MASTER'S DEGREE THESIS

FUZZY INFERENCE SYSTEM FOR EMG ASYMMETRY INDEX VALIDATION

Submitted in the accademic year 2019/2020

Biomedical Engineering

Polytechnic of Turin

Turin, Italy

Supervisors:

Gabriella Balestra

Cristina Castagneri

Samanta Rosati

Author:

Luca Colaci

Abstract

Human gait is not a periodic movement, its analysis will always present a typical intrasubject, intersubject variability. For this reason recent study enlighted the importance of taking in account a large number of gait cycles to be analyzed, making easier to compare parameters. From this point of view the detecting of gait deviations can be more accurate than visual assessment and can be used in different clinical cases. Although the existence of many protocols for recording gait signals, the selection of signal's features is the main issue when a machine learning algorithm is used to fit a designed model. In this study an organized data set composed by gait parameters observations, is processed through unsupervised and supervised learning algorithms in order to predict an asymmetry level assigned to each observation through the EMG asymmetry index proposed in a recent study. Each parameter used to build the dataset derives from sEMG signals recorded during trials performed in previous studies, in particular the electrical activity produced by four skeletal muscles of the subject's lower limbs, whom did not show any pathology that could affect walking task. In this study is possible to find the exploration of the input space and a design of a fuzzy logic controller by the tuning of fuzzy rules and membership function parameters.

Contents

Chapter 1

Introduction

Gait analysis

We all as modern human being are supposed to walk by mean of lower limbs, in a repetitive loss and recover of balance. Human walking is a low-cost in energy mode locomotion that involves both ability and adaptability in moving and it is probably the most practiced activity in the world.
Walking is linked with health, autonomy, and quality of life but also to self-worth, this is why gait analysis's diagnostic impact has been seen and characterized during time.

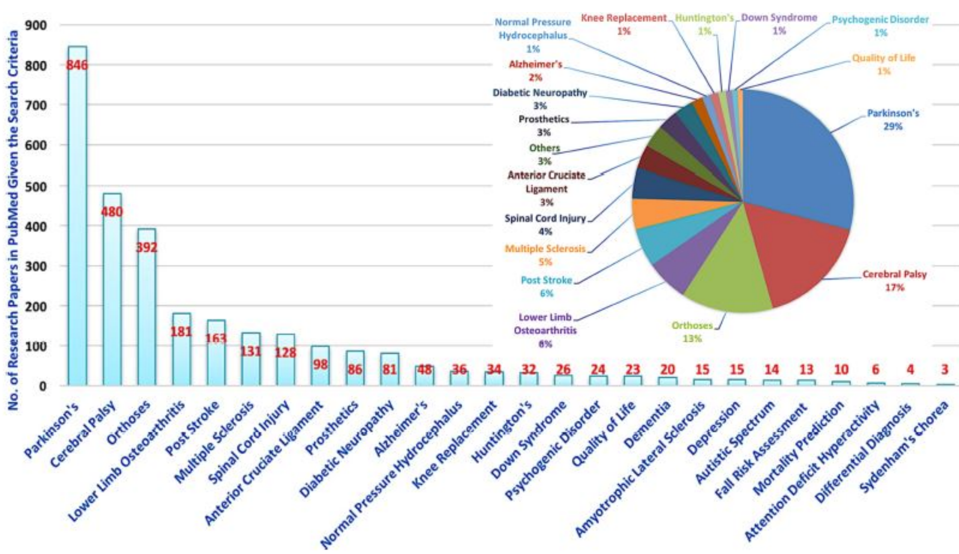


Figure 1.1: Research efforts in gait patologies from 1970 [1]

1.0.1 Evolution of human gait analysis

Aristotle (384-322 b.C) reflected on physics of movement and discussed on the nature of moving of animals in his work *De motu animalium*. His grandson, Erasistratus (310-250 b.C) has probably been the first anatomist and physician to discover that muscle could contract.

Also Galen (130-210 a.D) was attracted by physics of movement and in a very important historical context such Roman Empire, he gave a contribute to better understand human activities through basic anatomical knowledge.

Honorè de Balzac (1799 – 1850 a.D), a french novelist, decided to undertake his personal in-depth study based on observation, then called *Theory of walking*. In this study is possible to find a definition of gait, a classification of gait phases and a list of affecting factors, also many accurate statements all embellished with author's personal style, sarcasm and critical tone.

What we know so far about a gait cycle, is that it is considered as the movement realized between two foot strikes of the same foot. It is universally divided in two phases. The stance phase, which is the movement during the floor contact, is divided in three periods with changing of support: an initial double-limb support, then a single-limb stance and a second double-limb support. The swing phase, around the 40% of the cycle, is also divided in three periods: the initial swing, the mid swing and terminal swing.

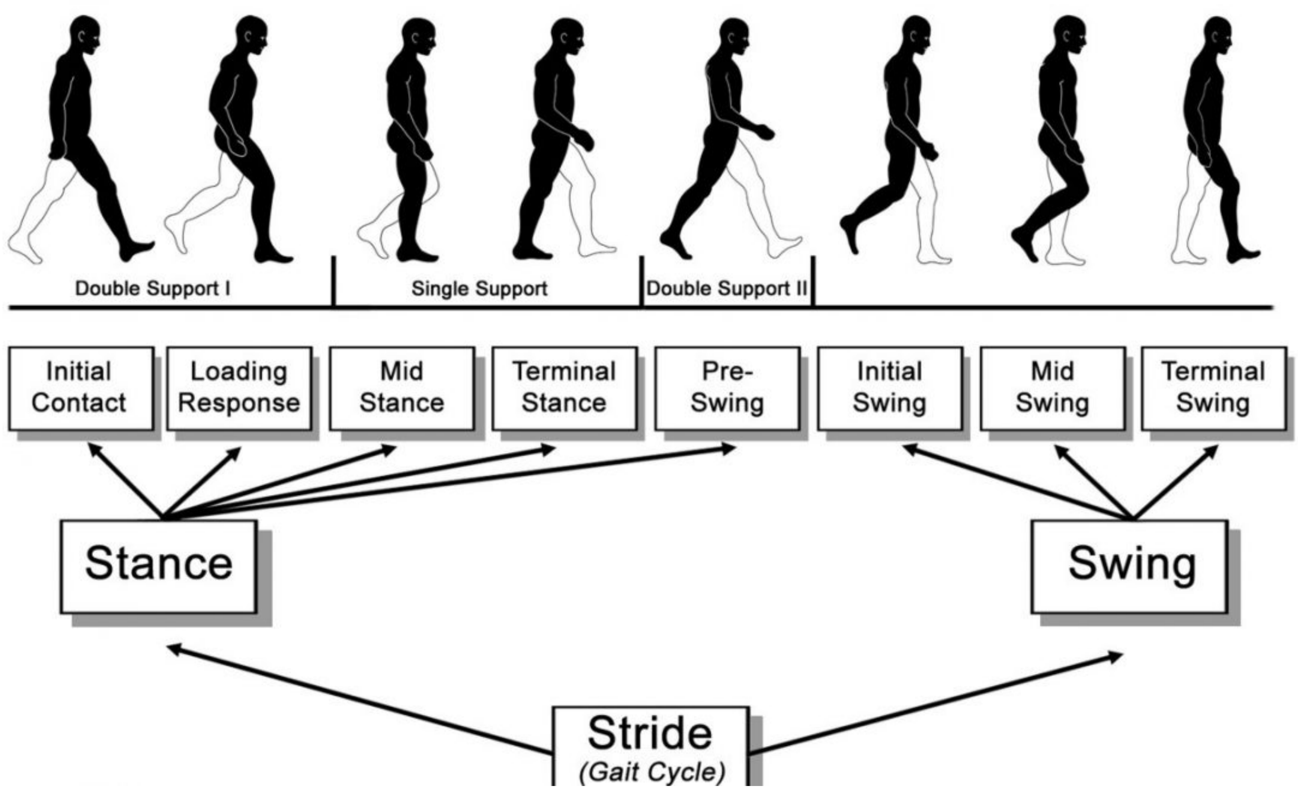


Figure 1.2: Gait cycle phases

From an anatomical point of view, after the initial double-limb support, when the opposite toe is leaving the ground, there are two possible movement: the moving forward of the knee with consequent eccentric activation of quadriceps, and the foot slap, with the consequent eccentric activation of tibialis anterior (TA). Similar activity of the two muscles allows the toe-off of the opposite foot and the reversal of fore-aft shear. In this last event the activity of intrinsic foot muscles makes it a rigid structure over which leg moves as an inherently unstable inverted pendulum, also through the contraction of gastrocnemius (LGS) and quadriceps, to initiate knee extension. During the opposite foot strike event the push off of the right leg is aided by LGS shortening until the toe off, which is characterized also by the quick changing of activation of iliopsoas muscle from eccentric to concentric, allowing the hip to move and contribute to the balancing.

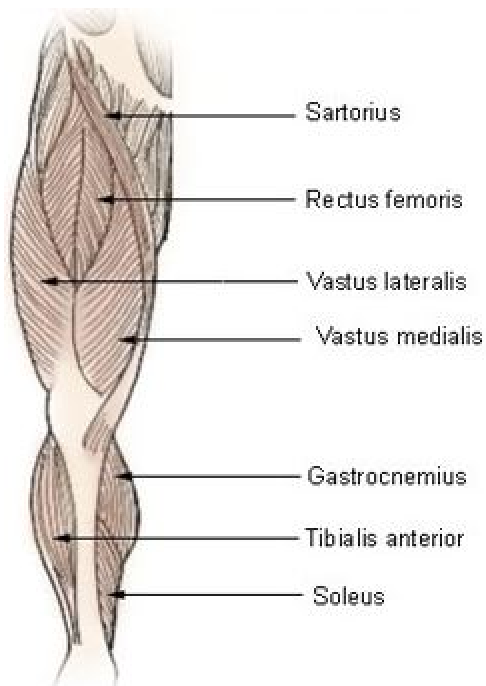


Figure 1.3: anatomy of lower limb

During the swing phase one should see the acceleration of the leg, due to the simultaneous work of several muscles as TA, that allows foot clearance, and rectus femori (RF) for knee extension, and hip flexor muscle. In the middle of the swing phase the inertial motion of the leg allows a low cost in energy activity, in fact few muscles contract, as lateral harmstring (LH), that takes also part in the consequent deceleration or terminal swing when the tibia is perpendicular to the ground, together with TA and quadriceps.

From a posturographical point of view, balancing is the ability to mantain the Center of Mass (COM) within the base of support (BOS). During gait the COM is shifted forward, the body falling is stopped by the non-weight-bearing-limb, which swings into its new position just in time. The combined vertical and horizontal motions of the COM of the body during gait describe a double sinusoidal curve. Saunders et. al in 1953 [2] described six biomechanic movements each of that if impaired could affect the COM's path smoothness, while whit the impairment of two or more of these determinants affect directly the efficacy of the walking task:

- pelvic rotation [3]
- pelvic obliquity
- knee flexion in stance [4]
- foot and ankle motion [5]
- lateral displacement of the pelvis [6]
- axial rotations of the lower extremities [7]

Another approach to gait analysis is described by Perry in 1994, highlighting four prerequisite of normal gait:

- stability of the weight-bearing foot during the stance phase [8]
- clearance of the non-weight-bearing foot throughout swing phase [9]
- appropriate pre-positioning of the foot for the next gait cycle, during terminal swing [10]
- appropriate step length [11]

features to which Gage et al. added energy conservation. [12, 13]

1.0.2 Laboratory gait analysis

Although gait abnormalities can be then assessed through a systematic observation of the gait features, measurement and laboratory analysis has been deeply improving for gait analysis. Researchers begun to elaborate on methodology for gait data collection and interpretation, in order to achieve some evidences that could support the technical and diagnostic efficacy of laboratory gait analysis. [14] The impossibility to determine why biomechanical consequences are visible in subjects with certain pathology through an observational analysis, imposes to integrate such investigation with measurements of kinetics or of muscular activity by dynamic electromyography. Chambers in [13] uses example of clinical case in which the misleading of the only observational analysis could hide different but efficient therapies: as a child with an equinovarus foot examined by a good physician by mean of his eyes will surely show a swing phase varus, and the physician should make his therapeutic decision on that, as a split posterior tendon transfer (SPLATT), all the etiology's alternatives for this problem, like the tibialis anterior spasticity during the heel off or the amount of pyramidal involvement in the activity are not considered, because in this case those are information that only by mean of specific instrumentation is possible to have.

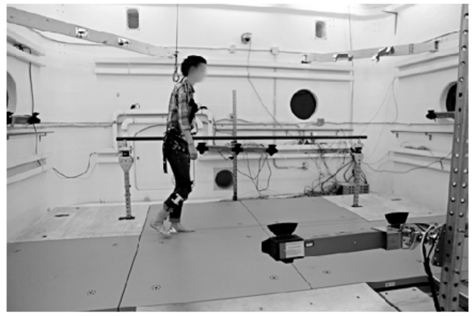


Figure 1.4: A patient underwent to a laboratory gait analysis

The main approaches in laboratory gait analysis are:

- kinematics: a camera positioned so that standardized reflecting skin markers together with a computer to process images, is a standard set used to measures the three dimension dynamic range of motion of a joint [15]
- kinetic: kinematic and anthropometric data can be integrated to a ground reaction force measurements in order to describe forces acting on a moving body [16]
- muscle activity: surface or fine wire EMG is used to measure the muscle impulses or muscle activations. [17]

Data model used in this study is a collection of parameters extracted from EMG signal acquired in order to investigate gait analysis according to the last approach of those listed. In this case there is not an integration of other approaches, just sEMG signals which pre and post processing are briefly explained in materials and methods chapter.

1.0.3 Study of muscle activity in gait analysis

Although many standard has been described in order to maximize the potentiality of use of EMG recording in the study of human activities [17, 18, 19], the monitoring of the muscle activity involved in locomotion can be still a challenging task, because as already said, has an intrinsically intra-subject variability, and such behaviour is assessed also in EMG signals and among healthy subjects.[14, 20] Also through the using sEMG, as a non invasive method that allows to quantify the muscle activity, there are some issueas:

- electrode placement [21]
- subcutaneous fat thickness [22]
- skin temperature and impedance [23]
- crass-talk [24]

The signal can be normalized to the maximum contraction for a specific muscle, [25, 26] with helping in reducing the intra-subject variability. By the way this have to be considered a difficult task with elderly people or symptomatic patients that could not perform maximal contractions because of pain and muscle inhibition or a risk of injury,[27] Knaflitz et al developed a double-threshold statistical method for the detection of sEMG signal, allowing a user-independent assessment of muscle activation intervals. It consists of selecting a first threshold z and observing m successive samples: if at least r0 (second threshold) out of successive samples are above z, the presence of the signal is acknowledged. Values of the three parameters z, r0, and m are selected to jointly minimize the false alarm probability value and maximize the detection probability for each specific signal-to-noise ratio. The setting of z is based on the assessment of background noise level, as a necessary input parameter. Furthermore, the double-threshold detector requires estimating the signal-to-noise ratio in order to fine tune r0. Has also been developed a statistical algorithm for the background noise level and signal-to-noise ratio estimation, necessary to run double-threshold algorithm. [28]

This approach allows to take in consideration the principal activation of a muscle, as the activation really necessary for the walking activity, and so if not are discarded those strides in which muscle activity pattern are infrequent. In a recent study the muscluar co-contractions were quantified by assessing the overlapping period among activations intervals of the considered muscles in the very same strides.[29] The collection and automated processing of a large number of gait cycles is the base of statistical gait analysis (SGA), a recent methodology which performs statistical characterization of gait analyzing spatial-temporal and sEMG-based parameters. Recently, SGA was further improved with the introduction of an algorithm called CIMAP (Clustering for Identification of Muscle Activation Patterns) thank to which strides can be collected and clustered according to their modality (i.e. number of activations for those strides is the same), defining principal activation as the intersection of the dendrogram cluster's prototypes.[30, 31] The main steps of CIMAP are:

- the definition of centroid for each cluster as a vector containing just 0 and 1 as the median value of discrete timings in which signal is detected, computed among element belonging to the same cluster.

- each element (i.e a vector representative the ON/OFF activation intervals for one stride) is firstly considered as a cluster centroid. After the computation of the farthest distances of every pair of elements in the considered clusters through a complete linkage method, in the end of each iteration, the two closest clusters are merged together. For distance measurements Manhattan and Chebyshev metrics are considered.

- the selection of the cutoff points based on the differences of inter-cluster distance. The first cut point is related to the first iteration in which the difference is higher than the average difference among the series. The second cut point is related to the first iteration in which the difference is higher than the sum of mean and standard deviation of the series. The third cut point is a moving average window of 5 points applied backward to the series to check when it stops decreasing monotonically. The best cutoff was identified using CUT_IND that takes into account both the intra-cluster variability and the number of cycles included in the representative cluster, for each dendrogram are so computed the three cutoff points and selected the one corresponding to the lowest *CUT_IND* value that can be expressed using:

$$CUT\_IND = \frac{\sum_{i=1}^{n} INTRA\_VAR_i / n}{\sum_{i=1}^{n} |C_i|} \quad (1)$$

where n is the number of representative clusters —C_i— is the number of cycles included in each cluster C_i, and *INTRA_VAR*i is the intra-cluster variability of the i-*th* cluster calculated by using:

$$INTRA\_VAR_i = \overline{dist(cycle_j, cycle_k)}, \forall jk \in C_i \quad (2)$$

where dist is the Manhattan distance.

- the identification of final result is based on the similarity between the centroids and the cluster elements, and the number of cycles included in the representative clusters. Specifically, for a single cluster the *CLUSTER_VAR* index is calculated as:

$$CLUST\_VAR_i = \sum_{j=1}^{p} dist(cycle_j, CLC_i) \Big/ p, \forall j \in C_i \quad (3)$$

  where p is the number of cycles included in the representative cluster *Ci*, *CLCi* is the cluster centroid and dist represents the Manhattan distance. The final value is computed as the mean value among all the clusters. The clustering result with the lowest value of the *CLUST_VAR* index was considered as the best result: low values of the index are associated to high intra-cluster similarity and/or high number of cycles included in representative clusters.

Resuming, there are several methods developed by researchers in order to extract useful information from highly correlated gait variables that comes from several fields as computer science, psychology, cognitive science, physics and engineering, to measure gait patterns, with such different factors that is still difficult to recommend one [32]. Here a statistical algorithm is presented, based on sEMG variables, using a fuzzy logic controller.

1.0.4 EMG asymmetry index

Gait asymmetry can be generally defined as the non-identical behaviour of limbs during walking. It can be detected in various clinical conditions, [33, 34, 35] meaning that for a specific experiments, control subject's gait parameters shows a symmetric behaviour, but Sadeghi et al. [36] reviewed an extensive literature and reported asymmetry of lower limbs also for able-bodied-gait, through the extraction of many different parameters as spatio-temporal, kinematic and kinetic. Gait asymmetry measurements can be either discrete or statistical and is always based on a specific gait symmetry definition, and for some researchers this factor is one reason of lack of reliability, together with the poor statistical differences between bilateral parameters. [32] Anyway the development of increasingly sensitive methods of measuring and analyzing asymmetry in movements, together with the possibility to use statistical method to compare them, continues to bring improvements to the assessment of gait symmetry. Recently a new EMG asymmetry index has been introduced [37], standing on CIMAP algorithm previously cited:

$$EMG\_ASYM\_INDEX = \sum\nolimits_{i=1}^{N} \frac{|R_i - L_i|}{N} \cdot 100\% \tag{4}$$

where R and L are the strings corresponding to the principal activations of right and left sides respectively and N is the number of elements used for representing the principal activations. The index value is equal to 0% when the two controlateral muscles are active at the same percent of the gait cycle, and is equal to 100% when, with reference to the same percent of the gait cycle, a muscle is active when controlateral is not.

Other indices that is possible to find in literature in general summarize and condense information arising from many parameters into a single indicator or score, for example Normalcy index (NI) or Gillette Gait Index (CGI). This type of index is based on the fact that using techniques from multivariate statistics it is possible to uncorrelate the discrete variables and calculate the distance in a new uncorrelated coordinate system, in fact the normalcy index is the square of this uncorrelated distance.[38] A more recent one, the Gait Deviation Index is defined as a scaled distance between the 15 gait feature scores for a subject and the average of the same 15 gait feature scores for a control group of typically developing (TD) children.[39]

1.0.5 Fuzzy logic in gait analysis

As already said the interaction in a complex non-linear fashion of gait variables is due to the intrinsic non-linear dynamics of human movement. In this framework summary statistics and waveform parameterizations often offers limited additional insight beyond that observable from bivariate plots of gait data. The complexity of human system makes conventional quantitative techniques of system analysis intrinsically unsuited, this stands on what for Zadeh might be called the *principle of incompatibility* that informally states that as the complexity of a system increases, until the precision begin to be a mutually exclusive characteristics, our ability to make significant and precise statements about its behaviour decreases. [40]

One of the main difference between classical and fuzzy set is that in classical set a data point can be comprise or not, while in fuzzy set a partial membership is available in multiple sets. This because the using of fuzzy logic involves the partitioning of the domain of a continuous variable into a small collection of fuzzy sets. [41]

Fuzzy logic is firstly introduced by the mathematician Lotfi A. Zadeh in 1964, because of his realization that more often classes of object of physical world do not have a defined criteria of membership, so in a lack of any existing mathematical framework that could cope with the complexity of biological or humanistic system he published fuzzy sets and later he extended the work on possibility theory into a formal system of mathematical logic and he introduced a new concept for applying natural language terms. [31] Zadeh assert that fuzzy logic is more than an addition to the methods of dealing with imprecision, uncertainty and complexity. It gives rather capability of "progression from perceptions to precisiated words, progression from unprecisiated words to precisiated words, progression from numbers to precisiated words and the computing with words through a normal language computation".[42] As already said membership in a fuzzy set is a matter of degree, it is precisiated through graduation, i.e association with a given fuzzy set A of a membership function, μ_A, a mapping from a space to a grade of membership space, with a value $\mu_A(u)$ representing the grade of u in A. This can be a name-based grade of membership when u is the name of an object, as for example the membership grade of Phil in a set called cold, $\mu_A(u)$ can rather be attribute-based when the grade of membership is a function of an attribute of u, for example the temperature, but it can also be a perception-based membership if u is a perception of an object, for example looking at the symptoms Phil's membership grade is 0.9. If fuzzy set is boundary oriented, differently to probability theory that is measure oriented in the sense of cardinality, fuzzy logic is both boundary and measure- oriented. It has four facets:

- Logical facet: is the fuzzy logic in its narrow sense, it may be viewed as a generalization of multivalued logic.
- Set theoretic facet: preceding fuzzy logic, it is focused on the the theory of fuzzy sets which have unsharpened boundaries, rather than on issues which relate to logical inference. In fact it is frequently referred as fuzzy mathematics.
- Epistemic facet: is focused on knowledge, meaning and imprecise information. In this facet a natural language is viewed as a system for describing perceptions. From FLe two branches diffused, the possibility theory and computational theory of perception.
- Relational facet: is focused on fuzzy relations and, more in generally, on fuzzy dependencies.

A possible application of fuzzy logic is the fuzzy control system(FCS). One of the first fuzzy system was presented by Mamdani in 1975, which applied a set of fuzzy rules supplied by experienced human operators to control a steam engine and boiler combination, but then was applied for many different industrial aims [43]. The first step in the design of a FCS is the fuzzufier, a system addressed to the processing of crisp inputs via the membership functions, i.e. given a crisp values, determines the grade of membership to each of the appropriate fuzzy sets via membership functions. Most of machine learning techniques can generate the knowledge partly owned by membership function definition, that normally can be acquired from human expertise A fuzzy rule is a statement in the form IF x is A THEN y is B where x and y are linguistic variables and A and B are linguistic values determined by fuzzy sets on the universe of discourses X and Y respectively, there can be many x_i and y_j with i and j the number of antecedent and consequent, corresponding to the IF and THEN part of the statement respectively. In this case antecedent are 5, which are then calculated and resolved in a single number using fuzzy set operations, while consequent is just the output with EMG_ASYM_INDEX's relative fuzzy sets. The basic operations on fuzzy sets are union, intersection and complement. The operator for evaluating the rule antecedent in this case is AND operator, this operator uses functions called t-norms to generalize an intersection operation of discrete fuzzy sets in the form $\mu_{A \cap U} = \min(\mu_A(x),\mu_B(x))$. t-norms is a binary operation T: $[0, 1]x[0, 1] \rightarrow [0, 1]$ which satisfy the property of commutavity, associativity, identity element and monotonicity.

There are three possible operation in Mamdani system:

- minimum intersection $\mu_{A \cap U} = \min(\mu_A(x),\mu_B(x))$. As an AND operator uses the minimum t-norm for the conjunction of the membership functions, meaning that the membership degree is the minimum of the value of the two overlapping membership functions. Other t-norms such product t-norm, Łukasiewicz t-norm and Nilpotent minimum are avaliable for the intersection.

- standard union $\mu_{A \cup U} = \max(\mu_A(x),\mu_B(x))$. As an OR operator uses the maximum t-conorms, meaning the maximum of the value of the two overlapping membership functions. Other possible t-conorms are bounded sum, algebraic sum and drastic union.

- standard complement μ_{A_C} = 1 - μ_A. As NOT operator is the degree to which x does not belong to A.

