Cardiovascular dysautonomias diagnosis using crisp and fuzzy decision trees: a comparative study

Ali Idri and Ilham Kadi

University Mohammed V, Rabat, Morocco

eHealth Informatics 2016

24-25 May, 2016, Vienna, Austria

Contents

Motivations

- In the Autonomic nervous system (ANS) unit of Avicenne university hospital in Morocco, a set of tests is performed to diagnose patients with cardiovascular dysautonomias

- The tests results are measured and analyzed manually by the specialists

- A classification model was developed using C4.5 algorithm and the ANS dataset to support cardiac doctors in their decisions

Objective

- Applying Fuzzy Decision Trees in order to provide a fuzzy classifier for cardiovasular dysautonomias.

- Provide a comparison between the results obtained using crisp Decision Trees and Fuzzy Decision Trees especially in terms of: **Easy to interpret and Accuracy rates.**

Background: Autonomic Nervous System

- **Autonomic Nervous System :**
  - is part of the nervous system that controls involuntary actions, such as the beating of the heart and the widening or narrowing of the blood vessels.
  - controls body temperature, digestion, metabolism, the production of body fluids, and other processes.
  - is frequently subject to malfunctions that are called dysautonomias.
  - This disorder can cause serious problems, including:
    - Blood pressure troubles
    - Heart problems
    - Trouble with breathing and swallowing

Background: Autonomic Nervous System

- Doctors can check for signs of dysautonomias by measuring blood pressure and heart rate with the following four tests:
  - **Deep breathing test:** measures changes in heart rate (HR) in response to a deep breath
  - **Hand grip test:** This is a manual effort contraction performed to determine changes in the blood pressure
  - **Mental stress test:** The patient performs mental arithmetic calculations
  - **Orthostatic test:** aims at measuring HR and BP variations in different positions: stand up and rest.

Background: Decision tree

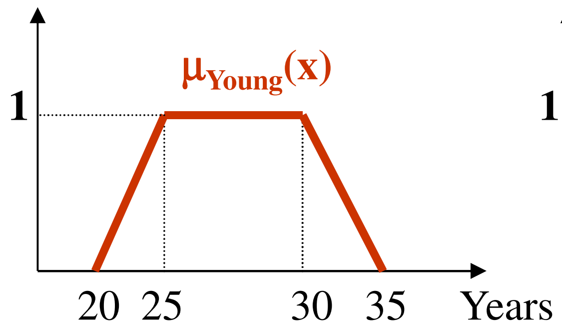
- Decision Tree (DT) algorithm is considered as one of the popular classification and regression techniques.

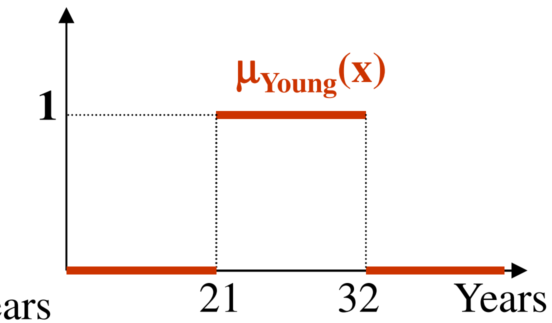
- It is an induction learning algorithm based on the training data, which has the advantages of simplicity, transparency and ability to extract decision rules.

- There are many algorithms to construct DTs such as ID3, C4.5 and CART.

Background: Fuzzy logic

- Fuzzy logic is an approach based on "degrees of truth“ logic rather than the classical "true or false" logic.
- The idea of fuzzy logic was first advanced by Zadeh in 1965.
- Fuzzy logic is well adapted for dealing with uncertain and/or incomplete knowledge, and for problems with multiple solutions.

