelligence (CI) or gradient descent (GD) theory can be utilised to optimise model structures. The back-error-propagation (BEP) algorithm is perhaps the most used technique for this modelling structure's GD-based optimisation [9]. The BEP is a gradient-based method that operates in the weighted space of an MSE cost function. Therefore in this chapter, the common parameters c_j^i and $[\sigma_i^1, \sigma_i^2]$ of the antecedent GT2 membership functions, and the weighting factors $[w_{l,\alpha_s}^i, w_{r,\alpha_s}^i]$ at each α-level of a GT2-RBFNN, are optimised by using the BEP algorithm.

6.4 NAUCK INDEX IN TYPE-2 FUZZY LOGIC SYSTEMS

The computation of the interpretability index (Nauck Index) for Type-2 Fuzzy Logic systems is presented here. As described in Section 6.2, the NI consists of three constituents: complexity ($comp$), average partition normalised index ($\overline{part}$) and average normalised coverage of fuzzy partition ($\overline{cov}$). From equation (6.2), the component $comp$ only depends on the number of classes, the number of rules, and the number of input variables. Therefore, the calculation of component $comp$ in Type-2 Fuzzy Logic systems are similar with the Type-1 Fuzzy Logic systems, where the same value of this component is used to obtain the Nauck's index for both upper membership functions (UMF) and lower membership functions (LMF).

Similarly, this situation is also applicable for the second component, $\overline{part}$ which only consists of the number of membership functions (or granules, or rules) and the number of input variables. Hence, the calculation of the partition index is also the same with the Type-1 systems.

The focus of this subsection is therefore to show the computation of the component coverage of fuzzy partition ($\overline{cov}$). It is proposed to have two values of interpretability index, for both UMF and LMF. The computation of coverage index is done on the 2-D domain of $\tilde{A}$, which is called the footprint of uncertainty (FOU) of $\tilde{A}$ or the $\alpha = 0$ plane[233], [234] (as in Interval type-2 Fuzzy sets) such that:

$$FOU(\tilde{A}) = \tilde{A}_0 \tag{6.25}$$

Table 6.1 shows an example of corresponding interval membership degree (or interval value fuzzy sets (IVFS)), LMF and UMF for each domain value X_j. By definition any IVFS is characterised by their LMF and UMF [234] i.e.:

$$\hat{A}(x) = (\bar{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{A}}(x)) \tag{6.26}$$

Table 6.1 is based on the membership functions shown in the Figure 6.5, with x_1 corresponds to $x = 0.1$, x_2 corresponds to $x = 0.2$ and x_{10} corresponds to $x = 1$.

For the i-th fuzzy set $\tilde{A}^i{}_j$ in the input variable X_j, a Gaussian membership function (MF) with a fixed mean c_j^i and an uncertain width that takes on values in $[\sigma_i^1, \sigma_i^2]$ is used [235] (see Figure 6.5).

Table 6.1: Corresponding interval membership degree, LMF and UMF for each domain value X_j

x_i	$\hat{A}(x)$	$\underline{\mu}_{\tilde{A}}(x)$	$\bar{\mu}_{\tilde{A}}(x)$
x_1	[0.03,0.135]	0.03	0.135
x_2	[0.135,0.32]	0.135	0.32
x_3	[0.41, 0.61]	0.41	0.61
x_4	[0.8, 0.88]	0.8	0.88
x_5	[1,1]	1	1
x_6	[0.8, 0.88]	0.8	0.88
x_7	[0.41, 0.61]	0.41	0.61
x_8	[0.135, 0.32]	0.135	0.32
x_9	[0.03, 0.135]	0.03	0.135
x_{10}	[0, 0.04]	0	0.04

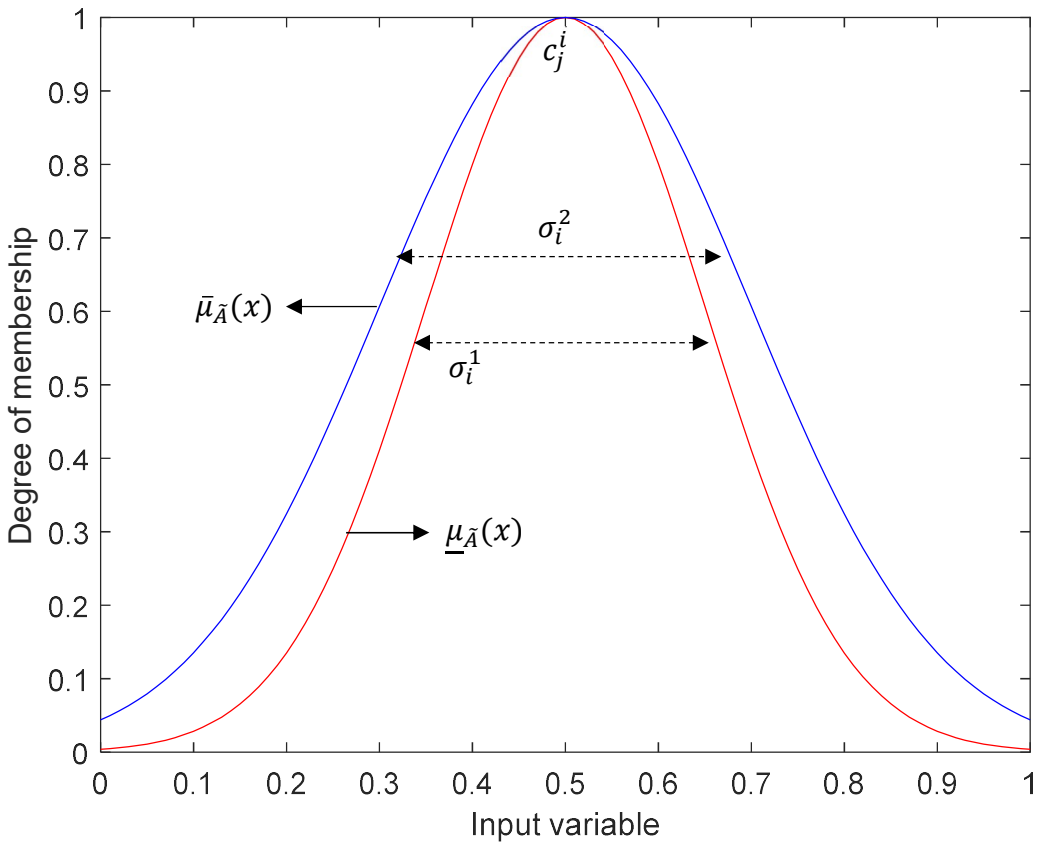


Figure 6.5: Interval type-2 fuzzy set with an uncertain width

To have two values for the Nauck's index (LMF and UMF), the value of $h_j(x)$ (the sum of membership degrees for the j-th input variable) can be calculated by using both upper and lower membership functions. Therefore, the $h_j(x)$ for the LMF can be described as:

$$\underline{h}_j(x) = \sum_{k=1}^{p_j} \underline{\mu}_{\tilde{A}^k{}_j}(x) \tag{6.27}$$

and $h_j(x)$ for the UMF is given by:

$$\overline{h_j}(x) = \sum_{k=1}^{p_j} \bar{\mu}_{\tilde{A}^k{}_j}(x) \tag{6.28}$$

in which p_j is the number of membership functions in the j-th input variable, $\underline{\mu}_{\tilde{A}}$ is the lower membership functions and $\bar{\mu}_{\tilde{A}}$ is the upper membership function.

To demonstrate the calculation of the coverage, a set of synthetic membership functions for an input variable (x) with three membership functions is shown in Figure 6.6 and 6.7. The standard deviations for the LMF and UMF $[\sigma_i^1, \sigma_i^2]$ are set to be 0.15 and 0.2, respectively, with fixed centres at $x = \{0,\ 0.5, 1\}$. Taking $x = 0.2$ and $x = 0.6$ as example, and the new equations (6.27) and (6.28), based on Matlab R2020a the value for coverage parameters are tabulated in Table 6.2.

Table 6.2: Coverage parameters for Type-2 Fuzzy Logic systems for both LMF and UMF

	$x = 0.2$	$x = 0.6$
$\underline{h}_j(x)$	0.5464	0.8293
$\overline{h_j}(x)$	0.9315	1.0178

Next, the rest of calculation is similar with the Nauck's index in Type-1 FL systems, i.e. equation (6.6) and (6.7). As we can see in Table 6.2, the value of $\hat{h}_j(x)$ is equal to all $\underline{h}_j(x)$ and $\overline{h_j}(x)$, except $\overline{h_j}(x)$ when $x = 0.6$. In this case (where $\overline{h_j}(x) > 1$) , it will activate the selection of $\hat{h}_j(x)$ as:

$$\hat{h}_j(x) = \frac{p_j - \overline{h_j}(x)}{p_j - 1} \tag{6.29}$$

which is based on equation (6.6). Repeating these steps for all domain values, the coverage for LMF and UMF for variable x are 0.7512 and 0.9617, respectively.

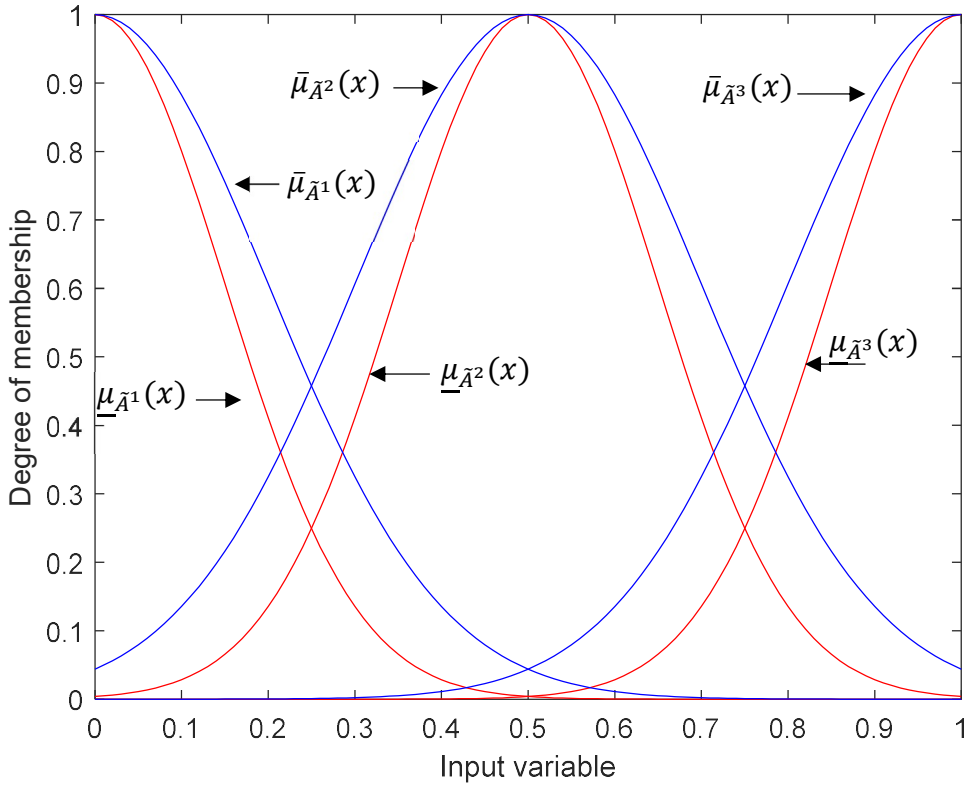


Figure 6.6: Type-2 membership functions

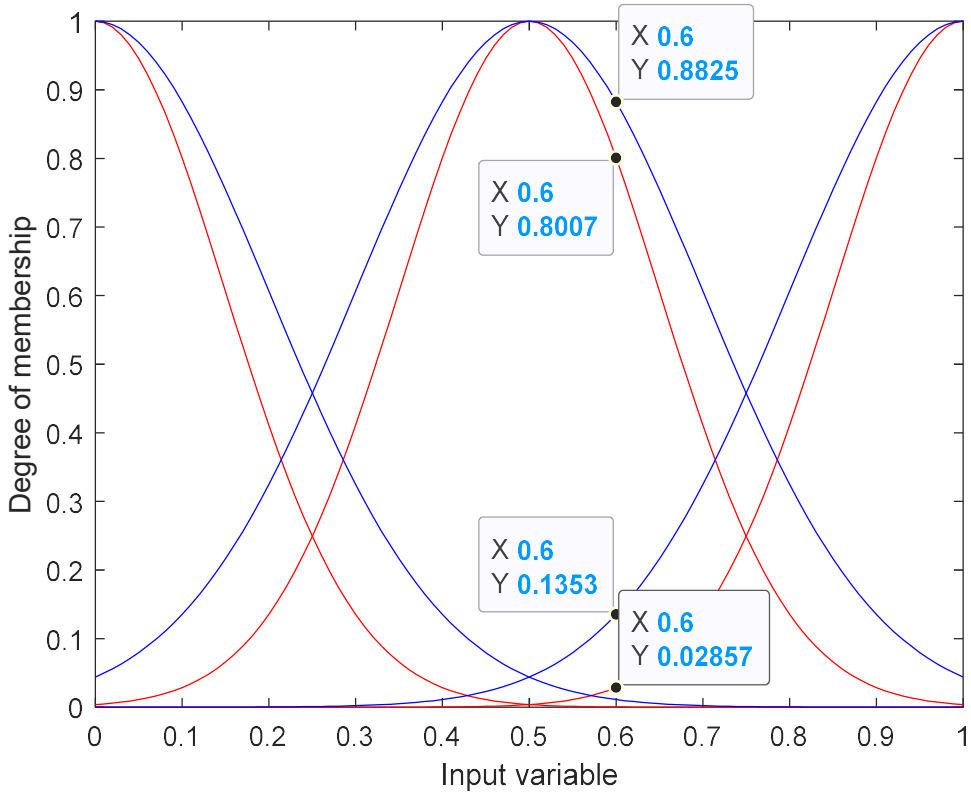


Figure 6.7: Value of membership degrees used in the computation of coverage parameters

6.5 EXPERIMENTAL RESULT

The weighted GrC (W-GrC) algorithm elaborated in Chapter 4 is repeated here with other classifiers: RBFNN and GT2-RBFNN. The analysis presented in this section covers both the system's accuracy and interpretability (with Nauck's index). The proposed method is validated using five classification datasets that vary in terms of size and dimensionality – Iris, Wine, Breast Cancer, Heart and Cardiotocography.