The knowledge base represented by rules is used to perform the symbolic processing of dynamic knowledge and the direct and inverse transformations between linguistic terms and numerical data. This symbolic processing is performed using the compositional rule of inference, which allows the determination of the activation of each rule consequent according to the degree of matching of the antecedent. The most common method of correlating the rule consequent with the truth value of the rule antecedent is to simply cut the consequent membership function at the level of the antecedent truth, this method is called clipping of correlation minimum. Zadeh's sup-min and Larsen's sup-prod are the most popular composition operators. The inverse mapping done in the first step brings to the defuzzification of the qualitative information coming out from the processing of dynamic knowledge, and basically has its main issue in finding a suitable many-to one mapping from a possibility distribution to the output space. The basic properties of a desirable defuzzification are consistency, section invariance and monotonicity. A consistent defuzzification method maps convex crisp sets to their centroid. It follows that the empty set and the universal set are both defuzzified to their centre and a fuzzy singleton is defuzzified to its sole non-zero truth element. Section invariance guarantees that the defuzzified value is uniquely dependent on the output space elements having a non-zero truth value. Outside this set, any modification of the fuzzy universe of discourse does not affect the defuzzified value. Monotonicity requires that, for any decrease or increase in the truth degree of a single-output space element, the defuzzification result remain unchanged, is moved away from or getting closer to that element. This implies that each single-output space element contributes to the final defuzzified value increasing with the degree of truth of that element.[44]

There are different defuzzification procedures in FL:

- Center-average defuzzifier:

$$Y_{COG} = \frac{\sum_a^b \mu_A x}{\sum_a^b \mu_A}, \tag{5}$$

  It finds the point, known as center of gravity, where a vertical line would slice the aggregate set into two equal masses. It is section invariant, monotonous and consistent, and its deterministic response curve is characterized by a smooth and continuous behaviour.

- bisector of area: The bisector method finds the vertical line that divides the fuzzy set into two sub-regions of equal area. It is sometimes, but not always, coincident with the centroid line. COG is a close variant of COA or center of area, meaning that if the output distribution is symmetrical, the two methods give identical results.

- middle, smallest or larger value of maximum: it is section invariant and monotonous, but a consequence of using this method is that the information not related to rules of maximal activation is ignored. In fact this values result the same in those aggregated fuzzy sets with one maximum value point.

Another type of fuzzy system is the Sugeno fuzzy system, introduced by Michio Sugeno in 1985. [45] The defuzzification process for a Sugeno system uses a weighted aerage or weighted sum of a few data points rather than compute a centroid of a two-dimensional area. Each rule generates two values: the rule output level z_i, which is either a constant value or a linear function of the input values $z_i = a_i x + b_i y + c_i$, where x and y are the values of input 1 and input 2, respectively and a_i, b_i and c_i are constant coefficients. For a zero-order Sugeno system, z_i is a constant ($a = b = 0$). The second value is w_i, rule firing strength derived from the rule antecedent $w_i = AndMethod(F_1(x), F_2(y))$, where $F_1(x)$ and $F_2(y)$ are the membership functions for input 1 and 2, respectively. The output of each rule is the weighted output level, which is the product of w_i and z_i. A first order Sugeno systems can be visualize thinking of each rule as defining the location of a moving singleton. That is, the singleton output spikes can move around in a linear fashion within the output space, depending on the input values. The rule firing strength then defines the size of the singleton spike. Because of the linear dependence of each rule on the input variables, the Sugeno method is ideal for acting as an interpolating supervisor of multiple linear controllers that are to be applied, respectively, to different operating conditions of a dynamic nonlinear system.

There are few example in literature of fuzzy systems tested to identify the type of gait. Rosati et al. recently developed an efficient fuzzy logic based index for assessing gait impairment as a result of the aggregation of two fuzzy systems, one based on gait phases, and another based on knee joint kinematics parameters. [44] This index is then validated by an Analytic Hierarchy Process (AHP) which is based on an hierarchical organization of evaluation criteria, a comparison of all elements taken in couple because is simpler for an expert to rely on a binary comparison, and finally a ranking of all possible alternatives based on the expressed judgement. This method differs from methods used to define other indices such as GGI or GDI because is not based on the calculation of the distance between a new subject impairment score and the avarage of the control population, in fact it applies an inference procedure. More specifically in Rosati work as in this work, that uses this method to validate EMG asymmetry index, only the construction of the not altered input MFs was based on the percentiles calculated across the population, that are more stable if one or more subjects are changed in the dataset. Moreover the using of principal component analysis for the GGI and the value decomposition for the GDI makes more complicated the understanding of which aspect or variable mostly contributed to the final result.

Chapter 2

Materials and Methods

Population

The data set is a collection of 5 parameters extracted from four population of muscles activity tracks of control patient's lower limb. During trials each subject experienced a self-selected speed walk, covering the distance corresponding to 100 gait cycles at least and wearing surface EMG probes.

The system of acquisition used during trials is STEP32 (Medical Technology, Italy), a multi-channel system that allowed to acquire surface EMG signals from 4 muscles of patient's lower limbs, bilaterally: gastrocnemious lateralis (GL), lateral harmstring (LH), rectus femori (RF) and tibialis anterior (TA). For instance the distinction of dominant side is for the right side in each case, so each muscle of those listed has been analyzed in term of different activity between dominant (D) and non-dominant (ND) sides for each subject.

The assigned class to each elements depends on the EMG asymmetry index described in the introduction chapter. An initial partition is made considering 3 classes, defined by values of the EMG asymmetry index:

- from 0 to 10 included, for the first class
- from 10 to 20 included, for the second class
- more than 20 for the third class.

The original pool is composed by 384 elements with an EMG asymmetry value given each one. Class balancing is the only operation made in this study in order to build the data set, and has been done through a manual selection of 80 observations for each class from the pool. After thi

In the Table 1.1 is shown the size of the original pool divided in three classes. According to the healthy behavior of the participant in the test and the partition made for the classes, the first class is the largest in term of number elements, and the third is the smallest. After the choice to have a perfect balance among classes, the smallest one imposed then the size of the data set, rounded to 80 elements for each class. In order to avoid overfitting the data set is finally divided in training set and test set, respectively the 75% and the 25% of the data set.

An observation is an array of 5 features chosen for the representation of the model, the formulation of each feature is presented in the next section.

Class	n elements	LGS	LH	RF	TA
First	176	56	40	36	44
Second	127	23	35	34	35
Third	81	17	21	26	17

Table 2.1: Raw data set

Class	n element	LGS	LH	RF	TA
First	60	18	15	9	18
Second	60	8	18	17	17
Third	60	15	15	16	14

Table 2.2: training set

Class	n elementi	LGS	LH	RF	TA
First	20	7	6	5	2
Second	20	4	4	7	5
Third	20	1	6	10	3

Table 2.3: test-set.

2.0.1 Features

The parameters used to develop the model are computed after a processing already seen in the introduction chapter. Resuming the strides are collected and clustered in order to discard the irrelevant ones, after this the identification and overlapping of prototypes among clusters bring to the definition of principal activations, that are vectors containing 0 and 1 depending to the activation timings. The EMG asymmetry index is then computed for each controlateral muscle couple, but in this framework is also possible to define and compute other parameters, these parameters are those used for the model as the FIS inputs:

- DIFF_AP: is the difference of the number of principal activation intervals between D and ND sides. So this number can be 0 when cluster prototypes brought to the same amount of principal activations for both sides, while it increase with the number of intervals discrepancy.
- DIFF_AS: secondary activations are interpreted as those involved in "correcting" the main gait patterns when the environment or the subject internal needs change for some

reason. Due to the complementary role, secondary activation are not present in the totality of the clusters prototypes. DIFF_AS is the difference of secondary activations timings between the collected clusters of D and ND sides, so this value can be 0 when no secondary activation are detected for both sides, or when in the two clusters there is the same number of secondary activations intervals.

- DIFF_DC: this parameter represents the difference of percentage of cycle in which there is activity between D and ND sides, it is computed as the difference of prototypes overlapping, so it can be 0 when controlateral muscle are active in the same instants, or for the same number of instants, while can be equal to 100 when muscle are active in a complementary way.

- ASI: as already said in introduction chapter, recent study addressed in methods for definition and detection of gait symmetry encouraged the using of statistical methods in order to compare existent asymmetry indexes, as fourth parameters here is so inserted an asymmetry index theorized by Schmidt et al. [46]

$$ASI\,(\%) = \left|\frac{2 \times (MA_L - MA_R)}{MA_L + MA_R}\right| \times 100\% \tag{1}$$

where MA_L and MA_R represent individual mean muscle activities, obtained during the complete gait cycle of the left and right limb respectively.

- SI: another index present in literature is defined by Burnett et al. in [47] as a ratio of root mean square amplitude during stance phase for ND and D limb.

$$SI = \left|\frac{(RMS_N D, stance)}{RMS_D, stance}\right| \tag{2}$$

here the right side is considered as the dominant limb, as a control population is considered.

Fuzzy logic controller design

In order to build the fuzzy inference system the MATLAB ® toolbox is used. For a more complex model could be very important, in terms of performances, the implementation of dimensionality reduction algorithm and feature selection, by the way the raw dataset is limited in amount of observations (raws) and features (columns) making inconvenient to think immediately to those processes.
The first thing done in this part of the study is setting the parameters of the FIS. After this there will be other four attempts of setting new parameters starting from the evidences of the previous attempts but using different strategies, anyway here the choice is to make not an automatic process for the tuning of parameters in order to find the optimal solution, but to approach to the problem in an heuristic way.

2.0.2 MFs design

There are several way to define a MF and its parameters:

- Horizontal Method: an horizontal membership function define a fuzzy set not in form of commonly used vertical membership function type $\mu = f_1(x)$ but in the horizontal form $x = f_2(\mu)$. It was elaborated by Andrzej Piegat and can be applied in case of all membership function-types. It is based on the introduction of another variable that is $\alpha_x \in [0, 1]$ called relative-distance-measure, that allows for determining of any point lying between two borders $x_L(\mu) and x_R(\mu)$. [48]

- Parwise Comparison: this method was introduced by Saaty and consists of comparing the strength by which two objects possess the quality being analyzed. A numerical scale is provided to express the relative strength of the property. Comparisons are repeated for all pairs of objects, after this a nonsymmetric full matrix of relative weights is formed. The degrees of belongingness are the components of the eigenvector corresponding to the maximum eigenvalue. [49]

- Clustering: a clustering algorithm is implemented to detect the clusters in the field database of the specific problem, then the clusters are agglomerated or divided to form the new clusters according to the proximity or the dissimilarity between objects. Secondly, the new clusters are projected into the domains of the control variables to form the intervals and the intervals could be used for determining the leftmost values and the rightmost values of the triangular membership functions. Moreover, the projections of the centroids of the clusters are the center values of the triangular membership functions. Finally, the membership functions of the fuzzy controller are established. [50]

- Genetic Algorithm: together with clustering is an automatic method used to define MFs. This method considers parameters of a MFs as for example the base points and center point, and after the translation in a binary value or gene and the consequent creation of a generation, it provide optimization by the calculation of fitness function, this after the process's steps: reproduction with the selection of two parent, crossover and mutation. [51]

- Boxplot: this method does not need a convergence as for the previous two, it basically uses the representation of a variable distribution to construct the MF. For each class the interval of values taken by the variable is visualized, while outliers are marked with special symbol. With this representation it is easy to see if there is a class in which some variable has special values. [52] Other researchers uses the comparison of multiple boxplots standard deviation and average to a threshold to decide if this method can be effective or not. [53]

For the design of MFs it has been started from the distribution of each feature's values among the training and test set, the fuzzy sets and the geometries of MFs are chosen based

on the percentiles calculated across the values distribution. The definiton of the limit of the fuzzy set is linked with distribution representation through boxplots, but also is strictly linked with the nature of the variable, i.e an index as SI or ASI has theoretical limit, so after this the outliers over that limit are brought to the limit value. This has not been necessary for the output even with presence of outliers, in fact after the definition of classes limits ex-ante, a column with [1,2,3] values are created and aggregated to the dataframe, and then compared with a column created in the same way after the output defuzzification.
Looking deeper at the variables boxplot of the training set, one can not appreciate a characteristic distribution among the three classes. In particular the DIFF_AP variable have 4 possible values going from 0 to 3, and its distribution in the second and third classes are the same apart from two outliers which value are considered as the right limit of the fuzzy set. Boxplot is unlikely the most accurate way to represent the distribution, in fact here is choice to attribute two triangular MFs as shown in the next figure, standing on what this variable means in terms of discrepancy in lower limb muscles activities. Being the fuzzy logic based on linguistic variable, here the difference in principal activations among the two sides is considered 'low' from 0 to 1.5, while an 'high' difference is considered when it is from 1.5 to 3, with a linear dependence of the membership degree.

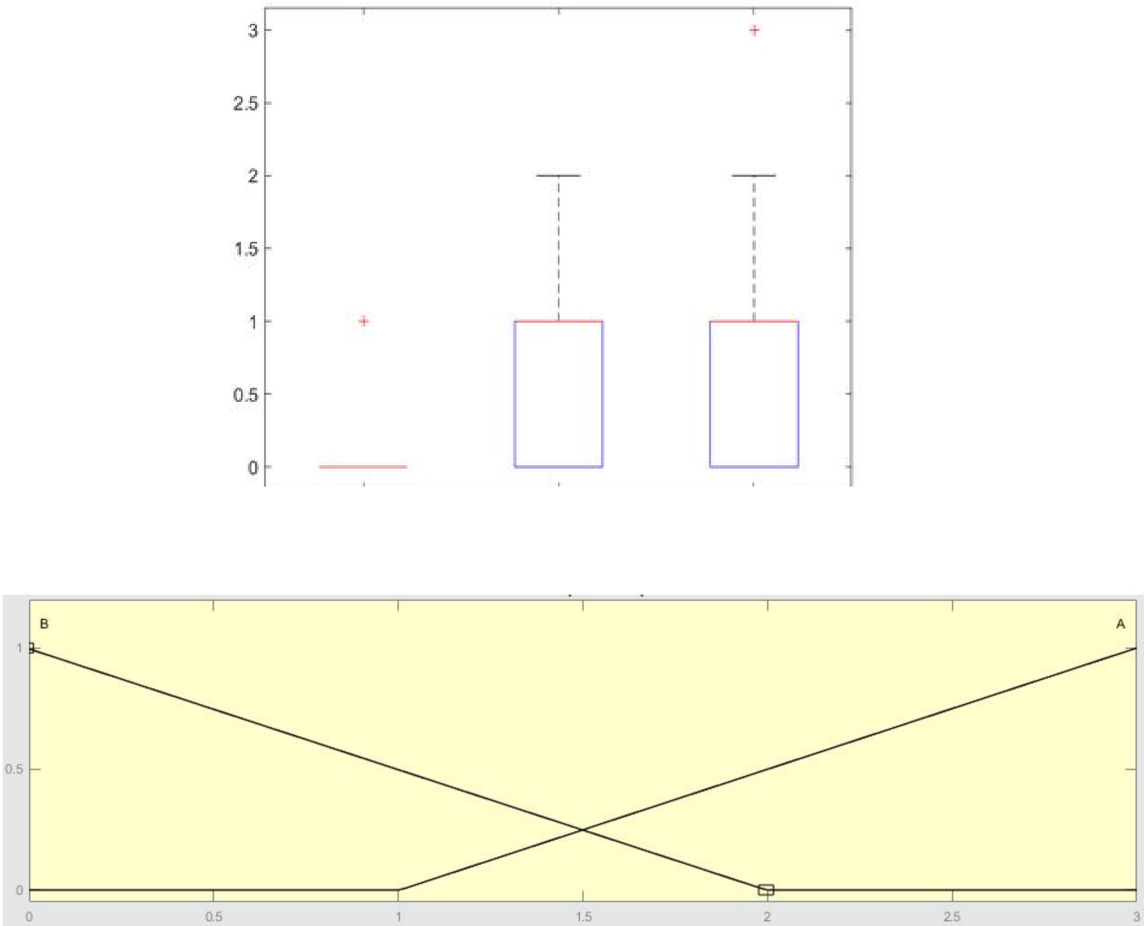


Figure 2.1: boxplot of first variable DIFF_AP among the classes, starting from the first class on the left, and relative MF

For the second variable DIFF_AS, the principle is the same but there is a certain possibility to discriminate three triangular MFs, because of the characteristic behaviour of the second class values. Anyway is decided to try two different conformations, with two or three triangular MF, choosing the one that give the best performance, that is the second one. The conformation with two MF is basically the equal to the one of previous variable, with the difference of taking the point of separation in 2.5. The presence of outliers here is indicative of the presence of a strong discrepancy in difference of number of secondary activations in cycle rows, but still empirically here is stated that an 'high' difference in number of secondary activations for a control subject can be limited to 6.

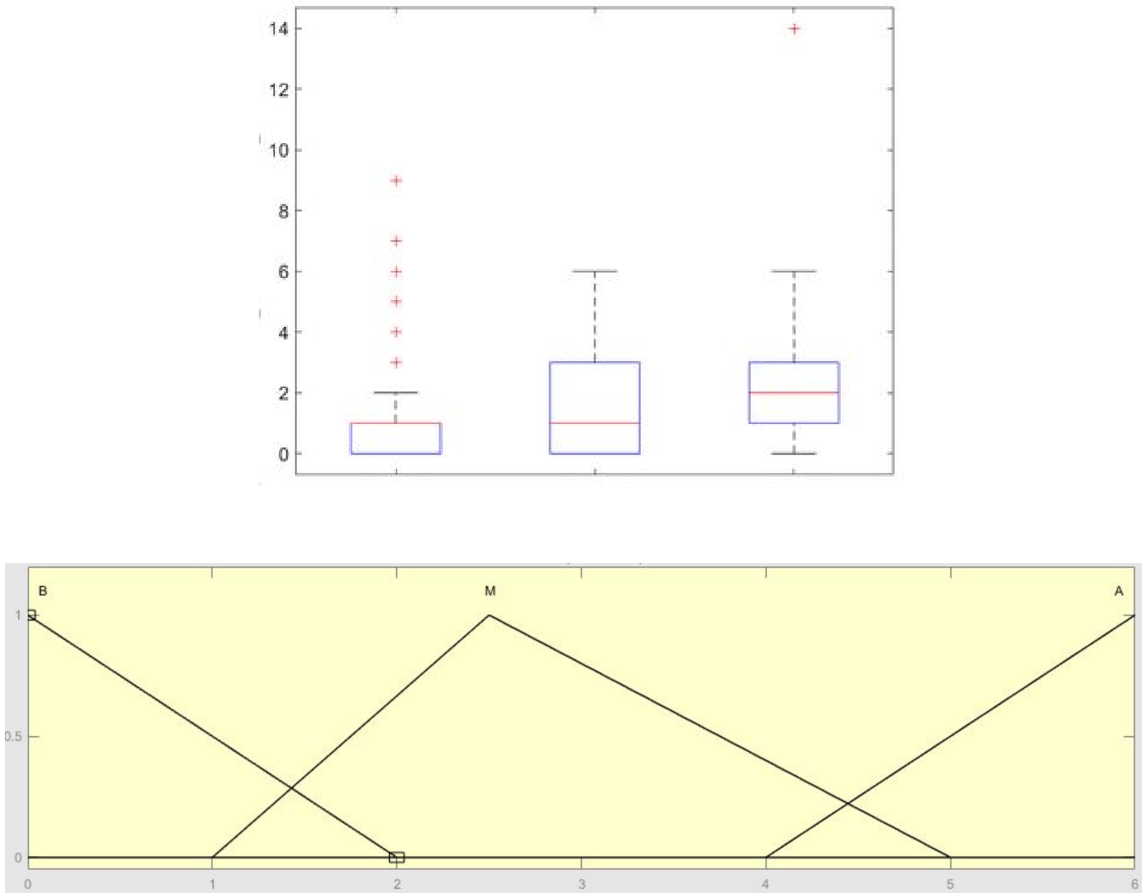


Figure 2.2: boxplot of second variable DIFF_AS among the classes, starting from the first class on the left, and relative MFs

The third variable has a different distribution among classes. It is in fact shown in figure how the upper whiskers of boxplots has a stair behaviour, while the bottom whiskers lies on the same value, that means that the third class's values present a more spread distribution. This makes a more stable indexing of the fuzzy sets, that is actable in a trapezoidal MFs geometry.

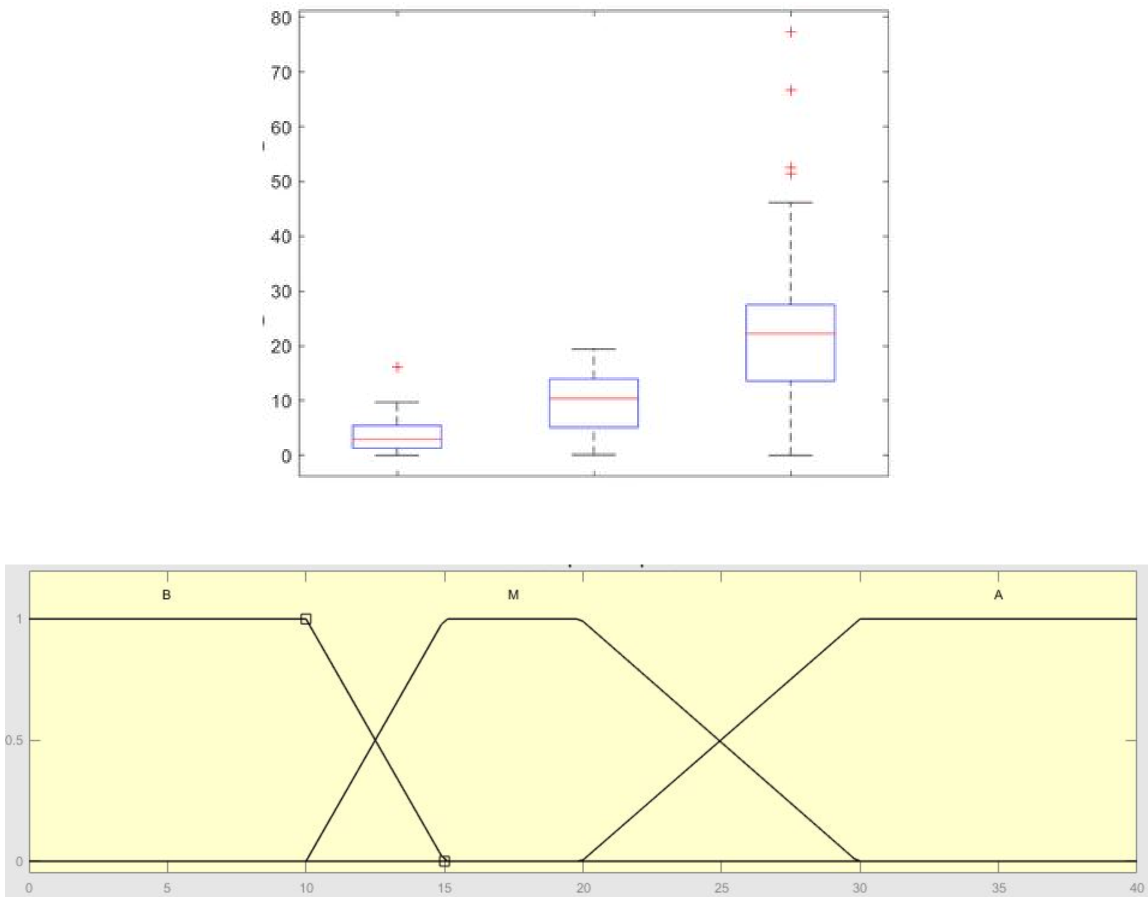


Figure 2.3: boxplot of the third variable DIFF_DC among the classes, starting from the first class on the left, and relative MFs

For the ASI variable once again is chosen to use a range of the fuzzy sets that is coherent with the nature of the variable, even if the upper whisker of the boxplot of the variable objects standing for the 3 class is over the limit that is 100. Also for this variable two conformations, one with two MF that intersect on 50 and the other shown in the figure, are compared. Here performances are evaluated, with an improving after the insertion of the MF expressing the membership degree of a middle range of ASI.