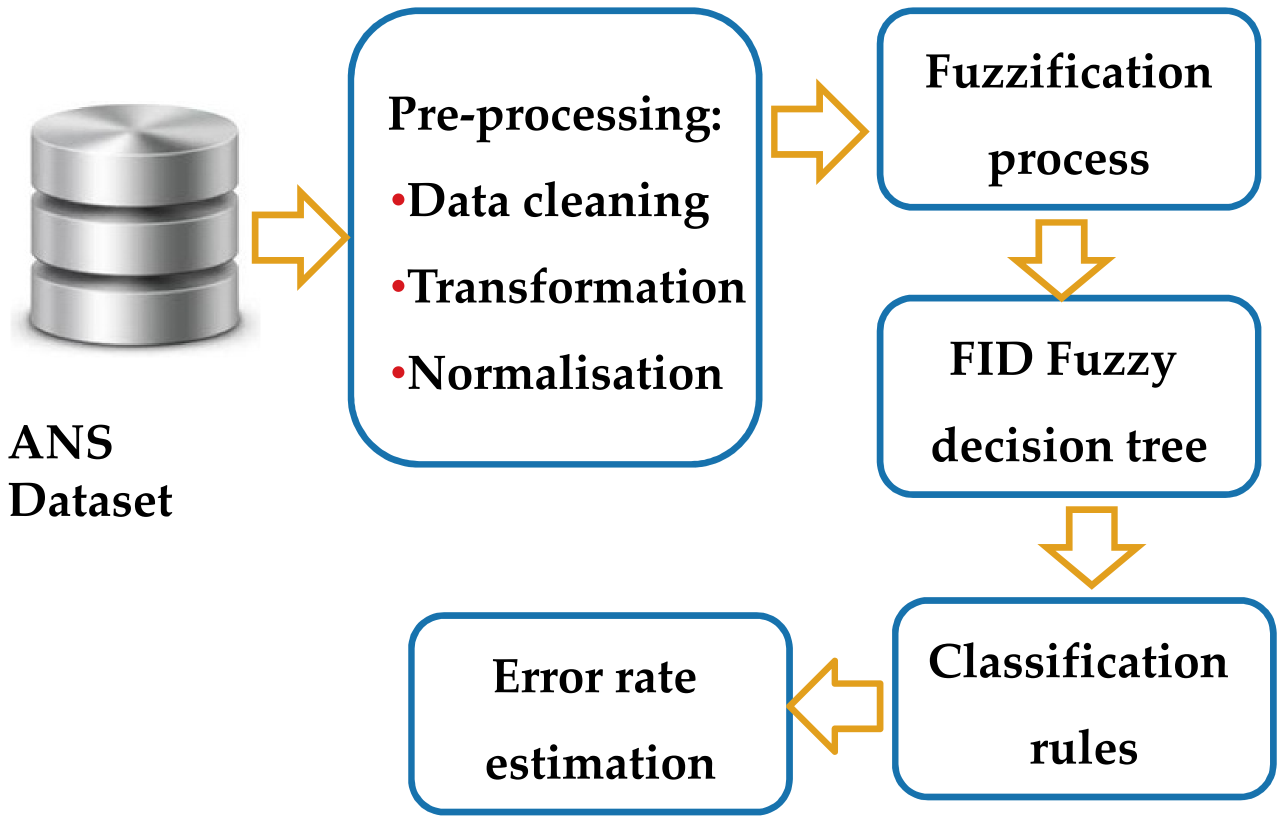




Background: Fuzzy Decision Tree

- FDTs combine the Decision Tree paradigm with the fuzzy sets theory.
- They are a generalization of crisp DTs to handle imprecise and uncertain attributes with numeric-symbolic values.
- The incorporation of fuzzy logic in building DTs requires a fuzzy partitioning for building fuzzy subsets of each input variable.
- In the fuzzy trees:
  - each node is associated with a linguistic variable
  - Each branch is associated with a fuzzy subset of this variable
  - Every path leading to a leaf of the tree will match with a fuzzy rule

Experimental design



Experimental design: medical dataset

- The dataset was collected from the ANS unit of the cardiology department of Avicenne hospital in Morocco
- It is the same one used in our previous research

- It contains:

  - The records of 178 patients
  - 66 features for each patient including general information and medical data
  - This dataset includes records of patients suffering from cardiovascular dysautonomias who went to ANS unit in the period between January 2013 and May 2014.

Experimental design

Details about input data and classes for all ANS tests.

