



HAL
open science

Composants à ondes élastiques de surface pour le filtrage à gabarits maîtrisés aux fréquences radios pour applications spatiales et professionnelles

Loïc Braun

► **To cite this version:**

Loïc Braun. Composants à ondes élastiques de surface pour le filtrage à gabarits maîtrisés aux fréquences radios pour applications spatiales et professionnelles. Acoustique [physics.class-ph]. Université de Franche-Comté, 2015. Français. NNT : 2015BESA2020 . tel-01508650

HAL Id: tel-01508650

<https://theses.hal.science/tel-01508650>

Submitted on 14 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SPIM

Thèse de Doctorat

UFC

école doctorale sciences pour l'ingénieur et microtechniques
UNIVERSITÉ DE FRANCHE-COMTÉ

THÈSE présentée par

CARINE CAROLINE CRIDELICH

pour obtenir le

Grade de Docteur de
l'Université de Franche-Comté

Spécialité : **Sciences pour l'ingénieur**

Influence of restraint systems during an automobile crash

Prediction of biomechanical injuries for frontal impact sled tests based on data mining

Thèse soutenue le 17 décembre 2015

VÉRONIQUE BERGE-CHERFAOUI	Rapporteur	Maître de conférences HDR, Université de Technologie de Compiègne
LATIFA OUKHELLOU	Rapporteur	Directrice de recherche, IFSTTAR
NICOLAS MÉGER	Examineur	Maître de conférences HDR, Université de Savoie, IUT d'Annecy
SERGE FENET	Examineur	Maître de conférences, Université Claude Bernard Lyon 1
NADINE PIAT	Examineur	Professeur des Universités, ENSMM, Besançon
HERVÉ GUYENNET	Examineur	Professeur des Universités, UFC, Besançon
NOURREDINE ZERHOUNI	Directeur de thèse	Professeur des Universités, ENSMM, Besançon
EMMANUEL RAMASSO	Co-directeur de thèse	Maître de conférences, ENSMM, Besançon



SPIM

Thèse de Doctorat



UFC

école doctorale **sciences pour l'ingénieur et microtechniques**
UNIVERSITÉ DE FRANCHE-COMTÉ

Influence of restraint systems during an automobile crash

Prediction of biomechanical injuries for frontal impact sled
tests based on data mining

■ **CARINE CAROLINE CRIDELICH**

ACKNOWLEDGMENTS

In this section, I would like to thank all my colleagues and managers who contributed to my thesis but also the persons who were indirectly involved, sometimes without knowing it.

First of all, my gratitude goes to my supervisor, Prof. Dr. Nouredine ZERHOUNI, for giving me the chance to write this project with some special conditions imposed by my actual situation (engineer post in Germany, thesis in parallel). And also many thanks to Dr. Emmanuel RAMASSO for his supervision and for having trusted me in spite of my rather little knowledge in Data Mining. But also for his devotion and his patience in spite of my frustrations.

The opportunity to write this thesis in an unknown field, the Data Analysis, came from Dr. Konrad ELSÄSSER, Senior Manager System Engineering Application & Simulation Global (TRW Automotive GmbH Alfdorf), who accepted that I worked on this subject in parallel to my current engineering post, but also (I hope) who supported me for believing in my chances to conduct this project to the end. I am thankful for the help and support offered by Dr. ELSÄSSER in his role of "permanent" supervisor.

Many thanks also to Christophe DIELEMANS (ENSMM Besançon), who helped me find a fitting French team for my thesis, to Ralf WAELDCHEN and Torsten SASSMANNSHAUSEN, respectively former and actual TRW Managers Systems Engineering, to Dr. Swen SCHAUB, TRW Senior Manager Engineering Strategy & Communication, for his helpful comments about the papers and to Manfred PREHN, TRW European Patent Counsel, for his help in the issuing of the contract and papers. I would also like to thank my German tutors, Dr. Bernd WEILER and Dr. Michaela KOLBE for their contribution and their advices.

And this thesis would not be finished without the help of Pablo JUESAS CANO, Mehdi EL HANOUNE and Kuevi EKUE-HETTAH (Fabrice for those who know him), who actively participated in the Matlab programming. A wink at Pablo who adjusted himself to my feminine wishes for the design of the Graphical User Interface. Also thanks to Tobias BEISSWENGER for doing the laborious work by raising all the parameters data of the sled tests.

I would also like to thank the E2/A assistant engineering, Tanja DITTBRENNER, with whom I had some laughs booking my hostel in Island. She succeeded in managing all the administrative problems that I gave her (strange invoices from Island, airline strike,...) And I do not agree with her, I am sure that these type of problems do not just happen to me !

Finally, I would like to thank Lorraine MC CORMACK-HERKOMMER, who proof read my thesis to correct all my English errors and to try to understand what I wanted to say. I hope I did not make too many spelling mistakes on this page of acknowledgments.

THESIS OUTLINE

To begin this thesis, I present in **Chapter 1: Introduction and research objectives** the general context of the research. We will see some generalities about passive safety and legislation which allow the evaluation of the safety car. After that, a presentation about the crash tests in general, then the dummies used in order to simulate the injury severity of the car occupants during a frontal impact crash. The dummy instrumentation that helps measure this injury severity will be also explained. The last section is focused on the requirements of the firm and thus of the thesis' objectives.

I then present the literature in **Chapter 2: State-of-the-art**. After a review of the current researches in automobiles crashes, the literature review is focused on the Knowledge Discovery in Databases (KDD) process and especially the clustering and classification methods. The literature review for these two methods has been focused on the automobile domain, especially safety systems and reduction of the injury severity during a road accident. It will help understand the state-of-the-art in data analysis, data mining methods, clustering and classification applying in the field of automobile safety.

I finally present the core of this thesis from **Chapters 3 until 6**.

As the benefits of the KDD has been described, I continue with the application of this process on the crash tests data through **Chapter 3: Sled tests classification through k-NN method**. Firstly, a short explication of the available data will be done. In fact, we have to differentiate the inputs (parameters of the restraints systems available during the crash) and the outputs (dummy biomechanical values measured through high tech sensors in order to evaluate the injury severity of the car occupant). The KDD process will be applied on the inputs, i.e. on the parameters of the airbags, seat belts and dummy position. These parameters have been chosen for their eventual influence on the outputs. A data mining method, the k nearest neighbour, has been also applied in order to find the nearest sled test included in the database to a new sled test. Section 2 regroups the validation for the different used algorithms. The first one helps replace some parameters missing values and the following validations are focused on for the k-nearest neighbour which allows to have a first idea of the expected score in EURO NCAP and FMVSS NR.208.

The next chapter, **Chapter 4: Clustering of the dummy biomechanical values**, is focused on the outputs, i.e. the dummy biomechanical values. In order to find correlations between the inputs and outputs, the outputs have been ranged into clusters. I present in Section 1 the outputs' clustering obtained on the passenger side for the EURO NCAP rating. The results for the driver side will be also resumed but the graphs are gathered in Appendix D. I continue with the next section, Section 2 which is based on the first section but for the FMVSS NR.208 legislation. This section also presents only one configuration for the passenger side (HIII 50th dummy belted) and the results for the others configurations and the driver side. The additional graphs can be found in Appendixes E and F.

Chapter 5: Parameters' influence through decision trees proposes an analysis of the parameters' influence for the two ratings: the EURO NCAP and the US NCAP. After a

data description, the methodology is discussed and the CART algorithm is applied, using the parameters as inputs and the clusters obtained in Chapter 4 as outputs. Only the relevant dummy parts are studied in this chapter, it means that only the results for the dummy chest for the EURO NCAP are presented and those for the dummy head and neck are resumed in Appendix G. The dummy femurs are for this test configuration not relevant because of the absence of a knee bag. All the four parts of the HIII 5th dummy influence the U.S. rating score and for this reason, is widely discussed in this Chapter.

Finally **Chapter 6: General conclusion** summarises the results obtained during this thesis and their benefits for the passive safety engineers in their daily work. Some perspectives with the application of Variational Bayesian methods are also discussed in order to improve the data clustering and the results. This project aims at being a starting point for the engineers in their objectives to improve the car safety through the amelioration of the restraint systems.