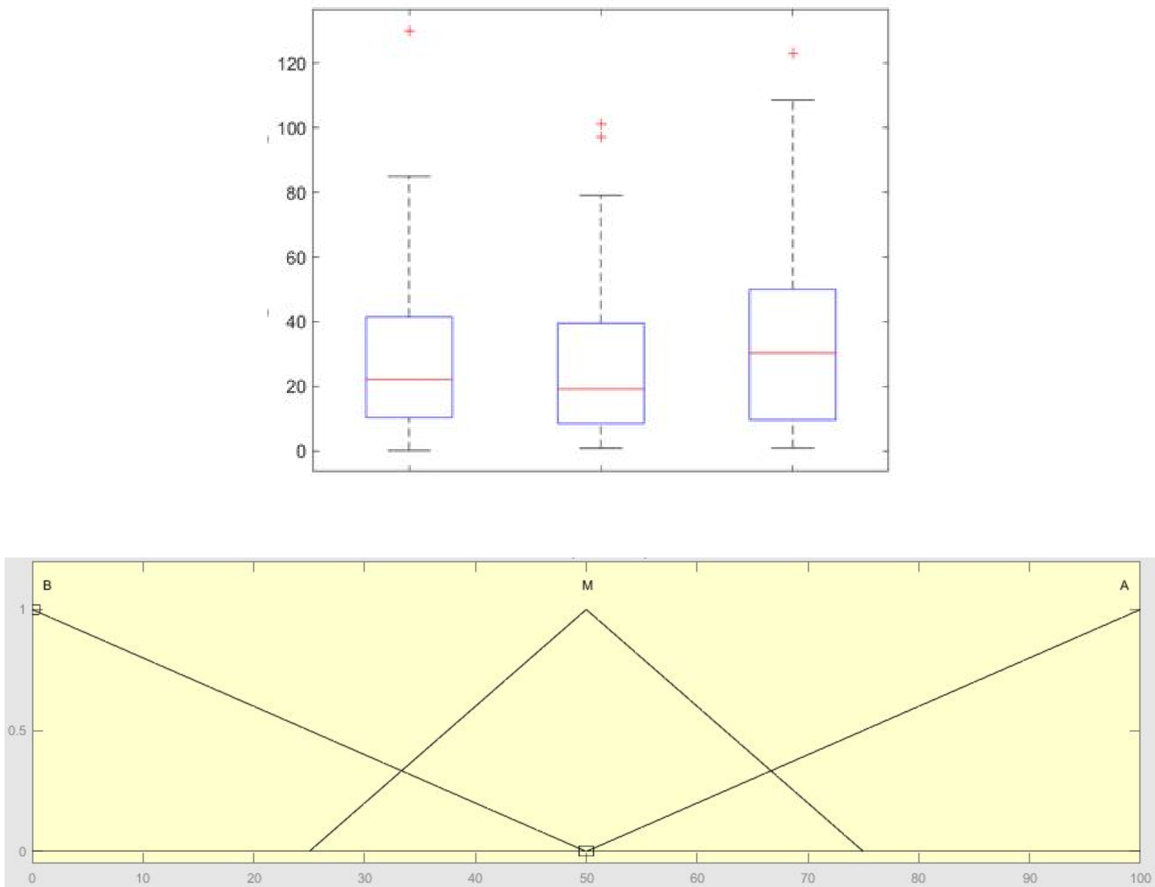
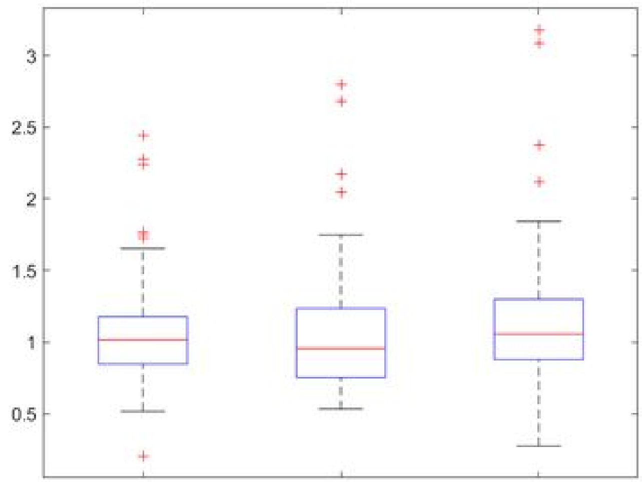


Figure 2.4: boxplot of the fourth variable ASI among the classes, starting from the first class on the left, and relative MFs

Finally the design of the SI variable MFs is conducted. This variable can expresses a negative or a positive asymmetry, giving information on which side's activity mostly affect the symmetry, in fact a perfect symmetric gait this index takes the value of 1, so the linguistic variable ASN and ASP stands for the negative asymmetry and ASP stands for the positive one. But this information is useless for the design of a fuzzy logic. Anyway the values distribution show how the median of the boxplots of the first class's values lies really close to 1 while the other two differ. Also for this variable the third class is spread on almost all the range. The limit of this fuzzy sets is necessary linked with the nature of the variable.



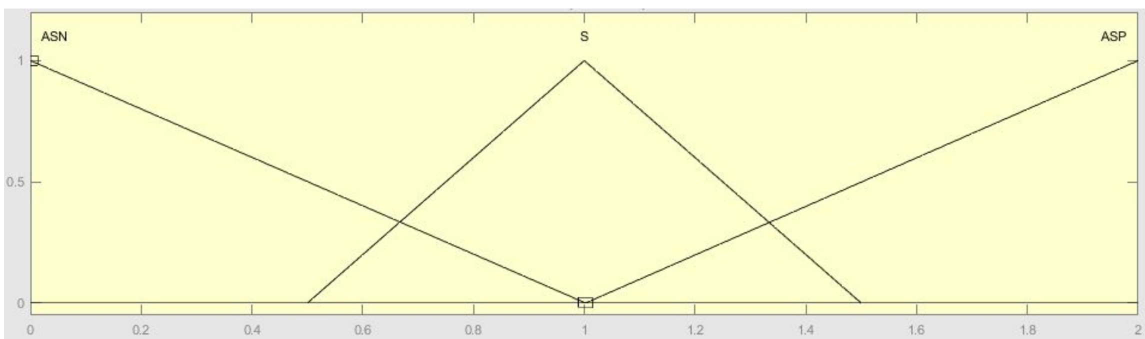


Figure 2.5: boxplot of the fifth variable SI among the classes, starting from the first class on the left, and relative MFs

2.0.3 Fuzzy rules design

Also for the extraction of rule there are several methods in literature. Mostly used are fuzzy clustering, genetic learning algorithm, particle swarm optimization and other supervised learning techniques, J. Wilson in [54] harmonizes the description of this algorithm with the possible application, and important criteria for rules quality assessment that find wide agreement. In particular he distinguish criteria that are:

- distinguishability of the space of membership function linked with a clear linguistic term
- coverage or completeness of the entirety of the universe of discourse
- complementarity of the membership to all rules, which sum should be one

This criteria allows to deal with complex problems with more accuracy but with a "low-level interpretability", linked with data representation, Other crucial aspects of human-centric modelling are those that allows to fully exploit fuzzy logic as a predictor, understanding how the model gives a particular prediction. This criteria are:

- simplicity of the model, that could mean low number of rule
- readability of the rule, that should be understandable to a human, for example putting the limit of conditions to around maximum of seven
- transparency of the meaning of the conclusion of the rules for a human reader.

The construction of fuzzy rules in this study is made through the implementation of hierarchical clustering algorithms, computed for each class. The principle is to partition the initial crisp data. With the aim to improve the creation of representative clusters a min-max scaling is done in order to have each value in the interval $[0, 1]$. One of the algorithm used is dendrogram, an algorithm that is used for its graphical representation of the clusters in which on the x-axis is expressed the cluster distance while on the y-axes is expressed the hierarchical level of aggregation, which also define the partition. In matlab is possible to implement this hierarchical clustering through the function *linkage*, with the possibility also to define the linkage method and the distance metric. The linkage creteria expresses which distance to take in consideration in the agglomeration of clusters, there are possible solutions, here is chosen to use the complete linkage method that is the largest distance between objects in the two clusters, caluclated using the cityblock metric:

$$d(r, s) = max(dist(x_{ri}, x_{sj})), i \in (1, ..., n_r), j \in (1, ..., n_s) \quad (?)$$

where x_{ri} and x_{sj} are the ith and the jth objects in cluster r and s respectively, while n_r and n_s are the number of objects in those cluster.
Another method used in this the centroid linkage method that uses the Euclidean distance between the centroids:

$$d(r, s) = \parallel c_r - c_s \parallel \quad (?)$$

Where c_r and c_s are the clusters centroids.
After the definition of a cut off point the centroid of each cluster is used for the creation of the rules. Centroids in this case are a 5 dimensional rows that represent the cluster depending on the method chosen. The number of rules are variable among the classes, The performances are evaluated through the plot of confusion matrix. After the evaluation of the performances a manual tuning is applied changing the number of MFs or the ranges.
When elements are clustered just the clusters with a certain number of elements and centroids that give a unique solution in the RB are considered.
Resuming, in this first attempt, the model is trained after some pre-processes linked with the design of fuzzy controller. A possible pipeline of pre-processing firstly states if the initial crisp dataset is scaled for the creation of rules, then which couple of linkage-metric method is used between centroid-euclidean and complete-cityblock for the hierarchical clustering. The pipeline that gives best performance so far is *scaling—complete-cityblock*. By the way this

process has to be putted in parallel with the fuzzy set definition, because only after this task a crisp centroid can be translated in the linguistic antecedent referring to a membership function activation. For this reason after changing in conformation of a specific fuzzy set, all the pipelines are implemented again and performances are evaluated. The performance of this first FIS on the training data shows how the system tend to classify in second class most of the objects of the first and third class,

FIS RULES		
CLASS	N° rules	N°clusters
first	6	8
second	7	10
third	11	11

8	3	
48	50	34
	2	12

Figure 2.6: sizes of rule sets obtained from hierarchical clustering in the upper image, and the confusion matrix with the performance of the first design of FIS, on the y-axes is possible to see the predicted class's objects while on the x-axes is possible to count the real class's objects

2.0.4 Five rules set

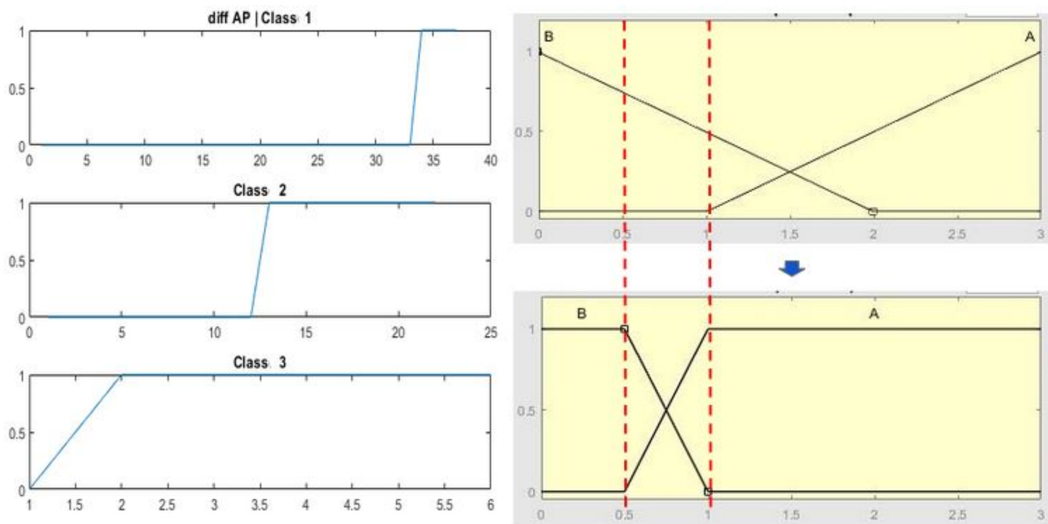
In this part of the study is implemented an elementary system with just five rules that could express the simplest combination to deal with the problem. The variables aren't dependent among them, by the way is simple to think that if an observation is composed by crisp value that if fuzzified activate the first MF 'low' for the first four variable and the 'S' standing for symmetry for the last variable, then the output of a rule with this antecedent should give a 'low' asymmetry as output. The same principle can be applied to the second rule consequent, and so on, looking at the MFs configuration for each variable and output the set represented in the next figure is composed. Obviously a system so made is not able to do a correct and complete classification the dataset, but this rule set want to be a base on which optimize input sets, and to which new rule will be added.

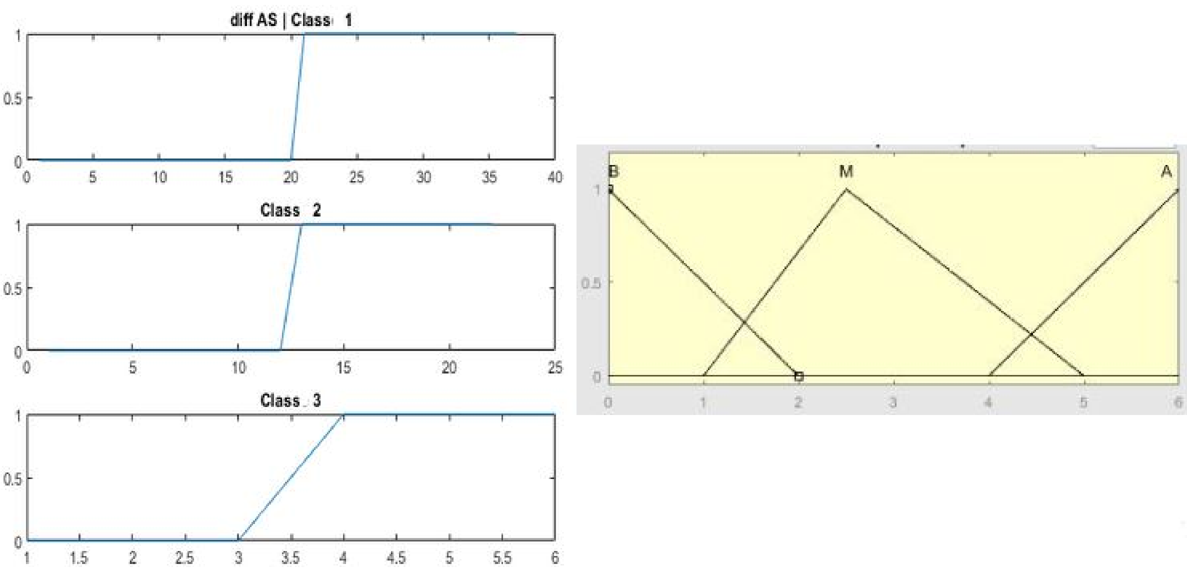
DIFF AP	DIFF AS	DIFF DC	ASI	SI	**OUTPUT**
B	B	B	B	S	**B**
A	M	M	M	ASN	**M**
A	M	M	M	ASP	**M**
A	A	A	A	ASN	**A**
A	A	A	A	ASP	**A**

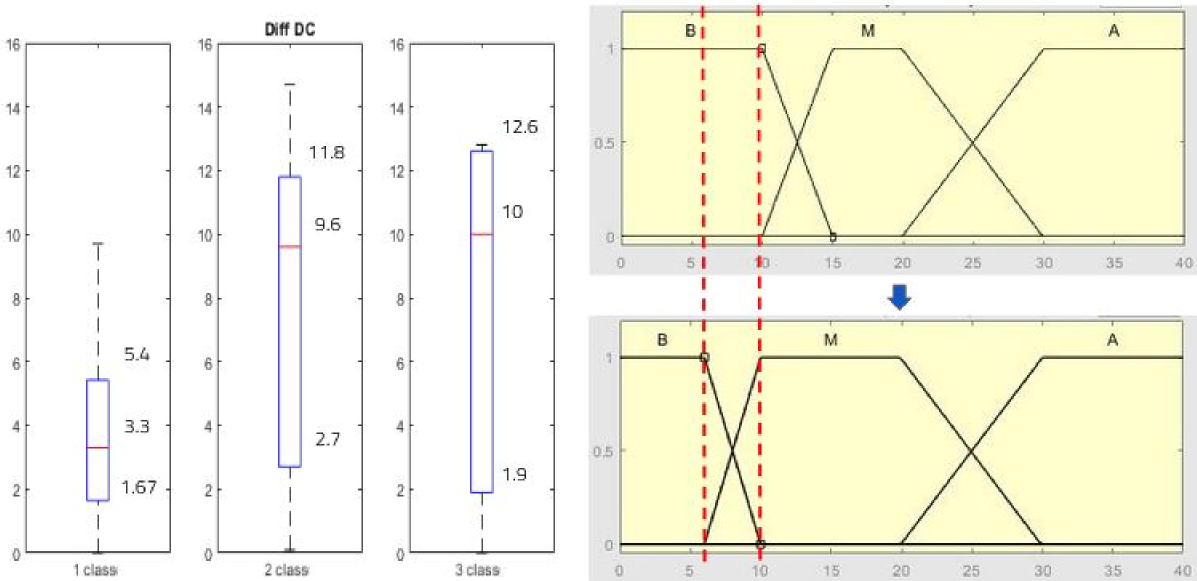
37	22	6
		2
		2

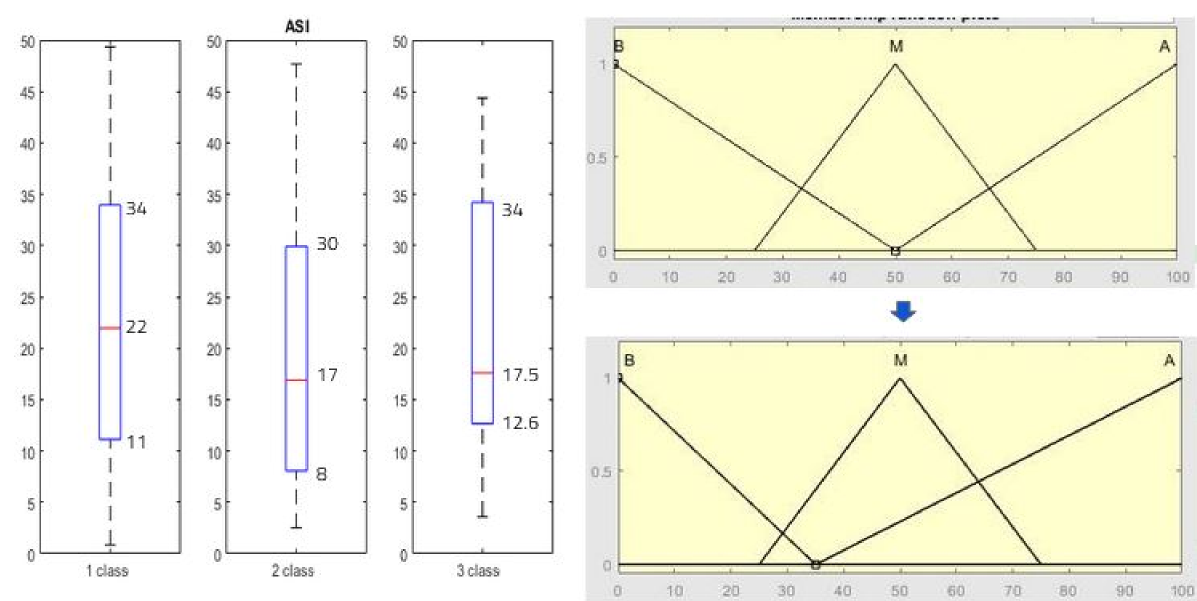
Figure 2.7: five rule set and relative performance with previous fuzzy sets, non-classified elements are 23 for the first class, 38 for the second and 50 for the third

This performance show a classification basically concentrated on the first class, so the distribution of the true positive elements of the first class and the elements of the second and third class classified in the first are represented in order to see what change can be operate to the MFs to improve the performances of classification.


diff AP | Class 1
Class 2
Class 3
B
A


diff AS | Class 1
Class 2
Class 3
B
M
A


Diff DC
1 class
2 class
3 class
5.4
3.3
1.67
11.8
9.6
2.7
12.6
10
1.9
B
M
A


ASI
1 class
2 class
3 class
34
22
11
30
17
8
34
17.5
12.6
B
M
A

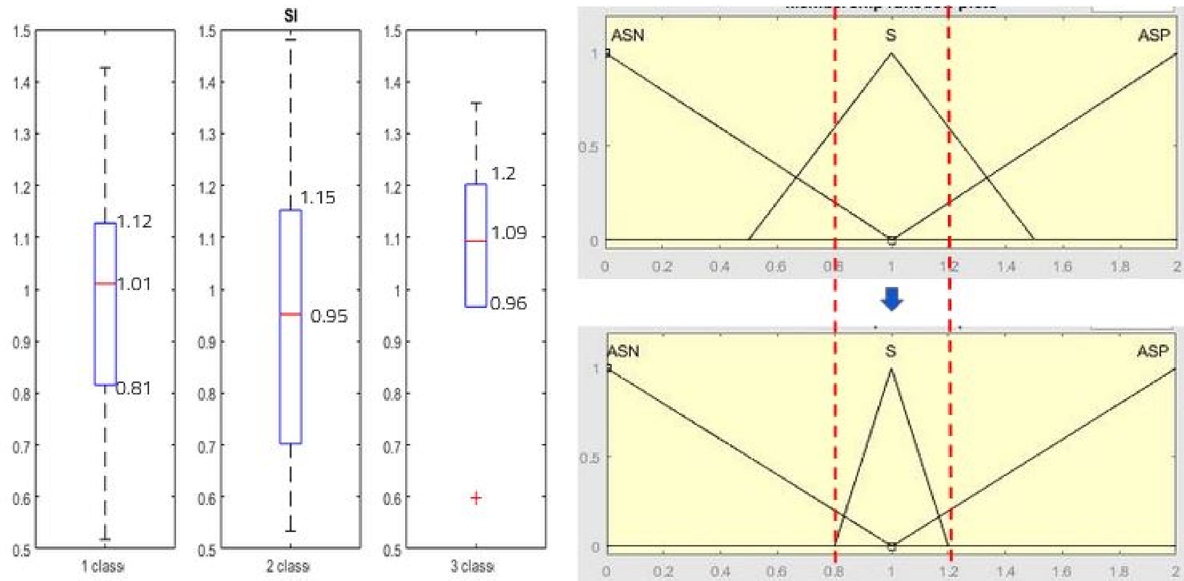


Figure 2.8: variable distribution among misclassified objects for each class and relative MFs modifications

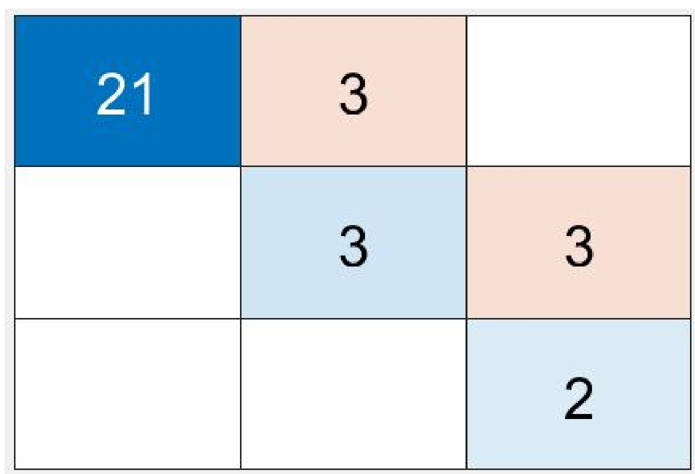


Figure 2.9: performances of the new system, with 39 non-classified objects for the first class, 54 for the second and 55 for the third

A more reasonable performance are obtained, but the non-classified object are still too much, a new process with the same method previously used is implemented to explore the distribution of the third class non-classified objects, which express the most asymmetric behaviour of the element in the dataset.