The feature weighting parameter is set to $\beta \geq 1$ (ranging from 2 to 10) in accordance with the recommendations in [181] in order to study how changing the value of β might affect the model's prediction performance. 10 trials are performed for each value of β with five information granules (or rules) being produced, except for Cardiotocography with 3 trials considering the size of the dataset (large number of instances). The system's interpretability (minimum number of rules) and capability to attain good accuracy are all taken into consideration while choosing the five information granules, and for consistency, the number of granules is the same across

all datasets. All feature weights start out at one (equal weighting) and are anticipated to change adaptively as the granulation process progresses. The objective of the simulation is to demonstrate the impact of W-GrC as compared to the standard GrC in terms of model accuracy and interpretability.

6.5.1 System's predictive accuracy

W-GrC's performance with various values of β in RBFNN and GT2-RBFNN is summarised in Tables 6.3 to 6.8. The results are compared to the standard GrC, which is referred to 'GrC' in the tables. RBFNN improves its classification accuracy for the Iris data at $\beta \in \{2,5,6,8,9,10\}$ with the maximum accuracy being 98.33%. Accordingly, W-GrC with GT2-RBFNN surpasses the standard GrC at $\beta \in \{2,3,4,5,7,9,10\}$, with a maximum accuracy of 98%, which happens at $\beta = 2$ and $\beta = 10$.

The significance of W-GrC can be seen in RBFNN at $\beta \in \{2,4,10\}$ in the Wine dataset (13 input features compared to 4 in Iris). Remarkably, the majority of values of in GT2-RBFNN produces accuracy that is greater than that of the standard GrC. It should be observed, nonetheless, that the standard GrC performs somewhat worse in GT2-RBFNN than in RBFNN.

It was clear from the Breast Cancer dataset that greater values of β ($\beta \geq 6$) are necessary for RBFNN to produce successful results. The accuracy is maximum when $\beta = 7$ with 98.16% and while standard GrC scored 97.43%. With the exception of when $\beta = 2$, the W-GrC outperforms the original GrC for GT2-RBFNN, following a nearly identical pattern in Wine.

In Heart, RBFNN performs more consistently than GT2-RBFNN, where W-GrC consistently outperforms the standard GrC in terms of accuracy, with the exception of the case where $\beta = 2$. In contrast, W-GrC with GT2-RBFNN achieves the greater accuracy at $\beta \in \{3,4,5,7,8\}$.

In the Cardiotocography dataset, which is the largest dataset in this study (with 2126 instances and 21 input features), GT2-RBFNN outperforms its counterpart RBFNN. GT2-RBFNN and RBFNN record the maximum predictive accuracies of 88.92% and 87.76%, respectively. In GT2-RBFNN, good impact can be observed for all values of β except when $\beta = 2$. The pattern in RBFNN is different, where better results are achieved at $\beta \in \{2,3,5,7,8,10\}$.

We can infer from the results that W-GrC performs better than the standard GrC. However, to attain higher predictive performance, careful selection of β (which is a hyperparameter for the algorithm) is essential. When a suitable parameter is selected, W-GrC is demonstrated to achieve the best classification accuracies. This is because, when information granules are formed, the more significant input features are given larger weights in each iteration.

Benchmarking this chapter's findings against those of similar works demonstrated that they are comparable to those of other researchers' models, notably those utilising neural networks and support vector machines (SVM). For the Iris data, neural network research [236] and [237] produced accuracies of 98.04% and 97.66%, respectively. These studies also reported accuracy rates of 96.72% and 98.66% for the Wine. The Radial Basis Function Network (RBFN) model created in [238] obtained 96.77% for the Breast Cancer dataset, whereas [239] achieved 98.51%. Although the results from the Heart are not as good as those from other datasets (87.46% in RBFNN and 84% in GT2-RBFNN), they are nonetheless on par with other works, such as [240] with accuracy levels of 86.25% and [241] with accuracy levels of 83%. For the Cardiotocography case study, the benchmarked accuracies are 84.2% [242], 86.38% [243] and 83.12% [244].

Finally, the results are benchmarked against Support Vector Machine (SVM), another widely used technique in machine learning. Iris results achieved in [245] and [246] for instance, recorded the accuracies of 96% and 98%, whereas Wine results presented in [246] and [247] were 98.73% and 100%, respectively. Breast Cancer and Heart case studies also showed comparable performances, as presented in [248] (Breast Cancer – 98.07%), [239] (Breast Cancer – 97.14%), [249] (Heart – 82.9%) and [250] (Heart – 86.89%). Finally, the performance with Cardiotocography is also benchmarked, such as 74.92% in [244], 86.59% in [251] and 84.38% in [243].

Table 6.3: Performance of W-GrC in Type-1 RBFNN with different β values (Iris and Wine)

	Iris			**Wine**		
	RMSE	**Acc. (%)**	**Standard deviation (%)**	**RMSE**	**Acc. (%)**	**Standard deviation (%)**
GrC	0.0763	97.67	2.74	0.0866	98.67	1.72
$\beta = 2.0$	0.0777	98.33	2.36	0.0726	99	1.61
$\beta = 3.0$	0.0909	96	3.06	0.0904	98.33	2.36
$\beta = 4.0$	0.0859	96.33	3.67	0.0714	99.33	1.41
$\beta = 5.0$	0.0732	98	2.33	0.0797	98.67	1.76
$\beta = 6.0$	0.0783	98.33	1.76	0.0868	97.33	5.84
$\beta = 7.0$	0.0735	97.67	1.61	0.0881	98.33	2.36
$\beta = 8.0$	0.0765	98	1.72	0.0854	98.33	3.6
$\beta = 9.0$	0.0785	98.33	1.76	0.0807	98.33	4.22
$\beta = 10.0$	0.0761	98	1.72	0.0768	99.67	1.05

Table 6.4: Performance of W-GrC in Type-1 RBFNN with different β values (Breast Cancer and Heart)

	Breast Cancer			**Heart**		
	RMSE	**Acc. (%)**	**Standard deviation (%)**	**RMSE**	**Acc. (%)**	**Standard deviation (%)**
GrC	0.1486	97.43	1.05	0.3339	84.75	4.99
$\beta = 2.0$	0.1676	96.91	1.95	0.3479	84.41	3.46
$\beta = 3.0$	0.1506	97.35	1.84	0.3336	86.95	2.89
$\beta = 4.0$	0.1628	96.84	1.04	0.3383	86.44	2.65
$\beta = 5.0$	0.151	97.35	1.11	0.3306	85.42	2.42
$\beta = 6.0$	0.154	97.5	1.11	0.3273	87.46	4.39
$\beta = 7.0$	0.1302	98.16	1.16	0.3263	85.08	4.51
$\beta = 8.0$	0.1473	97.5	1.35	0.3198	86.27	3.04
$\beta = 9.0$	0.1352	97.94	1.19	0.3233	85.42	3.12
$\beta = 10.0$	0.1599	97.72	1.07	0.3338	85.08	4.21

Table 6.5: Performance of W-GrC in Type-1 RBFNN with different β values (Cardiotocography)

	Cardiotocography		
	RMSE	**Acc. (%)**	**Standard deviation (%)**
GrC	0.1773	84.63	2.54
$\beta = 2.0$	0.1505	86.12	2.58
$\beta = 3.0$	0.1757	86.2	2.59
$\beta = 4.0$	0.1796	83.69	2.51
$\beta = 5.0$	0.1609	87.76	2.63
$\beta = 6.0$	0.178	84.39	2.53
$\beta = 7.0$	0.1733	85.49	2.56
$\beta = 8.0$	0.175	85.25	2.56
$\beta = 9.0$	0.1897	82.2	2.47
$\beta = 10.0$	0.1678	84.94	2.54

Table 6.6: Performance of W-GrC in General Type-2 RBFNN with different β values (Iris and Wine)

	Iris			**Wine**		
	RMSE	**Acc. (%)**	**Standard deviation (%)**	**RMSE**	**Acc. (%)**	**Standard deviation (%)**
GrC	0.1179	96.33	2.92	0.0971	96.33	3.67
$\beta = 2.0$	0.1101	98	1.72	0.1062	97	1.89
$\beta = 3.0$	0.1131	97.67	2.74	0.1033	97.33	2.11

$\beta = 4.0$	0.1183	96.67	2.72	0.0867	97.67	2.25
$\beta = 5.0$	0.1078	97.67	2.25	0.095	96.33	2.46
$\beta = 6.0$	0.1139	96.33	2.46	0.0853	97.67	2.74
$\beta = 7.0$	0.1094	97.33	2.63	0.088	99	1.61
$\beta = 8.0$	0.1122	95.33	2.81	0.0968	97.33	2.11
$\beta = 9.0$	0.1158	97.33	2.11	0.0933	98.67	1.72
$\beta = 10.0$	0.1101	98	1.72	0.1051	96.33	3.67

Table 6.7: Performance of W-GrC in General Type-2 RBFNN with different β values (Breast Cancer and Heart)

	Breast Cancer			**Heart**		
	RMSE	**Acc. (%)**	**Standard deviation (%)**	**RMSE**	**Acc. (%)**	**Standard deviation (%)**
GrC	0.1741	96.25	2.12	0.3059	81.69	4.06
$\beta = 2.0$	0.1824	96.18	1.58	0.3063	81	5.99
$\beta = 3.0$	0.1663	96.76	1.44	0.3072	83	3.78
$\beta = 4.0$	0.1451	97.21	1.3	0.31	84	2.65
$\beta = 5.0$	0.1685	96.76	1.21	0.3034	83	3.73
$\beta = 6.0$	0.1662	96.62	1.56	0.2997	80	2.75
$\beta = 7.0$	0.1574	97.35	1.39	0.305	81.86	4.66
$\beta = 8.0$	0.1608	96.91	1.69	0.3031	81.86	5.06
$\beta = 9.0$	0.1502	97.65	0.67	0.3009	81.69	5.23
$\beta = 10.0$	0.1654	96.47	1.83	0.31	80.68	4.67

Table 6.8: Performance of W-GrC in General Type-2 RBFNN with different β values (Cardiotocography)

	Cardiotocography		
	RMSE	**Acc. (%)**	**Standard deviation (%)**
GrC	0.1584	87.36	2.62
$\beta = 2.0$	0.1661	86.52	2.59
$\beta = 3.0$	0.1648	87.77	2.63
$\beta = 4.0$	0.1545	87.98	2.67
$\beta = 5.0$	0.1544	88.92	2.63
$\beta = 6.0$	0.159	87.51	2.65
$\beta = 7.0$	0.1561	88.45	2.65
$\beta = 8.0$	0.1585	88.35	2.66
$\beta = 9.0$	0.1584	88.56	2.63
$\beta = 10.0$	0.148	87.67	2.62

6.5.2 Interpretability index

In this work, interpretability and accuracy of the develop systems are given consideration in building the FL models. As presented in the previous section, the proposed W-GrC is shown to improve the RBFNN and GT2-RBFNN models' predictive performance. In this section, the impact on models' interpretability is considered.

Table 6.9 shows the Nauck's index for the W-GrC and the original GrC in both RBFNN and GT2-RBFNN. The NI for a dataset with W-GrC is taken as the average of all trials ($\beta = 2.0$ to $\beta = 10.0$) as in Section 6.5. It is shown that in both models (RBFNN and GT2-RBFNN), Improved predictive accuracy can be achieved with W-GrC without significantly sacrificing the model's interpretability. In all experiments,

the impact on Nauck's index was less than 10%, indicating minimum impact of W-GrC in the interpretability of FL models.

The interpretability in Type-1 RBFNN is consistently higher than its counterpart General Type-2 RBFNN. According to [252], one of the properties which potentially decrease the overall interpretability of T2 systems is due to the lack of agreed mechanism to derive the footprint of uncertainty (FOU). However, by offering a method for creating the footprint of uncertainty's borders that maintains shape coherency throughout the creation of type-2 sets, this can be prevented.