CONTENTS

1	Introduction and research objectives	1
1.1	Background on passive safety	1
1.1.1	Passive safety and accident's causes	1
1.1.2	Legislation and ratings	2
1.1.2.1	ECE regulations and Euro NCAP rating	4
1.1.2.2	FMVSS Nr.208 regulation and US NCAP rating	5
1.1.3	Crash tests dummies and injuries criteria	10
1.1.3.1	Description of the sled test facility and measured data	10
1.1.3.2	Crash test dummies and instrumentation	10
1.1.3.3	Description of the injuries criteria for the dummy body parts	13
1.1.3.4	Passive restraint systems	17
1.2	Research objectives	18
2	State-of-the-art	21
2.1	Methods for crash data analysis	21
2.1.1	Current researches in automobile crashes	21
2.1.1.1	Clustering methods	22
2.1.1.2	Classification methods	24
2.1.2	Synthesis of the research bibliography	26
2.2	The KDD process and data mining methods	26
2.2.1	Description of the KDD process	27
2.2.2	Unsupervised learning through the clustering	29
2.2.2.1	The Gustafson-Kessel algorithm as unsupervised learning	31
2.2.2.2	Fuzzy clustering	31
2.2.2.3	The GK algorithm	35
2.2.2.4	The Davies Bouldin Index	37
2.2.2.5	Generalities	37
2.2.2.6	Davies-Bouldin Index	39
2.2.3	Supervised learning through the classification	40

2.2.3.1	Background of the basic tree model	41
2.2.3.2	CART as supervised learning	43
2.2.3.3	Notations	43
2.2.3.4	The growing process	44
2.2.3.5	Splitting criteria and impurity measures	44
2.2.3.6	Stopping rules and evaluation of the tree	47
3	Sled tests classification through k-NN method	49
3.1	Application of the Knowledge Discovery in Databases	49
3.1.1	Step 1: Data selection	49
3.1.2	Step 2: Data pre-processing	49
3.1.3	Step 3: Data transformation	52
3.1.4	Step 4: Data Mining Method: k-nearest neighbours classification	53
3.1.5	Step 5: Data classification	53
3.2	Validation of the k-NN classification	54
3.2.1	Algorithm for missing data	54
3.2.2	The sled tests classification	56
3.2.3	Estimation of the Euro NCAP score with all known parameters	57
3.2.4	Estimation of the Euro NCAP score with unknown parameters	58
3.3	Discussion on the k-NN classification	59
4	Clustering of the dummy biomechanical values	61
4.1	Outputs' clustering Euro NCAP	61
4.1.1	Clustering for the HIII 50th dummy (passenger side)	61
4.1.1.1	Generalities	61
4.1.1.2	Dummy head: clusters' formation and evaluation	62
4.1.1.3	Dummy neck: clusters' formation and evaluation	64
4.1.1.4	Dummy chest: clusters' formation and evaluation	66
4.1.2	Clustering for the 50th dummy (driver side)	68
4.1.3	Conclusion for the Euro NCAP rating clustering	68
4.2	Outputs' clustering FMVSS Nr.208 sled tests	68
4.2.1	Clustering for the FMVSS Nr.208 legislation	69
4.2.2	Clustering for the HIII 50th dummy belted (passenger)	70
4.2.2.1	Dummy head: cluster's formation and evaluation	70
4.2.2.2	Dummy neck: cluster's formation and evaluation	72
4.2.2.3	Dummy chest: cluster's formation and evaluation	73

4.2.2.4	Dummy femurs: cluster's formation and evaluation	74
4.2.3	Clustering for the HIII 50th dummy unbelted (passenger)	74
4.2.4	Clustering for the HIII 5th dummy unbelted (passenger)	76
4.2.5	Clustering for the HIII 5th dummy belted (US NCAP rating, passenger)	77
4.2.6	Clustering on the driver side	79
4.3	Overview of the clustering results	80
5	Parameters' influence through decision trees	89
5.1	Data description and methodology	89
5.2	Euro ncap rating: analysis of the dummy chest	91
5.3	US ncap rating: hiii 5th belted dummy (passenger)	102
5.3.1	Dummy head	102
5.3.2	Dummy neck	109
5.3.3	Dummy chest	115
5.3.4	Dummy femurs	125
5.4	Conclusion	130
6	General conclusion	133
I	Annexes	155
A	Complement to chapter 1	157
B	Legal requirements for the ece legislation	163
B.1	Euro NCAP rating: attribution of the points	163
C	Panels of the Graphical User Interface	165
C.1	Configuration's panel	165
C.2	General panels	166
C.2.1	Panel for the driver and passenger airbag	166
C.2.2	Panel for the kneebag	166
C.2.3	Panel for the seat belt	167
C.2.4	Panel for the sled test's general parameters	168
C.2.5	Overview panel	169
C.3	Sled test's results	169
D	Complement to Chapter 4.1 (Euro NCAP rating, driver side)	171

D.1	Dummy head	171
D.2	Dummy neck	171
D.3	Dummy chest	172
D.4	Dummy femurs	172
E	Complement to Chapter 4.2 (FMVSS Nr.208, passenger side)	177
E.1	Passenger side : HIII 50th dummy unbelted	177
E.1.1	Dummy head	177
E.1.2	Dummy neck	179
E.1.3	Dummy chest	179
E.1.4	Dummy femurs	179
E.2	Passenger side : HIII 5th dummy unbelted	179
E.2.1	Dummy head	179
E.2.2	Dummy neck	179
E.2.3	Dummy chest	180
E.2.4	Dummy femurs	180
E.3	Passenger side : HIII 5th dummy belted (rating)	180
E.3.1	Dummy head	181
E.3.2	Dummy neck	181
E.3.3	Dummy chest	181
E.3.4	Dummy femurs	181
F	Complement to Chapter 4.2 (FMVSS Nr.208, driver side)	183
F.1	Driver side : HIII 50th dummy unbelted	183
F.1.1	Dummy head	183
F.1.2	Dummy neck	183
F.1.3	Dummy chest	184
F.1.4	Dummy femurs	184
F.2	Driver side : HIII 5th dummy unbelted	185
F.2.1	Dummy head	185
F.2.2	Dummy neck	186
F.2.3	Dummy chest	186
F.2.4	Dummy femurs	187
F.3	Driver side : HIII 50th dummy belted (rating)	187
F.3.1	Dummy head	188
F.3.2	Dummy neck	188

F.3.3	Dummy chest	189
F.3.4	Dummy femurs	189
F.4	Driver side : HIII 5th dummy belted	190
F.4.1	Dummy head	190
F.4.2	Dummy neck	190
F.4.3	Dummy chest	191
F.4.4	Dummy femurs	191
G	Decision trees Euro NCAP	193
G.1	Dummy head	193
G.2	Dummy neck	196

INTRODUCTION AND RESEARCH OBJECTIVES

Worldwide road accidents kill 1.2 million persons per year and hurt forty times more [80]. Government and international associations campaign since many years for safe roads and their actions have already proved their efficiency, without being able to influence the inequalities between countries. Indeed, more than 90% of the deaths take place accidentally in low-income or in developing countries, which have less developed active and passive safety systems compared to Europe or North America.

Since several years, the automobile safety is one of the priorities of all countries and is of increasing customer anxiety. The automobile manufacturers continually develop new restraint systems to help improve occupant protection during a vehicle collision. The active safety regroups all the elements connected to the vehicle and to the environment, which thanks to their presence or functionality can avoid the occurrence of an accident. This is always on call before the accident whilst the passive safety occurs during the accident and aims at minimizing the gravity of an accident. The development of the new safety systems involves the execution of crash tests based on the use of specific dummies. The test configuration depends on the legislation of the country. The injuries imposed on the dummies are estimated through the measurement of biomechanical criteria. 28 000 sled tests data, including frontal, side and rear impact crashes have been collected by automotive suppliers. A part of them will be used to discover correlations between all the sled tests. These correlations will help engineers to understand the relationships between the input's parameters (test configuration, restraint systems) and the outputs (injury severity) and also to reduce the daily amount of sled tests.

1.1/ BACKGROUND ON PASSIVE SAFETY

1.1.1/ PASSIVE SAFETY AND ACCIDENT'S CAUSES

The automobile is nowadays the most often used mean of transport. Safety had to be developed during the last years to reduce the accidents that this popularity engenders. In 2003, 6.613 people died in Germany in an automobile accident and, mainly thanks to the development of active and passive safety, this number has been reduced in 2011 by almost 42% (3.648 deaths) [73]. As shown in Fig. 1.1(a) [43], the percentage of accidents was decreasing during the years but also the proportion of accidents which caused not

damages, even if automobile buyers take safety into consideration before choosing their car (Fig. 1.1(b)).