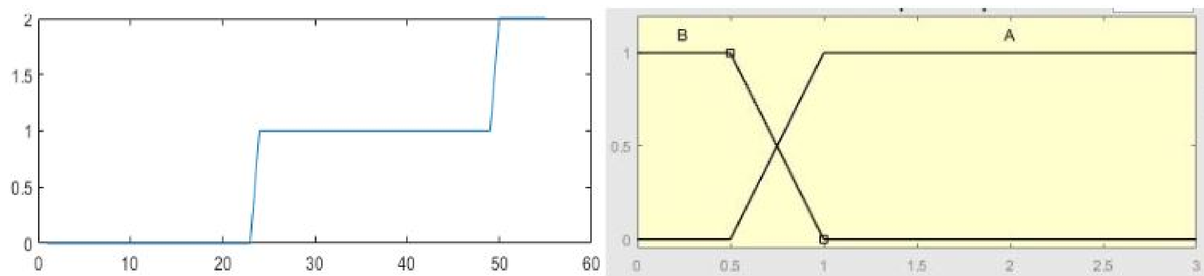


Figure 2.10: distribution of non-classified objects of the third class for the variable DIFF_AP

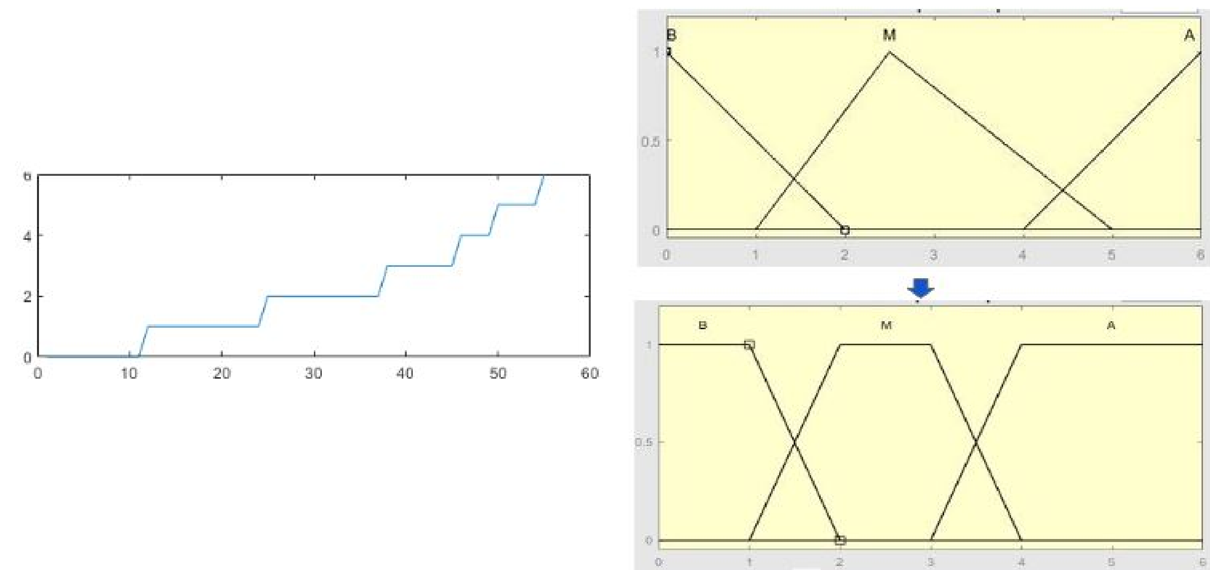


Figure 2.11: distribution of non-classified objects of the third class for the variable DIFF_AS

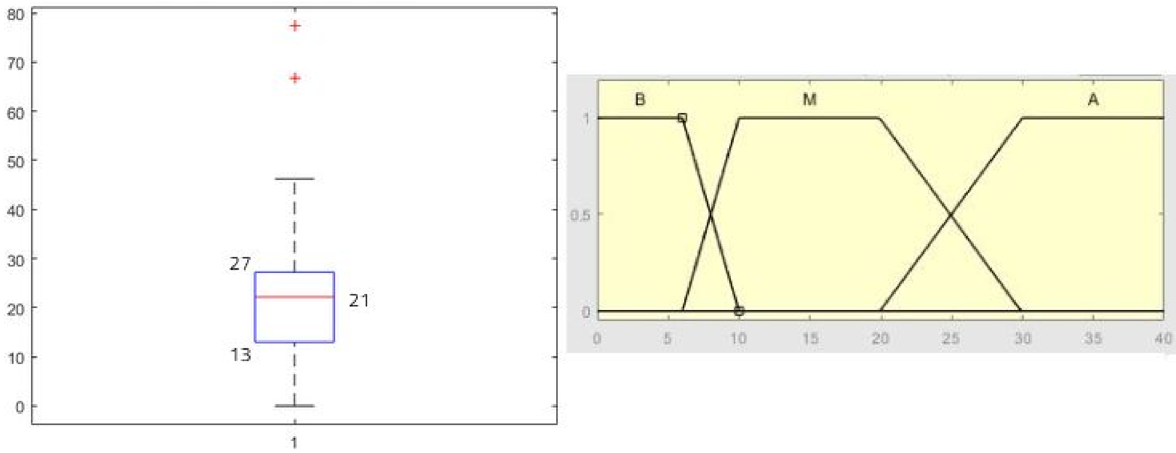


Figure 2.12: distribution of non-classified objects of the third class for the variable DIFF_DC

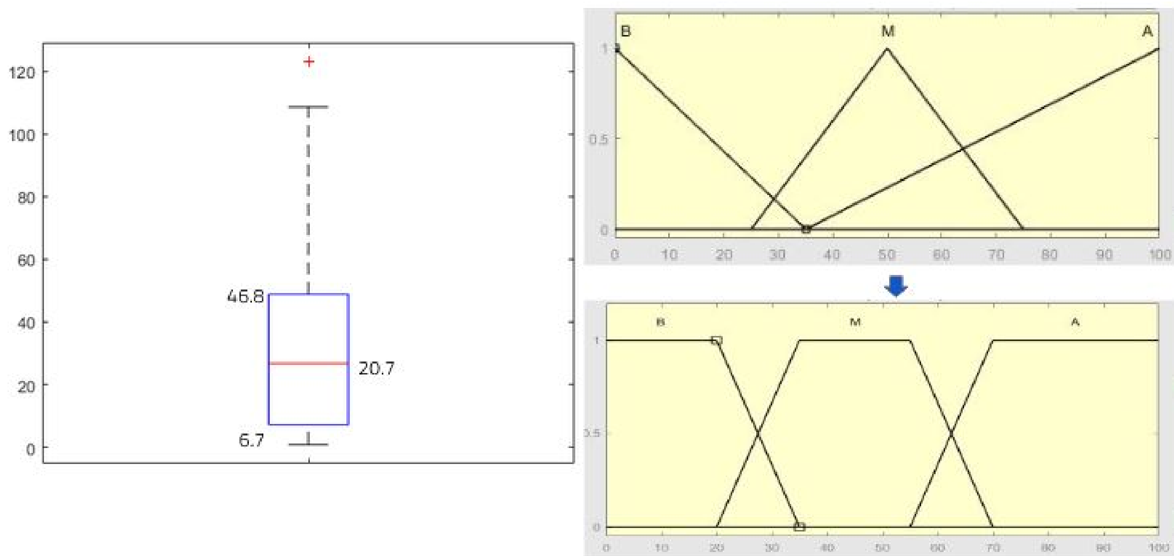


Figure 2.13: distribution of non-classified objects of the third class for the variable ASI

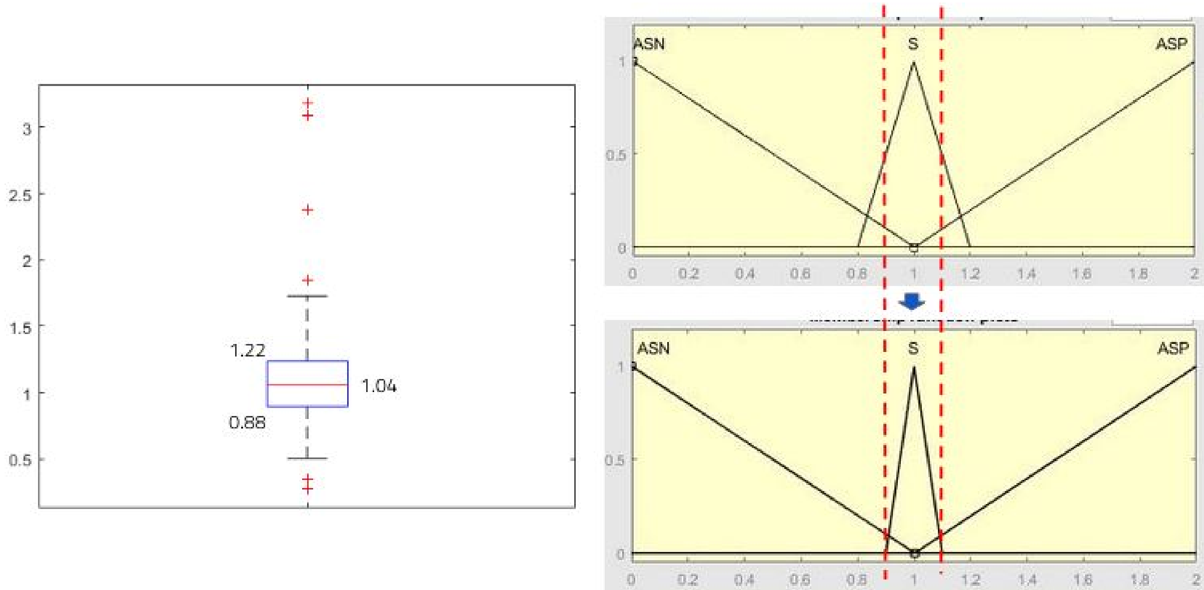


Figure 2.14: distribution of non-classified objects of the third class for the variable SI

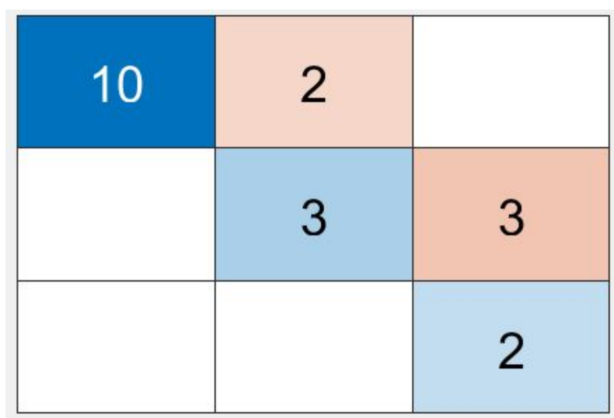


Figure 2.15: performances of the new system, with 50 non-classified objects for the first class, 55 for the second and 55 for the third

Starting again from the performances of the last classification, a new rule set is created and added to the base set of five rules, through the clustering of non-classified object of all the classes. The method used is still the hierarchical dendrogram clustering algorithm.

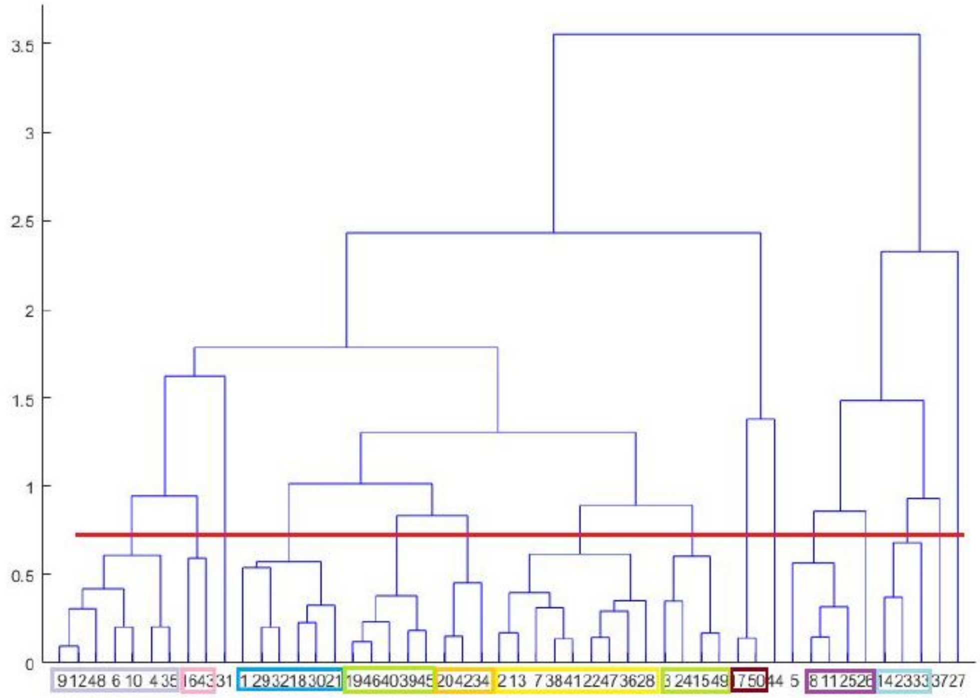


Figure 2.16: cut off of dendrogram of element belonging to the first class brought to the detection of 15 clusters and consequent extraction of 9 rules, because clusters with one element and repeted rules are not considered

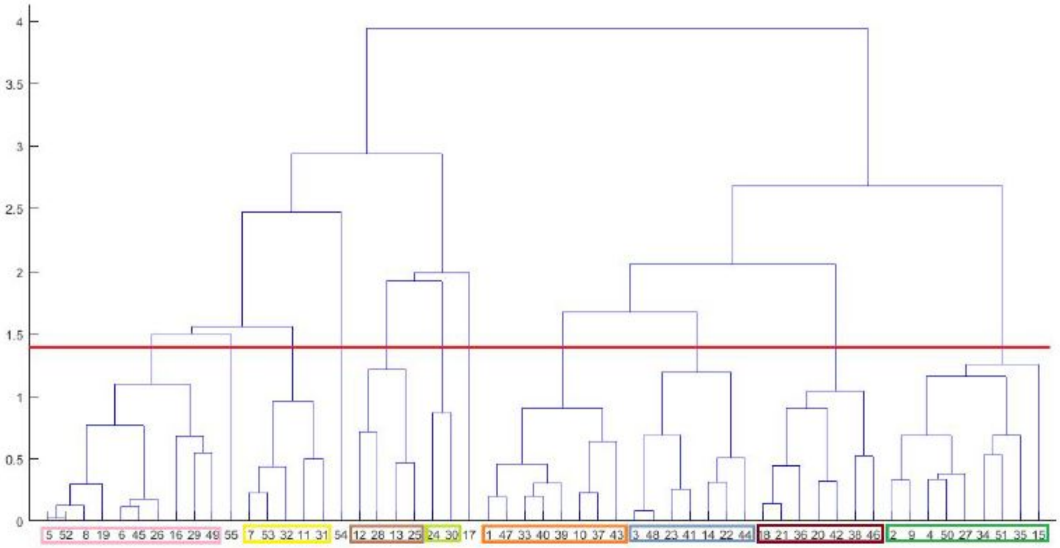


Figure 2.17: dendrogram of element belonging to the second class. The cut off brought to the detection of 11 clusters and the extraction of 7 rules

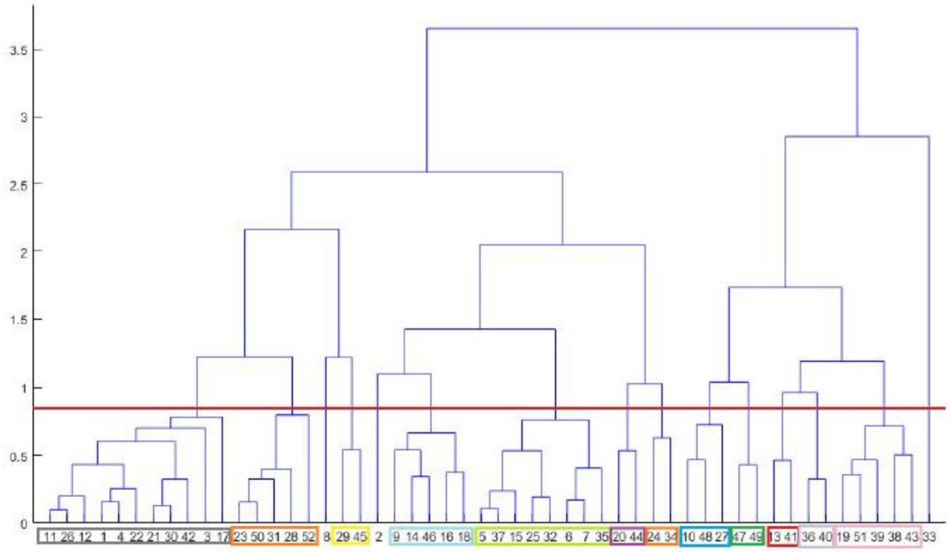


Figure 2.18: dendrogram of element belonging to the third class. The cut off brought to the detection of 11 clusters and the extraction of 7 rules

1

diff_AP	diff_AS	diff_DC	ASI	SI
B	B	B	M	ASN
B	M	B	B	S
A	B	B	B	S
B	B	B	M	ASP
A	A	B	B	S
B	A	B	B	S
B	B	B	B	ASN
B	B	B	M	S
B	B	B	A	ASP

2

diff_AP	diff_AS	diff_DC	ASI	SI
B	A	M	B	S
B	B	M	B	S
A	B	M	B	S
A	M	B	B	ASP
A	B	B	A	ASP
B	B	M	M	ASP
B	B	B	B	ASN

3

diff_AP	diff_AS	diff_DC	ASI	SI
B	B	B	M	ASP
B	B	A	M	S
A	A	A	B	S
A	A	A	B	ASN
A	B	M	A	ASN
A	B	M	M	ASP
B	A	M	M	ASP
B	B	M	A	S
A	M	B	M	ASN
A	M	B	B	ASP
A	B	M	B	ASP
B	M	M	B	S
B	M	A	B	S

Figure 2.19: rule set extracted from dendrogram of non-classified object of each class, markers highlight the equal rule extracted of different classes, that bring inaccuracy

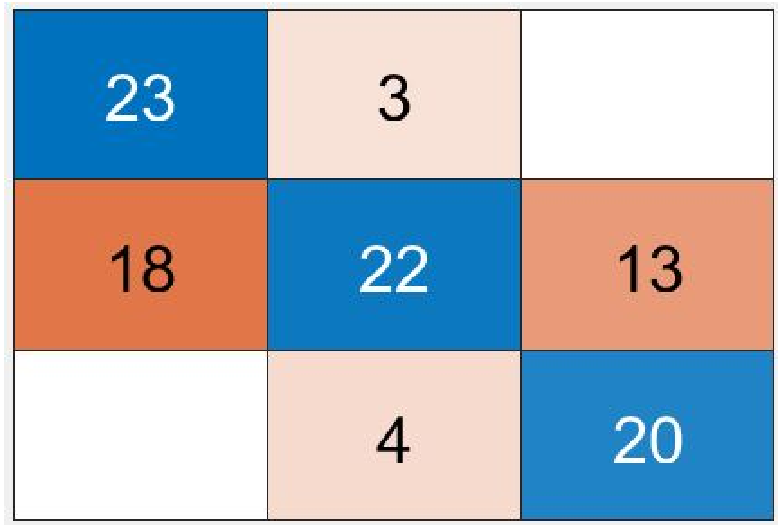
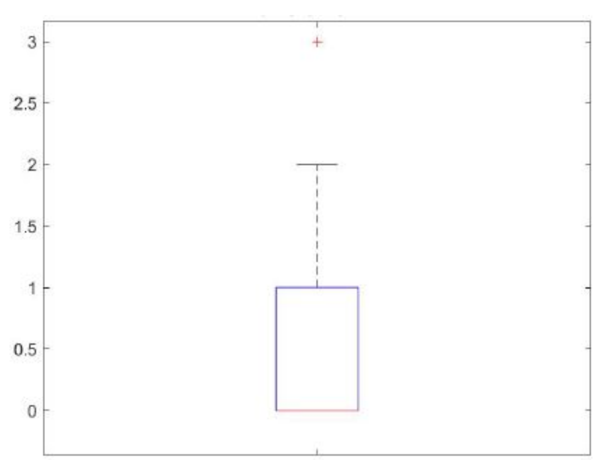


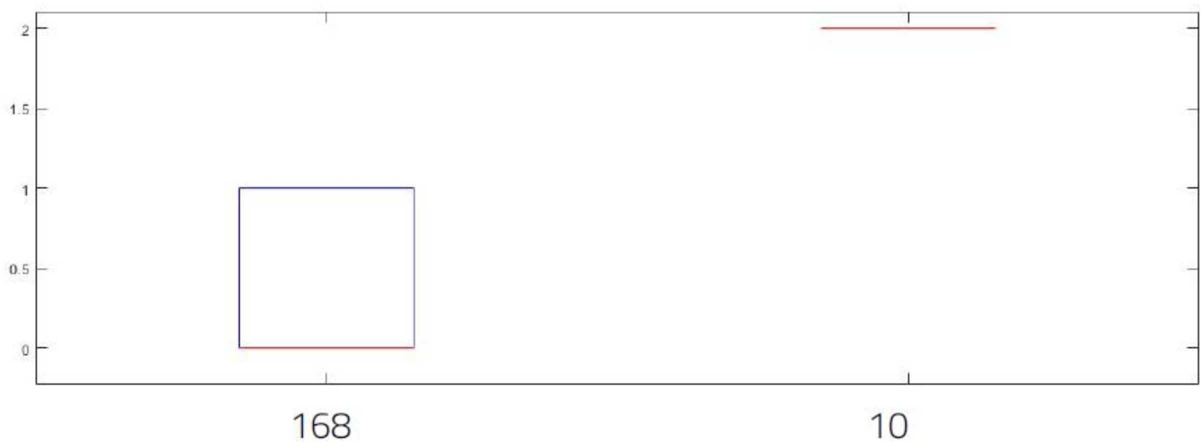
Figure 2.20: performances obtained from last system, on the y-axes there is the predicted class's element while on the x-axes there is the true class labeled object

2.0.5 Clustering of single feature

The risk in bias when using the value distribution for the creation of MFs addressed this study to the clustering of the single features in order to evaluate a possible dissimilarity with a number of clusters equal to the number of MFs. The first thing done here is a clustering of original dataset observation's variables, after this a clustering with the saturation of outliers on the limit of fuzzy set range is implemented for each variable, except for DIFF_AP variable for which the saturation to the upper whisker has been necessary. Finally another clustering is implemented, after outliers of initial distribution removal for each feature.




3
2.5
2
1.5
1
0.5
0
168
12


2
1.5
1
0.5
0
168
10

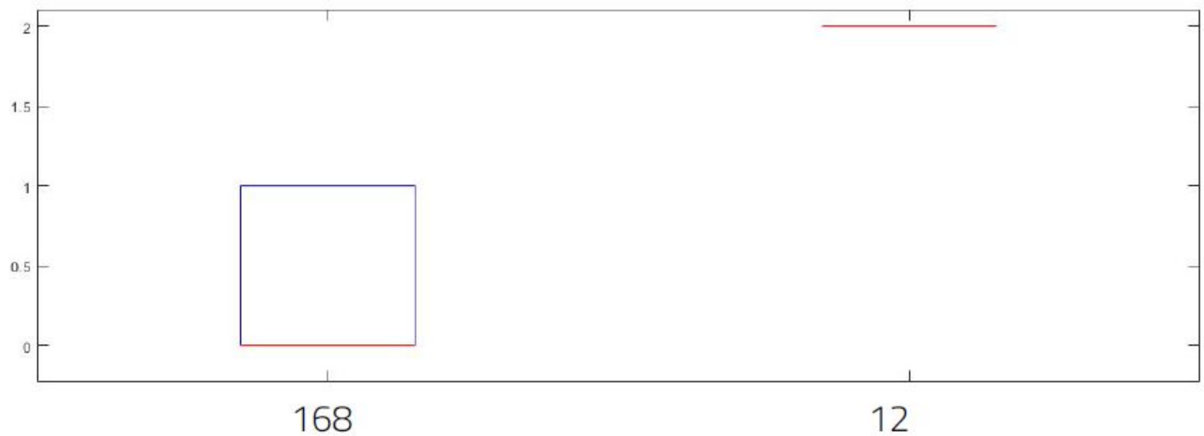


Figure 2.21: the first figure represent value distribution described through a boxplot of the first variable DIFF_AP; the second image represent the clustering result with cut off to two clusters using centroid and complete method that gives the same result, on the bottom the number of element per cluster is shown; the third figure shows the clustering result with cut off to two clusters using centroid and complete deleting outliers; the last figure shows clustering results with cut off to two clusters using centroid and complete bringing outliers to the limit value

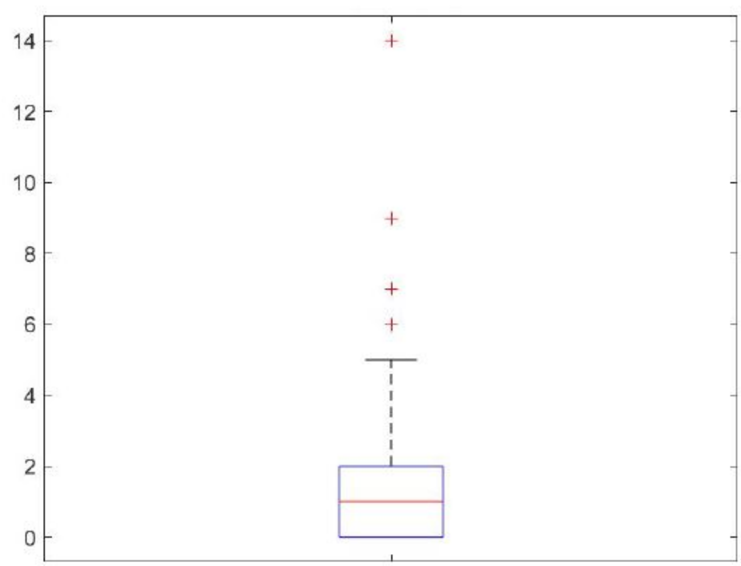
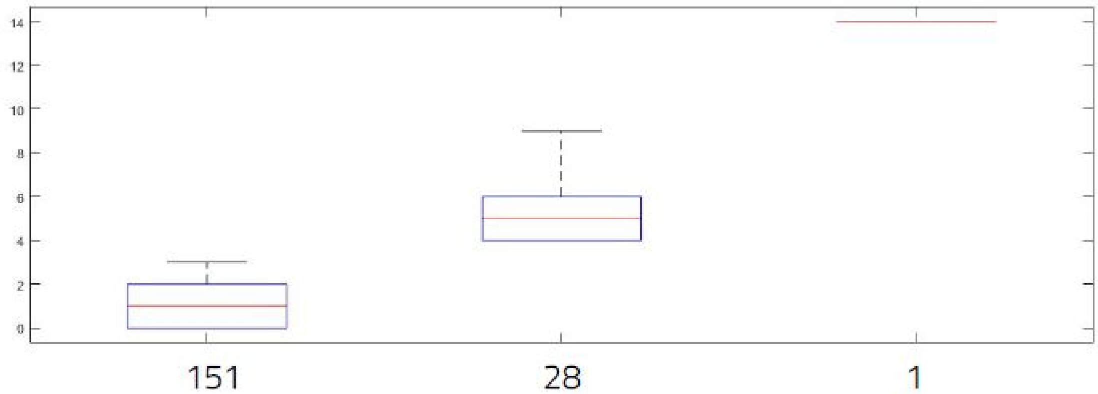
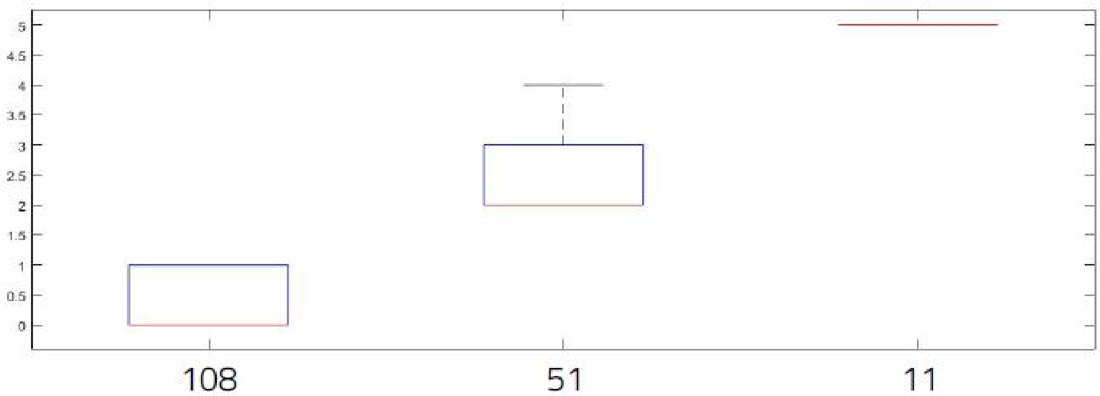
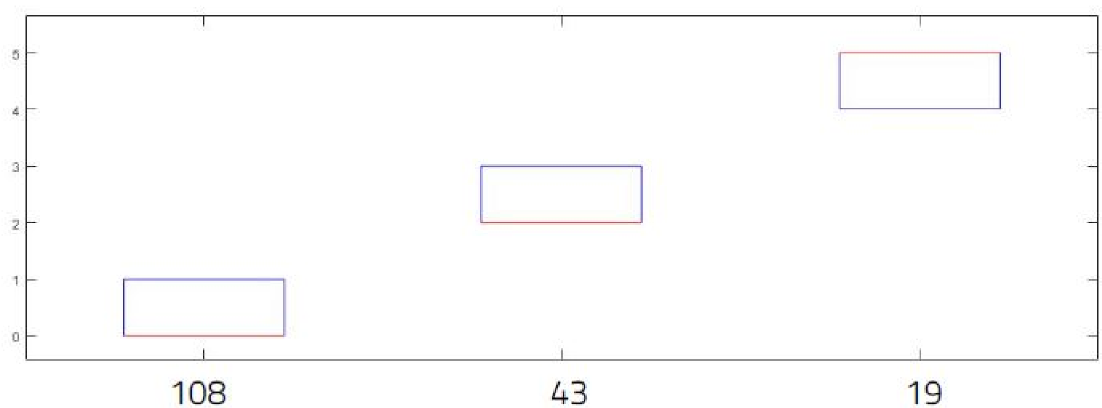