The measured NI is compared with other research that use similar datasets. However, in the literature, the comparison is still limited due to the use of different datasets for example Abalone [253], Thyroid type-2 [254], and Liver Disorder [255] . In [92], using 15 linguistic variables, the NI for Iris with Adaptive Dynamic Clustering Neuro-Fuzzy (ADCNF), Enhanced Neuro-Fuzzy (E23NF) and Transparent Neuro-Fuzzy ranged between 0.036 to 0.053, Breast Cancer (0.0067 to 0.072) and Heart (0.0027 to 0.0653). In [256], the NI for Wine with 49 total rule length (TRL) is 0.00153. This benchmark shows that the measured NI in this chapter are comparable with other related works in literature, hence concluding that the increase in performance (predictive accuracy) did not result in any significant interpretability loss.

Table 6.9: Comparison of the interpretability index

	RBFNN		**General Type-2 RBFNN**			
	W-GrC	**GrC**	**W-GrC**		**GrC**	
			UMF	**LMF**	**UMF**	**LMF**
Iris	0.0252	0.0239	0.0081	0.0132	0.0083	0.0135
Wine	0.0061	0.0065	0.0027	0.0045	0.0027	0.0044
Breast Cancer	0.0065	0.0067	0.0049	0.0052	0.0054	0.0055
Heart	0.0044	0.0044	0.0017	0.0019	0.0018	0.002
Cardiotocography	0.004	0.0039	0.0022	0.0025	0.0025	0.0029

6.6 SUMMARY

In this chapter, the interpretability measure known as Nauck's index (NI) is extended for the first time to Type-2 Fuzzy Logic systems. This is achieved via the computation of lower membership function (LMF) and upper membership functions (UMF). The weighted GrC (W-GrC) is also extended to neuro-fuzzy modelling structures based on Radial Basis Function (RBF). The relevance of the input features is represented by the feature weighting algorithm and these weights have an impact on the granulation as it develops. Based on data from the UCI machine learning repository, the effectiveness of W-GrC is shown in type-1 and type-2 FLSs, where both system's accuracy and interpretability are presented and analysed.

In both Type-1 and Type-2 RBFNN models, the experimental results show that W-GrC performs better against the conventional GrC in terms of the predictive performance when a suitable hyperparameter is chosen. This due to assigning the right weightage for the input variables according to their importance. Interestingly, the feature weights evolve throughout the iterative data granulation process, meaning that the weights are not predetermined or treated as constant value. Instead, it takes into account the current information granules, specifically the within-granule variances in assigning the new adaptive weights for each iteration.

Besides, the proposed interpretability index based on NI is successfully implemented on the Type-2 FL systems. The computation of NI for Type-2 FL systems is shown in detail, taking into account both upper membership function and lower membership function. The assessment of interpretability for the W-GrC shows that there is no major impact on the rulebase interpretability, indicating the potential of W-GrC in achieving high predictive accuracy while maintaining the system's interpretability.

Main results and methodology from this work is included in paper entitled "An evolving feature weighting framework for radial basis function neural network models". It was published in Expert Systems journal in November 2022. The results presented this journal was mainly on the performance of W-GrC in RBFNN and GT2-RBFNN classifiers.

For the future, this work can be extended to another quadrant of interpretability – semantics at the fuzzy partition. The Nauck's index used in this work covers another three quadrants of interpretability, which are complexity at the rule base level, complexity at the fuzzy partition level and semantics at the fuzzy partition level (based on the three components in the Nauck's index). Therefore, criteria in semantics at the rule-base level such as consistency, relevance and co-firing rules has the potential to be investigated and to be used as an index in measuring the interpretability of T2-FLS.

CHAPTER 7

CONCLUSIONS AND FUTURE RESEARCH

This chapter provides the conclusions of this thesis and gives recommendations for future research.

7.1 CONCLUSIONS

One of the advantages of using FL modelling is its capability in representing the knowledge in the form of if-then rules that imitates the human way of thinking. This can be achieved by using the idea of granular computing (GrC), for example using the iterative data granulation in extracting the information from the data, and using this information in building the FL rule-bases. The aims of this research work are to investigate methods to address uncertainty in GrC, methods to address variable importance in GrC, methods to address granular overlapping in GrC, and methods to evaluate the resulting impact on the interpretability. Several new computational frameworks are developed in this thesis to address a number of challenges in system interpretability, i.e. uncertainty, feature weighting and overlapping.

The issue of interpretability due to the data uncertainty is tackled with the aid of information theory. Shannon entropy is used to quantify the uncertainty during data granulation (as a conflict) between information granules due to similar process conditions with different outcomes. The entropy expresses the reluctance of two granules that are about to be combined. It therefore directs the granulation process to combine the granules with the least amount of uncertainty, hence yielding better representation of the systems through FL rule-bases. The GrC with entropy is demonstrated to perform better than the standard GrC and other clustering algorithms, obtaining the highest accuracies in Iris, Wine and Glass. Penalising the uncertainty in

the merging process promotes higher quality information granules, and therefore, FL rule-bases formed are more distinguishable and provide better predictions.

In an attempt to organise data collectively, one would naturally exclude, or assign less weights for the features that are less important to the task. In this thesis, a new iterative data granulation technique with evolving feature weighting, named as weighted GrC (W-GrC) is introduced in order to describe the significance of input features and utilise these weights to guide the information granulation process. Rather than assigning weight at the beginning of the granulation process and using constant weights throughout the iterations, the feature weights are calculated with reference to the variances in the current information granules and adaptively change in each iteration. With this embedded feature-weighting technique, the iterative data granulation is allowed to prioritise the more relevant input variables to have more influence in the merging process. Within Type-1 FLS framework, W-GrC is tested with datasets with various dimensionality or number of input variables (Iris – 4, Wine – 13 and Glass – 10), and the results show that W-GrC outperforms the standard GrC in terms of RMSE and predictive accuracy with appropriate choice of β. Furthermore, the final rule-bases developed with W-GrC show good distinguishability as compared with the final rule-bases developed with the conventional GrC. Improved predictive accuracies are also observed when W-GrC is applied in Type-1 and Type-2 RBFNN, in which the experiments are expanded to Breast Cancer, Heart and Cardiotocography datasets. While it has been demonstrated that W-GrC can produce greater accuracy, no significant degradation in the model's interpretability is observed (using Nauck's index), with impact less than 10% for all datasets.

Models' transparency and rules consistency may be affected due to the presence of granular overlapping. Therefore, a new iterative data granulation algorithm is developed to control the amount of overlapping among the granules. The new compatibility function incorporated the R-value, a metric that represents the ratio of overlapping areas among categories in a data cluster. By enabling an object to belong to one or more granules rather than just one, this strategy leads to the overlapping of the final granules. This enables the new GrC algorithm to control the degree of overlapping in the information granules, and provides capability in dealing with datasets with high overlapping. It is demonstrated that GrC with overlap measure performs better in Iris, and the benefit is more pronounced in more challenging datasets

(Wine and Glass). According to the findings, although having a greater density factor, the GrC with overlapping measure surpasses the traditional GrC in terms of classification accuracy. This demonstrates the strength of the overlapping GrC and its potential for creating FL rule-bases. This method convinces the compatibility search to limit the degree of granular overlapping, which would lessen model transparency and jeopardise the consistency of the rules.

Most of the applications of Nauck's index as the interpretability measure are within the Type-1 Fuzzy Logic systems. Therefore, a new formulation for interpretability index in Type-2 Fuzzy Logic systems is developed for the first time based on the Nauck's index. This is achieved via the computation of lower membership function (LMF) and upper membership functions (UMF). The proposed interpretability index based on Nauck's Index is successfully implemented on the Type-1 and Type-2 FL systems in which the weighted GrC (W-GrC) is extended to neuro-fuzzy modelling structures based on Radial Basis Function (RBF). The effectiveness of W-GrC is demonstrated in type-1 and type-2 FLSs using data from the UCI machine learning repository (Iris, Wine, Heart, Breast Cancer and Cardiotocography) where both the system's accuracy and interpretability are demonstrated and analysed. Indicating the possibility of W-GrC to achieve high predictive accuracy, the examination of interpretability for the W-GrC reveals that there is no significant impact on the rule-base interpretability.

7.2 RESEARCH LIMITATIONS AND FUTURE RESEARCH

Despite the promising new results and knowledge gained from the proposed frameworks in this thesis, further study is needed to create a transparent and interpretable method based on Fuzzy Logic and RBFNN that strikes a fair balance between accuracy and generalisation. Quantification of uncertainty is one way to achieve this, in which Information Theory is applied to numerically quantify the uncertainty that occurs during the data granulation level (captured via conflict). The computation of uncertainty in this thesis, however, makes use of histogram method to determine the probability, and hence the Shannon entropy. The size of the bins used to categorise the data may have an impact on the histogram's structure and patterns. The computation of Shannon entropy could be challenging since it involves the unknown probability density function. To tackle this, kernel density estimator could be a

prominent way to provide the approximation on the probability density. Besides, the study of uncertainty quantification can be extended to other types of uncertainty, such as nonspecificity and fuzziness.

One direction for future work in the W-GrC would be to combine the feature weighting technique with a feature selection /pruning algorithm. This would suggest a two-step approach so that the most relevant input features are selected first, and the weights are then readjusted accordingly. Even though maintaining the whole input features may have some advantages (i.e. preserving the information), this approach could impact the system's interpretability due to the high number of features. Rather than maintaining the whole input features (as demonstrated in this thesis), the input features can be reduced via a feature selection process. Therefore, this combination will promote higher simplicity (hence interpretability) of the system due to the feature reduction task.

Granular overlapping can have an impact on a model's transparency and rule consistency. In this research, the R-value is easily computed because the labels are known and available. However, it would not be straightforward in the case of unsupervised learning, where the data labels are unknown. In this case, one might have some partial knowledge of the clustering, like in the setting of semi-supervised clustering, which could be used to extend the presented framework in this thesis.

The interpretability studies in this chapter covered three quadrants of interpretability - Complexity at the rule base level, Complexity at the fuzzy partition level and Semantics at the fuzzy partition level. For future work, another quadrant of interpretability Semantics at the rule base level can be included, which covers criteria such as consistency, relevance and co-firing rules.

REFERENCES

[1] G. Panoutsos, "Linguistic Granules-Oriented Neural-Fuzzy Modelling for Knowledge Elicitation and Optimisation." PhD Thesis, The University of Sheffield, August 2007.

[2] G. Panoutsos, M. Mahfouf, "Granular Computing and Evolutionary Fuzzy Modelling for Mechanical Properties of Alloy Steels, *IFAC Proceedings Volumes*, vol. 38, no. 1, pp. 205-210, 2005.

[3] W. Pedrycz, "Granular computing: An introduction," *Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS*, vol. 3, pp. 1349–1354, 2001. [Online]. Available: 10.1109/nafips.2001.943745.

[4] S. Sakinah S. Ahmad, "Fuzzy Modeling through Granular Computing." PhD Thesis, Department of Electrical and Computer Engineering, University of Alberta, 2012.

[5] E. de la Rosa and W. Yu, "Data-Driven Fuzzy Modeling Using Deep Learning," ArXiv, abs/1702.07076, Feb. 2017. [Online]. Available: 10.1109/TSMC.2018.2812156.

[6] W. Pedrycz and A. Bargiela, "Granular clustering: A granular signature of data," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 32, no. 2. pp. 212–224, Apr. 2002. [Online]. Available: 10.1109/3477.990878.

[7] M. T. Lwin and M. M. Aye, "A Modified Hierarchical Agglomerative Approach for Efficient Document Clustering System," *American Scientific Research Journal for Engineering, Technology, and Sciences*, vol. 29, pp. 228-238, 2017.

[8] G. Panoutsos and M. Mahfouf, "Discovering Knowledge and Modelling Systems using Granular Computing and Neurofuzzy Structures," 2005.

[9] G. Panoutsos and M. Mahfouf, "A neural-fuzzy modelling framework based on granular computing: Concepts and applications," *Fuzzy Sets and Systems,* vol. 161, no. 21, pp. 2808–2830, Nov. 2010. [Online]. Available: 10.1016/j.fss.2010.06.004.

[10] A. Rubio-Solis, T. Salgado-Jimenez, L. G. Garcia-Valdovinos, L. Nava-Balanzar, R. A. Hernandez-Hernandez, and U. Martinez-Hernandez, "An evolutionary

general type-2 fuzzy neural network applied to trajectory planning in remotely operated underwater vehicles," in *IEEE International Conference on Fuzzy Systems*, Jul. 2020. [Online]. Available: 10.1109/FUZZ48607.2020.9177708.

[11] A. Baraka and G. Panoutsos, "Long-term learning for type-2 neural-fuzzy systems," Fuzzy Sets and Systems, vol. 368, pp. 59–81, Aug. 2019. [Online]. Available: 10.1016/j.fss.2018.12.014.

[12] A. Rubio-Solis and G. Panoutsos, "An ensemble data-driven fuzzy network for laser welding quality prediction," in *IEEE International Conference on Fuzzy Systems,* Aug. 2017. [Online]. Available: 10.1109/FUZZ-IEEE.2017.8015496.

[13] S. Khanmohammadi, N. Adibeig, and S. Shanehbandy, "An improved overlapping k-means clustering method for medical applications," *Expert Systems with Applications,* vol. 67, pp. 12–18, Jan. 2017. [Online]. Available: 10.1016/j.eswa.2016.09.025.

[14] A. E. Danganan and E. de Los Reyes, "Ehmcoke: An enhanced overlapping clustering algorithm for data analysis," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4, pp. 2212–2222, Aug. 2021. [Online]. Available: 10.11591/EEI.V10I4.2547.