ANS tests	Measured values	Description	Input attributes
Deep Breathing	Vagal response (VR_DB)	Vagal response measured using HR values in Deep Breathing test (DB)	Age VR
Hand Grip	Vagal response (VR_HG)	Vagal response measured using HR values in Hand Grip test (HG)	Age VR
	PSR α	Peripheral sympathetic response α measured using BP values in HG test	Age PSR
Mental stress	CSR α	Central sympathetic response α measured using BP values in Mental Stress test (MS)	Age CSR α
	CSR β	Central sympathetic response β measured using BP values in MS test	Age CSR β

Experimental design

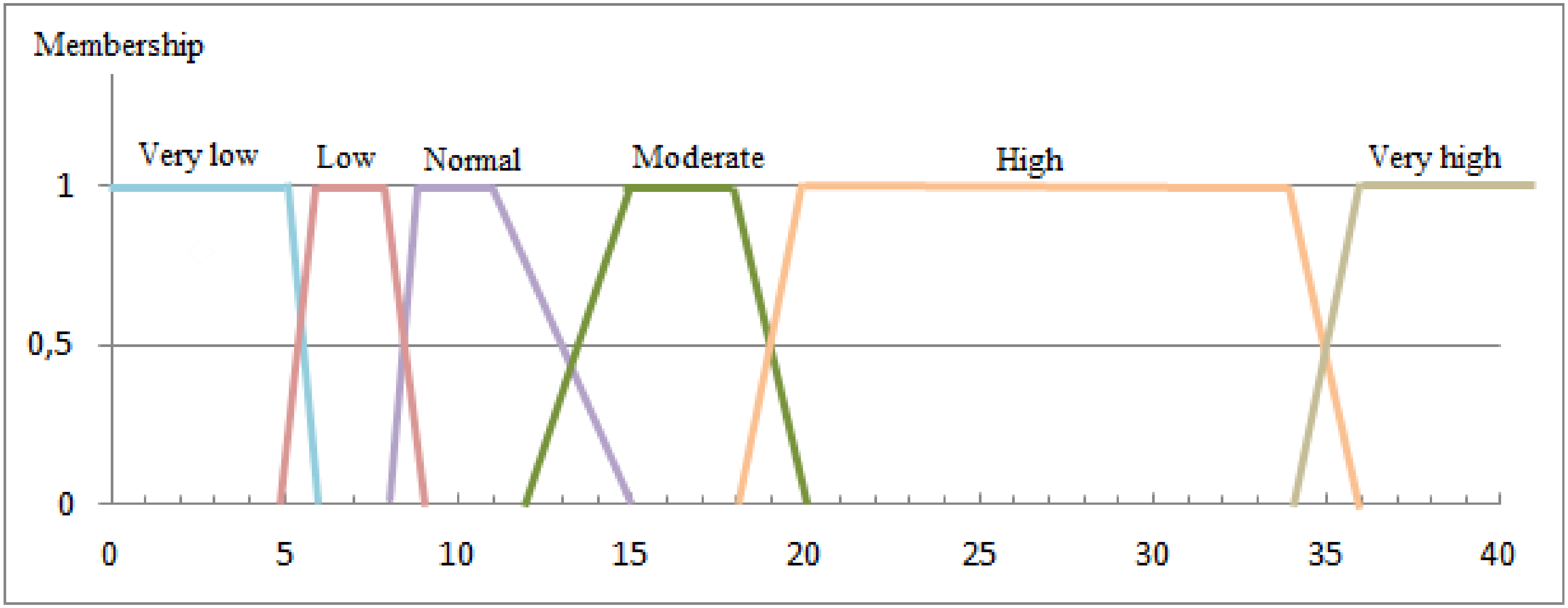
ANS tests	Measured values	Description	Input attributes
Orthostatic test	Vagal response (VR_Ort)	Vagal response measured using HR values in Orthostatic test (Ort)	Age VR
	SP_HR	heart rate measured in Orthostatic test using supine position (SP)	Age HRmin HRmax
	SP_BP	blood pressure measured in Orthostatic test using supine position	Age BPmin BPmax

Experimental design: Fuzzification process

- Fuzzification process aims at transforming numeric values to linguistic ones.
- Membership functions must be built to define the degree of membership of a numeric value to fuzzy sets of linguistic variables.
- For each linguistic value, a membership function is **empirically** defined.
- The choice of shapes of membership functions may be subjective. **Trapezoidal** shape for membership functions is used in this study.