Figure 2.22


14
12
10
8
6
4
2
0
151
28
1


5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
108
51
11


5
4
3
2
1
0
108
43
19

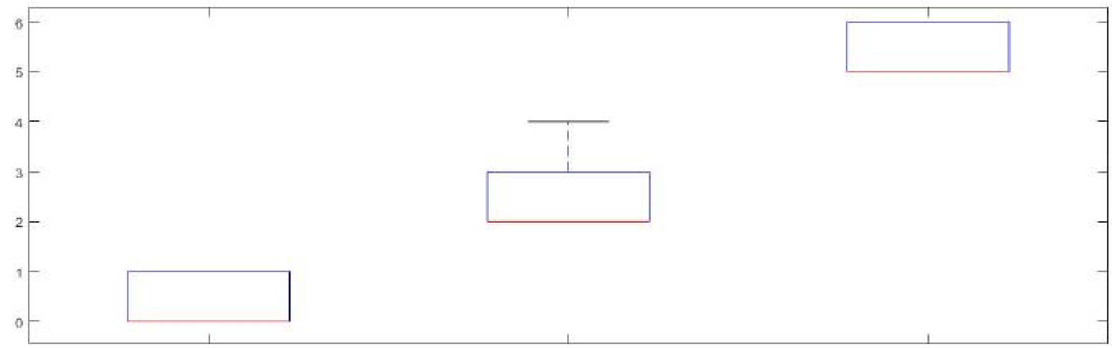
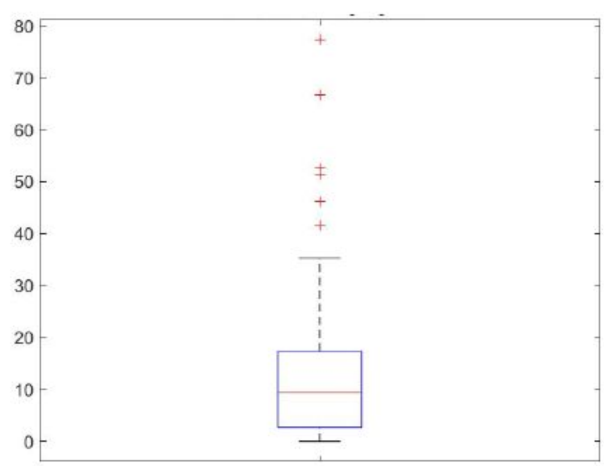


Figure 2.23: the first figure represent value distribution described through a boxplot of the second variable DIFF_AS; the second image represent the clustering result with cut off to three clusters using centroid and complete method that gives the same result, on the bottom the number of element per cluster is shown; the third figure shows the clustering result with cut off to three clusters using centroid method deleting outliers; the fourth figure shows the clustering result with cut off to three clusters using complete method deleting outliers; the last figure shows clustering results with cut off to three clusters using centroid and complete bringing outliers to the limit value (saturation)



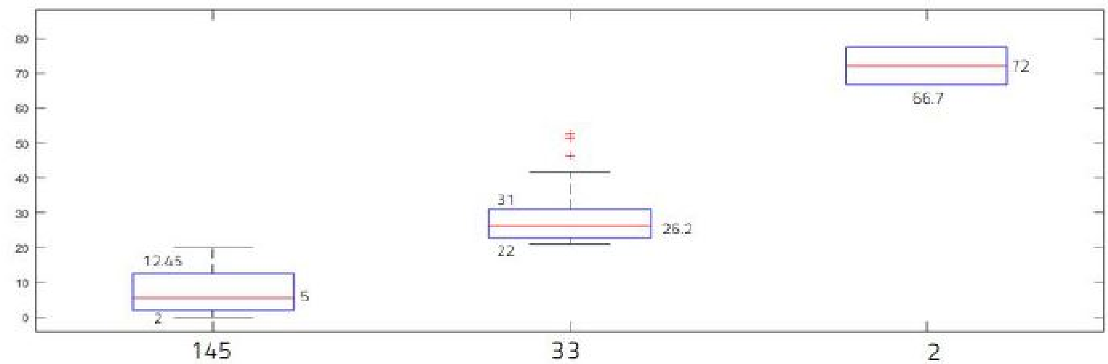
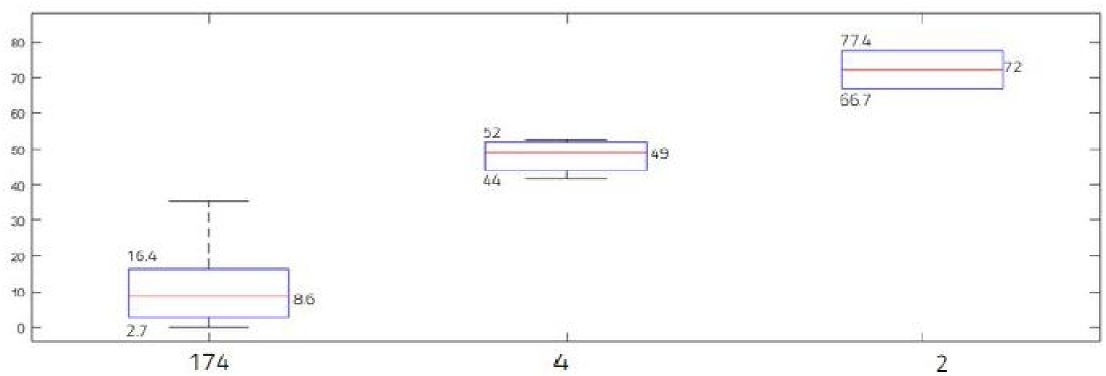
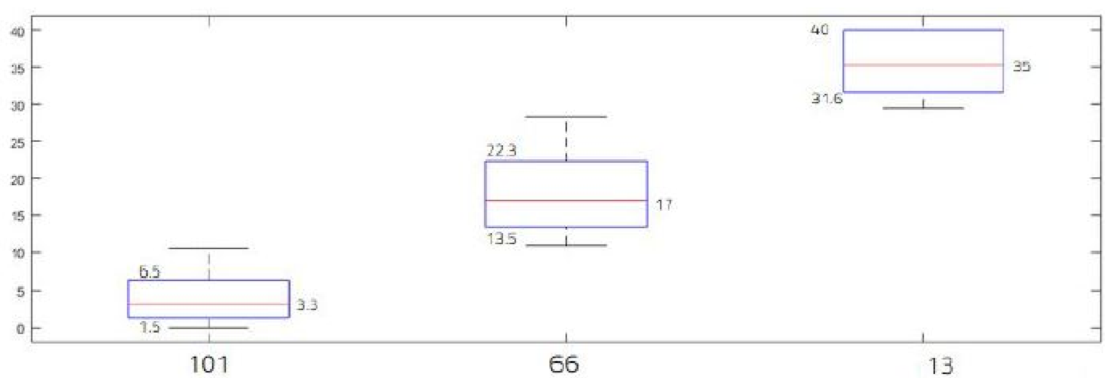

80
70
60
50
40
30
20
10
0
12.45
5
2
31
26.2
22
72
66.7
145
33
2

Figure 2.24


80
70
60
50
40
30
20
10
0
16.4
8.6
2.7
52
49
44
77.4
72
66.7
174
4
2


40
35
30
25
20
15
10
5
0
6.5
3.3
1.5
22.3
17
13.5
40
35
31.6
101
66
13

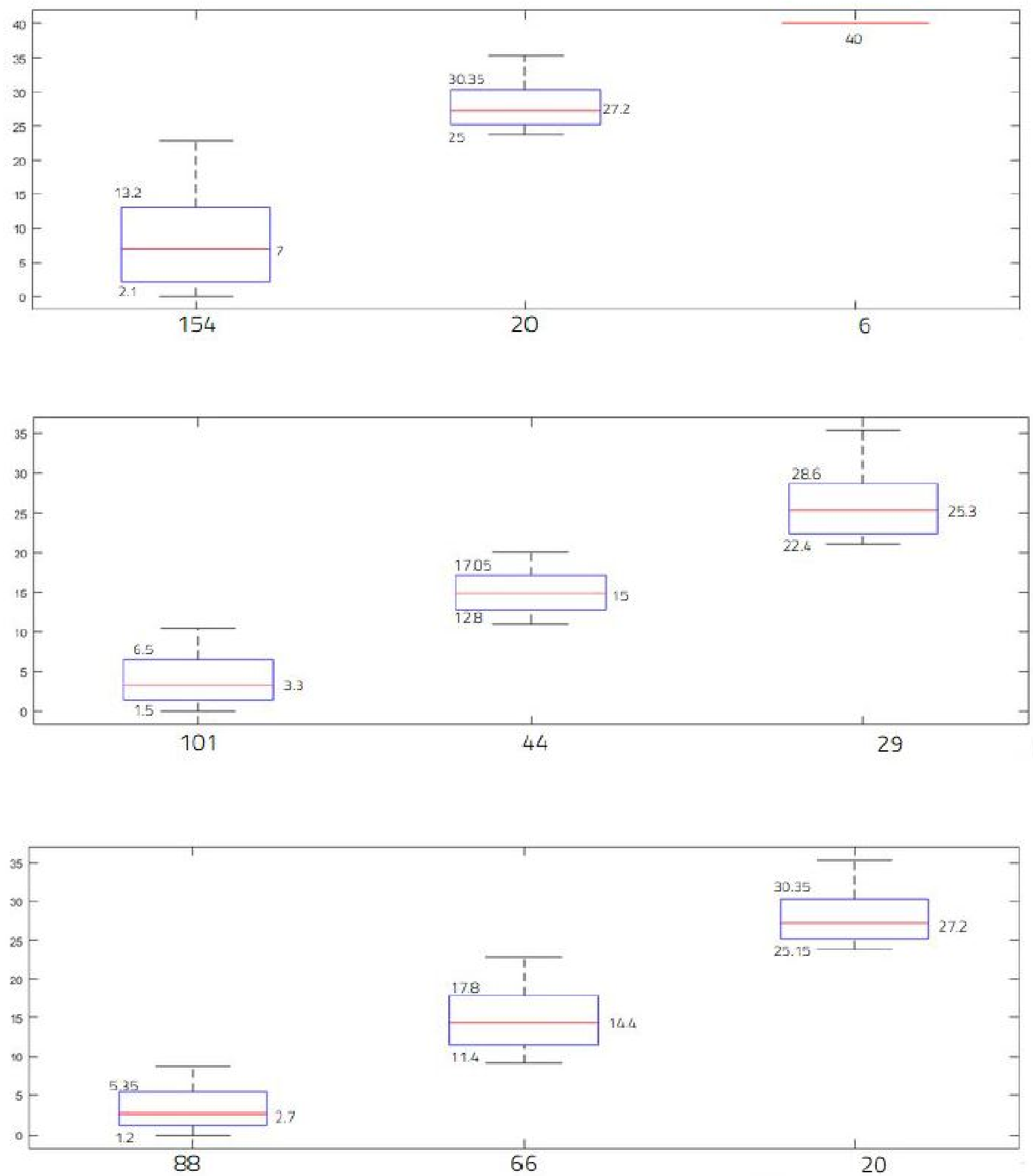
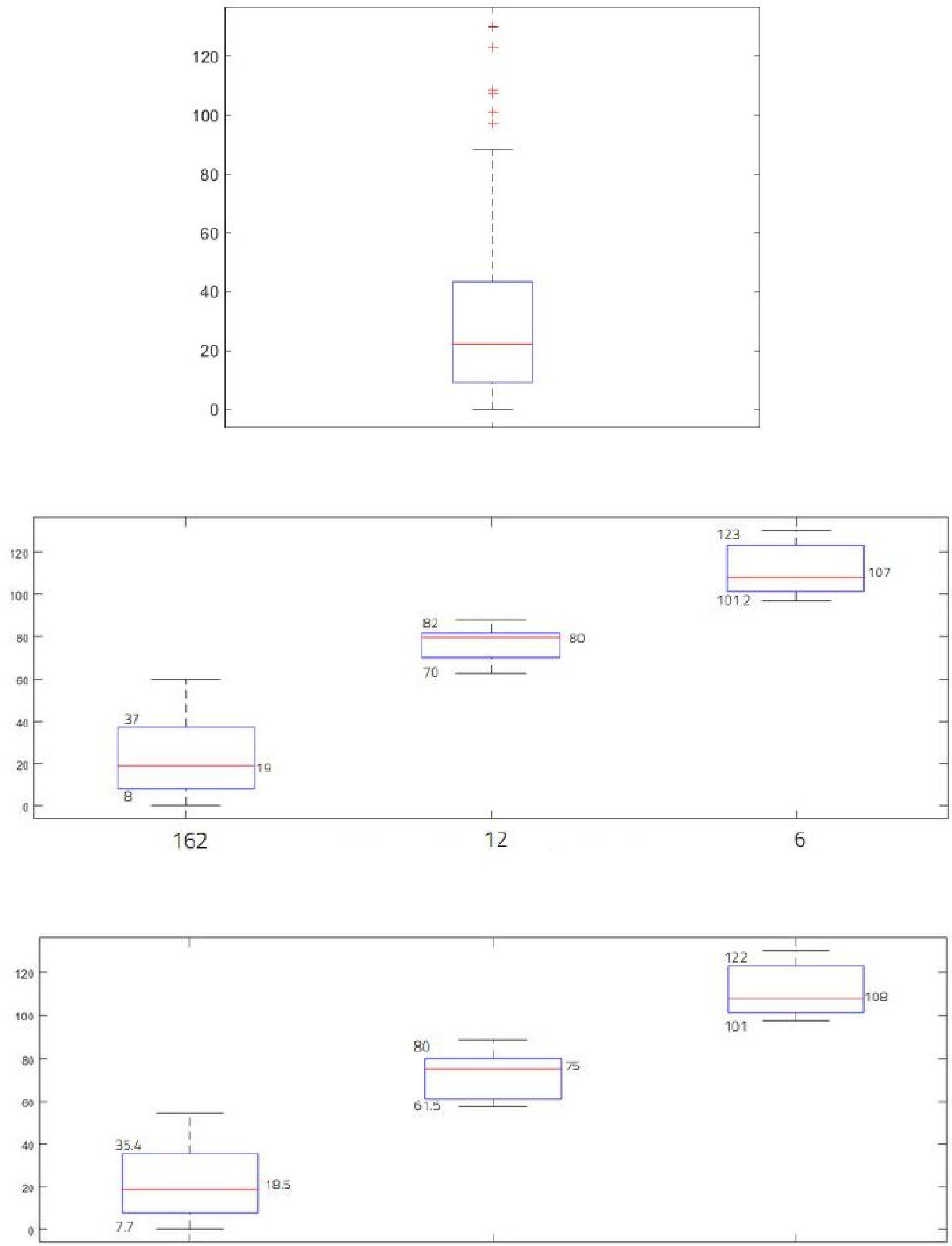
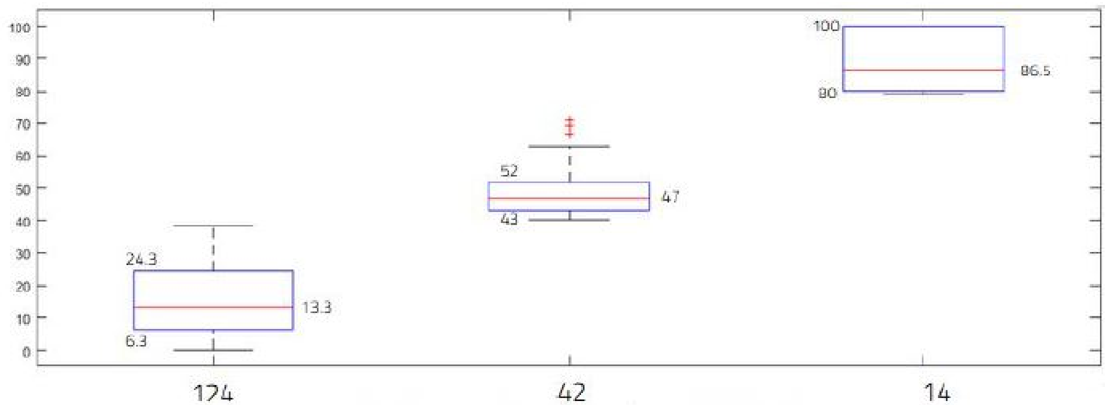


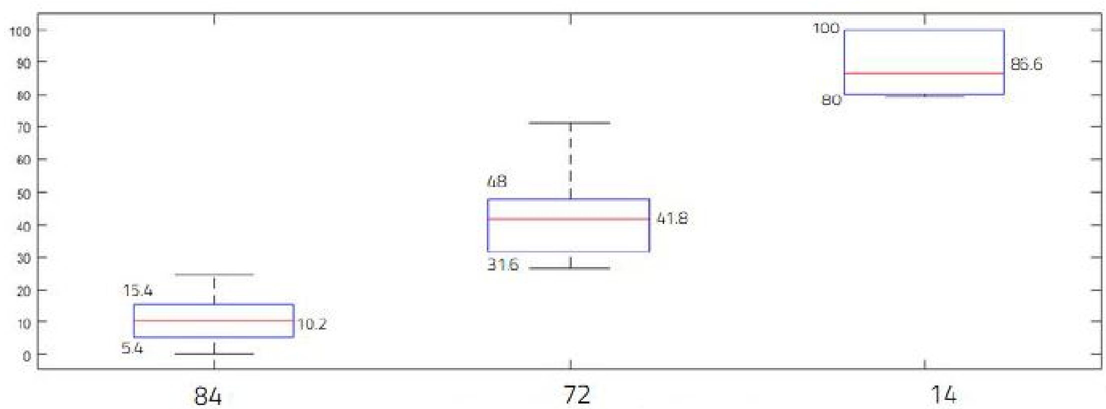
Figure 2.25: the first figure represent value distribution described through a boxplot of the third variable DIFF_DC; the second image represent the clustering result with cut off to three clusters using complete method, on the bottom the number of element per cluster is shown; the third figure shows the clustering result with cut off to three clusters using centroid method; the fourth figure shows the clustering result with cut off to three clusters using complete method with saturation of outliers; the fifth figure shows clustering results with cut off to three clusters using centroid with saturation of outliers; the sixth figure represent the clustering with method complete, deleting outliers from distribution; the last figure shows clustering result using centroid method and deleting outliers

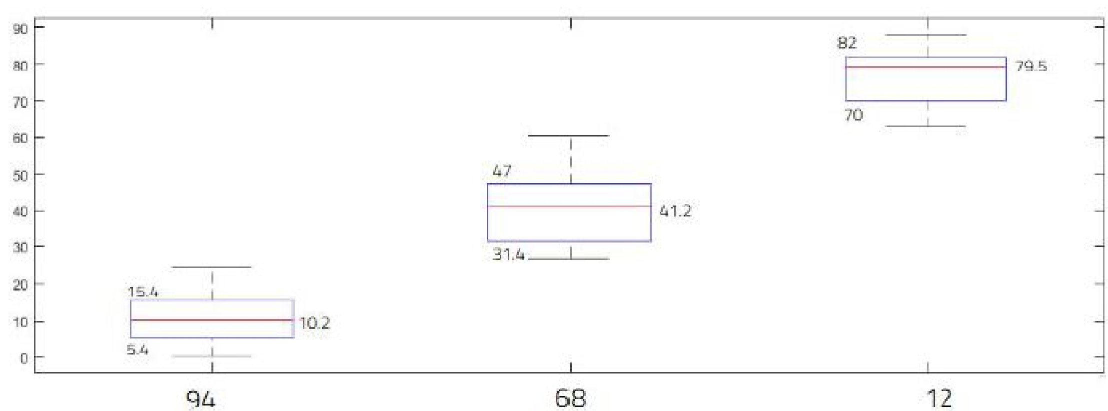


0
20
40
60
80
100
120
37
19
8
82
80
70
123
107
101.2
162
12
6
35.4
18.5
7.7
80
75
61.5
122
108
101