[15] J. J. Whang, Y. Hou, D. F. Gleich and I. S. Dhillon, "Non-Exhaustive, Overlapping Clustering," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 11, pp. 2644-2659, 1 Nov. 2019. [Online]. Available: 10.1109/TPAMI.2018.2863278.

[16] T. R. Razak, J. M. Garibaldi, C. Wagner, A. Pourabdollah and D. Soria, "Interpretability and Complexity of Design in the Creation of Fuzzy Logic Systems — A User Study," *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 420-426, 2018. [Online]. Available: 10.1109/SSCI.2018.8628924.

[17] J. Casillas, P. Martínez, and A. D. Benítez, "Learning consistent, complete and compact sets of fuzzy rules in conjunctive normal form for regression problems," *Soft computing*, vol. 13, no. 5, pp. 451–465, 2009. [Online]. Available: 10.1007/s00500-008-0361-5.

[18] A. K. Shukla and A. Yadav, “Interpretability Assessment in Fuzzy Rule Based Systems,” *International Journal of Scientific & Engineering Research,* vol. 5, no. 7, 2014, [Online]. Available: http://www.ijser.org

[19] T. R. Razak, J. M. Garibaldi, C. Wagner, A. Pourabdollah, and D. Soria, “Interpretability indices for hierarchical fuzzy systems,” in *IEEE International Conference on Fuzzy Systems,* Aug. 2017. [Online]. Available: 10.1109/FUZZ-IEEE.2017.8015616.

[20] N. Amaitik, “The Basics of Fuzzy Systems Technology: A Complete Tutorial,” 2020. [Online]. Available: https://www.researchgate.net/publication/345503118

[21] J. Kerr-Wilson, “Fuzzy Rule-Based Systems: Design, Analysis, and Applications,” PhD Thesis, Department of Electrical and Computer Engineering, University of Alberta, 2019.

[22] A. Mustafa, “Soft-Computing and Human-Centric Approaches for Modelling Complex Manufacturing Systems,” PhD Thesis, Department of Automatic Control and Systems Engineering, The University of Sheffield, 2017.

[23] J. M. Belman-Flores, D. A. Rodríguez-Valderrama, S. Ledesma, J. J. García-Pabón, D. Hernández, and D. M. Pardo-Cely, “A Review on Applications of Fuzzy Logic Control for Refrigeration Systems,” *Applied Sciences*, vol. 12, no. 3, 2022. [Online]. Available: 10.3390/app12031302.

[24] M. O. Okwu and A. N. Nwachukwu, “A review of fuzzy logic applications in petroleum exploration, production and distribution operations,” *Journal of Petroleum Exploration and Production Technology*, vol. 9, no. 2. Springer Verlag, pp. 1555–1568, Jun. 01, 2019. [Online]. Available: 10.1007/s13202-018-0560-2.

[25] J. Y. Kim, “Introduction to Fuzzy Set Theory and The Hyperplane Separation Theorem of Fuzzy Convex Sets,” 2019.

[26] S. Kaur, “"An Overview of Fuzzy Sets", *International Journal of Emerging Technologies and Innovative Research,* vol.5, no. 7, pp. 404-410, July-2018, [Online]. Available: http://www.jetir.org/papers/JETIR1807065.pdf.

[27] L. A. Zadeh, “Fuzzy sets,” *Information and Control*, vol 8, pp. 338-353, 1965.

[28] A. Setiawan, E. R. Arumi, and P. Sukmasetya, “Fuzzy Membership Functions Analysis for Usability Evaluation of Online Credit Hour Form,” *Journal of Engineering Science and Technology,* vol. 15, pp. 3189-3203, 2020.

[29] A. V. Raj and G. Sathiyapriya, “Intuitionistic Bell Shape Fuzzy Number,” *International Journal of Applied Engineering Research,* vol. 14, no. 3, pp. 79-82, 2019.

[30] L. Wang, “A Course in Fuzzy Systems and Control,” Prentice-Hall, Englewood Cliffs, 1997.

[31] L. Iliadis, S. Skopianos, S. Tachos, and S. Spartalis, “A Fuzzy Inference System Using Gaussian Distribution Curves for Forest Fire Risk Estimation,” *IFIP Advances in Information and Communication Technology,* vol 339, 2010. [Online]. Available: https://doi.org/10.1007/978-3-642-16239-8_49.

[32] P. L. Gentili, “Implementing Fuzzy Sets and Processing Fuzzy Logic Information by Molecules,” *Proceedings 2022,* vol. 81, no. 94, pp. 1-5, 2022. [Online]. Available: https://doi.org/10.3390/proceedings2022081094

[33] R. Amini and S. C. Ng, “Comparison of Artificial Neural Network, Fuzzy Logic and Adaptive Neuro-Fuzzy Inference System on Air Pollution Prediction,” *Journal of Engineering Technological Advances,* Vol. 2, No. 1, pp. 14-22, 2017.

[34] G. J. Klir and Bo. Yuan, Fuzzy sets and fuzzy logic : theory and applications. Prentice Hall, 1995.

[35] J. M. Mendel, "Fuzzy logic systems for engineering: a tutorial," in *Proceedings of the IEEE,* vol. 83, no. 3, pp. 345-377, March 1995. [Online]. Available: 10.1109/5.364485.

[36] E. H. Mamdani and S. Assilian, “An experiment in linguistic synthesis with a fuzzy logic controller,” *International Journal of Man-Machine Studies,* vol. 7, issue 1, pp. 1-13, 1975. [Online]. Available: https://doi.org/10.1016/S0020-7373(75)80002-2.

[37] M. Sugeno and G. T. Kang, “Structure identification of fuzzy model,” *Fuzzy Sets and Systems,* vol. 28, pp. 15-33, 1988.

[38] E. Vlamou and B. Papadopoulos, “Fuzzy logic systems and medical applications,” *AIMS Neuroscience*, vol. 6, no. 4. AIMS Press, pp. 266–272, 2019. [Online]. Available: 10.3934/Neuroscience.2019.4.266.

[39] E. Ontiveros, P. Melin, and O. Castillo, “Designing hybrid classifiers based on general type-2 fuzzy logic and support vector machines,” *Soft Computing,* vol. 24, pp. 18009–18019, 2020. [Online]. Available: https://doi.org/10.1007/s00500-020-05052-x

[40] H. Torres-Salinas, J. Rodríguez-Reséndiz, E. E. Cruz-Miguel, and L. A. Ángeles-Hurtado, “Fuzzy Logic and Genetic-Based Algorithm for a Servo Control System,” *Micromachines,* vol. 13, no. 4, Apr. 2022. [Online]. Available: 10.3390/mi13040586.

[41] H. Fattahi, A. Agah, and N. Soleimanpourmoghadam, “Multi-Output Adaptive Neuro-Fuzzy Inference System for Prediction of Dissolved Metal Levels in Acid Rock Drainage: a Case Study,” *Journal of AI and Data Mining,* vol. 6, pp. 121-132, 2018.

[42] Z. J. Viharos and K. B. Kis, “Survey on Neuro-Fuzzy Systems and their Applications in Technical Diagnostics and Measurement,” *Measurement,* vol. 67, 2015. [Online]. Available: 10.1016/j.measurement.2015.02.001.

[43] R. Jafari, S. Razvarz, A. Gegov, and B. Vatchova, “A Survey on Applications of Neuro-Fuzzy Models,” in *2020 IEEE 10th International Conference on Intelligent Systems*, *IS 2020 - Proceedings,* pp. 148–152, 2020. [Online]. Available: 10.1109/IS48319.2020.9200185.

[44] K. Siminski, “FuBiNFS – fuzzy biclustering neuro-fuzzy system,” *Fuzzy Sets and Systems,* vol. 438, pp. 84–106, Jun. 2022. [Online]. Available: 10.1016/j.fss.2021.07.009.

[45] A. Y. Egwoh, E. A. Onibere, and P. O. Odion, “Application of Neuro-fuzzy System : A Literature Review,” *IJCSNS International Journal of Computer Science and Network Security*, vol. 18, no.12, pp. 1-6, December 2018.

[46] J. R. Castro, M. A. Sanchez, C. I. Gonzalez, P. Melin, and O. Castillo, “A New Method for Parameterization of General Type-2 Fuzzy Sets,” *Fuzzy Information*

and Engineering, vol. 10, no. 1, pp. 31–57, Jan. 2018. [Online]. Available: 10.1080/16168658.2018.1509519.

[47] J. T. Rickard, J. Aisbett, and G. Gibbon, “Fuzzy subsethood for fuzzy sets of type-2 and generalized type-n,” *IEEE Transactions on Fuzzy Systems,* vol. 17, no. 1, pp. 50–60, 2009. [Online]. Available: 10.1109/TFUZZ.2008.2006369.

[48] K. Mittal, A. Jain, K. S. Vaisla, O. Castillo, and J. Kacprzyk, “A comprehensive review on type 2 fuzzy logic applications: Past, present and future,” *Engineering Applications of Artificial Intelligence,* vol. 95, Oct. 2020. [Online]. Available: 10.1016/j.engappai.2020.103916.

[49] J. M. Mendel, R. I. 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. [Online]. Available: 10.1109/TFUZZ.2006.879986.

[50] J. M. Mendel, "Type-2 fuzzy sets and systems: an overview," in *IEEE Computational Intelligence Magazine*, vol. 2, no. 1, pp. 20-29, February 2007. [Online]. Available: 10.1109/MCI.2007.380672.

[51] A. Baraka, G. Panoutsos, and S. Cater, “Perpetual Learning Framework based on Type-2 Fuzzy Logic System for a Complex Manufacturing Process,” in *IFAC-PapersOnLine,* vol. 49, no. 20, pp. 143–148, 2016. [Online]. Available: 10.1016/j.ifacol.2016.10.111.

[52] P. Chandra, D. Agarwal, and P. K. Shukla, “Improved Fuzzy Modeling of Thyroid Disease Detection using Interval Type 2 Fuzzy Techniques,” *International Journal of Engineering and Advanced Technology*, vol. 9, no. 6, pp. 220–223, Aug. 2020. [Online]. Available: 10.35940/ijeat.C5931.089620.

[53] S. Qin, C. Zhang, T. Zhao, W. Tong, Q. Bao, and Y. Mao, “Dynamic High-Type Interval Type-2 Fuzzy Logic Control for Photoelectric Tracking System,” *Processes*, vol. 10, no. 3, Mar. 2022. [Online]. Available: 10.3390/pr10030562.

[54] Q. F. Liao and D. Sun, “Interaction Measures for Control Configuration Selection Based on Interval Type-2 Takagi-Sugeno Fuzzy Model,” *IEEE Transactions on Fuzzy Systems,* vol. 26, no. 5, pp. 2510–2523, Oct. 2018. [Online]. Available: 10.1109/TFUZZ.2018.2791929.

[55] A. Sakalli, T. Kumbasar, and J. M. Mendel, “Towards Systematic Design of General Type-2 Fuzzy Logic Controllers: Analysis, Interpretation, and Tuning,” *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 2, pp. 226–239, Feb. 2021. [Online]. Available: 10.1109/TFUZZ.2020.3016034.

[56] M. Almaraashi, R. John, A. Hopgood, and S. Ahmadi, “Learning of Interval and General Type-2 Fuzzy Logic Systems using Simulated Annealing: Theory and Practice,” *Information Sciences,* vol. 360, 2016. [Online]. Available: 10.1016/j.ins.2016.03.047.

[57] C. Wagner and H. Hagras, “Toward general type-2 fuzzy logic systems based on zSlices,” *IEEE Transactions on Fuzzy Systems,* vol. 18, no. 4, pp. 637–660, Aug. 2010. [Online]. Available: 10.1109/TFUZZ.2010.2045386.

[58] J. M. Mendel, F. Liu, and D. Zhai, “α-Plane representation for type-2 fuzzy sets: Theory and applications,” *IEEE Transactions on Fuzzy Systems,* vol. 17, no. 5. pp. 1189–1207, 2009. [Online]. Available: 10.1109/TFUZZ.2009.2024411.

[59] H. Hamrawi, S. Coupland, and R. John, “A novel alpha-cut representation for type-2 fuzzy sets,” in *2010 IEEE World Congress on Computational Intelligence*, WCCI 2010, 2010. [Online]. Available: 10.1109/FUZZY.2010.5584783.

[60] C. Wagner and H. Hagras, “Toward general type-2 fuzzy logic systems based on zSlices,” *IEEE Transactions on Fuzzy Systems,* vol. 18, no. 4, pp. 637–660, Aug. 2010. [Online]. Available: 10.1109/TFUZZ.2010.2045386.

[61] J. M. Mendel, “General type-2 fuzzy logic systems made simple: A tutorial,” *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 5, pp. 1162–1182, Oct. 2014. [Online]. Available: 10.1109/TFUZZ.2013.2286414.

[62] W. Pedrycz, A. Gacek, and X. Wang, “A Hierarchical Approach to Interpretability of TS Rule-Based Models,” *IEEE Transactions on Fuzzy Systems,* vol. 30, no. 8, pp. 2861–2869, Aug. 2022. [Online]. Available: 10.1109/TFUZZ.2021.3094661.

[63] C. Castiello, A. M. Fanelli, M. Lucarelli, and C. Mencar, “Interpretable fuzzy partitioning of classified data with variable granularity,” *Applied Soft Computing*

Journal, vol. 74, pp. 567–582, Jan. 2019. [Online]. Available: 10.1016/j.asoc.2018.10.040.