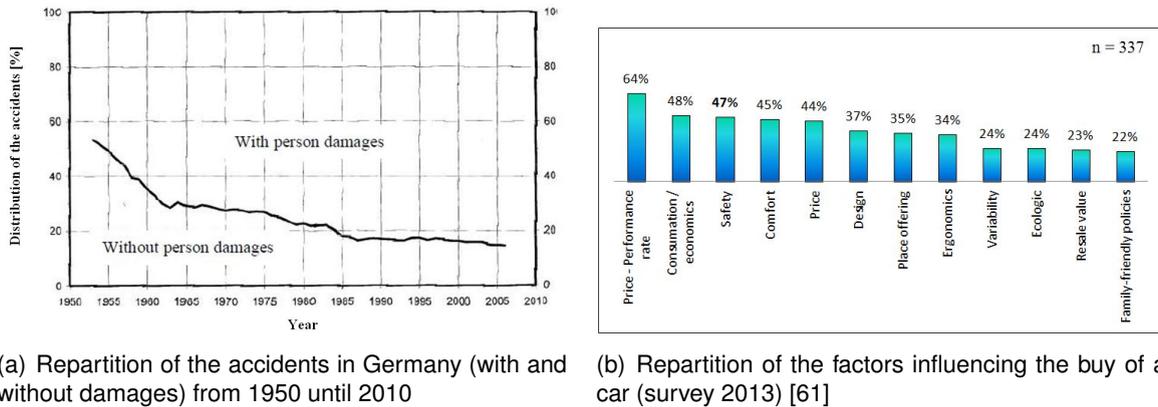


Figure 1.1: Reduction of accidents and consideration of the car safety

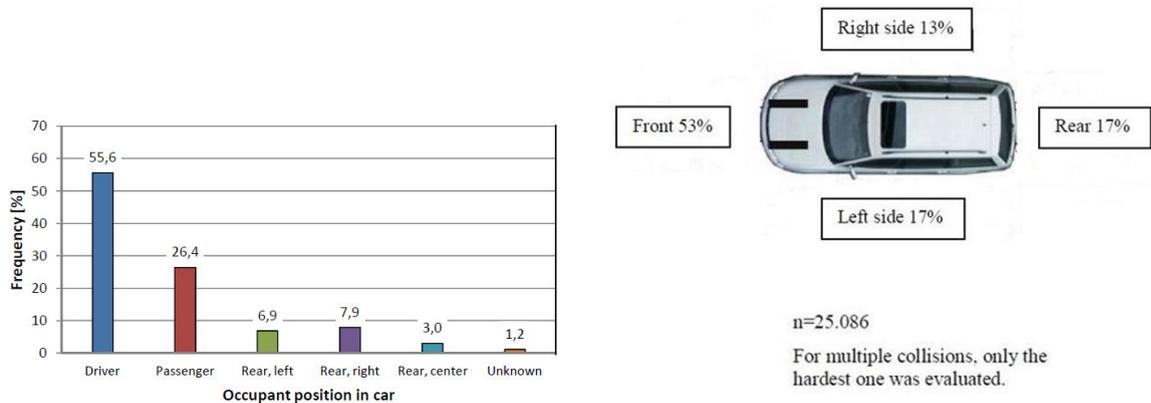
The reasons explaining an automobile accident are multiple, but driver behaviour is often the main cause. Passive safety aims at "repairing" human errors. It was established (Fig. A.1) that driver mistakes by turn or priority mistakes are the main causes responsible of an accident. Seat belt and airbags cannot avoid an accident (this is the objectives of active safety) but aim at protecting car occupants during the crash. Without passive safety systems, humans errors would have irreversible effects. Fig. 1.2 presents the repartition of collision's impact and most exposed occupants to injuries (i.e. the driver and passenger seated on the first row). According to Johannsen [65], the most common impact direction on the German roads is the frontal collision, with 53% of the total number of accidents (Fig. 1.2(b)). The remaining 47% are separated between right side, left side and rear evenly. The data was collected by the German In Depth Accident Study (GIDAS) between July 1999 and June 2013. The driver and passenger are the most sensitive occupants (Fig. 1.2(a)) because of the predominance of frontal impact crashes. There are good arguments to focus this doctoral research on frontal impact sled tests, especially on the driver and passenger occupants.

1.1.2/ LEGISLATION AND RATINGS

In the last thirty years, lots of new legislation and programmes have been established in order to test the car safety, but also to support the customer's choice with the help of released passive safety results.

Legislation and ratings have to be differentiated. The legislation has to be completed to enable a new car model to go on the market. It regroups many test configurations for frontal, side and pedestrian safety (Fig. 1.3). The most famous legislation is the Economic Commission for Europe (ECE) regulations and for the American market the Federal Motor Vehicle Safety Standards (FMVSS).

All new car models must pass certain safety tests before they are sold. The test configuration (dummies, initial velocity, direction of the impact crash) is defined by the legislation of the country and depends on the restraint systems that have to be validated. Addi-



(a) Repartition of the accidents' frequency according to the occupant position in car

(b) Repartition of the collision's impacts - Dominance of frontal impact crashes

Figure 1.2: Preponderance of frontal impact crashes with driver and passenger as most exposed car occupants

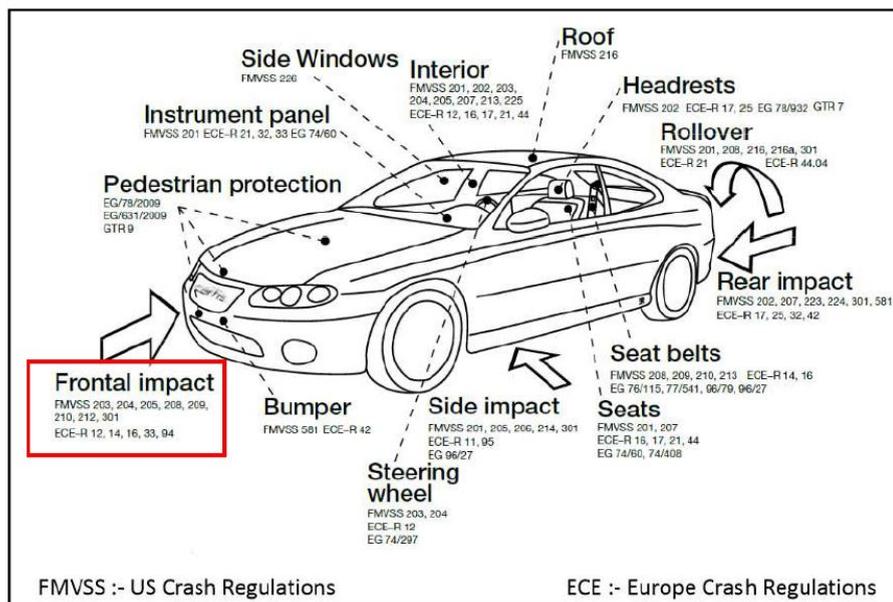


Figure 1.3: Main crash regulations in Europe and USA

tionally customer ratings are executed independently of the customer and legislation to provide a minimum safety for new cars. European New Car Assessment Program (EURO NCAP) evaluates the car safety for Europe, US NCAP for the U.S.A. and LATIN NCAP for South America. New Car Assessment Program (NCAP) is a non government program based on the "Allgemeine Deutsche Automobil-Club" (ADAC) programm which regroups methods of tests, automobile designs and others tests relating to the automobile. The first one NCAP has been created in America by the United States National Highway Traffic Safety Administration (NHTSA) in 1979. Its methods have been continually improved during the years by adapting it to the new technologies and restrictions. ADAC is yet

an active member of the EURO NCAP, as other famous associations like the "Fédération Internationale de l'Automobile" (FIA). Each rating stipulates the test setup (rigid barrier, overlap), impact velocity and dummy type. This independent assessment evaluates the occupants damages and informs the customer on the car safety level (star evaluation).

1.1.2.1/ ECE REGULATIONS AND EURO NCAP RATING

The United Nations Economic Commission for Europe (UNECE or ECE) regulations are European vehicle requirements initiated by the legislation for car safety, lighting and environment protection. Five main rules regulate frontal impact crashes. ECE-R94 is in charge of the occupant protection for a frontal collision which is defined as an impact between two cars. This collision is simulated through an automobile crash against a deformable barrier with an overlap of 40%. The impact velocity is fixed at 56 km/h and the driver and passenger occupants must be HIII 50th dummies belted. This dummy corresponds to the supposed median adult male drivers (based on U.S. data).

Since 1997, EURO NCAP is a programme based in Brussels, Belgium which is in charge in Europe to evaluate the passive and active safety on cars and which attributes a stars number corresponding to the obtained score in the test. Table 1.1 resumes the safety performance's definition according to EURO NCAP [72].

Table 1.1: Definition of the safety performance for the EURO NCAP rating

	Definition of the safety performance according to EURO NCAP
★★★★★	Overall good performance in crash protection. Well equipped with robust crash avoidance technology
★★★★	Overall good performance in crash protection; additional crash avoidance technology may be present
★★★	Average to good occupant protection but lacking crash avoidance technology
★★	Nominal crash protection but lacking crash avoidance technology
★	Marginal crash protection

Since 2009, EURO NCAP gives one overall star rating per vehicle. This score takes four sections into consideration:

- ◆ **Section I:** Adult occupant protection
- ◆ **Section II:** Child occupant protection
- ◆ **Section III:** Pedestrian protection
- ◆ **Section IV:** Safety assistance

For the offset frontal impact crash, the EURO NCAP rating test depicts that the driver and the passenger have to be a HIII 50th Percentile dummy. The impact velocity is 64 km/h against a deformable barrier (with an overlap of 40%), the crash velocity for the rating is thus higher as for the ECE-R94 regulation. Fig. 1.4 resumes the crash configuration for ECE regulation and EURO NCAP rating.

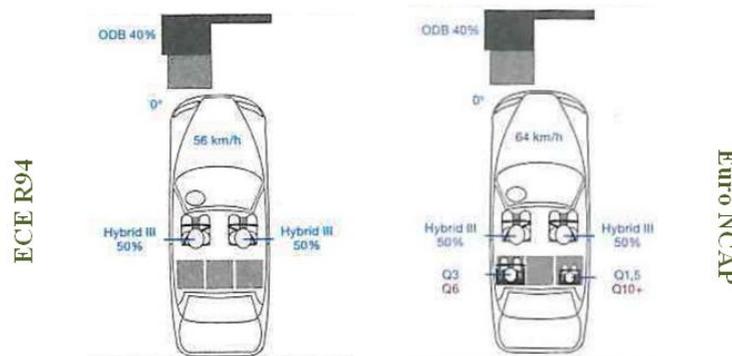


Figure 1.4: European regulation and EURO NCAP rating for frontal impact

For each section of the rating, a weight factor is applied on the score. These weights are continually updated from time to time. Table 1.2 summarises the weight factors for the period 2013-2015 [59] and confirms the importance of the adult occupant protection in EURO NCAP rating.

Table 1.2: Weight factors for the EURO NCAP sections

Year	2013	2014	2015
Section I: Adult occupant protection	50%	40%	40%
Section II: Child occupant protection	20%	20%	20%
Section III: Pedestrian protection	20%	20%	20%
Section IV: Safety assistance	10%	20%	20%

Table 1.3 explains the overall calculation of the EURO NCAP rating score including all four sections with their weights from 2014 [75]. It specifies for example that for adult occupant protection (Section I), the offset frontal impact crash represents 16 points out of 38. The Section I is additionally weighted at 40% for the whole rating score calculation. The rating score is then deduced from the four sections score.