Figure 2.26


100
90
80
70
60
50
40
30
20
10
0
24.3
13.3
6.3
52
47
43
100
86.5
80
124
42
14


100
90
80
70
60
50
40
30
20
10
0
15.4
10.2
5.4
48
41.8
31.6
100
86.6
80
84
72
14


90
80
70
60
50
40
30
20
10
0
15.4
10.2
5.4
47
41.2
31.4
82
79.5
70
94
68
12

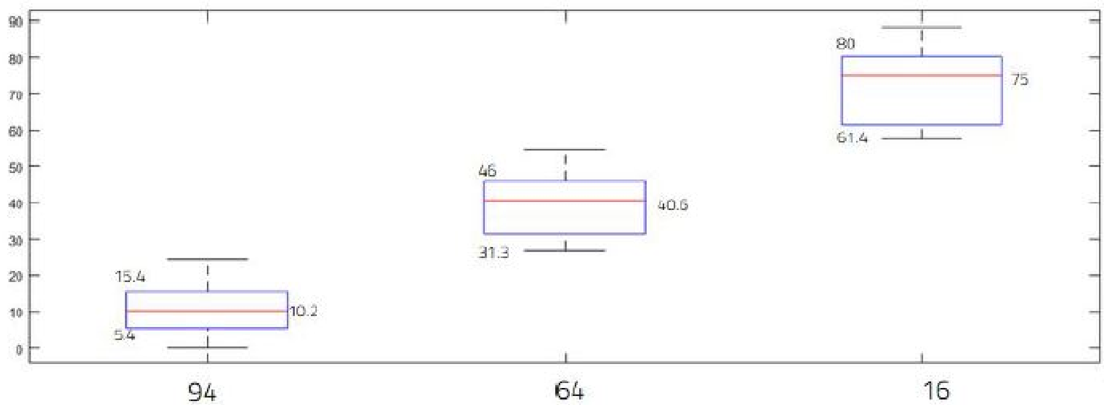
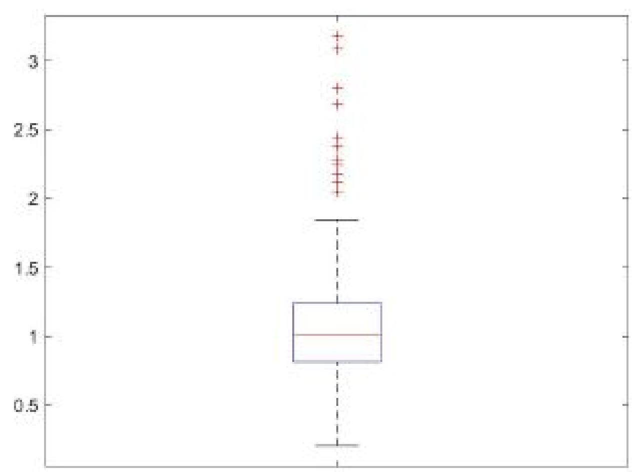
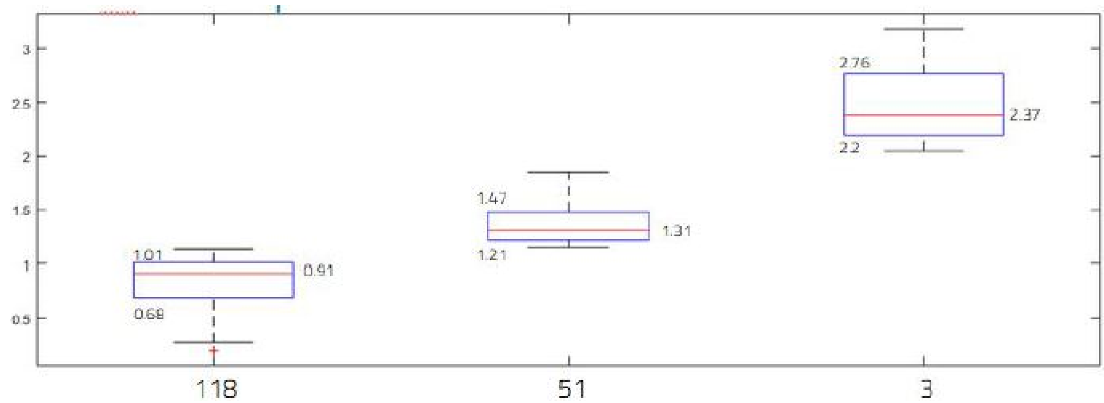
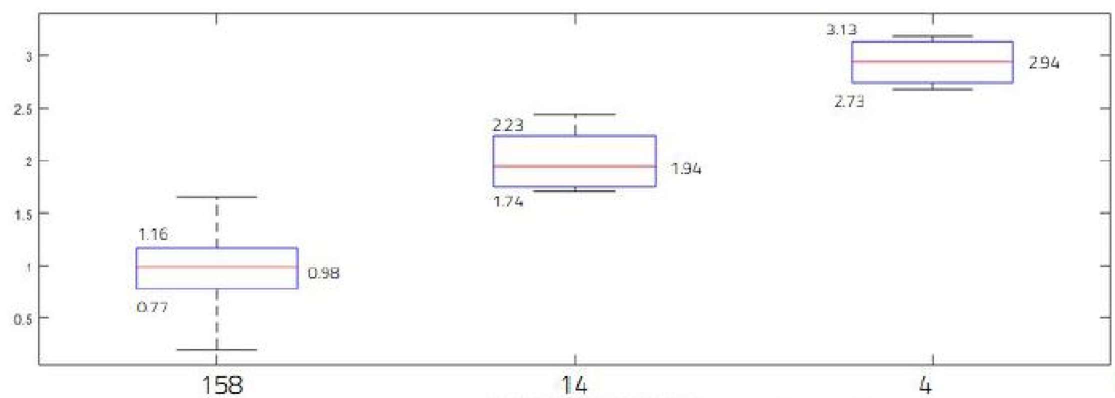
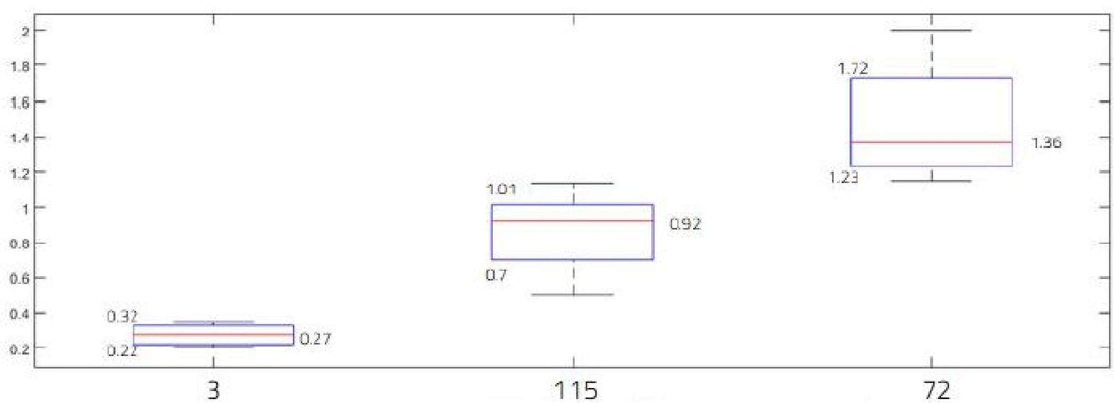


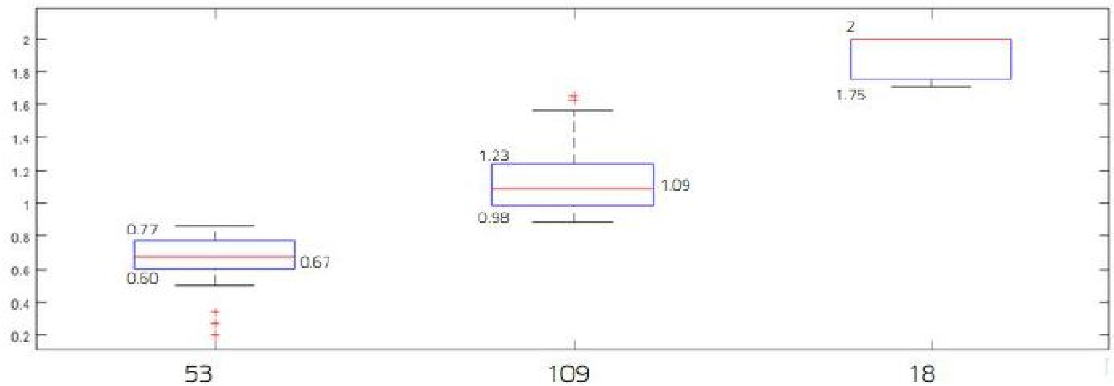
Figure 2.27: the first figure represent value distribution described through a boxplot of the fourth variable ASI; the second image represent the clustering result with cut off to three clusters using complete method, on the bottom the number of element per cluster is shown; the third figure shows the clustering result with cut off to three clusters using centroid method; the fourth figure shows the clustering result with cut off to three clusters using complete method with saturation of outliers; the fifth figure shows clustering results with cut off to three clusters using centroid with saturation of outliers; the sixth figure represent the clustering with method complete, deleting outliers from distribution; the last figure shows clustering result using centroid method and deleting outliers

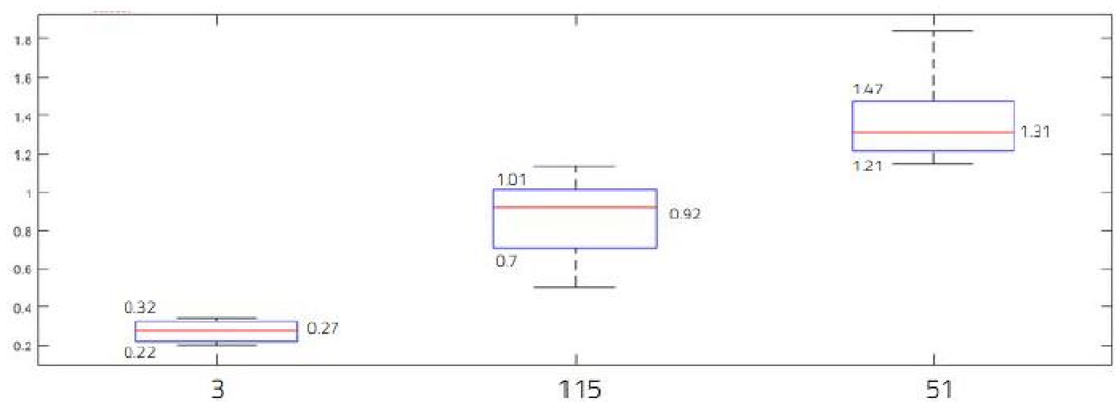



3
2.5
2
1.5
1
0.5
2.76
2.37
2.2
1.47
1.31
1.21
1.01
0.91
0.68
118
51
3


3
2.5
2
1.5
1
0.5
3.13
2.94
2.73
2.23
1.94
1.74
1.16
0.98
0.77
158
14
4


2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
1.72
1.36
1.23
1.01
0.92
0.7
0.32
0.27
0.22
3
115
72


2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0.77
0.67
0.60
1.23
1.09
0.98
2
1.75
53
109
18


1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0.32
0.27
0.22
1.01
0.92
0.7
1.47
1.31
1.21
3
115
51

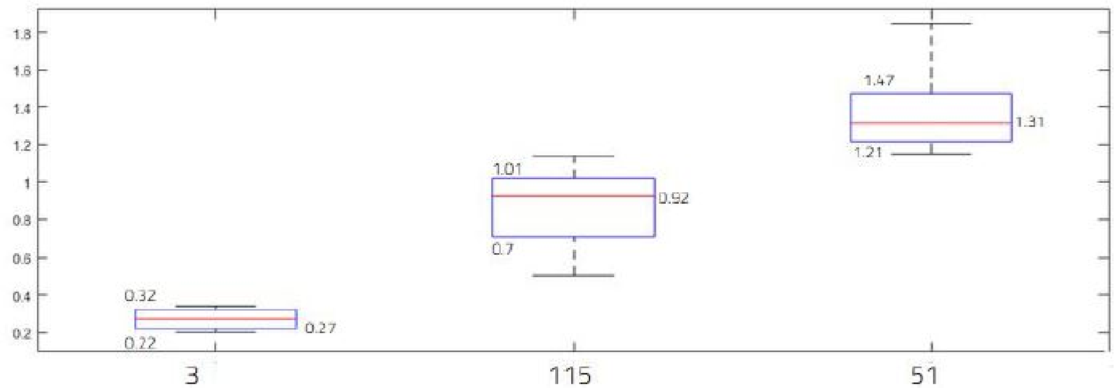


Figure 2.28: the first figure represent value distribution described through a boxplot of the fourth variable SI; the second image represent the clustering result with cut off to three clusters using complete method, on the bottom the number of element per cluster is shown; the third figure shows the clustering result with cut off to three clusters using centroid method; the fourth figure shows the clustering result with cut off to three clusters using complete method with saturation of outliers; the fifth figure shows clustering results with cut off to three clusters using centroid with saturation of outliers; the sixth figure represent the clustering with method complete, deleting outliers from distribution; the last figure shows clustering result using centroid method and deleting outliers

The modification of MFs are made considering the distribution as result of the clustering implemented using the method complete, with the saturation of outliers. For DIFF_DC a different conformation is considered raising the number of MFs, with a cut off that brings to the detection of six clusters. For ASI the possible conformations that are compared in term of performances are with a cut off of two or four clusters, that means the presence of two or four MFs. After performance evaluation the last choice is taken as good. At this point actually a possible misleading behaviour of the last variable SI is noticed, as the first and the third MFs basically express the same situation, here is decided to transpose the symmetry value from 1 to 0, scaling all the other values to a maximum of 1 that is the situation of total asymmetry.

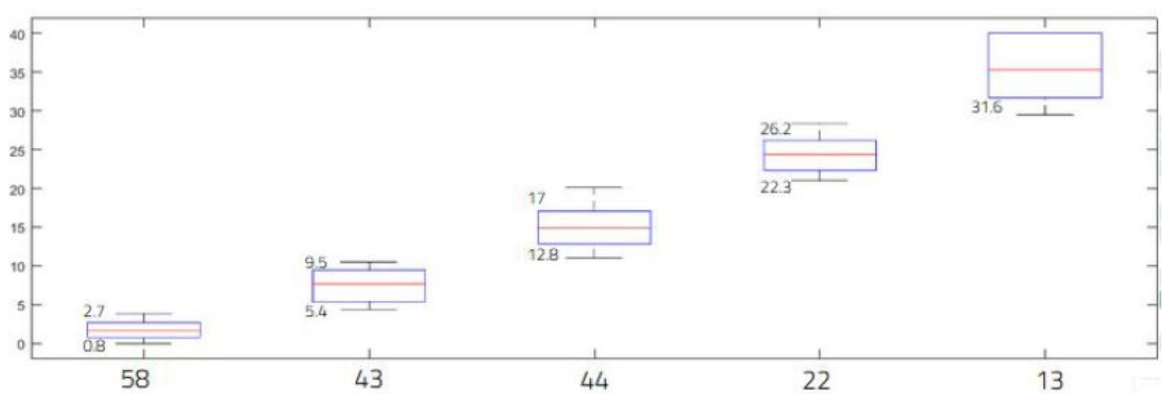


Figure 2.29: new cut off for DIFF_DC and ASI dendrogram respectively

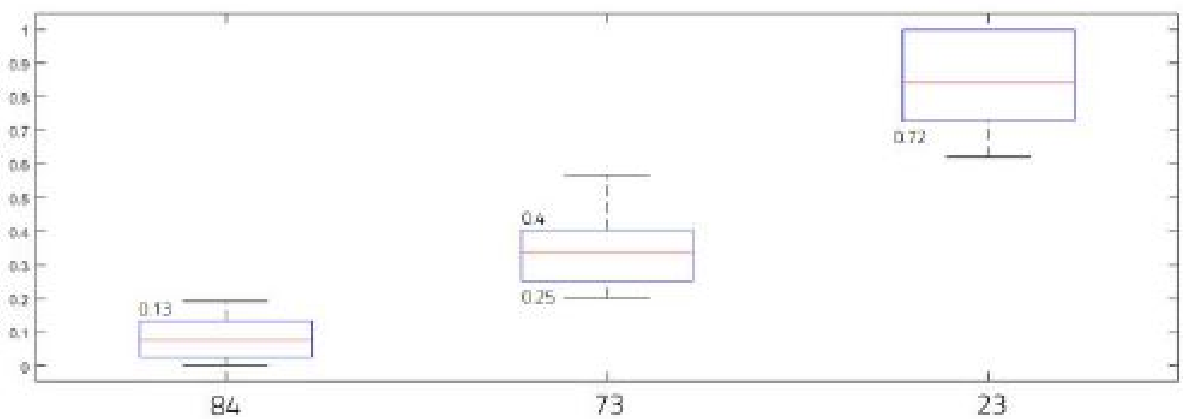
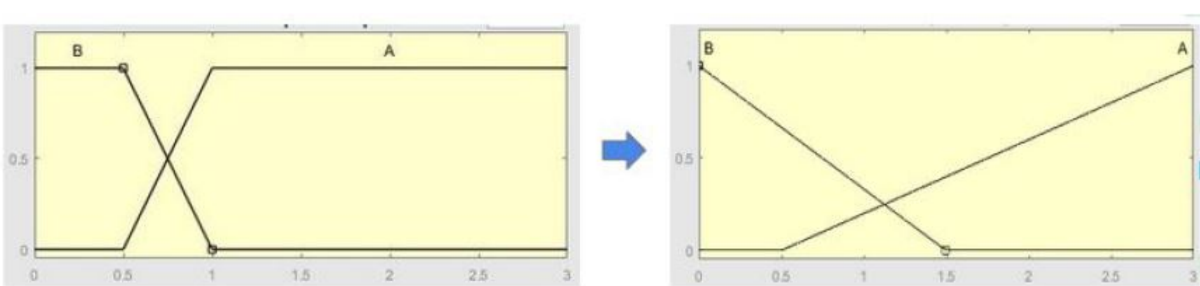


Figure 2.30: new cut off for DIFF_DC, ASI and SI dendrogram respectively



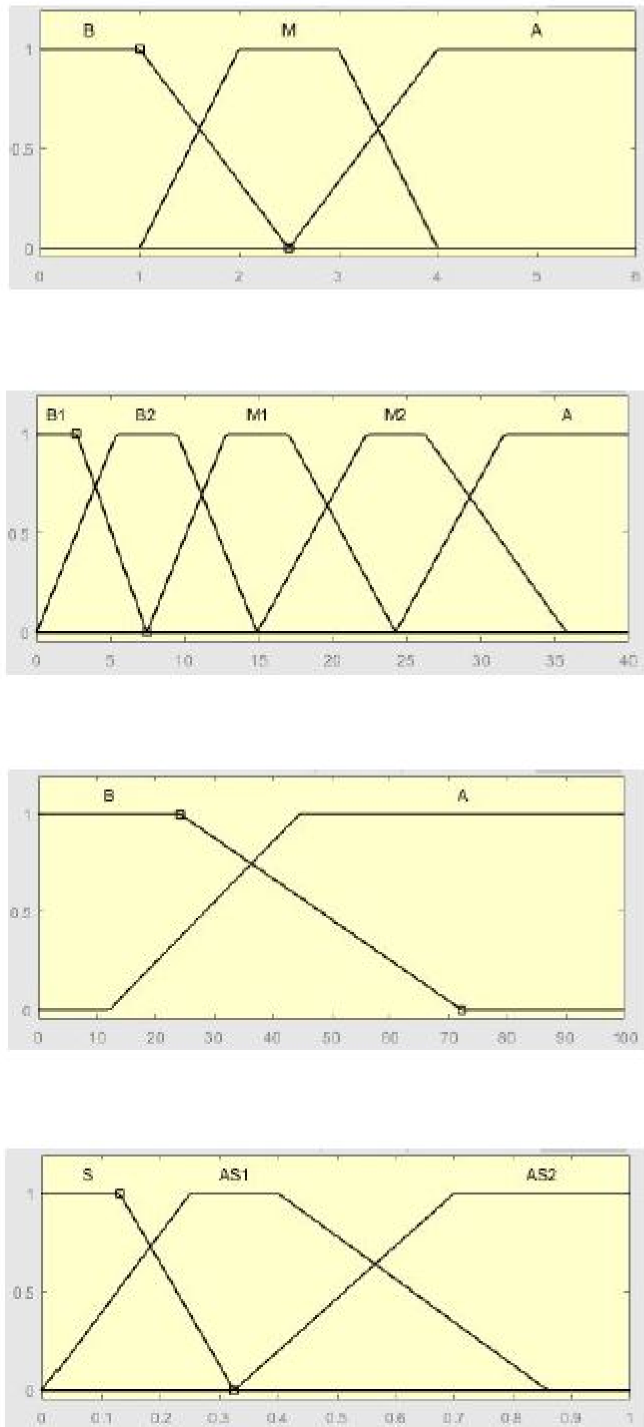


Figure 2.31: modification of the five variables, DIFF_AP, DIFF_AS, DIFF_DC, ASI and SI after clustering method application

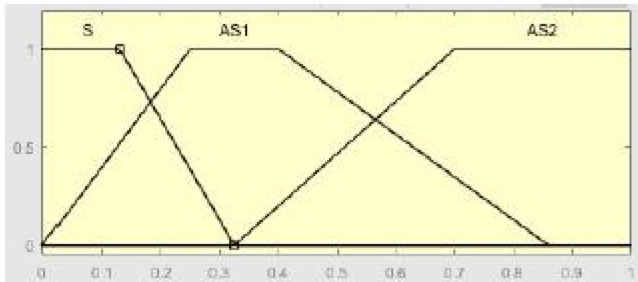


Figure 2.32: performances of the new system on training dataset, non-classified elements drastically decreased to 3 for the second and 10 for the third class

After every modification of MF a new rule set is necessary, because the possible antecedent linguistic variable are changed from the previous system design. A clustering of the dataset is implemented using the method already seen, with the unique difference that now at least 3 elements clusters are considered. Matlab with fuzzy logic toolbox gives the possibility to build a tab that represent rule firing for each observation, just putting desired output before calling *evalfis* function. Thank to this instrument has been possibile to detect which rule brings a certain inaccuracy. In particular for rules 6 and 14, that have different consequent, performances are compared in case of absence of each of one. Actually looking at the rule firing for the rule 14, is possible to see that rule 14 is active for just three elements in the dataset so is not a considerable rule. So the performances are evaluated in absence of both the rule or in only presence of the rule 6. Also the defuzzification method, never changed from the centroid method, has been compared with MOM method, but performance still are far from being satisfactory.

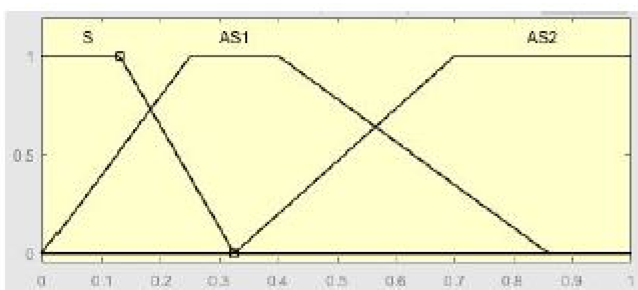


Figure 2.33: New set of rule obtained after MFs modification

Is necessary to introduce the intra-cluster variability measure, calculated as the average 'cityblock' distance of elements of a certain cluster. Additionally MFs of the FIS output are changed from a triangular geometry MFs which intersection lies on 6.67 and 13.34 to a more coherent output whit trapezoidal MFs intersecting in 10 and 20 exactly as for class limits. Elements that stands in around these limits are not considered so far, in order to reduce possible noise, and another rule set is then extracted.

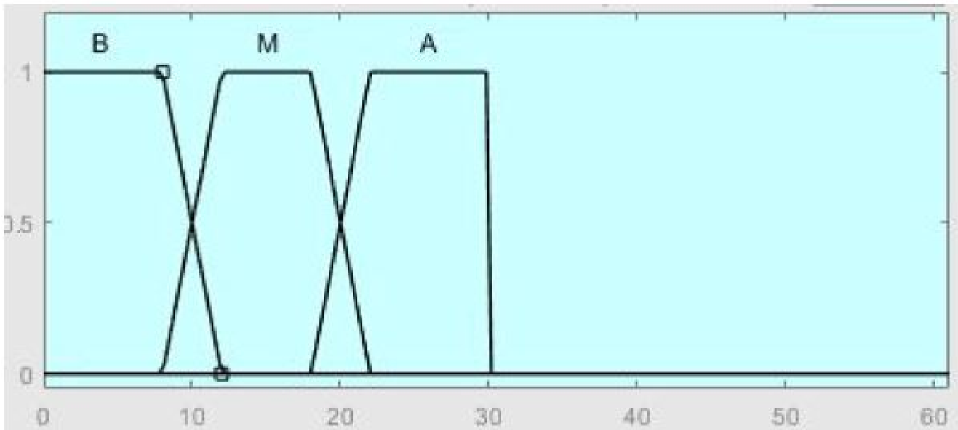
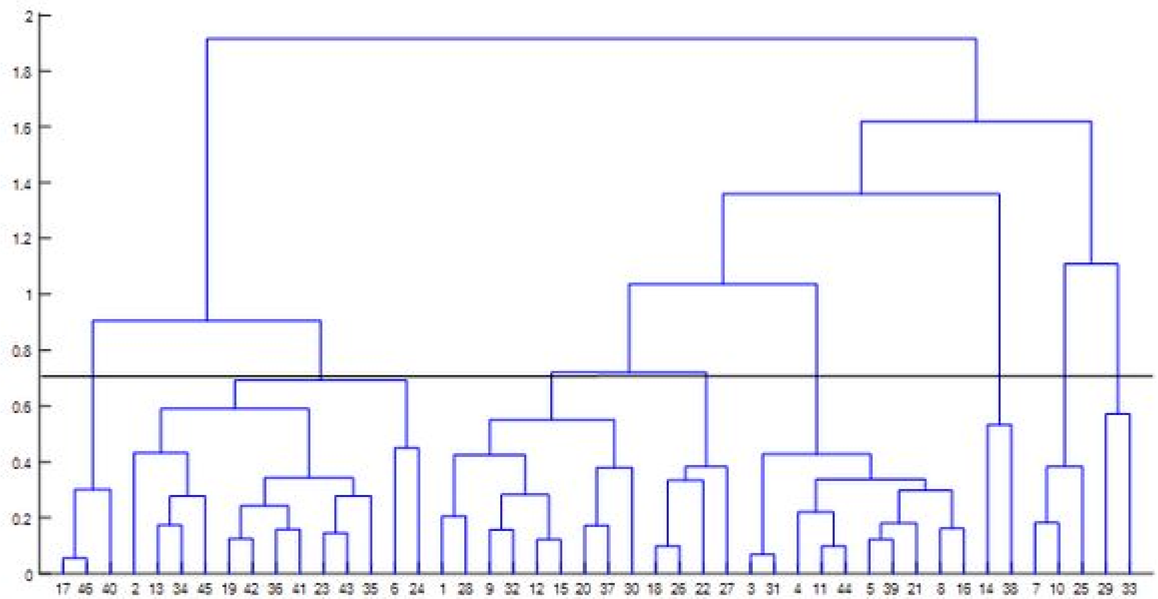
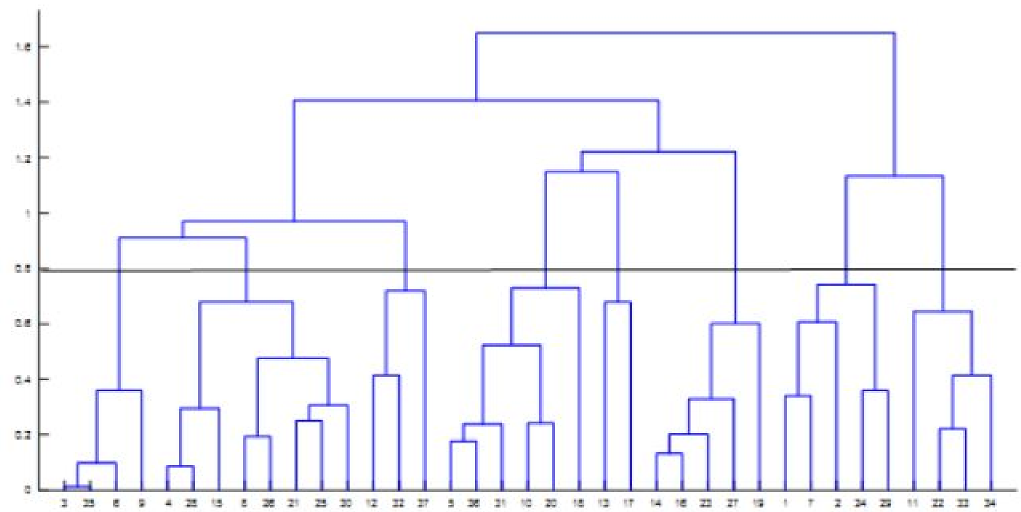