[64] J. M. Alonso, L. Magdalena, and G. González-Rodríguez, "Looking for a good fuzzy system interpretability index: An experimental approach," *International Journal of Approximate Reasoning,* vol. 51, no. 1, pp. 115–134, Jan. 2009. [Online]. Available: 10.1016/j.ijar.2009.09.004.

[65] D. Pekaslan, C. Chen, C. Wagner, and J. M. Garibaldi, "Performance and Interpretability in Fuzzy Logic Systems – Can We Have Both?" *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pp. 571-584, 2020. [Online]. Available: 10.1007/978-3-030-50146-4_42.

[66] P. Carmona and J. L. Castro, "FuzzyFeatureRank. Bringing order into fuzzy classifiers through fuzzy expressions," Fuzzy Sets and Systems, vol. 401, pp. 78–90, Dec. 2020. [Online]. Available: 10.1016/j.fss.2020.03.003.

[67] M. J. Gacto, R. Alcalá, and F. Herrera, "Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures," Information Sciences, vol. 181, no. 20, pp. 4340–4360, Oct. 2011. [Online]. Available: 10.1016/j.ins.2011.02.021.

[68] V. Margot and G. Luta, "A New Method to Compare the Interpretability of Rule-Based Algorithms," *AI 2021,* vol. 2, pp. 621-635, 2021. [Online]. Available: https://doi.org/10.3390/ai2040037.

[69] C. Messien, "Optimization of Fuzzy Systems by Evolutionary Algorithms."

[70] J. Hao, S. Luo, and L. Pan, "Rule extraction from biased random forest and fuzzy support vector machine for early diagnosis of diabetes," *Sci Rep*, vol. 12, no. 1, Dec. 2022. [Online]. Available: 10.1038/s41598-022-14143-8.

[71] H. Jiang, R.-C. Chen, Q.-E. Liu, and S.-W. Huang, "Fuzzy Rules Reduction Based on Sparse Coding," *International Journal of Applied Science and Engineering,* vol. 16, p. 215, 2019. [Online]. Available: 10.6703/IJASE.201911_16(3).215.

[72] D. Hein, S. Udluft, and T. A. Runkler, "Generating interpretable fuzzy controllers using particle swarm optimization and genetic programming," in *GECCO 2018*

Companion - Proceedings of the 2018 Genetic and Evolutionary Computation Conference Companion, pp. 1268–1275, 2018. [Online]. Available: 10.1145/3205651.3208277.

[73] P. Hajek, "Interpretable Fuzzy Rule-Based Systems for Detecting Financial Statement Fraud. In: MacIntyre, J., Maglogiannis, I., Iliadis, L., Pimenidis, E. (eds) *Artificial Intelligence Applications and Innovations. AIAI 2019. IFIP Advances in Information and Communication Technology,* vol 559, 2019. [Online]. Available: https://doi.org/10.1007/978-3-030-19823-7_36.

[74] F. Cheong and R. Lai, "Constraining the optimization of a fuzzy logic controller using an enhanced genetic algorithm," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 30, no. 1, pp. 31–46, 2000. [Online]. Available: 10.1109/3477.826945.

[75] F. Guo et al., "A Concise TSK Fuzzy Ensemble Classifier Integrating Dropout and Bagging for High-Dimensional Problems," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 8, pp. 3176–3190, Aug. 2022, [Online]. Available: 10.1109/TFUZZ.2021.3106330.

[76] P. V. de Campos Souza, E. Lughofer, and H. Rodrigues Batista, "An Explainable Evolving Fuzzy Neural Network to Predict the k Barriers for Intrusion Detection Using a Wireless Sensor Network," *Sensors,* vol. 22, no. 14, Jul. 2022. [Online]. Available: 10.3390/s22145446.

[77] A. Gegov, D. Sanders, and B. Vatchova, "Aggregation of Inconsistent Rules for Fuzzy Rule Base Simplification," *International Journal of Knowledge-based and Intelligent Engineering Systems,* vol. 21, pp. 135-145, 2017. [Online]. Available: 21. 135-145. 10.3233/KES-170358.

[78] A. Dvořák and M. Štěpnička, "Mamdani-Assilian rules: with or without continuity?" *CEUR Workshop Proceedings* (CEUR-WS.org)", [Online]. Available: https://ifm.osu.eu/antonin-dvorak/10001/.

[79] V. F. Popescu and M. S. Pistol, "Fuzzy logic expert system for evaluating the activity of university teachers," *International Journal of Assessment Tools in Education,* pp. 991–1008, Nov. 2021. [Online]. Available: 10.21449/ijate.1025690.

[80] M. I. Rey, M. Galende, M. J. Fuente, and G. I. Sainz-Palmero, “Multi-objective based Fuzzy Rule Based Systems (FRBSs) for trade-off improvement in accuracy and interpretability: A rule relevance point of view,” *Knowledge Based Systems,* vol. 127, pp. 67–84, Jul. 2017. [Online]. Available: 10.1016/j.knosys.2016.12.028.

[81] L.-C. Dutu et al., “A Fast and Accurate Rule-Base Generation Method for Mamdani Fuzzy Systems,” *IEEE Transactions on Fuzzy Systems,* vol. 26, no. 2, pp. 715–733, 2018. [Online]. Available: 10.1109/TFUZZ.2017.2688349ï.

[82] W. Pedrycz, A. Gacek, and X. Wang, “A Hierarchical Approach to Interpretability of TS Rule-Based Models,” *IEEE Transactions on Fuzzy Systems,* vol. 30, no. 8, pp. 2861–2869, Aug. 2022. [Online]. Available: 10.1109/TFUZZ.2021.3094661.

[83] C. Fuchs, S. Spolaor, M. S. Nobile, and U. Kaymak, “A Graph Theory Approach to Fuzzy Rule Base Simplification,” in *Communications in Computer and Information Science,* 2020, vol. 1237 CCIS, pp. 387–401. [Online]. Available: 10.1007/978-3-030-50146-4_29.

[84] M. A. H. Ali, S. Mekhilef, N. Yusoff, and B. A. Razak, “Laser simulator logic: A novel inference system for highly overlapping of linguistic variable in membership functions,” *Journal of King Saud University - Computer and Information Sciences,* 2022. [Online]. Available: 10.1016/j.jksuci.2022.07.017.

[85] M. Shanmugapriya, H. Khanna Nehemiah, R. S. Bhuvaneswaran, K. Arputharaj, and J. Jabez Christopher, “SimE: A Geometric Approach for Similarity Estimation of Fuzzy Sets,” *Research Journal of Applied Sciences, Engineering and Technology,* vol. 13, no. 5, pp. 345–353, Sep. 2016. [Online]. Available: 10.19026/rjaset.13.2952.

[86] D. Wang et al., “An interpretable neural fuzzy inference system for predictions of underpricing in initial public offerings,” *Neurocomputing,* vol. 319, pp. 102–117, Nov. 2018. [Online]. Available: 10.1016/j.neucom.2018.07.036.

[87] H. Megherbi, A. C. Megherbi, and K. Benmahammed, “On accommodating the semantic-based interpretability in evolutionary linguistic fuzzy controller,”

Journal of Intelligent and Fuzzy Systems, vol. 36, no. 4, pp. 3765–3778, 2019. [Online]. Available: 10.3233/JIFS-18637.

[88] M. Lloret-Climent, J.-A. Nescolarde-Selva, and S. Pérez-Gonzaga, “Coverage and invariability in fuzzy systems,” *International Journal of General Systems,* vol. 43, pp. 96-104, 2014. [Online]. Available: 10.1080/03081079.2013.862240.

[89] D. D. Nauck, "Measuring interpretability in rule-based classification systems," *The 12th IEEE International Conference on Fuzzy Systems,* FUZZ '03, pp. 196-201 vol.1, 2003. [Online]. Available: 10.1109/FUZZ.2003.1209361.

[90] T. R. Razak, J. M. Garibaldi, C. Wagner, A. Pourabdollah, and D. Soria, “Toward a Framework for Capturing Interpretability of Hierarchical Fuzzy Systems - A Participatory Design Approach,” *IEEE Transactions on Fuzzy Systems,* vol. 29, no. 5, pp. 1160–1172, May 2021. [Online]. Available: 10.1109/TFUZZ.2020.2969901.

[91] F. Sabrina, S. Sohail, F. Farid, S. Jahan, F. Ahamed, and S. Gordon, “An Interpretable Artificial Intelligence Based Smart Agriculture System,” *Computers, Materials and Continua,* vol. 72, no. 2, pp. 3777–3797, 2022. [Online]. Available: 10.32604/cmc.2022.026363.

[92] H. R. Singh, D. Saroj, and K. Biswas, “Transparent Neuro-fuzzy model for Linguistic variables selection and rule-based classification,” *International Journal of Pure and Applied Mathematics*, vol. 118, no. 19, pp. 85–100, 2018.

[93] T. R. Razak et al., “An Exploratory Study of Hierarchical Fuzzy Systems Approach in A Recommendation System,” *Jurnal Intelek,* vol. 14, 2019. [Online]. Available: 10.24191/ji.v14i2.233.

[94] P. K. Shukla and S. P. Tripathi, “A new approach for tuning interval type-2 fuzzy knowledge bases using genetic algorithms,” *Journal of Uncertainty Analysis and Applications*, vol. 2, no. 1, Dec. 2014. [Online]. Available: 10.1186/2195-5468-2-4.

[95] H. Liu, C. Liu, and C. Wu, “A Framework of Granular Computing Clustering Algorithms,” *International Journal of Hybrid Information Technology,* vol. 8, no. 12, pp. 225–230, Dec. 2015. [Online]. Available: 10.14257/ijhit.2015.8.12.16.

[96] K. Xu, W. Pedrycz, Z. Li, “Granular computing: An augmented scheme of degranulation through a modified partition matrix,” *Fuzzy Sets and Systems,* vol. 440, pp. 131-148, 2022. [Online]. Available: https://doi.org/10.1016/j.fss.2021.06.001.

[97] L. A. Zadeh, “Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic,” *Fuzzy Sets and Systems,* vol. 90, pp. 111-127, 1997.

[98] L. Livi and A. Sadeghian, “Granular computing, computational intelligence, and the analysis of non-geometric input spaces,” *Granular Computing,* vol. 1, no. 1, pp. 13–20, Mar. 2016. [Online]. Available: 10.1007/s41066-015-0003-0.

[99] Y. Y. Yao, “Information granulation and rough set approximation,” International Journal of Intelligent Systems, vol. 16, pp. 87-104, 2001. [Online]. Available: 10.1002/1098-111X(200101)16:1<87::AID-INT7>3.0.CO;2-S.

[100] H. Liu, A. Gegov, and M. Cocea, “Rule-based systems: a granular computing perspective,” *Granular Computing,* vol. 1, no. 4, pp. 259–274, Dec. 2016. [Online]. Available: 10.1007/s41066-016-0021-6.

[101] A. Martino, L. Baldini, and A. Rizzi, “On Information Granulation via Data Clustering for Granular Computing-Based Pattern Recognition: A Graph Embedding Case Study,” *Algorithms,* vol. 15, no. 5, May 2022. [Online]. Available: 10.3390/a15050148.

[102] S. Zhou, D. Liu, S. Cao, X. Liu, and Y. Zhou, “An application status review of computational intelligence algorithm in GSHP field,” *Energy and Buildings,* vol. 203. Elsevier Ltd, Nov. 15, 2019. [Online]. Available: 10.1016/j.enbuild.2019.109424.

[103] N. Primeau, R. Falcon, R. Abielmona, and E. M. Petriu, “A review of computational intelligence techniques in wireless sensor and actuator networks,” *IEEE Communications Surveys and Tutorials,* vol. 20, no. 4, pp. 2822–2854, Oct. 2018. [Online]. Available: 10.1109/COMST.2018.2850220.

[104] M. Asim, Y. Wang, K. Wang, and P.-Q. Huang, "A Review on Computational Intelligence Techniques in Cloud and Edge Computing," in *IEEE Transactions*

on Emerging Topics in Computational Intelligence, vol. 4, no. 6, pp. 742-763, Dec. 2020. [Online]. Available: 10.1109/TETCI.2020.3007905.

[105] S. Butenkov, A. Zhukov, A. Nagorov, and N. Krivsha, "Granular Computing Models and Methods Based on the Spatial Granulation," in *Procedia Computer Science,* vol. 103, pp. 295–302, 2017. [Online]. Available: 10.1016/j.procs.2017.01.111.

[106] B. Ghiasi, H. Sheikhian, A. Zeynolabedin, and M. H. Niksokhan, "Granular computing-neural network model for prediction of longitudinal dispersion coefficients in rivers," *Water Science and Technology,* vol. 80, no. 10, pp. 1880–1892, Nov. 2020. [Online]. Available: 10.2166/wst.2020.006.

[107] Y. Liu, J. Zhao, W. Wang, and W. Pedrycz, "Prediction Intervals for Granular Data Streams Based on Evolving Type-2 Fuzzy Granular Neural Network Dynamic Ensemble," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 4, pp. 874–888, Apr. 2021. [Online]. Available: 10.1109/TFUZZ.2020.2966172.