For the frontal occupants, the maximal points that can be reached by the driver and the passenger is 16 points each, allocated evenly between the four dummy regions : head/neck, chest, knee/femur/pelvis and leg/foot (Table 1.4). Only the half score (i.e. 8 points) is considered for the overall score calculation, which explains the 16 points (driver and passenger) of the Section I adult occupant protection (Table 1.3).

The ECE-R94 regulation and the EURO NCAP rating control the car safety intended for the European market. Their equivalence for the U.S.A. are the FMVSS regulation (in particular the section 208 for occupant crash protection) and the US NCAP rating, which represent an additional challenge for the passive safety engineers because of the variety of dummies type and test configurations.

1.1.2.2/ FMVSS NR.208 REGULATION AND US NCAP RATING

Federal Motor Vehicle Safety Standard (FMVSS) are the regulations provided by the U.S. legislation, specifying the design, construction and safety requirements for the automo-

Table 1.3: Calculation of the overall EURO NCAP rating score (2014)

	Section I: Adult occupant protection	Section II: Child occupant protection	Section III: Pedestrian protec- tion	Section IV: Safety assistance
	Offset frontal impact : 16 Side impact (MDB) : 8 Side impact (Pole) : 8 Whiplash front : 2 Whiplash rear : 1 AEB city : 3	Dyn. tests frontal: 16 Dyn. tests side : 8 Installation of CRS : 12 Vehicle : 13	Head impact: 24 Leg impact : 6 Upper leg impact : 6	Seat belt reminder : 3 Speed assistance system: 3 ESC LDW / LKD : 1 AEB interurban : 3
Max. score	38	49	36	13
Normalised score (2)	Actual score / 38	Actual score / 49	Actual score / 36	Actual score / 13
Weighting (2014)	40%	20%	20%	20%
Weighted score (4)	Normalised score*40%	Normalised score*20%	Normalised score*20%	Normalised score*20%
Rating	Min. normalised score (2) by box for the respective star rating			
★★★★★	80%	75%	60%	65%
★★★★	70%	60%	50%	55%
★★★	50%	30%	40%	30%
★★	30%	25%	20%	20%
★	20%	15%	10%	10%

Table 1.4: EURO NCAP rating score repartition for the driver and passenger

Dummy region (worst performing region)	Maximal of points	
	Driver	Passenger
Head / Neck	4 points	4 points
Chest	4 points	4 points
Knee / Femur/ Pelvis	4 points	4 points
Leg / Foot	4 points	4 points
Total score	16 points	16 points
Half score for EURO NCAP calculation	8 points	8 points
Offset frontal impact	16 points	

bile. The FMVSS NR.208 regulation is focused on the occupant crash protection. The tests are executed with and without seat belts (Fig. 1.5) and the HIII 50th and HIII 5th dummies are used as frontal car occupant (both dummies tested on both sides). The HIII 50th dummy represents the median adult male and the HIII 5th dummy corresponds with females, which are larger than the supposed smallest 5th. The velocity impact for the belted test is 56 km/h (35 mph) and for the unbelted configuration, 40 km/h (25 mph), both against a full-width rigid barrier.

The US NCAP rating is based on this regulation. As for the EURO NCAP rating, it enables the evaluation of the car through a stars evaluation. The U.S. rating depicts that the

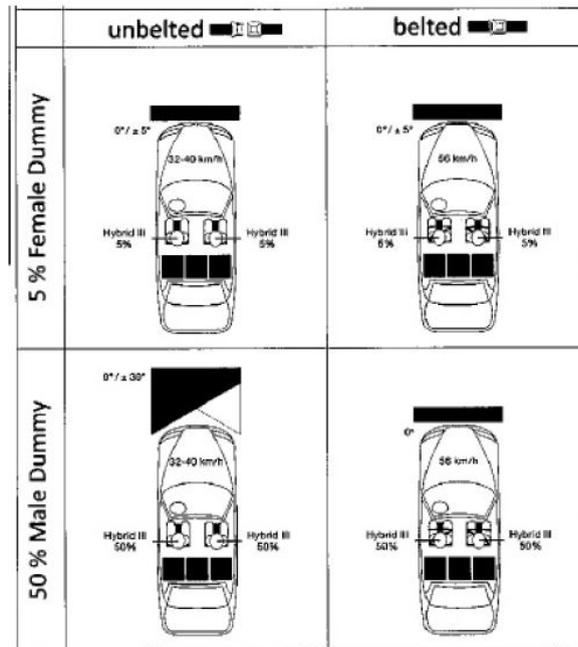


Figure 1.5: The four FMVSS NR.208 test configurations (in-position and full-width test)

driver has to be a HIII 50th dummy and the passenger a HIII 5th dummy. Both have to be belted, i.e. the impact velocity against the full-width rigid barrier is fixed at 56 km/h (35 mph), instead of 64 km/h against a deformable barrier (overlap of 40%) for the EURO NCAP rating. The stars number is defined according to the Relative Risk Score (RRS) (Table 1.5). This parameter is based on the combined probability P_{comb} , which evaluates the whole injury severity of the occupant by taking the single probabilities of the dummy part into account. The single probabilities are based on the Abbreviated Injury Scale (AIS), a coding system created to describe the injury severity and to symbolize the risk (from 1 to 6) to life associated with the injury. For example, AIS 1 regroups minor injuries and AIS 6 fatal injuries. AIS 1+ regroups all risk injuries from the minor one. Generally AIS 3+ injury risk curves are taken into account (it corresponds to serious to fatal injuries), except for the femur forces which use AIS 2+ (moderate to fatal injuries).

The injury criterion measured by the dummy sensors during the crash enables the calculation of the single injury probability which is expressed differently for the HIII 50th dummy (Table 1.6) and HIII 5th dummy (Table 1.7) [42]. The single probabilities of the head are calculated with the cumulative normal distribution defined as:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt \quad (1.1)$$

The combined probability P_{comb} is calculated according to Eq. 1.2 (valid for the driver and passenger side):

$$P_{comb} = 1 - (1 - P_{head}) \cdot (1 - P_{neck}) \cdot (1 - P_{chest}) \cdot (1 - P_{femur}) \quad (1.2)$$

Table 1.5: Star rating according to the injury probability and/or the Relative Risk Score (RRS)

	Probability	RRS
★★★★★	$P_{comb} < 0,10$	$RRS < 0,667$
★★★★	$0,10 \leq P_{comb} < 0,15$	$0,667 \leq RRS < 1,0$
★★★	$0,15 \leq P_{comb} < 0,20$	$0,10 \leq RRS < 1,33$
★★	$0,20 \leq P_{comb} < 0,40$	$1,33 \leq RRS < 2,667$
★	$0,40 \leq P_{comb}$	$2,667 \leq RRS$

Table 1.6: Injury criteria and probabilities for the HIII 50th dummy (driver side)

Dummy part	Injury criterion	Single injury probability
Head	HIC 15 (-)	$P_{head}(AIS\ 3+) = \Phi\left(\frac{\ln(HIC15)-7,45231}{0,73998}\right)$ where Φ is the cumulative normal distribution (Eq. 1.1)
Neck	N_{ij} and tension/compression (in kN)	$P_{NeckNij}(AIS\ 3+) = \frac{1}{1+\exp(3,2269-1,9688*N_{ij})}$ $P_{NeckTension}(AIS\ 3+) = \frac{1}{1+\exp(10,9745-2,375*NeckTension)}$ $P_{NeckCompression}(AIS\ 3+) = \frac{1}{1+\exp(10,9745-2,375*NeckCompression)}$ $\Rightarrow P_{Neck} = \max(P_{NeckNij}, P_{NeckTension}, P_{NeckCompression})$
Chest	Deflection (in mm)	$P_{chest}(AIS\ 3+) = \frac{1}{1+\exp(10,5456-1,568*(ChestDeflection)^{0,4612})}$
Femur	Force left and right (in kN)	$P_{femur}(AIS\ 2+) = \frac{1}{1+\exp(5,7949-0,5196*(FemurForce))}$

Table 1.7: Injury criteria and probabilities for the HIII 5th dummy (passenger side)

Dummy part	Injury criterion	Single injury probability
Head	HIC 15 (-)	$P_{head}(AIS\ 3+) = \Phi\left(\frac{\ln(HIC15)-7,45231}{0,73998}\right)$ where Φ is the cumulative normal distribution (Eq. 1.1),
Neck	N_{ij} and tension/compression (in kN)	$P_{NeckNij}(AIS\ 3+) = \frac{1}{1+\exp(3,2269-0,7619*N_{ij})}$ $P_{NeckTension}(AIS\ 3+) = \frac{1}{1+\exp(10,958-3,770*NeckTension)}$ $P_{NeckCompression}(AIS\ 3+) = \frac{1}{1+\exp(10,958-3,770*NeckCompression)}$ $\Rightarrow P_{Neck} = \max(P_{NeckNij}, P_{NeckTension}, P_{NeckCompression})$
Chest	Deflection (in mm)	$P_{chest}(AIS\ 3+) = \frac{1}{1+\exp(10,5456-1,7212*(ChestDeflection)^{0,4612})}$
Femur	Force left and right (in kN)	$P(AIS\ 2+) = \frac{1}{1+\exp(5,7949-0,7619*(FemurForce))}$

Once P_{comb} is determined, the ratio called Relative Risk Score is deduced from it, set to 15%. This percentage is imposed by NHTSA and allows an eventual adaptation of the values' conditions for reaching 5 stars.