Figure 2.34: New output for FIS consequent, the intersection and the geometry of the MFs are changed



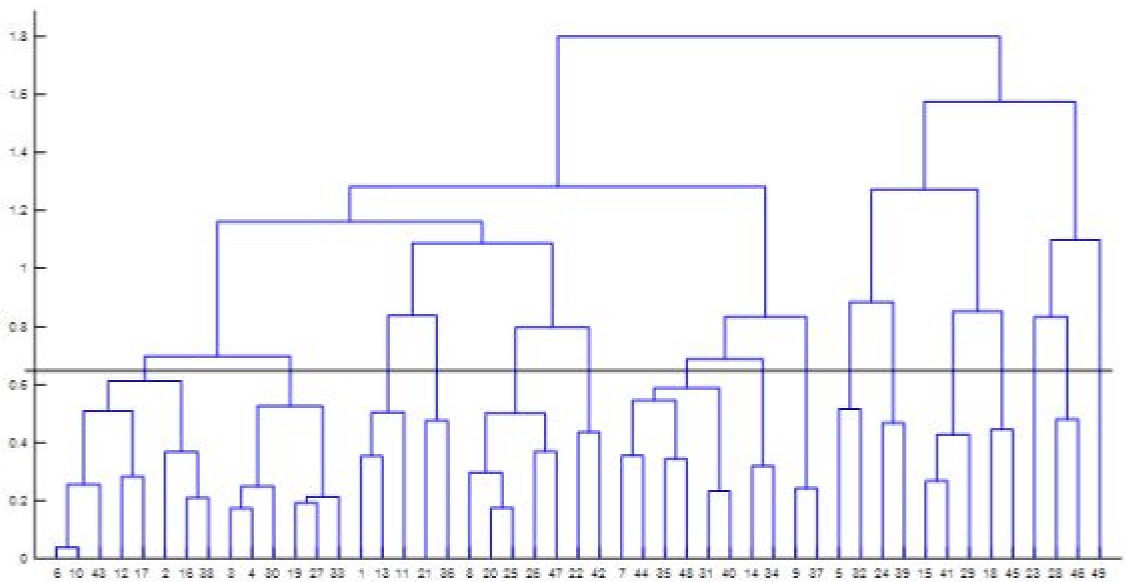
	N cluster	r					VAR INTRA
CLASS. 1	cluster 1	B	B	B2	B	S	14.73
	cluster 2	B	B	B2	B	S	12.78
	cluster 3	B	B	B1	A	AS2	14.13
	cluster 4	B	B	B1	A	AS1	13.65
	cluster 5	B	B	B1	B	S	9.97
	cluster 6	B	B	B1	B	S	6.13

Figure 2.35: Dendrogram of the first class, with a cut off on 8 clusters 6 rules are extracted



CLASS 2	cluster 1	B	B	B2	B	S	10.73
	cluster 2	B	B	M1	B	S	15.62
	cluster 3	B	B	M1	A	AS1	26.13
	cluster 4	B	A	M1	B	S	13.37
	cluster 5	B	M	B2	B	S	20.69
	cluster 6	B	B	M1	A	AS1	10.17
	cluster 7	B	B	B1	B	AS1	19.07

Figure 2.36: Dendrogram of the second class, with a cut off on 8 clusters 7 rules are extracted



CLASS 3	cluster 1	B	B	M1	A	AS1	22.80
	cluster 2	B	B	M2	B	S	12.38
	cluster 3	B	M	M2	B	AS1	25.01
	cluster 4	B	B	M1	B	S	13.53

Figure 2.37: Dendrogram of the second class, with a cut off on 16 clusters 4 rules are extracted. in this case also clusters with 3 elements are discarded

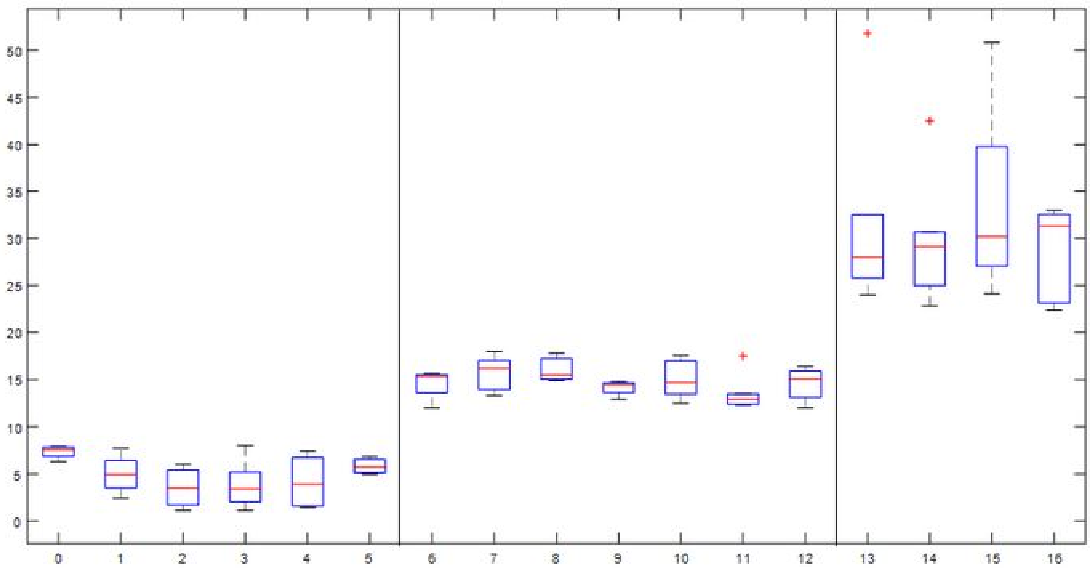


Figure 2.38: Distribution of EMG_ASYM_INDEX among the effective clusters of each class

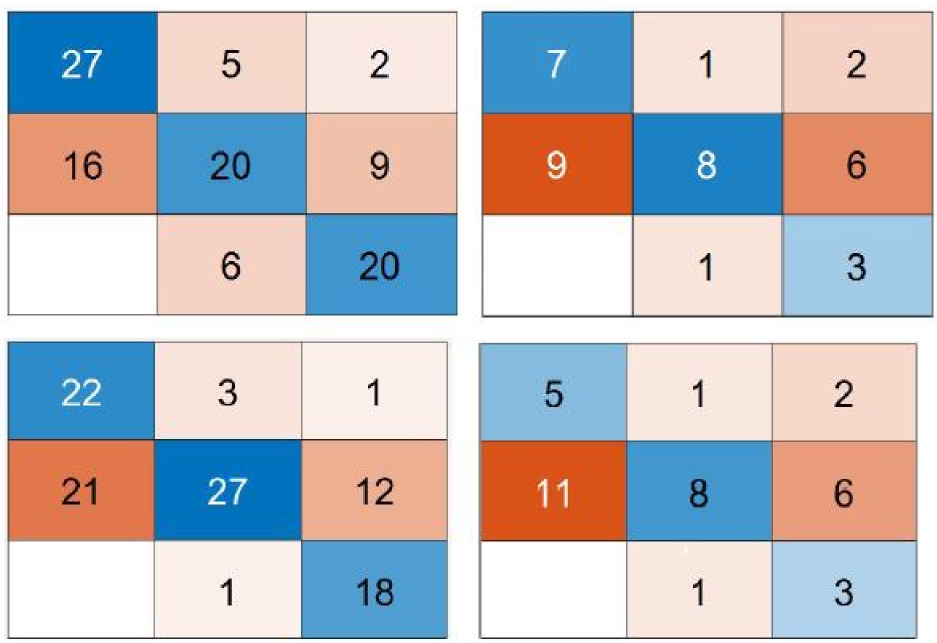


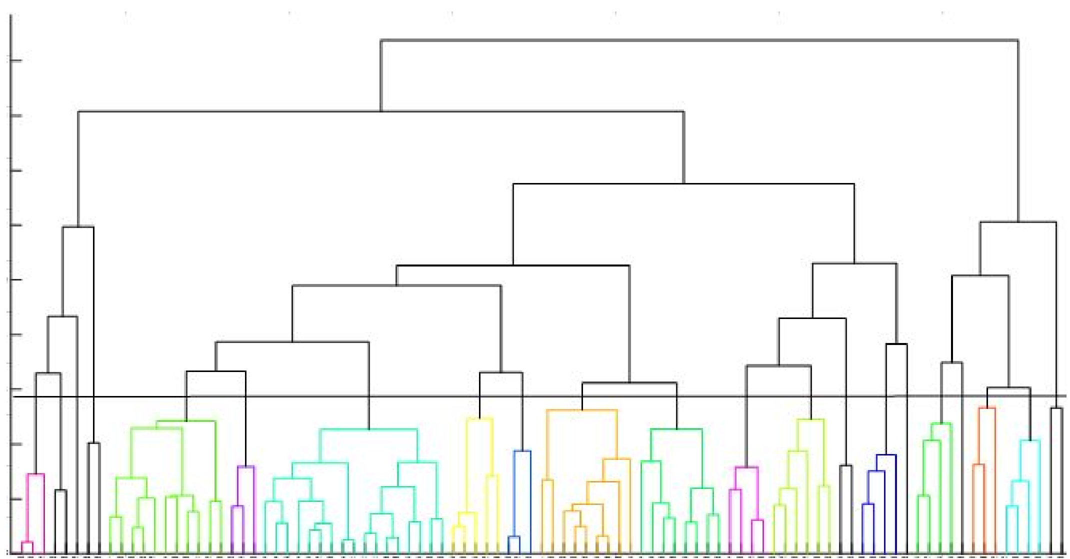
Figure 2.39: Train and test set performances, respectively on the right and on the left of the figure. The first row is indicative of MOM defuzzification method, while the second row is indivative of centroid defuzzification method

2.0.6 Class partition modification

The last step in this FIS design is the exploration of new classes partition. In particular a two classes partition and a five classes partition is implemented. The first partition, sees in the first class all the observation which EMG_ASYM_INDEX is equal or less than 15, while the second and last class is composed by those element which EMG_ASYM_INDEX is over this limit. For the second partition is considered the assigning 5 classes among elements, in particular:

- from 0 to 3 included, for the first class
- from 3 to 10 included, for the second class
- from 10 to 15 included, for the third class
- from 15 to 120 included, for the fourth class
- more than 20 for the fifth class.

Basically a new rule set is extracted using the same fuzzy sets but changing the number of classes, that unavoidably influences the number of clusters element, if one want to have a reasonable number of rules for each class.



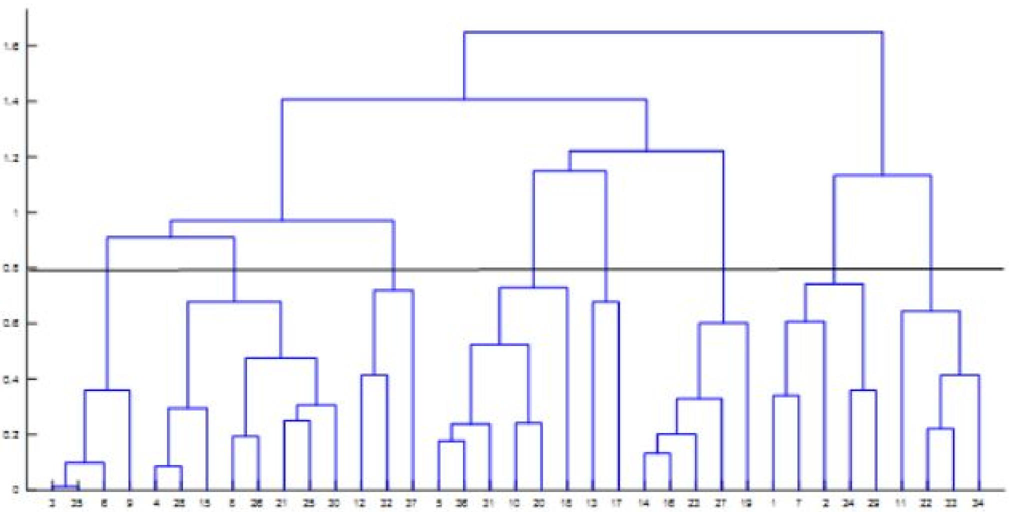
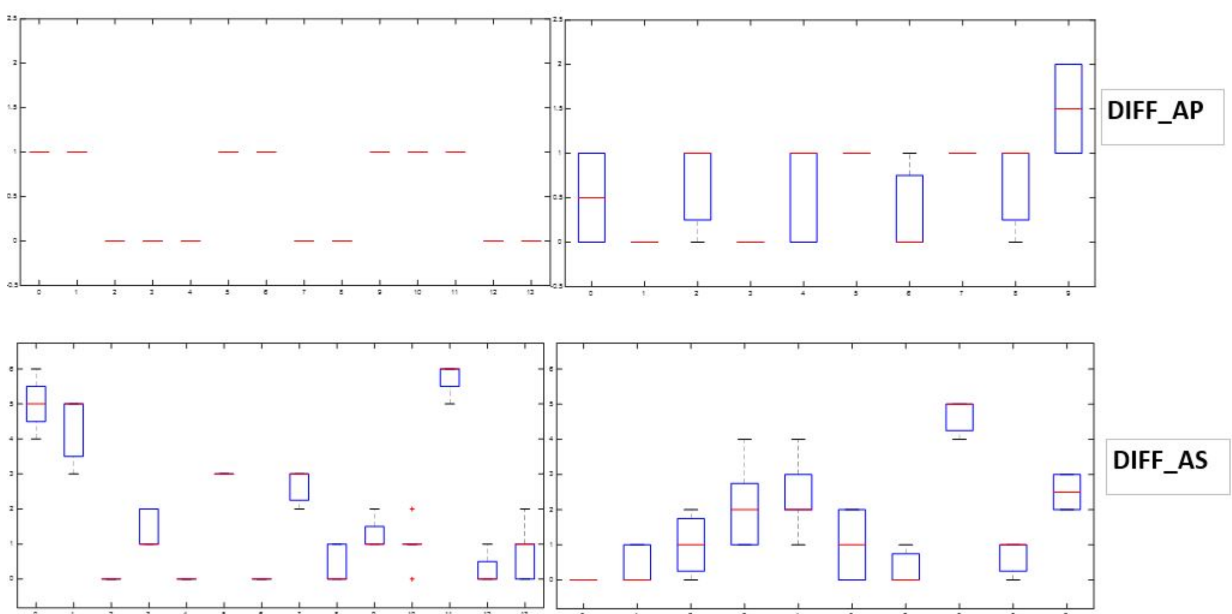
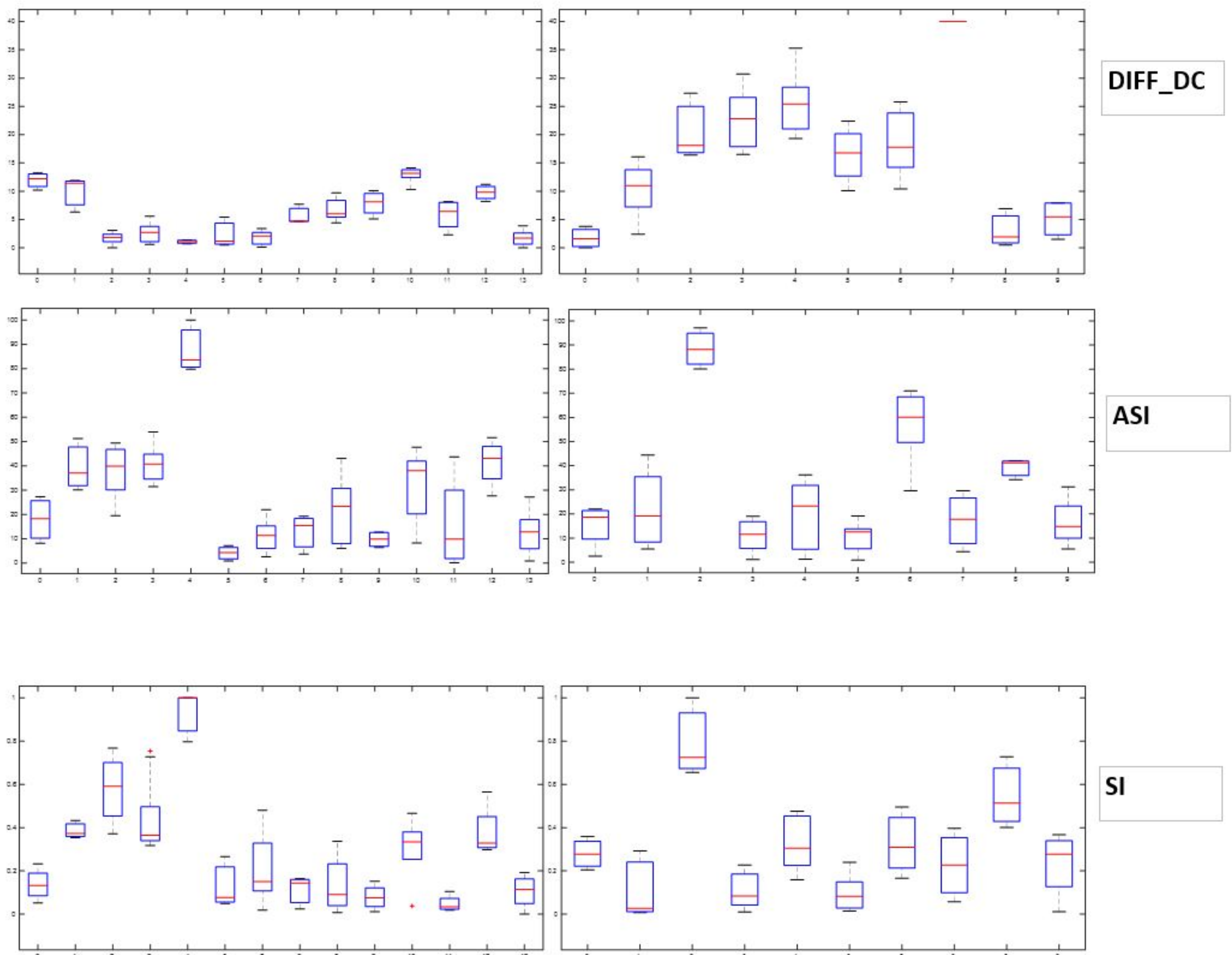


Figure 2.40: dendrograms of first and second classes




DIFF_DC
ASI
SI

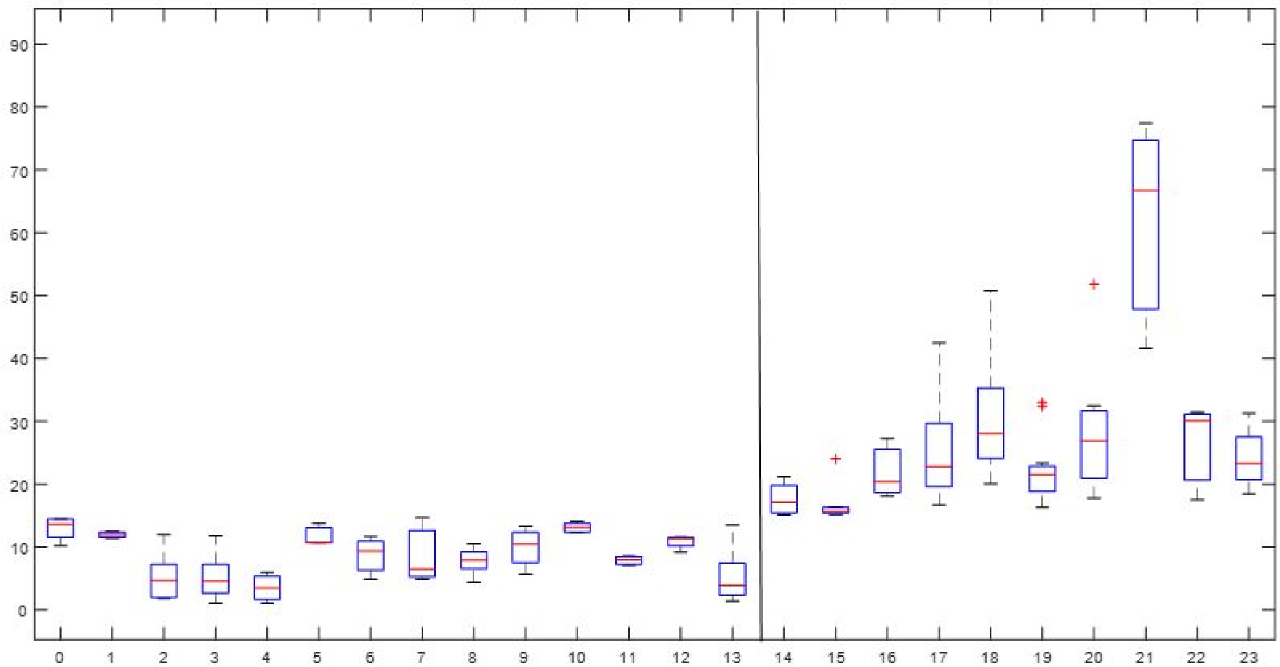


Figure 2.41: Cluster distribution for each variable in each class

B	A	M1	B	S
B	A	M1	A	AS1
B	B	B1	A	AS2
B	B	B1	A	AS1
B	B	B1	A	AS2
B	M	B1	B	S
B	B	B1	B	S
B	M	B2	B	S
B	B	B2	B	S
B	B	B2	B	S
B	B	M1	A	AS1
B	A	B2	B	S
B	B	B2	A	AS1
B	B	B1	B	S

B	M	M2	B	AS1
B	B	B2	B	S
B	B	M1	A	AS2
B	B	M2	B	S
B	M	M2	B	AS1
B	M	M1	B	S
B	B	M1	A	AS1
B	A	A	B	AS1
B	B	B1	A	AS1
B	M	B2	B	AS1

Figure 2.42: rules set extracted from clustering of elements belonging to the two possible classes

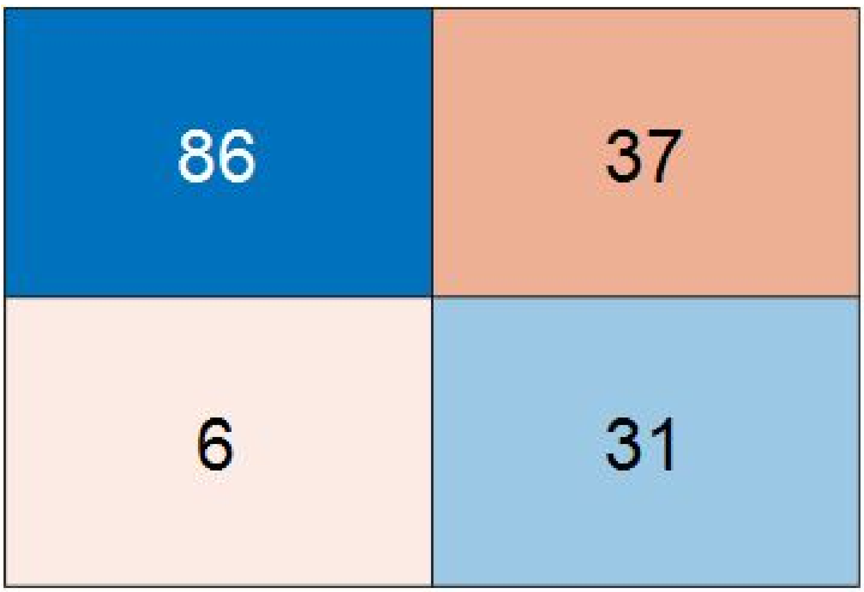
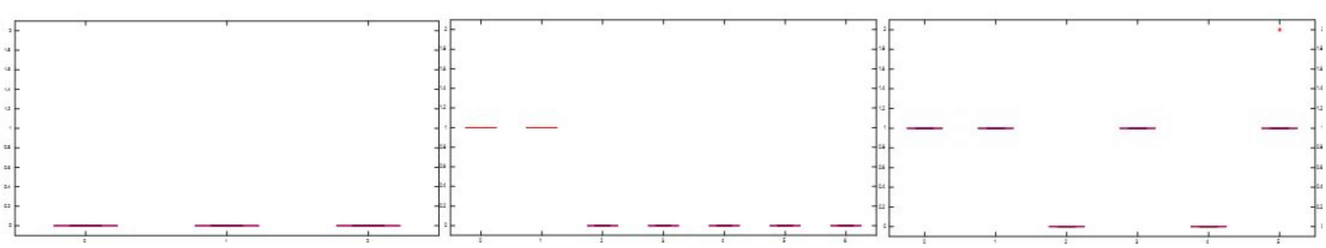