[108] L. Decker, D. Leite, F. Viola, and D. Bonacorsi, "Comparison of Evolving Granular Classifiers applied to Anomaly Detection for Predictive Maintenance in Computing Centers," *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS),* Apr. 2020, [Online]. Available: http://arxiv.org/abs/2005.04156

[109] X. Hou, Y. Chen, J. Lu, K. Wu, and N. Zeng, "PGNN: A Parallel Granular BP Neural Network for Classiication PGNN: A Parallel Granular BP Neural Network for Classification," 2022. [Online]. Available: 10.21203/rs.3.rs-1901755/v1.

[110] G. Felix, G. Npoles, R. Falcon, R. Bello, and K. Vanhoof, "Performance analysis of granular versus traditional neural network classifiers: Preliminary results," in *CIVEMSA 2018 - 2018 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications*, 2018. [Online]. Available: 10.1109/CIVEMSA.2018.8439971.

[111] H. Liu, A. Gegov, and M. Cocea, "Rule-based systems: a granular computing perspective," *Granular Computing,* vol. 1, no. 4, pp. 259–274, Dec. 2016. [Online]. Available: 10.1007/s41066-016-0021-6.

[112] C. U. Yeom and K. C. Kwak, "A Design and Optimization of a CGK-Based Fuzzy Granular Model Based on the Generation of Rational Information Granules," *Applied Sciences* (Switzerland), vol. 12, no. 14, Jul. 2022. [Online]. Available: 10.3390/app12147226.

[113] D. Shan, W. Lu, and J. Yang, "Interval Granular Fuzzy Models: Concepts and Development," *IEEE Access*, vol. 7, pp. 24140–24153, 2019. [Online]. Available: 10.1109/ACCESS.2019.2899830.

[114] K. Xu, W. Pedrycz and Z. Li, "Augmentation of the reconstruction performance of Fuzzy C-Means with an optimized fuzzification factor vector," *Knowledge-Based Systems,* vol. 222, 2021. [Online]. Available: https://doi.org/10.1016/j.knosys.2021.106951.

[115] M. A. Sanchez, J. R. Castro, V. Ocegueda-Miramontes, and L. Cervantes, "Hybrid learning for general type-2 TSK fuzzy logic systems," *Algorithms,* vol. 10, no. 3, 2017. [Online]. Available: 10.3390/a10030099.

[116] Q. Guo, Z. Yin, and P. Wang, "An Improved Three-Way K-Means Algorithm by Optimizing Cluster Centers," *Symmetry* (Basel), vol. 14, no. 9, Sep. 2022. [Online]. Available: 10.3390/sym14091821.

[117] Y. C. Wang, X. Q. Tang, and H. Xu, "Clustering Ensemble Technology Based on Granular Computing to Extract Cervical Cancer Predictors," *Mathematical Problems in Engineering 2022*, 2022. [Online]. Available: 10.1155/2022/4465007.

[118] X. Zhang, X. Shen, and T. Ouyang, "Extension of DBSCAN in Online Clustering: An Approach Based on Three-Layer Granular Models," *Applied Sciences,* vol. 12, no. 19, p. 9402, Sep. 2022. [Online]. Available: 10.3390/app12199402.

[119] T. Ouyang and X. Shen, "Online structural clustering based on DBSCAN extension with granular descriptors," Information Sciences (N Y), vol. 607, pp. 688–704, Aug. 2022. [Online]. Available: 10.1016/j.ins.2022.06.027.

[120] M. Z. Muda and G. Panoutsos, "An Entropy-Based Uncertainty Measure for Developing Granular Models," in *2020 7th International Conference on Soft*

Computing and Machine Intelligence, ISCMI 2020, pp. 73–77, 2020. [Online]. Available: 10.1109/ISCMI51676.2020.9311589.

[121] A. Rubio-Solis, P. Melin, U. Martinez-Hernandez, and G. Panoutsos, "General Type-2 Radial Basis Function Neural Network: A Data-Driven Fuzzy Model," *IEEE Transactions on Fuzzy Systems,* vol. 27, no. 2, pp. 333–347, 2019. [Online]. Available: 10.1109/TFUZZ.2018.2858740.

[122] W. Pedrycz, "From Fuzzy Models to Granular Fuzzy Models," *International Journal of Computational Intelligence Systems,* vol. 9, pp. 35-42, 2016. [Online]. Available: 10.1080/18756891.2016.1180818.

[123] E. Hüllermeier, "From knowledge-based to data-driven fuzzy modeling," *Informatik-Spektrum,* vol. 38, 2015. [Online]. Available: 10.1007/s00287-015-0931-8.

[124] S. L. Chiu, "Extracting Fuzzy rules from Data for Function Approximation and Pattern Classification," *Fuzzy Information Engineering: A Guided Tour of Applications*, 1997.

[125] M. Al-Shammaa and M. F. Abbod, "Automatic generation of fuzzy classification rules using granulation-based adaptive clustering," *2015 Annual IEEE Systems Conference (SysCon) Proceedings,* pp. 653-659, 2015. [Online]. Available: 10.1109/SYSCON.2015.7116825.

[126] P. Salgado and P. J. Garrido, "Fuzzy clustering of fuzzy systems," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics,* vol. 3, pp. 2368–2373, 2004. [Online]. Available: 10.1109/icsmc.2004.1400683.

[127] W. Pedrycz and H. Izakian, "Cluster-Centric Fuzzy Modeling," *IEEE Transactions on Fuzzy Systems,* vol. 22, no. 6, pp. 1585–1597, Dec. 2014. [Online]. Available: 10.1109/TFUZZ.2014.2300134.

[128] R.R.Yager, D. P. Filev, Approximate clustering via the mountain method, IEEE *Transactions on Systems Man and Cybernetics*, vol. 24(8), pp. 1279-1284, 1994.

[129] A. Baraka, Ali, G. Panoutsos, M. Mahfouf, and S. Cater, "A Shannon entropy-based conflict measure for enhancing granular computing-based information

processing," *Proceedings - 2014 IEEE International Conference on Granular Computing, GrC 2014.* [Online]. Available: 13-18. 10.1109/GRC.2014.6982799.

[130] A. R. Solis and G. Panoutsos, "Granular computing neural-fuzzy modelling: A neutrosophic approach," *Applied Soft Computing Journal,* vol. 13, no. 9, pp. 4010–4021, 2013. [Online]. Available: 10.1016/j.asoc.2012.09.002.

[131] L. A. Zadeh, "Is there a need for fuzzy logic?" Information Sciences (N Y), vol. 178, no. 13, pp. 2751–2779, Jul. 2008. [Online]. Available: 10.1016/j.ins.2008.02.012.

[132] M. Z. Rodriguez et al., "Clustering algorithms: A comparative approach," *PLoS One,* vol. 14, no. 1, Jan. 2019. [Online]. Available: 10.1371/journal.pone.0210236.

[133] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: An overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 2, no. 1, pp. 86–97, Jan. 2012. [Online]. Available: 10.1002/widm.53.

[134] A. Bargiela and W. Pedrycz, "From Numbers to Information Granules: A Study in Unsupervised Learning and Feature Analysis," *The Springer International Series in Engineering and Computer Science,* vol. 717, pp. 75–112, 2002. [Online]. Available: 10.1142/9789812778147_0004.

[135] A. Bargiela and W. Pedrycz, "Classification and clustering of granular data," in *Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS,* vol. 3, pp. 1696–1701, 2001. [Online]. Available: 10.1109/nafips.2001.943807.

[136] Y. Y. Yao, "Information granulation and rough set approximation," International Journal of Intelligent Systems, vol. 16, pp. 87-104, 2001. [Online]. Available: 10.1002/1098-111X(200101)16:1<87::AID-INT7>3.0.CO;2-S.

[137] J. He, P. Wang, and Z. Li, "Uncertainty measurement for a set-valued information system: Gaussian Kernel method," *Symmetry* (Basel), vol. 11, no. 2, Feb. 2019. [Online]. Available: 10.3390/sym11020199.

[138] J. Yao, “Information granulation and granular relationships," *2005 IEEE International Conference on Granular Computing,* vol. 1, pp. 326-329, 2005. [Online]. Available: 10.1109/GRC.2005.1547296.

[139] Croucher Foundation. and C. International Computer Science Conference (4th : 1997 : Hong Kong, Proceedings : Asia Pacific Software Engineering Conference and International Computer Science Conference : December 2-5, 1997, Hong Kong. IEEE Computer Society Press, 1997.

[140] G. J. Klir and Bo. Yuan, Fuzzy sets and fuzzy logic : theory and applications. Prentice Hall PTR, 1995.

[141] J. Liang, Z. Shi, D. Li, and M. J. Wierman, “Information entropy, rough entropy and knowledge granulation in incomplete information systems,” International Journal of General Systems, vol. 35, no. 6, pp. 641–654, Dec. 2006. [Online]. Available: 10.1080/03081070600687668.

[142] X. Zhang, W. Hu, and F. Yang, “Detection of Cause-Effect Relations Based on Information Granulation and Transfer Entropy,” *Entropy,* vol. 24, no. 2, Feb. 2022. [Online]. Available: 10.3390/e24020212.

[143] Y. Qian and J. Liang, “Combination Entropy and Combination Granulation in Incomplete Information System,” *Lecture Notes in Computer Science,* pp. 184-190, 2006. [Online]. Available: 10.1007/11795131_27.

[144] H. Zhang, “Feature Selection Using Approximate Conditional Entropy Based on Fuzzy Information Granule for Gene Expression Data Classification,” *Front Genet*, vol. 12, Mar. 2021. [Online]. Available: 10.3389/fgene.2021.631505.

[145] M. A. Sanchez, J. R. Castro, O. Castillo, O. Mendoza, A. Rodriguez-Diaz, and P. Melin, “Fuzzy higher type information granules from an uncertainty measurement,” *Granular Computing,* vol. 2, no. 2, pp. 95–103, Jun. 2017. [Online]. Available: 10.1007/s41066-016-0030-5.

[146] D. Xu and Y. Tian, “A Comprehensive Survey of Clustering Algorithms,” *Annals of Data Science,* vol. 2, no. 2, pp. 165–193, Jun. 2015. [Online]. Available: 10.1007/s40745-015-0040-1.

[147] A. Saxena et al., “A review of clustering techniques and developments, *Neurocomputing,* vol. 267, pp. 664-681, 2017. [Online]. Available: https://doi.org/10.1016/j.neucom.2017.06.053.

[148] Z. Zhang, F. Murtagh, S. van Poucke, S. Lin, and P. Lan, “Hierarchical cluster analysis in clinical research with heterogeneous study population: Highlighting its visualization with R,” *Annals of Translational Medicine,* vol. 5, no. 4, Feb. 2017. [Online]. Available: 10.21037/ATM.2017.02.05.

[149] M. Kiruthika, D. S. Sukumaran, and R. Scholar, “A Survey on Partitioning and Hierarchical based Data Mining Clustering Techniques,” 2018.

[150] F. Murtagh and P. Contreras, “Algorithms for hierarchical clustering: An overview,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 2, no. 1, pp. 86–97, Jan. 2012. [Online]. Available: 10.1002/widm.53.

[151] A. C. Benabdellah, A. Benghabrit, and I. Bouhaddou, “A survey of clustering algorithms for an industrial context,” in *Procedia Computer Science,* vol. 148, pp. 291–302, 2019. [Online]. Available: 10.1016/j.procs.2019.01.022.

[152] L. A. Zadeh, “Fuzzy sets and information granularity,” *Advances in Fuzzy Sets Theory and Applications,* 1979.

[153] L. A. Zadeh, “Fuzzy logic = computing with words," in *IEEE Transactions on Fuzzy Systems, vol*. 4, no. 2, pp. 103-111, May 1996. [Online]. Available: 10.1109/91.493904.

[154] L. A. Zadeh, “Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems,* vol. 90, Issue 2, pp. 111-127, 1997.

[155] T. Y. Lin, “Granular Computing: Fuzzy Logic and Rough Sets,” In: Zadeh, L.A., Kacprzyk, J. (eds) Computing with Words in Information/Intelligent Systems, *Studies in Fuzziness and Soft Computing,* vol 33. Physica, Heidelberg, 1999. [Online]. Available: https://doi.org/10.1007/978-3-7908-1873-4_9.

[156] W. Pedrycz, “Granular computing: An introduction,” in *Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS*, vol. 3, pp. 1349–1354, 2001. [Online]. Available: 10.1109/nafips.2001.943745.

[157] A. R. Solis, “Uncertainty and Interpretability Studies in Soft Computing with an application to Complex Manufacturing Systems,” PhD Thesis, Department of Automatic Control and Systems Engineering, The University of Sheffield, UK, 2014.

[158] G. J. Klir and Bo. Yuan, Fuzzy sets and fuzzy logic : Theory and applications. Prentice Hall PTR, 1995.

[159] B. Jiang, J. Pei, Y. Tao, and X. Lin, “Clustering uncertain data based on probability distribution similarity,” *IEEE Transactions on Knowledge and Data Engineering,* vol. 25, no. 4, pp. 751–763, 2013. [Online]. Available: 10.1109/TKDE.2011.221.

[160] H. E. Lee, K. H. Park, and Z. Z. Bien, “Iterative fuzzy clustering algorithm with supervision to construct probabilistic fuzzy rule base from numerical data,” I*EEE Transactions on Fuzzy Systems, vol*. 16, no. 1, pp. 263–277, Feb. 2008. [Online]. Available: 10.1109/TFUZZ.2007.903314.