Each measured criterion influences the rating score differently. Fig. 1.6 represents the single probability of each dummy part and their influence on the rating for the HIII 50th dummy (driver side) and for the HIII 5th dummy (passenger side) (graphical logarithmic representation through the curves $-100 * \log(1 - P_{dummypart})$).

The FMVSS NR.208 regulation as well as the US NCAP and EURO NCAP ratings are considered in this study. Because of a lower velocity in comparison with the rating, the ECE-R94 regulation is not taken into consideration. Dummies enable the validation of these regulations with the evaluation of the occupant injury severity measured with high tech sensors.

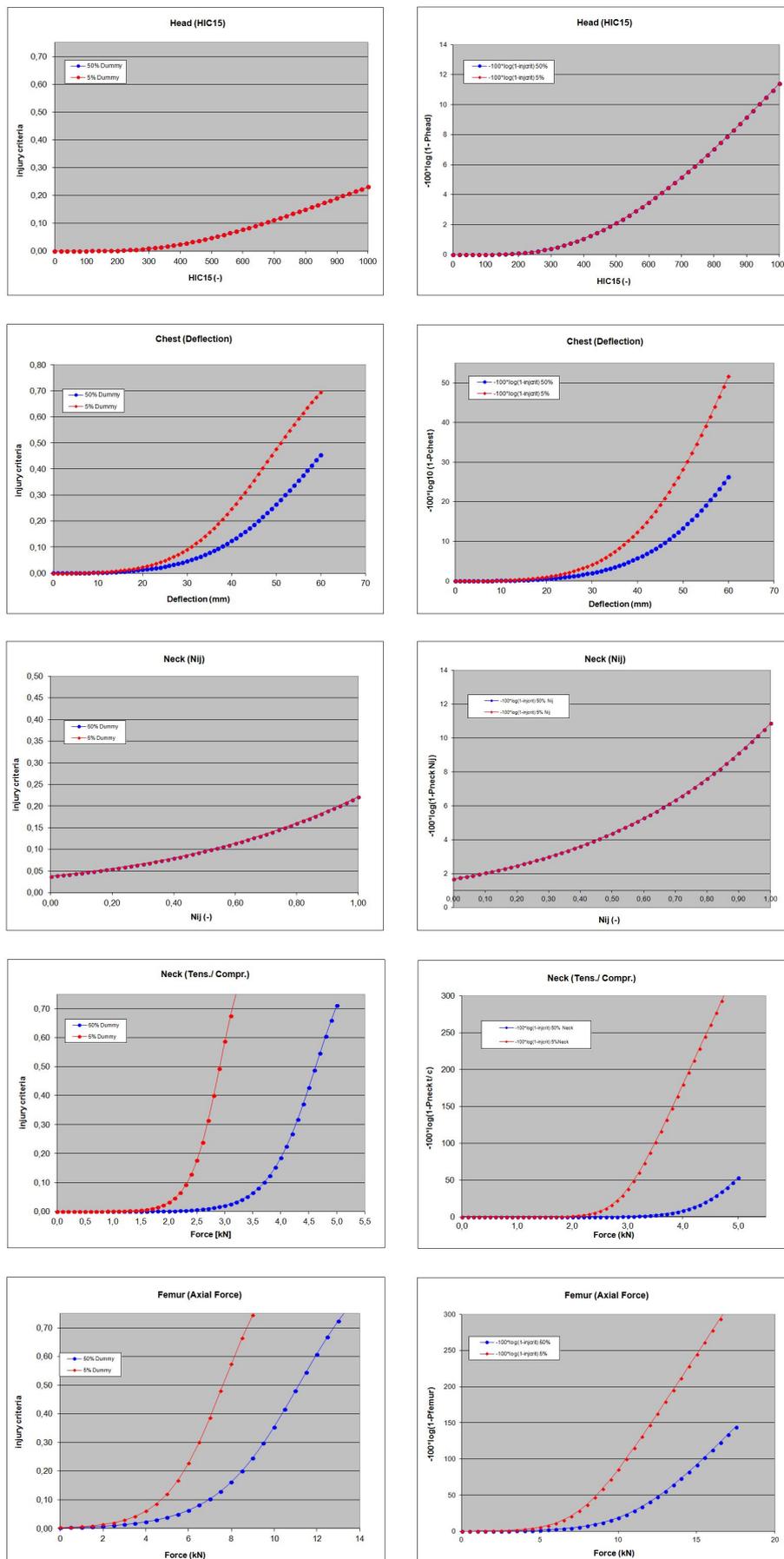


Figure 1.6: Injury criteria and their logarithmic representation using for the U.S. rating score calculation

1.1.3/ CRASH TESTS DUMMIES AND INJURIES CRITERIA

The level of injury severity of the car occupant is simulated with the help of sled test facilities and calibrated devices. These high tech dummies are composed of sensors which measure in real time the biomechanical values during the frontal sled test.

1.1.3.1/ DESCRIPTION OF THE SLED TEST FACILITY AND MEASURED DATA

For the development and dynamical testing of frontal airbags as well as seat belt systems, the automotive suppliers are equipped with a servo hydraulic HyG-Sled-facility from MTS Systems Corporation. The sled facility with the test setup is accelerated by a hydraulic shock piston from the standstill on the given speed. Besides, a defined acceleration is considered. The required impulse energy to reach the aimed acceleration is provided by compression and storage of nitrogen gas in hydro pneumatic pressure warehouses. The hydro system liquid is passed on to the three-stage servo-ventilation system which steers the oil flow in dependence of the wished acceleration to the concentric ordered working cylinder. The shock piston in the working cylinder accelerates the sled test platform with the test setup applied to it. For the simulation of a frontal impact, a massive reinforced buck of a certain vehicle type is fixed on the sled test facility with all system components, required dummies and measurement instrumentation (Fig. 1.7(a)). This type of facility, which acceleration can reach 110 g with a weight of the buck of maximal 2500 kg, is intended to reproduce a realistic, repetitive, and as a rule non-destructive way, the effects of a real automotive crash. With the given acceleration-time, the buck is implemented against the original direction of the traffic of the real vehicle. The facility is composed of:

- ◆ a jack of push with a force between 1500 and 4000 kN (or more) equipped with a servo valve of which the last floor has a flow of more than 100 000 L/min (Fig. 1.7(b));
- ◆ a trolley of shocks as a support of the reinforced buck which slips on a rail equipped with an autonomous braking system;
- ◆ a track with rails from 30 to 40 meters long, necessary for the braking of the buck.

The dummies used in the sled test are retained by the restraint systems, the dummy biomechanical values are recorded and all data and videos are documented and stored.

Since 2000, a large amount of data coming from the frontal sled test facility have been collected by some automotive suppliers. The number of available data is estimated at approximately 28,000 data (2001-2014).

1.1.3.2/ CRASH TEST DUMMIES AND INSTRUMENTATION

The Anthropomorphic Test Devices (ATD), or crash test dummies, are used by automotive companies as occupant to test the safety of any type of vehicle. There are different types of dummies, according to the kind of collision being checked. The most famous are:

- ◆ Hybrid: the first type of ATDs to be created. It is used for frontal impact crashes;

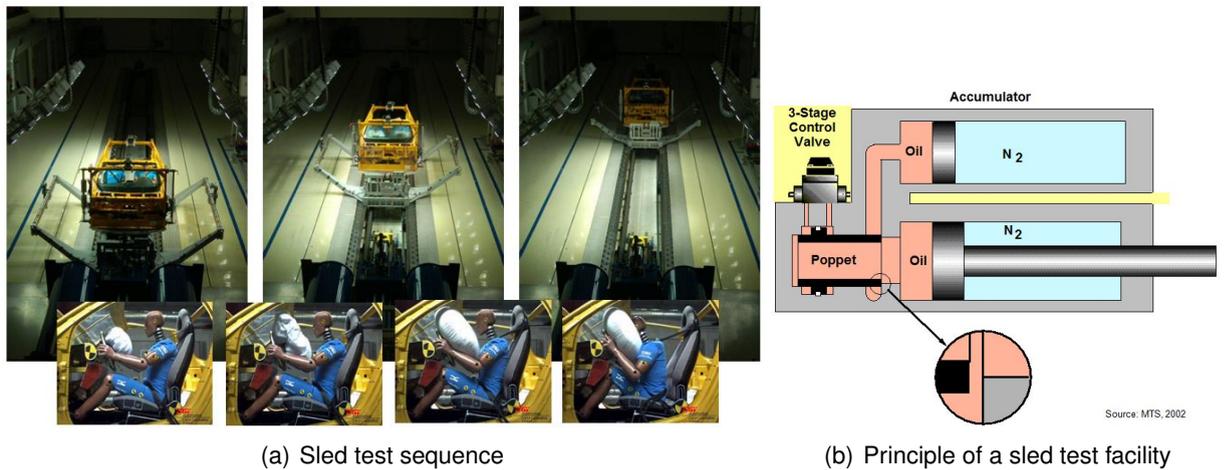


Figure 1.7: Description of a sled test facility (HyG)

- ◆ SID: Side Impact Dummy. As its name suggests, it is used for side collisions;
- ◆ BioRID: Biofidelic Rear Impact Dummy. It is used for rear impact crashes;
- ◆ CRABI: Child Restraint Air Bag Interaction. This 12-month-old child is used to test forward and rear facing car seats.