Results

Fuzzy sets defined for VR HG attribute



Experimental design: FID

- FID fuzzy decision tree is a program which generates a fuzzy logic based decision tree from continuous and/or discrete example data.
- FID has three major components
  - Partitioning continuous attributes,
  - Building an explicit tree,
  - Inducing knowledge from the tree.
- FID includes a pruning algorithm to avoid over-fitting, which balances a tradeoff between the size of the tree and its predictive accuracy.

Experimental design

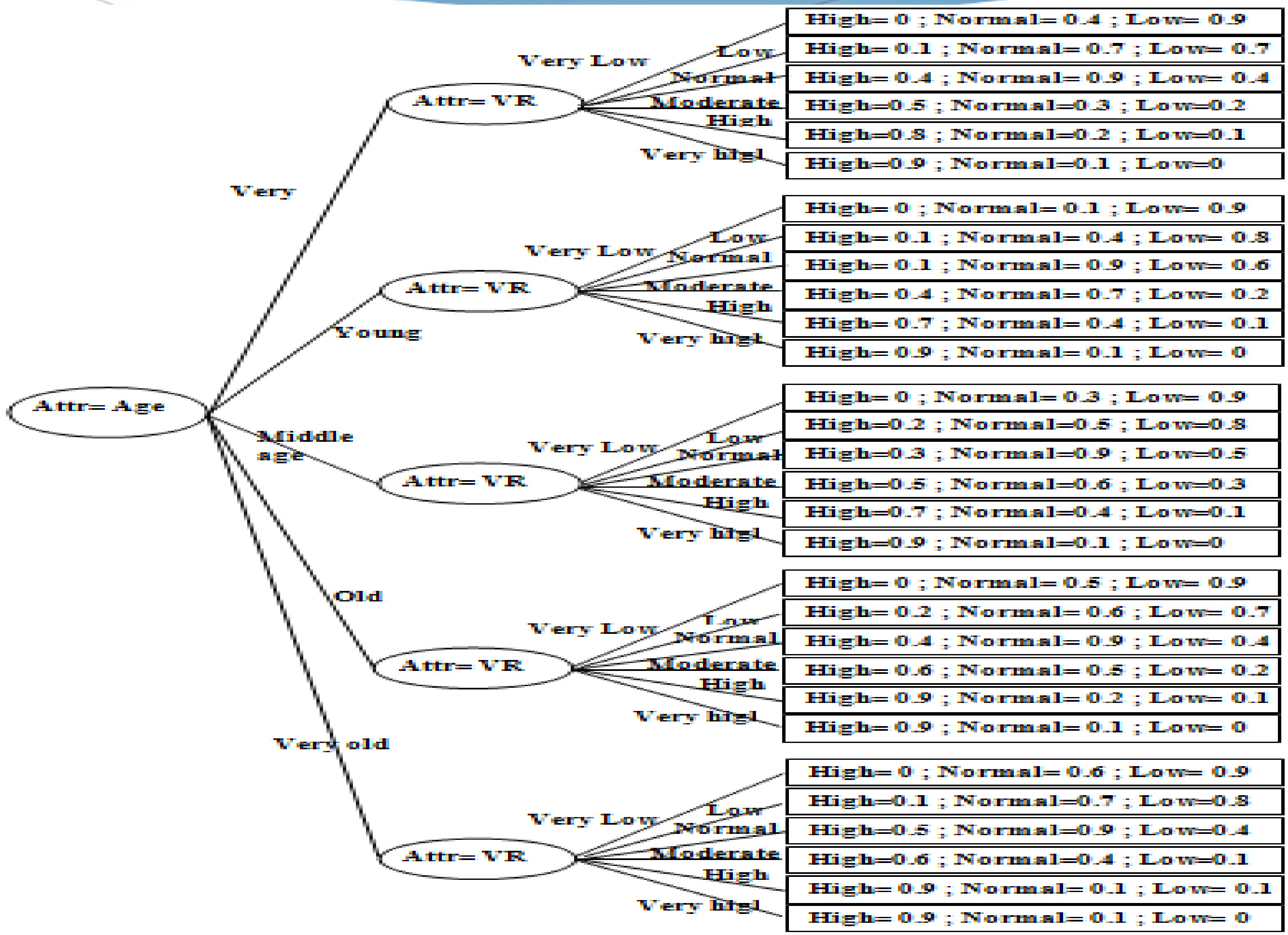
- For each ANS test, a set of preliminary conclusions is deduced for each ANS test
- For each ANS test, one , two or three conclusions were generated.
- Three classes (conclusions) were identified based on the expert's knowledge namely:
  - low,
  - normal
  - high.