Figure 2.43: performances of 2 classes partition



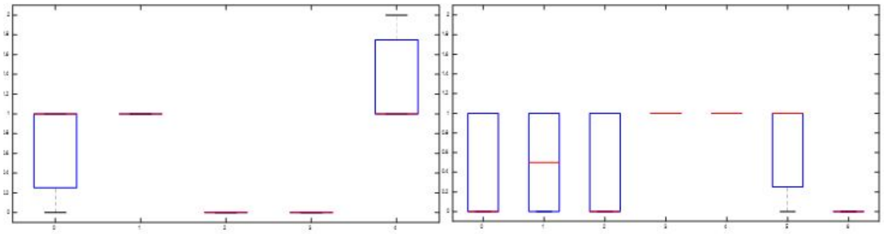


Figure 2.44: The distribution of elements belonging to clusters in which the class objects are aggregated, for the first variable DIFF_AP

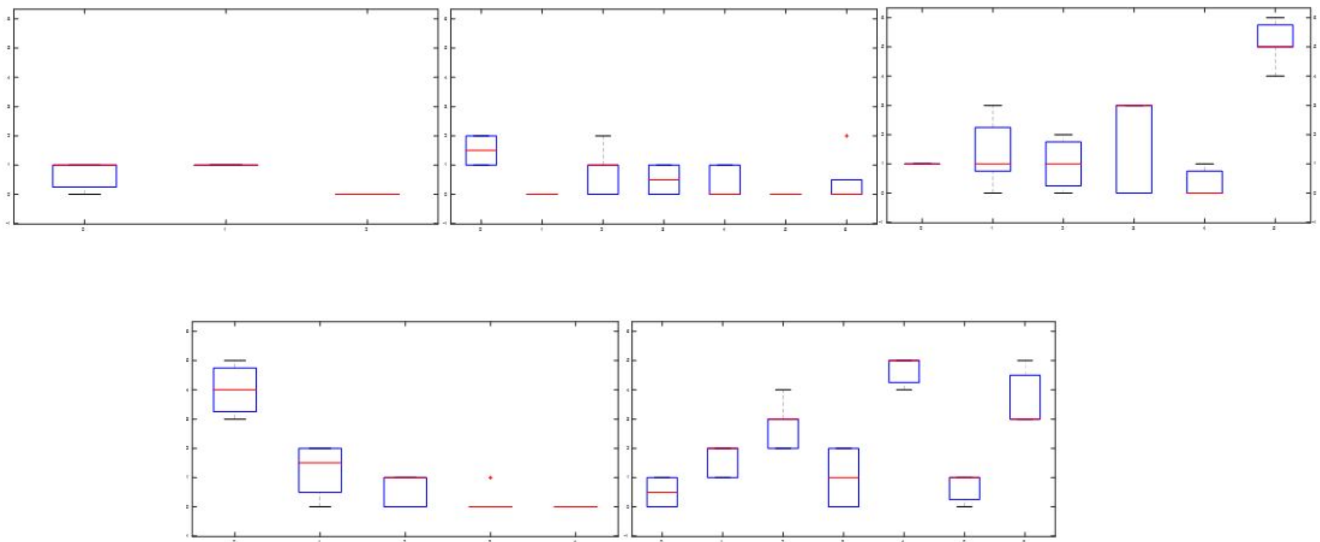
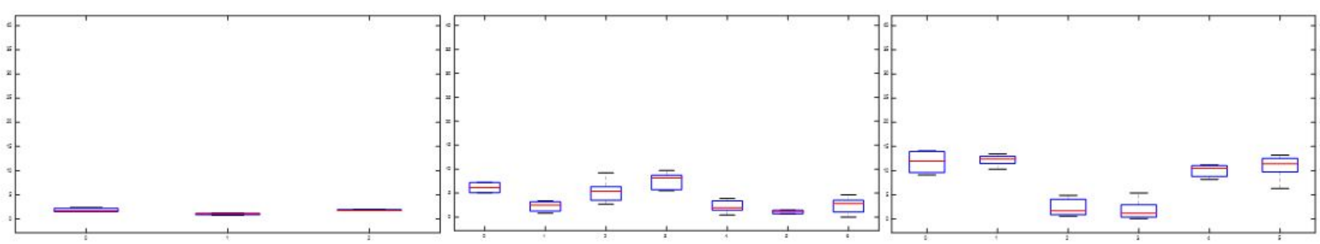


Figure 2.45: The distribution of elements belonging to clusters in which the class objects are aggregated, for the first variable DIFF_AS



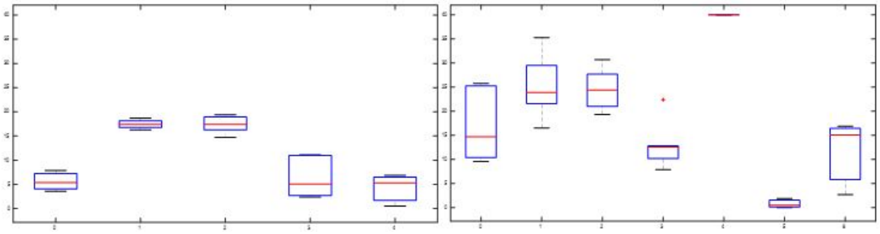


Figure 2.46: The distribution of elements belonging to clusters in which the class objects are aggregated, for the first variable DIFF_DC

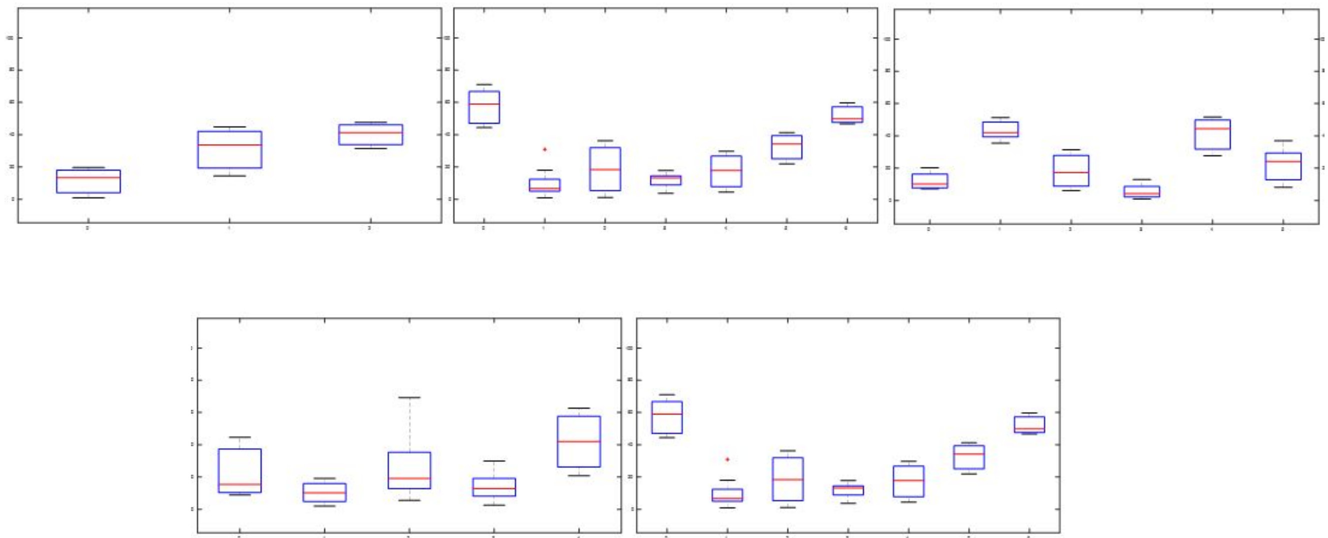
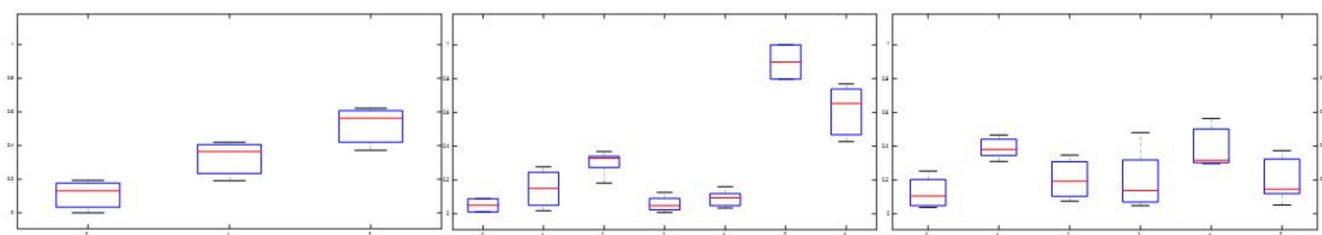


Figure 2.47: The distribution of elements belonging to clusters in which the class objects are aggregated, for the first variable ASI



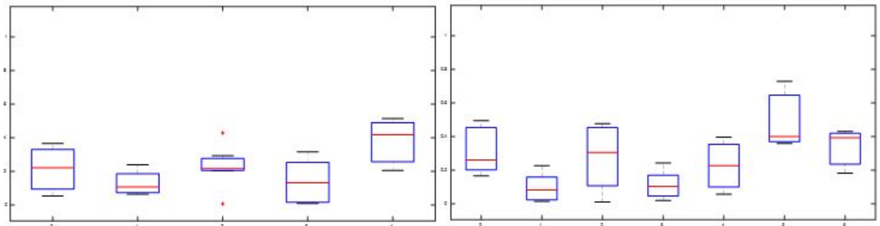


Figure 2.48: The distribution of elements belonging to clusters in which the class objects are aggregated, for the first variable DIFF_DC

B	B	B1	B	S
B	B	B1	B	AS1
B	B	B1	A	AS1

B	B	B1	B	S
B	B	B2	B	AS1
B	B	B2	B	S
B	B	B1	B	S
B	B	B1	A	AS2
B	M	B1	B	S
B	A	B2	B	S

B	B	M1	B	S
B	B	M1	A	AS1
B	B	B1	B	AS1
B	M	B1	B	S
B	B	B1	A	AS1
B	A	M1	B	S

B	M	B1	B	AS1
B	M	M1	B	S
B	B	M1	B	AS1
B	B	B2	B	S
B	B	B2	A	AS1

B	B	M1	A	AS1
B	M	M2	B	S
B	M	M2	B	AS1
B	B	M1	B	S
B	A	A	B	AS1
B	B	B1	B	AS1
B	M	M1	A	AS1

Figure 2.49: Set of rule starting from clustering of the first class, to the fifth. Rule with the same antecedent are marked with the same color

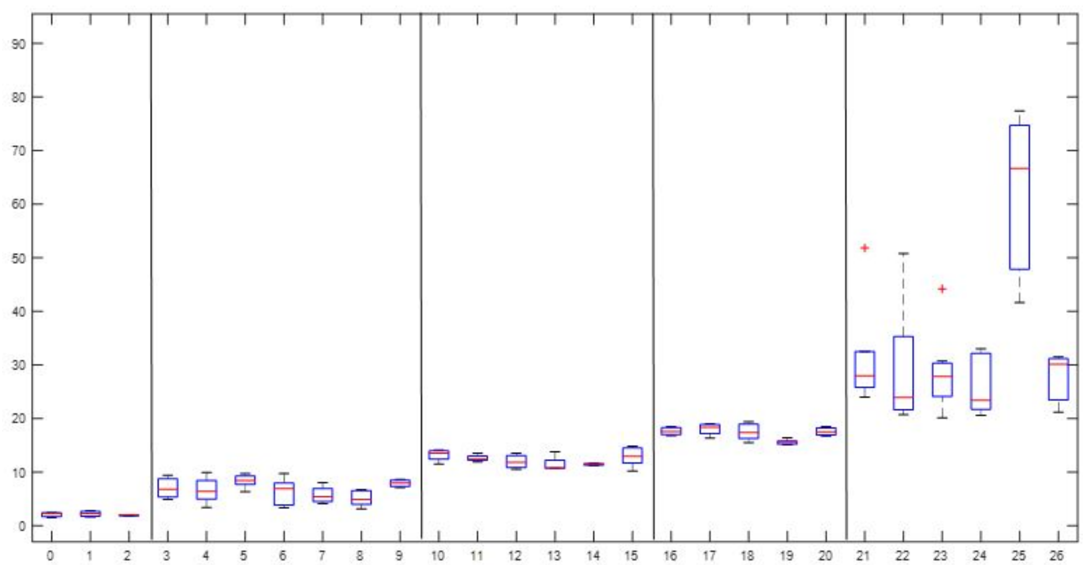


Figure 2.50: EMG_ASYM_INDEX distribution among the clusters of element belonging to each class

10	30	9	2	1
	11	5	5	3
3	5	12	7	11
	1	4	8	29

Figure 2.51: performances of 5 classes partition, there are still non-classified object, in particular 5 for the third class, 3 for the fourth and 16 for the 16 class

Chapter 3

Conclusion

The arising of gait laboratory analysis has given an important contribute in assessing patologies related with walking tasks, because it gives an important solution for the variability that gait carries on itself. The iterative process of design of a FIS is really connected to the problem that has been studied. Here a dataset composed by parameter coming from EMG signal are aggregated and classified by a fuzzy logic controller. The fact that information can be approximated in fuzzy sets and rule base is a strength point for fuzzy logic most of all in high-dimensional systems, but in this case the lack of accuracy is probably connected to the restriction imposed by the use of linguistic variable. Even if Mamdani-type FRBS is in development since its theorization in 1975, only recently in literature is possible to see application's of this method going deeper in the exploit of the fuzzy lofic for very complex problem. Surely a trend is the using of learning algorithm, with the possibility to cover big spaces and valorize the information granulation. In this study anyway hasn't been possible to converge to an optimal solution, probably because the model used is not still complete or because there are dependence that since now are not measurable or known among features. Anyway the aim was to better understand the problem features, that has been possible thank to a limited dimensionality and an easy interpretable formulation of variables, coming from recent studies. A next work could be the integration of this data model with other information on the patient condition that underwent on the trial, still trying to keep the model simplest as possible for the suitability of the fuzzy logic design. Another work can be the implementation of a learning algorithm, on this purpose genetic algorithm is widely used. Even if the number of observation is limited a local optimum can be a satisfactory solution for this problem.

Bibliography

[1] Shanshan Chen et al. "Toward pervasive gait analysis with wearable sensors: A systematic review". In: *IEEE journal of biomedical and health informatics* 20.6 (2016), pp. 1521–1537.

[2] Verne T Inman, Howard D Eberhart, et al. "The major determinants in normal and pathological gait". In: *JBJS* 35.3 (1953), pp. 543–558.

[3] D Casey Kerrigan et al. "Quantification of pelvic rotation as a determinant of gait". In: *Archives of physical medicine and rehabilitation* 82.2 (2001), pp. 217–220.

[4] Shinya Ogaya et al. "Muscle contributions to knee extension in the early stance phase in patients with knee osteoarthritis". In: *Gait & posture* 58 (2017), pp. 88–93.

[5] Jacqueline Romkes and Katrin Schweizer. "Immediate effects of unilateral restricted ankle motion on gait kinematics in healthy subjects". In: *Gait & posture* 41.3 (2015), pp. 835–840.

[6] Manabu Nankaku et al. "Gait analysis of patients in early stages after total hip arthroplasty: effect of lateral trunk displacement on walking efficiency". In: *Journal of Orthopaedic Science* 12.6 (2007), pp. 550–554.

[7] Jacob Apkarian, Stephen Naumann, and Bonnie Cairns. "A three-dimensional kinematic and dynamic model of the lower limb". In: *Journal of biomechanics* 22.2 (1989), pp. 143–155.

[8] Ilse Jonkers, Caroline Stewart, and Arthur Spaepen. "The complementary role of the plantarflexors, hamstrings and gluteus maximus in the control of stance limb stability during gait". In: *Gait & posture* 17.3 (2003), pp. 264–272.

[9] Amy R Wu and Arthur D Kuo. "Determinants of preferred ground clearance during swing phase of human walking". In: *Journal of Experimental Biology* 219.19 (2016), pp. 3106–3113.

[10] Blynn Shideler, Simon Taylor, and Rezaul Begg. *Real-time biofeedback rehabilitation tool guiding and illustrating foot placement for gait training.* US Patent App. 15/660,039. Dec. 2017.

[11] Yoong Ping Lim, Yi-Chung Lin, and Marcus G Pandy. "Effects of step length and step frequency on lower-limb muscle function in human gait". In: *Journal of biomechanics* 57 (2017), pp. 1–7.

[12] JR Gage, PA DeLuca, and TS Renshaw. "Gait analysis: principle and applications with emphasis on its use in cerebral palsy." In: *Instructional course lectures* 45 (1996), p. 491.

[13] Henry G Chambers and David H Sutherland. "A practical guide to gait analysis". In: *JAAOS-Journal of the American Academy of Orthopaedic Surgeons* 10.3 (2002), pp. 222–231.

[14] Tishya A.L Wren et al. "Efficacy of clinical gait analysis: A systematic review". eng. In: *Gait Posture* 34.2 (2011), pp. 149–153. ISSN: 0966-6362.

[15] Ganesh Roy et al. "Kinect Camera Based Gait Data Recording and Analysis for Assistive Robotics-An Alternative to Goniometer Based Measurement Technique". In: *Procedia Computer Science* 133 (2018). International Conference on Robotics and Smart Manufacturing (RoSMa2018), pp. 763–771. ISSN: 1877-0509. DOI: `https://doi.org/10.1016/j.procs.2018.07.121`. URL: `http://www.sciencedirect.com/science/article/pii/S1877050918310743`.

[16] Rosati S et al. "Basographic gait impairment score: A fuzzy classifier based on foot-floor contact parameters". eng. In: (2014).

[17] John V Basmajian. *Muscles alive : their functions revealed by electromyography / John V. Basmajian, Carlo De Luca*. eng. 5th ed. Baltimore: Williams and Wilkins, 1985.

[18] Aldo O Perotto. *Anatomical guide for the electromyographer: the limbs and trunk*. Charles C Thomas Publisher, 2011.

[19] DA Winter. "Anthropometry". In: *Biomechanics and motor control of human movement* (1990).

[20] Carlo J De Luca. "The use of surface electromyography in biomechanics". In: *Journal of applied biomechanics* 13.2 (1997), pp. 135–163.

[21] Rigmor Jensen, Anders Fuglsang-Frederiksen, and Jes Olesen. "Quantitative surface EMG of pericranial muscles. Reproducibility and variability". In: *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 89.1 (1993), pp. 1–9.

[22] Stuart M McGill. "Electromyographic activity of the abdominal and low back musculature during the generation of isometric and dynamic axial trunk torque: implications for lumbar mechanics". In: *Journal of Orthopaedic Research* 9.1 (1991), pp. 91–103.

[23] GERT-ÅKE HANSSON et al. "Electromyographic fatigue in neck/shoulder muscles and endurance in women with repetitive work". In: *Ergonomics* 35.11 (1992), pp. 1341–1352.

[24] Jørgen Winkel and Kurt Jørgensen. "Significance of skin temperature changes in surface electromyography". In: *European journal of applied physiology and occupational physiology* 63.5 (1991), pp. 345–348.

[25] Jaynie F Yang and DA Winter. "Electromyographic amplitude normalization methods: improving their sensitivity as diagnostic tools in gait analysis." In: *Archives of physical medicine and rehabilitation* 65.9 (1984), pp. 517–521.

[26] Gregory J Lehman and Stuart M McGill. “The importance of normalization in the interpretation of surface electromyography: a proof of principle”. In: *Journal of manipulative and physiological therapeutics* 22.7 (1999), pp. 444–446.

[27] KB Veiersted. “The reproducibility of test contractions for calibration of electromyographic measurements”. In: *European journal of applied physiology and occupational physiology* 62.2 (1991), pp. 91–98.

[28] Valentina Agostini and Marco Knaflitz. “An algorithm for the estimation of the signal-to-noise ratio in surface myoelectric signals generated during cyclic movements”. In: *IEEE transactions on biomedical engineering* 59.1 (2011), pp. 219–225.

[29] Francesco [Di Nardo] et al. “Normative EMG patterns of ankle muscle co-contractions in school-age children during gait”. In: *Gait Posture* 46 (2016), pp. 161–166. ISSN: 0966-6362. DOI: https://doi.org/10.1016/j.gaitpost.2016.03.002. URL: http://www.sciencedirect.com/science/article/pii/S0966636216000746.

[30] Samanta Rosati et al. “Muscle activation patterns during gait: A hierarchical clustering analysis”. In: *Biomedical Signal Processing and Control* 31 (2017), pp. 463–469. ISSN: 1746-8094. DOI: https://doi.org/10.1016/j.bspc.2016.09.017. URL: http://www.sciencedirect.com/science/article/pii/S1746809416301434.

[31] Balestra G. *Lecture notes in Classification and Interpretation of Biomedical Data*. 2016.

[32] Matthew Kirk Seeley. “Is Functional Asymmetry a Valid Explanation for Able-bodied Gait Asymmetries?” PhD thesis. University of Kentucky, 2006.

[33] Louis N Awad et al. “Walking speed and step length asymmetry modify the energy cost of walking after stroke”. In: *Neurorehabilitation and neural repair* 29.5 (2015), pp. 416–423.

[34] A De Stefano et al. “Effect of gait cycle selection on EMG analysis during walking in adults and children with gait pathology”. In: *Gait & posture* 20.1 (2004), pp. 92–101.

[35] Sarah J Mattes, Philip E Martin, and Todd D Royer. “Walking symmetry and energy cost in persons with unilateral transtibial amputations: matching prosthetic and intact limb inertial properties”. In: *Archives of physical medicine and rehabilitation* 81.5 (2000), pp. 561–568.

[36] Heydar Sadeghi et al. “Symmetry and limb dominance in able-bodied gait: a review”. In: *Gait & posture* 12.1 (2000), pp. 34–45.

[37] C Castagneri et al. “Asymmetry index in muscle activations”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.4 (2019), pp. 772–779.

[38] LM Schutte et al. “An index for quantifying deviations from normal gait”. In: *Gait & posture* 11.1 (2000), pp. 25–31.

[39] Michael H Schwartz and Adam Rozumalski. “The Gait Deviation Index: a new comprehensive index of gait pathology”. In: *Gait & posture* 28.3 (2008), pp. 351–357.

[40] Lotfi A Zadeh. "Outline of a new approach to the analysis of complex systems and decision processes". In: *IEEE Transactions on systems, Man, and Cybernetics* 1 (1973), pp. 28–44.

[41] Lotfi A Zadeh. "Fuzzy sets". In: *Information and control* 8.3 (1965), pp. 338–353.

[42] L.A. Zadeh. "Fuzzy logic". English. In: *Computational Complexity: Theory, Techniques, and Applications*. Vol. 9781461418009. Springer New York, 2013, pp. 1177–1200. ISBN: 9781461418009.

[43] Li Xin Wang and Hao Ying. "Adaptive fuzzy systems and control: design and stability analysis". In: *Journal Of Intelligent and Fuzzy Systems-applications in Engineering and Technology* 3.2 (1995), p. 187.

[44] Duc Truong Pham and Marco Castellani. "Action aggregation and defuzzification in Mamdani-type fuzzy systems". In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 216.7 (2002), pp. 747–759.

[45] Michio Sugeno. *Industrial applications of fuzzy control*. Elsevier Science Inc., 1985.

[46] Samanta Rosati et al. "Muscle activation patterns during gait: A hierarchical clustering analysis". eng. In: *Biomedical Signal Processing and Control* 31 (2017), pp. 463–469. ISSN: 1746-8094.

[47] David R Burnett et al. "Symmetry of ground reaction forces and muscle activity in asymptomatic subjects during walking, sit-to-stand, and stand-to-sit tasks". In: *Journal of Electromyography and Kinesiology* 21.4 (2011), pp. 610–615.

[48] Andrzej Piegat and Marek Landowski. "Horizontal membership function and examples of its applications". In: *International Journal of Fuzzy Systems* 17.1 (2015), pp. 22–30.

[49] Jean-Lou Chameau and Juan Carlos Santamarina. "Membership functions I: Comparing methods of measurement". In: *International Journal of Approximate Reasoning* 1.3 (1987), pp. 287–301.

[50] Hui Cao et al. "A Clustering-analysis-based membership functions formation method for fuzzy controller of ball mill pulverizing system". In: *Journal of Process Control* 23.1 (2013), pp. 34–43. ISSN: 0959-1524. DOI: https://doi.org/10.1016/j.jprocont.2012.10.011. URL: http://www.sciencedirect.com/science/article/pii/S0959152412002417.

[51] Michael A Lee and Hideyuki Takagi. "Integrating design stage of fuzzy systems using genetic algorithms". In: *[Proceedings 1993] Second IEEE International Conference on Fuzzy Systems*. IEEE. 1993, pp. 612–617.

[52] Karina Gibert and Alejandra Pérez-Bonilla. "Fuzzy Box-plot based Induction Rules. Towards automatic generation of classes-interpretations." In: *EUSFLAT Conf.* Citeseer. 2005, pp. 524–529.

[53] P Dinesh Asanka and Amal Shehan Perera. "Defining fuzzy membership function using box plot". In: *International journal of research in computer applications and robotics* 5.11 (2017), pp. 1–10.

[54] Jeremy Kerr-Wilson. “Fuzzy Rule-Based Systems: Design, Analysis, and Applications”. In: (2019).