For frontal impact crashes, three types of adult Hybrid dummies have been determined based on the characteristics of the people working in the American army:

- ◆ Hybrid III 95th Percentile Large Adult Male: it corresponds to the supposed males larger than 95th of the male drivers;
- ◆ Hybrid III 50th Percentile Middle Adult Male: median adult male;
- ◆ Hybrid III 5th Percentile Small Adult Female: it corresponds with females, which are larger than the supposed smallest 5th.

Table 1.8: Characteristics of the three Hybrid III Percentile dummies (weight and size)

	Hybrid III 5th Percentile	Hybrid III 50th Percentile	Hybrid III 95th Percentile
Weight	50 kg	77 kg	100 kg
Size	152 cm	175 cm	188 cm

All these dummies are calibrated devices which simulate the physical properties of a human. The high tech sensors measure the injury potential during a sled test like the velocity of impact, deceleration rates of various body parts, impact forces, etc. A modern dummy has over 200 sensors.

These sensors are able to acquire lots of data and are regrouped into three types of instrumentation:

- ◆ Accelerometers (Fig. 1.8(a)): they measure the rate at which speed the body part moves during the collision, in a particular direction. They are available for the dummy head, chest, pelvis, legs and eventually feet;
- ◆ Load sensors (Fig. 1.8(b)): they measure the amount of force applied on different body parts during the collision and convert this input mechanical force into an electrical output signal (i.e. force-time curves);
- ◆ Motion sensors (Fig. 1.8(c)): they are fitted in the dummy chest to measure its deflection during a collision.

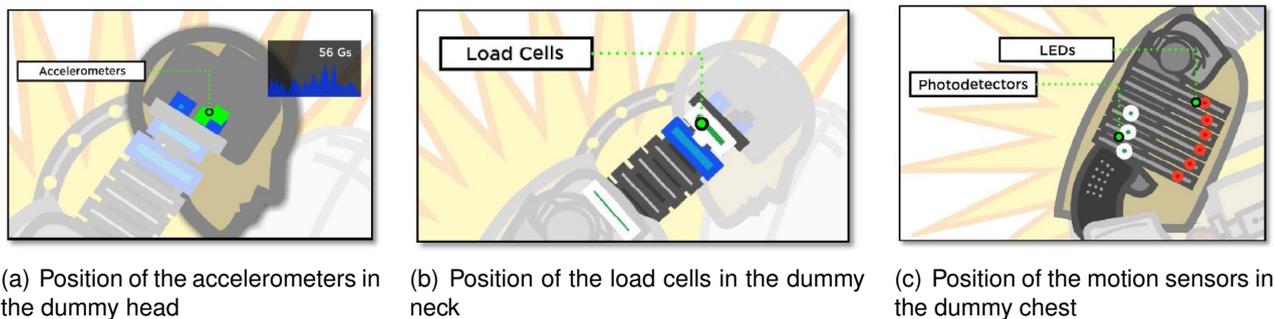


Figure 1.8: Dummy sensors

Table 1.9 and Table 1.10 present the different sensors available for respectively HIII 5th and HIII 50th dummies constructed by the company "Humanetics":

Table 1.9: Instrumentation's description of the HIII 5th dummy

Location	Description	Channels
Head	3 Accelerometers in a triax pack	Ax, Ay, Az (used for HIC)
Neck	Six-Axis Upper Neck Load Cell Five-Axis Lower Neck Load Cell	Fx, Fy, Fz, Mx, My, Mz Fx, Fy, Fz, Mx, My
Thorax	3 Accelerometers in a triaxial array Chest Displacement Potentiometer Std Equipment Fixe-Axis Thoracic Spine Load Cell	Ax, Ay, Az Fx, Fy, Fz, Mx, My
Lumbar Spine	Fixe-Axis Lumbar Spine Load Cell	Fx, Fy, Fz, Mx, My
Pelvis	3 Accelerometers or 1 Triax pack A.S.I.S. Load Cell (Iliac Wings)	Ax, Ay, Az Fx, My (per side)
Femur	Uniaxial Femur Load Cell Six-Axis Femur Load Cell Knee Shear Displacement	Fx (per leg) Fx, Fy, Fz, Mx, My, Mz (per leg) Dx (per knee)
Lower legs	Biaxial Knee Clevis Load Cells Four Axis Upper Tibia Load Cells Four Axis Lower Tibia Load Cells	Fx, Fz (per leg) Fx, Fz, Mx, My (per leg) Fx, Fz, Mx, My (per leg)

All these sensors - but also other measurements like the acceleration of the buck - are connected to one or two mini-Data Analysis Units (mini-DAU) which are able to capture each 16 gigabytes of data. The computers can record 20.000 data points per second for every sensor during a full 300-millisecond-long sled test but this recording can record up to 70.000 data points (Fig. 1.9(a)).

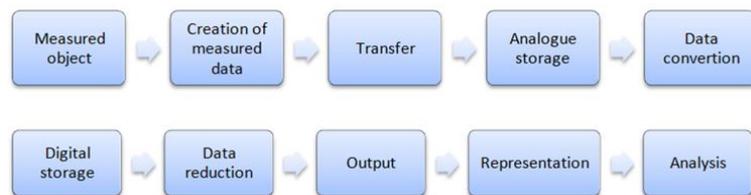
Table 1.10: Instrumentation’s description of the HIII 50th dummy

Location	Description	Channels
Head	3 Accelerometers in a triax pack Up to 15 Accelerometers	Ax, Ay, Az (used for HIC) 5X Ax, Ay, Az Head Rotation
Neck	Six-Axis Upper Neck Load Cell Six-Axis Lower Neck Load Cell	Fx, Fy, Fz, Mx, My, Mz Fx, Fy, Fz, Mx, My, Mz
Clavicle	Biaxial Load Cell (left and right)	Fx, Fz
Humerus	Four-Axis Load Cell (left and right)	Fx, Fy, Mx, My
Thorax	3 Accelerometers in a triaxial array Chest Displacement Transducer Four-Axis Rib/Spine Load Cells Fixe-Axis Thoracic Spine Load Cell	Ax, Ay, Az (Chest Accel) Dx (Stnd. Equipment) Fx, Fy, Fz, My Fx, Fy, Fz, Mx, My
Lumbar Spine	Three-Axis Lumbar Spine Load Cell	Fx, Fz, My
Pelvis	3 Accelerometers (or triax pack) Submarining Load bolts	Ax, Ay, Az Fx (3 per side)
Femur	Uniaxial Femur Load Cell or Six-Axis Upper Femur Load Cell	Fx (per leg) Fx, Fy, Fz, Mx, My, Mz (per leg)
Knee	Knee Displacement	Dx (per knee)
Lower legs	Biaxial Knee Clevis Load Cell Four Axis Upper Tibia Load Cell Four Axis Lower Tibia Load Cell	Fz (per leg) Fx, Fz, Mx, My (per leg) Fx, Fz, Mx, My (per leg)
Ankle	Fixe-Axis Load Cell	Fx, Fy, Fz, Mx, My (per leg)
Toe	Toe Load Cell	Fz (per foot)

After measuring the injuries through the sensors, the data is converted and stored digital, which enables the representation and vizualisation for the analysis of the biomechanical values (Fig. 1.9(b)).



(a) Dummy connection with mini-DAU



(b) Preparation of the measured data for the analysis

Figure 1.9: Data measurement and transformation

These measurements are analysed in order to determine the level of injury severity of the car occupant.

1.1.3.3/ DESCRIPTION OF THE INJURIES CRITERIA FOR THE DUMMY BODY PARTS

With the help of the data coming from the GIDAS database, the German In-Depth Accident Study, Florian Kramer [65] studied the injuries on each body part of the occupant during a collision. The head is the body part which is the most exposed, as shown in Fig. 1.10, with 33,4% of rugged people who had head injuries.

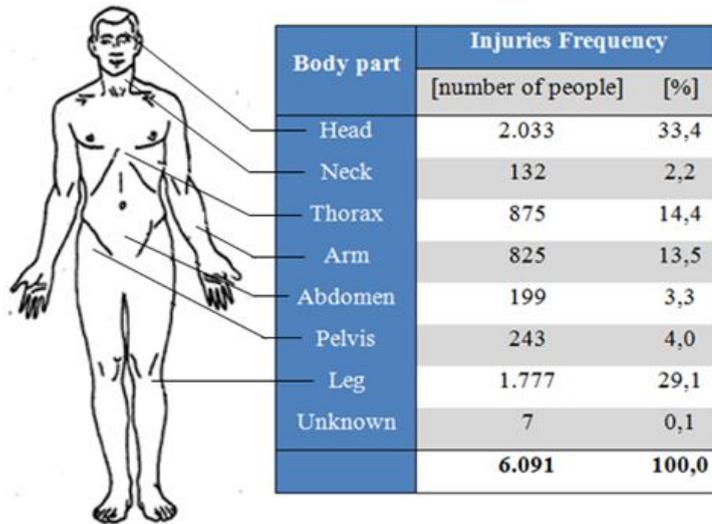


Figure 1.10: Injuries frequency for each dummy part

The neck, chest and lower extremities are also very sensitive parts and are measured during any sled test. The injuries suffered by the dummy parts are calculated directly from the measurements coming from the dummy sensors.