8 fuzzy decision trees were generated and tested

Results

- The FID 3.4 version was used in order to generate the FDTs. It is one of the FID programs
- The data set was randomly partitioned into training and independent test sets using the 10-fold cross-validation process.
- The data set was split into ten equal sized blocks with similar class distributions
- The error rate metric was used to evaluate the accuracy of the generated classifiers

Example of an FDT



Results: learning phase

Comparison of mean error rate obtained by crisp DTs and FDTs on training sets

ANS tests	Phase	Learning phase			
		Mean error rate		Std deviation	
		Crisp DT	FDT	Crisp DT	FDT
Deep Breathing	Vagal response	3.05%	**2.69%**	1.97	0.89
Hand Grip	Vagal response	3.72%	**2.99%**	1.53	1.13
Mental stress	PSR α	1.11%	**1.02%**	0.89	1.22
	CSR α	1.02%	1.13%	0.76	0.91
	CSR β	0.34%	**0.11%**	1.88	0.83
Orthostatic	Vagal response	0.27%	0.41%	0.91	0.97
	SP_FC	1.68%	**1.19%**	1.21	1.42
	SP_TA	2.14%	**1.83%**	2.36	1.65

Results: test phase

Comparison of mean error rate obtained by crisp DTs and FDTs on test sets

ANS tests	Phase	Test phase			
		Mean error rate		Std deviation	
		Crisp DT	FDT	Crisp DT	FDT
Deep Breathing	Vagal response	2.43%	**1.75%**	2.33	1.62
Hand Grip	Vagal response	2.98%	**2.07%**	1.18	1.05
	PSR α	3.21%	**2.72%**	2.19	1.79
Mental stress	CSR α	0.93%	1.31%	0.88	0.84
	CSR β	1.99%	**1.08%**	1.21	1.58
Orthostatic	Vagal response	1.07%	1.91%	0.91	0.99
	SP_FC	7.81%	**5.02%**	0.94	1.03
	SP_TA	2.31%	**1.94%**	0.80	0.78

Results

- Using the 10 fold cross-validation method, 10 experiments were carried out for each ANS test.
- A total of 80 experiments were performed.
- For the training sets, FDTs recorded low error rates comparing to crisp DT in 92% of cases (74 of 80 trials).
- Regarding the test sets, in 89% of cases (71 of 80 trials), the error rates obtained by FDTs were lower than those obtained by crisp DTs.

Results: rule interpretation

- The numerical data of the dataset were transformed into linguistic values.
- The extracted rules included linguistic forms instead of numerical ones which made them more readable and interpretable.
- Example:

> - Rule 1: IF Age="Young" AND VR_HG="High" THEN Class="High";
> - Rule 2: IF 25<Age<=35 AND VR_HG=25 THEN Class="High"

- Rule 1 presents a classification rules extracted from the generated FDT regarding Hand Grip test
- Rule 2 presents a classification rules extracted from the crisp DTs regarding the same test and expressing the same decision.
- The classification rules generated by FDTs are easy to understand by the experts.

Conclusion and Future work

- FDT techniques were applied on a dataset extracted from the ANS unit of Avicenne hospital.
- This work compared the results of crisp and fuzzy DTs as classifiers for cardiovascular dysautonomias.
- The generated FDTs were proved to be easy to interpret and accurate than those obtained using crisp DTs.

- The small size of the dataset used which requires performing more evaluation tests over large data sets.
- Constructing fuzzy sets of inputs using fuzzy clustering techniques (fuzzy C-means, fuzzy k-modes, etc.)

Thank you