[161] J. M. Cadenas, M. C. Garrido, R. Martínez, and A. Muñoz-Ledesma, “Classification based on Neighborhood from Datasets with Low Quality Data,” 2015.

[162] J. Abellán and É. Bossé, “Critique of Recent Uncertainty Measures Developed Under the Evidence Theory and Belief Intervals,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 50, no. 3, pp. 1186–1192, Mar. 2020. [Online]. Available: 10.1109/TSMC.2017.2770128.

[163] N. R. Pal, “On quantification of different facets of uncertainty*,” Fuzzy Sets and Systems,* vol. 107, Issue 1, pp 81–91, 1999. [Online]. Available: https://doi.org/10.1016/S0165-0114(98)00005-0.

[164] E. Hüllermeier, “Uncertainty in Clustering and Classification,” pp. 16-19, 2010. [Online]. Available: 10.1007/978-3-642-15951-0_6.

[165] Ö. E. Oruç, E. Kuruoglu, and A. Gündüz, “Entropy Applications for Customer Satisfaction Survey in Information Theory,” *Frontiers in Science,* vol. 1, no. 1, pp. 1–4, Aug. 2012. [Online]. Available: 10.5923/j.fs.20110101.01.

[166] L. Yin and Y. Deng, “Toward uncertainty of weighted networks: An entropy-based model,” *Physica A: Statistical Mechanics and its Applications,* vol. 508, pp. 176–186, Oct. 2018. [Online]. Available: 10.1016/J.PHYSA.2018.05.067.

[167] Q. Jiang, Y. Shen, H. Li, and F. Xu, “New fault recognition method for rotary machinery based on information entropy and a probabilistic neural network,” *Sensors* (Switzerland), vol. 18, no. 2, Feb. 2018. [Online]. Available: 10.3390/s18020337.

[168] L. Liu, S. Wang, D. Liu, Y. Zhang, and Y. Peng, “Entropy-based sensor selection for condition monitoring and prognostics of aircraft engine,” *Microelectronics* Reliability, vol. 55, no. 9–10, pp. 2092–2096, Aug. 2015. [Online]. Available: 10.1016/J.MICROREL.2015.06.076.

[169] S. Bobek and P. Misiak, “Uncertain decision tree classifier for mobile context-aware computing,” in *Lecture Notes in Computer Science*, vol. 10842 LNAI, pp. 276–287, 2018. [Online]. Available: 10.1007/978-3-319-91262-2_25.

[170] F. Gullo, G. Ponti, A. Tagarelli, and S. Greco, “An information-theoretic approach to hierarchical clustering of uncertain data,” Information Sciences (N Y), vol. 402, pp. 199–215, Sep. 2017. [Online]. Available: 10.1016/J.INS.2017.03.030.

[171] M. Aghagolzadeh, H. Soltanian-Zadeh, and B. N. Araabi, “Information Theoretic Hierarchical Clustering,” *Entropy,* vol. 13, no. 2, pp. 450–465, Feb. 2011. [Online]. Available: 10.3390/e13020450.

[172] M. Chau, R. Cheng, B. Kao, and J. Ng, “Uncertain Data Mining: An Example in Clustering Location Data,” *Lecture Notes in Computer Science,* vol. 3918, pp. 199-204, 2006. [Online]. Available: 10.1007/11731139_24.

[173] H. Sun, R. Chen, Y. Qin, and S. Wang, “Holo-Entropy Based Categorical Data Hierarchical Clustering,” *Informatica* (Netherlands), vol. 28, no. 2, pp. 303–328, 2017. [Online]. Available: 10.15388/Informatica.2017.131.

[174] T. M. Cover and J. A. Thomas, “Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing), John Wiley & Sons, Inc., 2005. [Online]. Available: 10.1002/047174882X.

[175] A. S. Shirkhorshidi, S. Aghabozorgi, and T. Ying Wah, “A Comparison study on similarity and dissimilarity measures in clustering continuous data,” *PLoS One,* vol. 10, no. 12, Dec. 2015. [Online]. Available: 10.1371/journal.pone.0144059.

[176] K. H. Knuth, “Optimal data-based binning for histograms and histogram-based probability density models,” *Digital Signal Processing: A Review Journal,* vol. 95, Dec. 2019. [Online]. Available: 10.1016/j.dsp.2019.102581.

[177] A. Riid and J. S. Preden, “Design of Fuzzy Rule-based Classifiers through Granulation and Consolidation,” *Journal of Artificial Intelligence and Soft Computing Research,* vol. 7, no. 2, pp. 137–147, Apr. 2017. [Online]. Available: 10.1515/jaiscr-2017-0010.

[178] J. Niu, D. Chen, J. Li, and H. Wang, “Fuzzy rule based classification method for incremental rule learning,” *IEEE Transactions on Fuzzy Systems*, pp. 1–1, Nov. 2021. [Online]. Available: 10.1109/tfuzz.2021.3128061.

[179] N. T. Nguyen, B. Trawinski, and R. Kosala, “Intelligent Information and Database Systems,” 2015. [Online]. Available: http://www.springer.com/series/1244.

[180] S. Butenkov, A. Zhukov, A. Nagorov, and N. Krivsha, “Granular Computing Models and Methods Based on the Spatial Granulation,” *in Procedia Computer Science,* vol. 103, pp. 295–302, 2017. [Online]. Available:: 10.1016/j.procs.2017.01.111.

[181] J. Z. Huang, M. K. Ng, H. Rong, and Z. Li, “Automated variable weighting in k-means type clustering," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 657-668, May 2005. [Online]. Available: 10.1109/TPAMI.2005.95.

[182] R. C. de Amorim, “Feature Relevance in Ward’s Hierarchical Clustering Using the Lp Norm,” Journal of Classification, vol. 32, no. 1, pp. 46–62, Apr. 2015. [Online]. Available: 10.1007/s00357-015-9167-1.

[183] R. Cordeiro De Amorim, “Learning feature weights for K-Means clustering using the Minkowski metric,” PhD Thesis, Department of Computer Science and Information Systems, Univ. of London, UK, 2011.

[184] W. Fouad, A. Badr, and I. Farag, “AIFSA: A New Approach for Feature Selection and Weighting,” ICIEIS 2011, Part II, CCIS 252, pp. 596–609, 2011.

[185] D. Panday, R. Cordeiro de Amorim, and P. Lane, “Feature weighting as a tool for unsupervised feature selection,” *Information Processing Letters*, vol. 129, pp. 44–52, Jan. 2018, [Online]. Available: 10.1016/j.ipl.2017.09.005.

[186] J. Tang, S. Alelyani, and H. Liu, “Feature Selection for Classification: A Review,” Data Classification: Algorithms and Applications, 2014.

[187] G. Siva Shankar, P. Ashokkumar, R. Vinayakumar, U. Ghosh, W. Mansoor, and W. S. Alnumay, “An Embedded-Based Weighted Feature Selection Algorithm for Classifying Web Document,” *Wireless Communications and Mobile Computing,* 2020. [Online]. Available: 10.1155/2020/8879054.

[188] S. Paul and S. Das, “Simultaneous feature selection and weighting - An evolutionary multi-objective optimization approach,” *Pattern Recognition Letters,* vol. 65, pp. 51–59, Nov. 2015. [Online]. Available: 10.1016/j.patrec.2015.07.007.

[189] M. A. Tahir, A. Bouridane, and F. Kurugollu, “Simultaneous feature selection and feature weighting using Hybrid Tabu Search/K-nearest neighbor classifier,” *Pattern Recognit Letters,* vol. 28, no. 4, pp. 438–446, Mar. 2007. [Online]. Available: 10.1016/j.patrec.2006.08.016.

[190] I. Niño-Adan, D. Manjarres, I. Landa-Torres, and E. Portillo, “Feature weighting methods: A review,” *Expert Systems with Applications,* vol. 184. Elsevier Ltd, Dec. 01, 2021. [Online]. Available: 10.1016/j.eswa.2021.115424.

[191] B. Zou, L. Wang, M. Zeng, K. Yue, and F. Wei, “A Novel LASSO-based feature weighting selection method for microarray data classification, vol. 5, 2015. [Online]. Available: 10.1049/cp.2015.0795.

[192] H. Zhou, X. Wang, and Y. Zhang, “Feature selection based on weighted conditional mutual information,” *Applied Computing and Informatics,* 2020. [Online]. Available: 10.1016/j.aci.2019.12.003.

[193] T. Mora, Marcelo and B. Jhosep, “Mutual Information: A way to quantify correlations,” *Revista Brasileira de Ensino de Física,* vol. 44, 2022. [Online]. Available: 10.1590/1806-9126-rbef-2022-0055.

[194] M. Alkhodari, D. K. Islayem, F. A. Alskafi, and A. H. Khandoker, “Predicting hypertensive patients with higher risk of developing vascular events using heart rate variability and machine learning,” *IEEE Access,* vol. 8, pp. 192727–192739, 2020. [Online]. Available: 10.1109/ACCESS.2020.3033004.

[195] H. Fernando and J. Marshall, “What lies beneath: Material classification for autonomous excavators using proprioceptive force sensing and machine learning,” Automation in Construction, vol. 119, Nov. 2020. [Online]. Available: 10.1016/j.autcon.2020.103374.

[196] M. L. McHugh, “The Chi-square test of independence,” *Biochem Med* (Zagreb), vol. 23, no. 2, pp. 143–149, Jun. 2012. [Online]. Available: 10.11613/BM.2013.018.

[197] Y. Wang, S. Cang and H. Yu, “Mutual information inspired feature selection using kernel canonical correlation analysis,” *Expert Systems with Applications: X,* vol. 4, 2019. [Online]. Available: https://doi.org/10.1016/j.eswax.2019.100014.

[198] M. Dousthagh et al. “Feature Weighting Using a Clustering Approach,” *International Journal of Modeling and Optimization,* vol. 9, no. 67, 2019. [Online]. Available: 10.7763/IJMO.2019.V9.686.

[199] H. Xing, H. Ming, D. Tian, B. Hu, “A Novel Support Vector Machine With Its Features Weighted by Mutual Information,” *Proceedings of the International Joint Conference on Neural Networks,* pp. 315 – 320, 2008, [Online]. Available: 10.1109/IJCNN.2008.4633810.

[200] G. Cleuziou, L. Martin, and C. Vrain, “PoBOC: An Overlapping Clustering Algorithm, Application to Rule-Based Classification and Textual Data,”

Proceedings of the 16th Eureopean Conference on Artificial Intelligence, pp. 440-444, 2004.

[201] B. B. Án and D. V. Nõ, "Survey of Overlapping Clustering Algorithms," *Computacion y Sistemas,* vol. 24, no. 2, pp. 575–581, 2020. [Online]. Available: 10.13053/CyS-24-2-3391.

[202] C. E. ben N'cir and N. Essoussi, "On the extension of k-means for overlapping clustering average or sum of clusters' representatives?," in *IC3K 2013; KDIR 2013 - 5th International Conference on Knowledge Discovery and Information Retrieval and KMIS 2013 - 5th International Conference on Knowledge Management and Information Sharing, Proc.,* pp. 208–213, 2013. [Online]. Available: 10.5220/0004626502080213.

[203] A. Adam and H. B. Kuleuven, "Dealing with overlapping clustering: A constraint-based approach to algorithm selection," *CEUR Workshop Proceedings,* vol. 1455, pp. 43-54, 2015.

[204] G. Cleuziou, "A Generalization of k-Means for Overlapping Clustering," 2007.

[205] A. A. Aroche-Villarruel, J. Carrasco-Ochoa, J. Fco Martínez-Trinidad, J. Arturo Olvera-López, and A. Pérez-Suárez, "LNCS 8495 - Study of Overlapping Clustering Algorithms Based on Kmeans through FBcubed Metric," Lecture Notes in Computer Science, vol 8495, 2014.

[206] C.-E. Ben, N.' Cir, and N. Essoussi, "Overlapping Patterns Recognition with Linear and Non-Linear Separations using Positive Definite Kernels," *International Journal of Computer Applications,* vol. 56, no. 9, 2012. [Online]. Available: http://mlkd.csd.auth.gr/multilabel.html.

[207] A. A. Aroche-Villarruel, J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad, J. A. Olvera-López, and A. Pérez-Suárez, "Study of overlapping clustering algorithms based on Kmeans through FBcubed metric," in *Lecture Notes in Computer Science*, vol. 8495, pp. 112–121, 2014.[Online]. Available: 10.1007/978-3-319-07491-7_12.

[208] F. Gama, S. Segarra, and A. Ribeiro, "Hierarchical Overlapping Clustering of Network Data Using Cut Metrics," *IEEE Transactions on Signal and Information*

Processing over Networks, Nov. 2016, [Online]. Available: http://arxiv.org/abs/1611.01393.

[209] I. Jeantet, Z. Miklos, D. Gross-Amblard, and Z. Miklós, "Overlapping Hierarchical Clustering (OHC)," *Lecture Notes in Computer Science*, vol 12080, 2020. [Online]. Available: https://hal.inria.fr/hal-02452729.

[210] S. Oh, "A new dataset evaluation method based on category overlap," *Computers in Biology and Medicine,* vol. 41, no. 2, pp. 115–122, Feb. 2011. [Online]. Available: 10.1016/j.compbiomed.2010.12.006.