Head: the injuries on the dummy head are estimated with the help of the criterion "Head Injury Criterion" (HIC) [39], which is the standardized maximum integral value of the head acceleration:

$$HIC_{(t_2-t_1)} = \sup_{t_1, t_2} \left\{ \left(\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} a(t).dt \right)^{\frac{5}{2}} (t_2 - t_1) \right\} \quad (1.3)$$

$$a = \sqrt{a_x^2 + a_y^2 + a_z^2} \quad (1.4)$$

The European and American legislation impose a maximal value for the HIC:

- ◆ EURO NCAP: HIC36 < 1000. The length of the time interval (t_2-t_1) is here 36 ms;
- ◆ FMVSS NR.208 : HIC15 < 700. The length of the time interval (t_2-t_1) is here 15 ms.

The acceleration (a_x, a_y, a_z) are measured by the three head accelerometers. The US legislation and EURO NCAP rating take the head acceleration a_{3ms} under consideration, which is the maximum 3ms-average of the resulting acceleration:

$$a_{3ms} = \frac{1}{\Delta t} \int_{t_1}^{t_2+\Delta t} a(t).dt \quad \Delta t = 3ms \quad (1.5)$$

Fig. 1.11 shows the risk's function for the head during a frontal impact, with the limit between reversible and irreversible damages [43]. It concludes that from HIC = 1000, the risk to have irreversible damages is 50%.

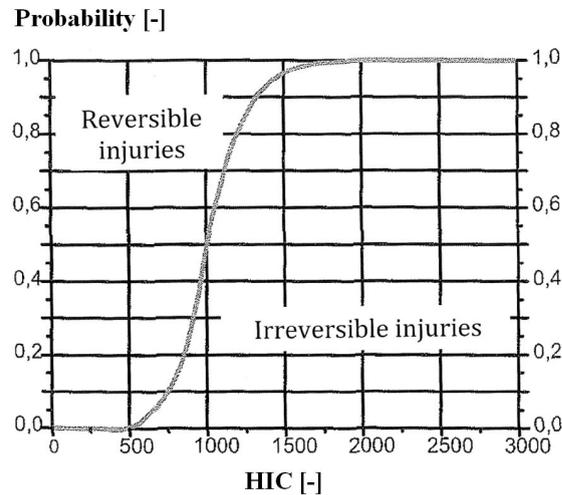


Figure 1.11: Repartition of the damages' gravity for the head (HIC criterion)

Neck: the injuries on the dummy neck are estimated through the criterion "Normalized Neck Injury Criterion" (N_{ij}) [39]. This is a combination of the forces and bending moment around a lateral axis at the transition from the head to the neck:

$$N_{ij} = \frac{F_z}{F_{zc}} + \frac{M_{OCy}}{M_{yc}} \quad (1.6)$$

with

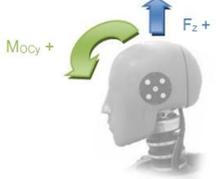
- ◆ F_z force at the transition from head to neck;
- ◆ F_{zc} critical force;
- ◆ M_{OCy} total moment = $M_y - (D \cdot F_x)$ with D the distance between the force sensor axis and the condyle axis;
- ◆ M_{yc} critical moment.

The neck criterion is thus defined by four possible neck movements (Table 1.11):

- ◆ N_{TE} : tension - extension (F_z positive, M_{OCy} negative);
- ◆ N_{TF} : tension - flexion (F_z positive, M_{OCy} positive);
- ◆ N_{CE} : compression - extension (F_z negative, M_{OCy} negative);
- ◆ N_{CF} : compression - flexion (F_z negative, M_{OCy} positive).

Chest: for this body part, there are two criteria to consider: the acceleration and the deflection. The chest acceleration, measured by the accelerometers only for the FMVSS NR.208 test, may not exceed the value of 60 g. The maximal value of the chest deflection depends on the regulation and the dummy type. For the EURO NCAP rating, the Viscous Criterion (VC) is also a parameter to be considered. This value corresponds to the

Table 1.11: Description of the four neck movements using for the points calculation in FMVSS NR.208 regulation

Tension - Extension	Tension - Flexion	Compression - Extension	Compression - Flexion
			

maximum crush of the momentary product of the thorax deformation speed and thorax deformation (Eq. 1.7):

$$VC = \text{Scaling factor} \cdot \frac{Y_{CFC180}}{Defconst} \cdot \frac{dY_{CFC180}}{dt} \quad (1.7)$$

with

- ◆ Y [m] : thoracic deformation;
- ◆ dY_{CFCxxx}/dt [m/s] : deformation velocity, $CFCxxx$ corresponds to the filter class applied on the data;
- ◆ $Defconst$ [mm] : dummy constant that is depth or width of half the rib cage.

As for the dummy head, Kramer [43] proposed a representation of the chest acceleration delimiting reversible and irreversible damages (Fig. 1.12). We can remark that the maximal value for the acceleration imposed by the FMVSS NR.208 (60 g) corresponds to 50% of the cases to have irreversible damages during a frontal impact.

Lower extremities: the Tibia Index (TI) and tibia compression, which is directly measured by the sensors, are used in order to evaluate the damages on the dummy legs. The TI is only taken into account for the EURO NCAP rating and is determined by the following formula (Eq. 2.4):

$$TI = \left| \frac{M_R}{(M_C)_R} \right| + \left| \frac{F_z}{(F_C)_z} \right| \quad (1.8)$$

$$M_R = \sqrt{(M_x)^2 + (M_y)^2} \quad (1.9)$$

with

- ◆ M_x : bending moment [Nm] around the x-axis;
- ◆ M_y : bending moment [Nm] around the y-axis;
- ◆ $(M_C)_R$: critical bending moment [Nm];
- ◆ F_z : axial compression [kN] in z-direction;

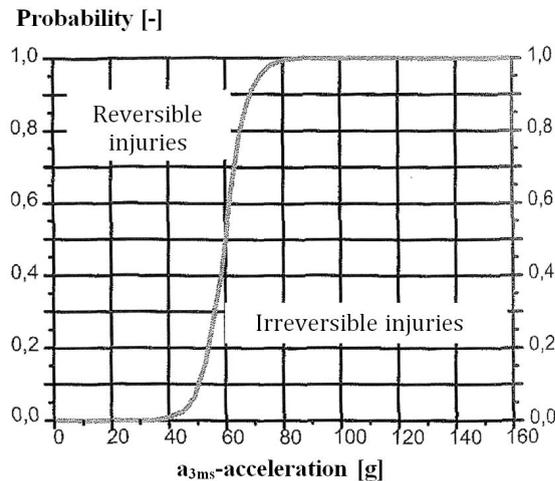


Figure 1.12: Repartition of the damages' gravity for the chest through its acceleration

- ◆ $(F_C)_z$: critical compression force [kN] in z-direction.

The high tech dummies enable the evaluation of the occupant damages during a frontal sled test. Their measured biomechanical values are continually improved by modifying passive safety systems such as airbags and seat belts.

1.1.3.4/ PASSIVE RESTRAINT SYSTEMS

The passive safety occurs during the accident and aims at minimizing its gravity. The main passive safety systems for the frontal occupants are the airbags including kneebag and the seat belts. Each restraint system is adapted according to the customer's wishes and legislation. For that, some parameters can be modified such as the volume of the airbag or the type of forces applied on the dummy (constant, degressive, regressive, switchable). An expert elicitation conducted with passive safety experts enabled to select 57 parameters on the driver side and 46 parameters on the passenger side which seem to have a significant influence on the dummy biomechanical values. Input parameters selected by the experts and considered in this research for studying their influence on the biomechanical values are regrouped into 5 categories:

- ◆ Driver airbag: volume, vent holes, tethers, inflator, steering wheel and column;
- ◆ Passenger airbag: divided panel, volume, vent holes, tethers, inflator;
- ◆ Knee bag: volume, tethers, inflator;
- ◆ Seat belt: pretensioners (retractor, buckle and anchor), angle shoulder/D-ring, webbing on spool, height adjustment, dynamic locking tongue;
- ◆ General: dummy temperature, dummy neck constellation, pulse evaluation, car classification.

Sohr [49] resumed in his publication "Benefit of adaptive occupant restraint systems" the benefits of advanced passive technologies on the biomechanical values of the occupant. His study is focused on the US NCAP rating requirements and based on valid occupant crash simulations. He showed that the integration of advanced restraint systems can reduce considerably the biomechanical criteria of the occupant. On the driver side, the introduction of an adaptive seat belt load limiter with double pretensioning, a knee airbag and an advanced airbag shape leads to the reduction of the biomechanical values of more than 60% for the US NCAP rating requirements. The crash pulse also influences the injury severity of the driver.

On the passenger side, as on the driver side, the knee airbag has a positive influence on the head and chest and its use can also reduce the chest deflection up to 40%. The adaptive seat belt load limitation and an adaptive airbag system load lead to the reduction of the head criterion by 70%. The Japanese firm Takata developed a so-called "Twinbag", a bag with two parallel pockets connected with a horizontal tube, which aims at reducing the neck injury probability (N_{ij}) up to 30%. The American firm TRW Automotive also developed a new generation of passenger airbag called "Dual Contour Bag" which can adjust its volume according to the dummy type and which reduces the N_{ij} values and the chest deflection, mainly for the HIII 5th dummy. As on the driver side, the harder the crash pulse is, the higher the biomechanical values are.

Sohr summarized the benefits of the advanced passive safety very well.

This first chapter resumed the context of the doctoral researches. Passive restraint systems (inputs) and dummy biomechanical values (outputs) have been presented and the relationships between both inputs and outputs are studied, independently of the customer and vehicle.

1.2/ RESEARCH OBJECTIVES

These researches are based on an industrial demand coming from a passive safety supplier.