[211] Z. Li, J. Qin, X. Zhang, and Y. Wan, "Addressing class overlap under imbalanced distribution: An improved method and two metrics," *Symmetry* (Basel), vol. 13, no. 9, Sep. 2021. [Online]. Available: 10.3390/sym13091649.

[212] J. Lee, N. Batnyam, and S. Oh, "RFS: Efficient feature selection method based on R-value," *Computers in Biology and Medicine*, vol. 43, no. 2, pp. 91–99, Feb. 2013. [Online]. Available: 10.1016/j.compbiomed.2012.11.010.

[213] E. B. Fatima, O. Boutkhoum, E. M. Abdelmajid, F. Rustam, A. Mehmood, and G. S. Choi, "Minimizing the Overlapping Degree to Improve Class-Imbalanced Learning under Sparse Feature Selection: Application to Fraud Detection," *IEEE Access,* vol. 9, pp. 28101–28110, 2021. [Online]. Available: 10.1109/ACCESS.2021.3056285.

[214] Z. Zhu and M. Zhang, "K-Nearest Neighbors (KNN) Classification with Different Distance Metrics," *Fundamentals of Data Science,* pp. 1-14, 2020.

[215] F. Azad and P. Kr Shukla, "A Review on the Development of Fuzzy Classifiers with Improved Interpretability and Accuracy Parameters," *Journal of Informatics Electrical and Electronics Engineering,* vol. 02, pp. 1–9, 2021. [Online]. Available: 10.54060/JIEEE/002.02.020.

[216] P. Chandra, D. Agarwal, and P. K. Shukla, "Improved Fuzzy Modeling of Thyroid Disease Detection using Interval Type 2 Fuzzy Techniques," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 6, pp. 220–223, Aug. 2020. 10.35940/ijeat.C5931.089620.

[217] Q. Litao, W. Weijia and L. Bill, “Learning Accurate and Interpretable Decision Rule Sets from Neural Networks,” *Association for the Advancement of Artificial Intelligence,* pp. 1-9, 2021.

[218] L. A. Lucas, T. M. Centeno, and M. R. Delgado, “Towards interpretable general type-2 fuzzy classifiers,” in *ISDA 2009 - 9th International Conference on Intelligent Systems Design and Applications,* pp. 584–589, 2009. [Online]. Available: 10.1109/ISDA.2009.28.

[219] P. K. Tripathi, “Interpretability and accuracy issues in evolutionary multi-objective fuzzy classifiers,” *International Journal of Soft Computing and Networking,* vol. 1, no. 55, 2016. [Online]. Available: 10.1504/IJSCN.2016.077043.

[220] P. K. Shukla and S. P. Tripathi, “A review on the Interpretability-Accuracy Trade-Off in Evolutionary Multi-Objective Fuzzy Systems (EMOFS),” *Information* (Switzerland), vol. 3, no. 3. pp. 256–277, 2012. [Online]. Available: 10.3390/info3030256.

[221] A. Baraka and G. Panoutsos, “Long-term learning for type-2 neural-fuzzy systems,” *Fuzzy Sets and Systems,* vol. 368, pp. 59–81, Aug. 2019. [Online]. Available: 10.1016/j.fss.2018.12.014.

[222] Y. Yang, P. Wang, and X. Gao, “A Novel Radial Basis Function Neural Network with High Generalization Performance for Nonlinear Process Modelling,” *Processes,* vol. 10, no. 1, Jan. 2022. [Online]. Available: 10.3390/pr10010140.

[223] D. Wu, C. T. Lin, J. Huang, and Z. Zeng, “On the Functional Equivalence of TSK Fuzzy Systems to Neural Networks, Mixture of Experts, CART, and Stacking Ensemble Regression,” *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 10, pp. 2570–2580, Oct. 2020. [Online]. Available: 10.1109/TFUZZ.2019.2941697.

[224] J. S. R. Jang, C.-T. Sun, J.-S. R. Jang, and C. T. Sun, “Functional Equivalence between Radial Basis Function Networks and Fuzzy Inference Systems,” *IEEE Trans Neural Netw,* vol 4, no. 1, 1996. [Online]. Available: https://www.researchgate.net/publication/2824946.

[225] A. Hryniowski and A. Wong, “DeepLABNet: End-to-end Learning of Deep Radial Basis Networks,” *Journal of Computational Vision and Imaging Systems*, vol. 5, nol 1, 2020.

[226] C. S. K. Dash, A. K. Behera, S. Dehuri, and S. Cho, "Radial basis function neural networks: a topical state-of-the-art survey," *Open Computer Science*, vol. 6, no. 1, pp. 33-63, 2016. [Online]. Available:10.1515/comp-2016-0005.

[227] E. T. Abdullah and A. Aytekin. (2019), “Application of Radial Basis Function Networks with Feature Selection for GDP Per Capita Estimation Based on Academic Parameters,” *Computer Systems Science and Engineering*, vol. 34, pp. 145-150. [Online]. Available: 10.32604/csse.2019.34.151.

[228] N. Nawaz, S. Harun, A. Sobri , R. Othman,and A. Heryansyah, “Application of radial basis function neural networks for modeling rainfall-runoff processes: A case study of Semenyih River catchment, Malaysia, *Chiang Mai J. Sci*., vol. 43, no. 6, pp. 1358-1367.

[229] N. Steele and J. Godjevac, “Radial Basis Function Artificial Neural Networks and Fuzzy Logic,” *Periodica Polytechnica,* vol. 42, no. 1, pp. 155-172, 1998.

[230] W. W. Tan and D. Wu, “Design of type-reduction strategies for type-2 fuzzy logic systems using genetic algorithms,” *Studies in Computational Intelligence,* vol. 66, pp. 169–187, 2007. [Online]. Available: 10.1007/978-3-540-72377-6_7.

[231] D. Wu and J. M. Mendel, “Enhanced Karnik-Mendel algorithms,” *IEEE Transactions on Fuzzy Systems,* vol. 17, no. 4, pp. 923–934, 2009. [Online]. Available: 10.1109/TFUZZ.2008.924329.

[232] A. Amini and N. Nikraz (2016), “Proposing Two Defuzzification Methods based on Output Fuzzy Set Weights,” *International Journal of Intelligent Systems and Applications*. vol. 2, pp. 1-12, 2016. [Online]. Available: 10.5815/ijisa.2016.02.01.

[233] J. M. Mendel, F. Liu, and D. Zhai, “α-Plane representation for type-2 fuzzy sets: Theory and applications,” *IEEE Transactions on Fuzzy Systems,* vol. 17, no. 5. pp. 1189–1207, 2009. [Online]. Available: 10.1109/TFUZZ.2009.2024411.

[234] H. Hamrawi, "Type-2 Fuzzy Alpha-Cuts," in *IEEE Transactions on Fuzzy Systems,* vol. 25, no. 3, pp. 682-692, June 2017. [Online]. Available: 10.1109/TFUZZ.2016.2574914.

[235] C. F. Juang and C. Y. Chen, “Data-driven interval type-2 neural fuzzy system with high learning accuracy and improved model interpretability,” *IEEE Trans Cybernetics*, vol. 43, no. 6, pp. 1781–1795, Dec. 2013. [Online]. Available: 10.1109/TSMCB.2012.2230253.

[236] Y. Zhou, Y. Niu, Q. Luo, and M. Jiang, “Teaching learning-based whale optimization algorithm for multi-layer perceptron neural network training,” *Mathematical Biosciences and Engineering,* vol. 17, no. 5, pp. 5987–6025, Sep. 2020. [Online]. Available: 10.3934/MBE.2020319.

[237] S. Saha and L. Ramanathan, “Building and Improving Artificial Neural Network Classifier,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 9, pp. 2742–2747, Jul. 2019. [Online]. Available: 10.35940/ijitee.i8311.078919.

[238] V. Chaurasia, S. Pal, and B. B. Tiwari, “Prediction of benign and malignant breast cancer using data mining techniques,” *Journal of Algorithms & Computational Technology,* vol. 12, no. 2, pp. 119–126, Jun. 2018. [Online]. Available: 10.1177/1748301818756225.

[239] M. M. Islam, M. R. Haque, H. Iqbal, M. M. Hasan, M. Hasan, and M. N. Kabir, “Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques,” *SN Computer Science*, vol. 1, no. 5, Dec. 2020. [Online]. Available: 10.1007/s42979-020-00305-w.

[240] C. Zhang, J. Qu, W. Li, and L. Zheng, “Predicting Cardiovascular Events by Machine Learning,” in *Journal of Physics: Conference Series*, vol. 1693, no. 1. Dec. 2020. [Online]. Available: 10.1088/1742-6596/1693/1/012093.

[241] S. Begum, A. Siddique, and R. Tiwari, “A Study for Predicting Heart Disease using Machine Learning,” *Turkish Journal of Computer and Mathematics Education*, Vol.12 No.10 (2021), pp. 4584-4592, 2021.

[242] E. Çapa, (2018), "Cardiotocography Data Set Classification with Extreme Learning Machine**,"** *International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES'18),* pp. 224-230, 2018.

[243] K. ParvathSuman, A. Raj, R.Balasubramanian, "Performance evaluation of machine learning techniques for fetal hypoxia prediction," *Journal of Information and Computational Science*, Vol. 9, Issue 8, pp. 809-820, 2019.

[244] B. Amin, M. Gamal, A. A. Salama, I. M. El-Henawy and K. Mahfouz, "Classifying Cardiotocography Data based on Rough Neural Network" *International Journal of Advanced Computer Science and Applications (IJACSA),* 10(8), 2019. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2019.0100846.

[245] S. Singaravelan, R. Arun, D. A. Shunmugam, K. R. Soundar, R. Mayakrishnan, and D. Murugan, "Analysis of Classification Algorithms on Different Datasets," Review of Innovation and Competitiveness," *Review of Innovation and Competitiveness,* vol. 4, no. 2, pp. 41-54, 2018. [Online]. Available: 4. 41-54. 10.32728/ric.2018.42/3.

[246] J. A. Nasiri and H. Shakibian, "Probabilistic twin support vector machine for solving unclassifiable region problem," *International Journal of Engineering, Transactions A: Basics,* vol. 35, no. 1, Jan. 2022. [Online]. Available: 10.5829/ije.2022.35.01a.01.

[247] F. Deng, S. Guo, R. Zhou, and J. Chen, "Sensor Multifault Diagnosis with Improved Support Vector Machines," *IEEE Transactions on Automation Science and Engineering*, vol. 14, no. 2, pp. 1053–1063, Apr. 2017. [Online]. Available: 10.1109/TASE.2015.2487523.

[248] Nurhayati, F. Agustian, and M. D. I. Lubis, "Particle Swarm Optimization Feature Selection for Breast Cancer Prediction," in *2020 8th International Conference on Cyber and IT Service Management, CITSM 2020*, Oct. 2020. [Online]. Available: 10.1109/CITSM50537.2020.9268865.

[249] W. Panup, W. Ratipapongton, and R. Wangkeeree, "A Novel Twin Support Vector Machine with Generalized Pinball Loss Function for Pattern

Classification," *Symmetry* (Basel), vol. 14, no. 2, Feb. 2022. [Online]. Available: 10.3390/sym14020289.

[250] N. N. Khanom, F. Nihar, S. S. Hassan, and L. Islam, "Performance Analysis of Algorithms on Different Types of Health Related Datasets," in *Journal of Physics: Conference Series,* vol. 1577, no. 1. Jul. 2020. [Online]. Available: 10.1088/1742-6596/1577/1/012051.

[251] S. Prasetyo, P. Prastyo, and S. Arti (2021, March 31), "A Cardiotocographic Classification using Feature Selection: A comparative Study," *JITCE (Journal of Information Technology and Computer Engineering),* vol. 5, no. 1, pp. 25-32, 2021. [Online]. Available: https://doi.org/https://doi.org/10.25077/jitce.5.01.25-32.2021.

[252] P. D'Alterio, J. M. Garibaldi, R. I. John, and A. Pourabdollah, "Constrained Interval Type-2 Fuzzy Sets," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 5, pp. 1212–1225, May 2021. [Online]. Available: 10.1109/TFUZZ.2020.2970911.

[253] P. Chandra, D. Agarwal, and P. K. Shukla, "Interpretability and Accuracy Analysis of Fuzzy Rule Based System Designed for Abalone," *International Journal of Recent Technology and Engineering (IJRTE),* vol. 8, no. 5, pp. 1335–1340, Jan. 2020. [Online]. Available: 10.35940/ijrte.E6131.018520.

[254] P. Chandra, D. Agarwal, and P. K. Shukla, "Improved Fuzzy Modeling of Thyroid Disease Detection using Interval Type 2 Fuzzy Techniques," *International Journal of Recent Technology and Engineering (IJRTE),* vol. 9, no. 6, pp. 220–223, Aug. 2020. [Online]. Available: 10.35940/ijeat.C5931.089620.

[255] A. K. Shukla and A. Yadav, "Interpretability Assessment in Fuzzy Rule Based Systems," *International Journal of Scientific & Engineering Research,* vol. 5, no. 7, 2014, [Online]. Available: http://www.ijser.org.

[256] J. Alonso and L. Magdalena, "An Experimental Study on the Interpretability of Fuzzy Systems," *Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference,* pp. 125-130, 2009.