Currently each passive restraint system (airbag, seat belt) are developed for each customer independently of one another, and the relationship between restraint system parameters and results have not yet been studied. After taking the customer demands into account (type of restraint system, presence or not of a kneebag, cost, etc.), the engineers develop the desired passive technologies according to the legislation and/or ratings. The biomechanical values of the dummy are measured and the results have been analysed. The sled tests are iterated until the results are either under the legal limits as well as the customer's limits or sufficient for getting the maximal rating score. These development steps are resumed in Fig. 1.13(a).

These researches aim at reducing the number of repeated sled tests through the optimisation of the passive safety configuration, and in particular through the discovering of relationships between the inputs parameters of the restraint systems and the outputs (biomechanical values). For that, data coming from the existing sled tests are studied independently of the vehicle and customer. A database containing all relevant data (airbag's parameters, seat belt parameters, dummy position) has been constructed for each dummy and legislation/ratings, thereby 5 databases have been created for each occupant side:

- ◆ EURO NCAP;
- ◆ FMVSS NR.208 with a belted HIII 5th dummy;
- ◆ FMVSS NR.208 with a belted HIII 50th dummy;
- ◆ FMVSS NR.208 with an unbelted HIII 5th dummy;
- ◆ FMVSS NR.208 with an unbelted HIII 50th dummy.

The researches are focused on the discovering of pattern / muster between the restraint system parameters and the dummy biomechanical values of each dummy part. The found pattern will help create an algorithm which will work before running a sled test, present the parameter's influence and finally propose a new solution based on the maximisation of the score. The engineers will thus be able to adapt the sled test's configuration before running it by learning on the already sled tests. It help reach better results while reduce the number of sled tests (which involves the reduction of time and cost). A representation of the new process including the discovered pattern / musters is proposed in Fig. 1.13(b).

In order to find such a model, a previous work on the crash data should be carried out. Firstly, in comparison with the researches on the road safety, it seems to be difficult to obtain a large among of data because of the lack of information for earlier sled tests. This reduced number of data causes a modification of the first thinking and produces some difficulties to work with methods specifying a large database. Moreover, once the sled tests have been selected, the first task will consist in preparing it, i.e. in researching the missing values or in replacing it principally. A big work should be thus done first on the data preparation and conversion in order to have anyway a enough big database containing clear information. The aim is here to obtain any useful information out of them. After preparing and cleaning the database, two tasks have to be differentiated: the first one will be focused on the dummy biomechanical values, i.e. the outputs. A discretization of the data in order to gather it into classes (clustering methods) could eventually help define the level of injury severity that could be then useful for the second step. The next task is focused on the inputs of the model, i.e. the parameters of the restraints systems (airbags, knee bag, seat belt), especially on the links between the inputs and the outputs. Because of the gathering of the outputs into classes, classification methods seem to be appropriate. The created model will help the engineers understand the relationships between some parameters and the test results. They can thus take into account these links before choosing / modifying a parameter. This learning on previous data will increase the understanding of the restraint systems and their influence on the car occupant.

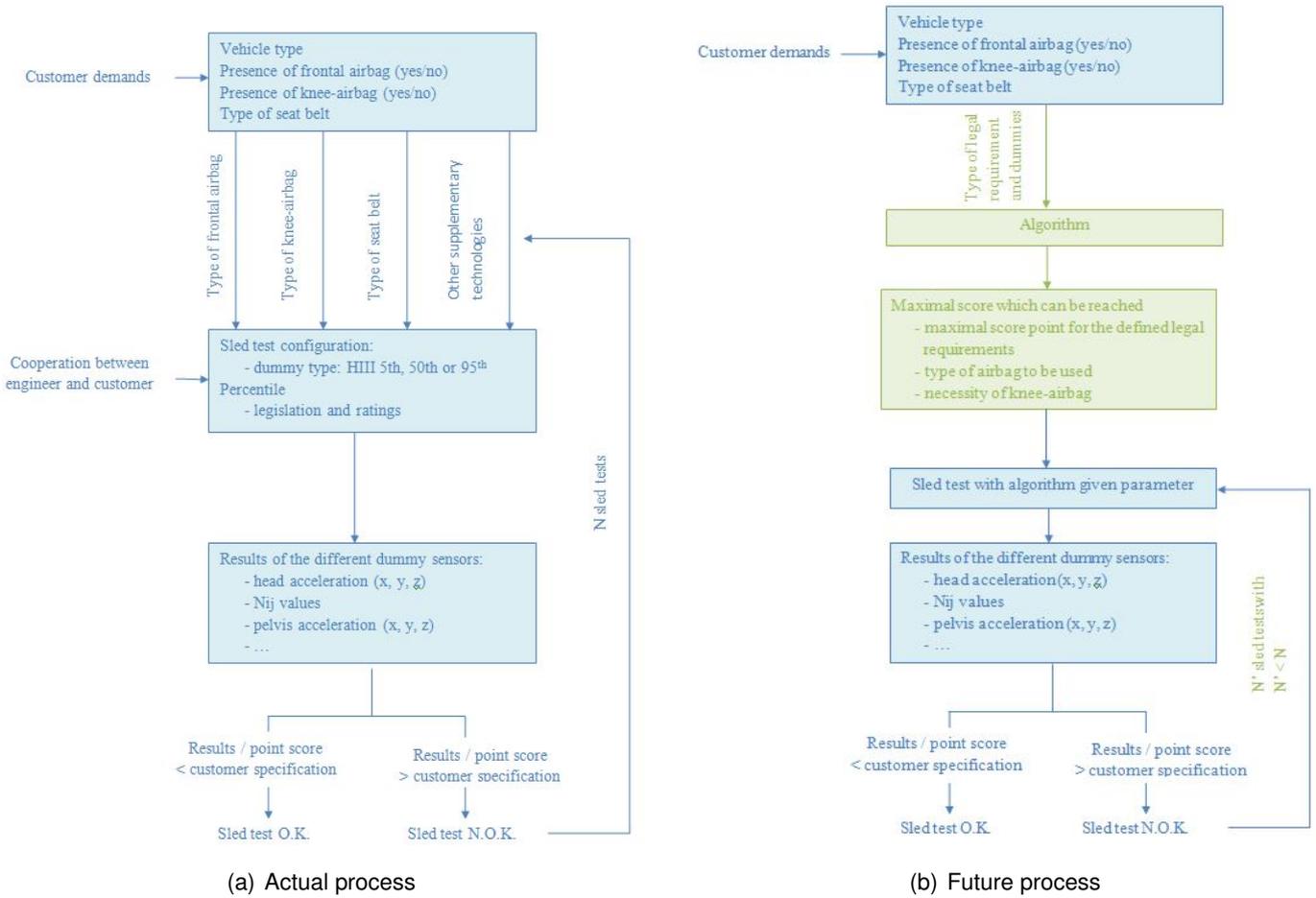


Figure 1.13: Actual and future processes for the development of passive systems

STATE-OF-THE-ART

In this chapter, a literature review focused on automobile crashes is presented. It aims at understanding the current state-of-the-art for crash data and defining the best appropriate methods/algorithms for the database used in this doctoral thesis.

2.1/ METHODS FOR CRASH DATA ANALYSIS

2.1.1/ CURRENT RESEARCHES IN AUTOMOBILE CRASHES

Many researches are focused on the prediction of collisions or of injury severity according to external factors (lighting, seat belt wearing, etc.). The common used techniques are data mining methods which help find out relationships between factors and crash severity.

Data mining technology can be used to predict and describe the driver's behaviour in order to understand and reduce the injury severity during an automobile crash. Many studies are being conducted in order to evaluate and to control the injury severity with the help of data mining methods. The most popular methods in this domain are the decisions trees and the neural networks. Fayyad et al. [13] enumerated some data mining methods:

1. Decision trees and rules: the decision tree is a binary tree in which a parameter is compared to some constants at each node;
2. Non linear regression and classification methods: the classification method uses a set of parameters to characterise each object. They regroup the techniques for prediction like neural networks or nearest-neighbour classifiers;
3. Example-based methods: it is a method which consists of using representative examples from the database to approximate a model;
4. Probabilistic graphical dependency models: the probabilistic model uses a graph which denotes the conditional dependence structure between random variables;
5. Relational learning models: also called "inductive logic programming", it uses the more flexible pattern language of logic ;
6. rough set theory: this theory is based on the study of the formal approximation of a set which gives the lower and upper approximations of the original set.

7. Theory of association rules : it allows the discovery of correlations and patterns in a database;
8. Clustering: grouping of objects in a same group (called cluster) which present more similarities between them rather than with others objects belonging to another cluster.

This literature review is separated according to the data mining techniques: the unsupervised learning (grouping of set objects and discovery of relationships between objects) through the clustering and the supervised learning (grouping of predefined classes and classification of new object affiliated to these classes) through the data classification. Both of them can be here applied on the input's parameters (classification) and on the outputs (clustering).

2.1.1.1/ CLUSTERING METHODS

The clustering is one of the data mining techniques that aims at describing and finding similarities between data. This technique can be applied on the dummy biomechanical values (outputs) in order to form groups of values (clusters) which present some similarities and then to classify the input's parameters.

Mohamed et al. [67] applied two partitioning clustering methods (i.e. which divide the data into k clusters) on data from crashes involving pedestrian and vehicle. The data has been collected in New York City, U.S.A. and in Montreal, Canada. For the first city, they applied the latent class analysis as unsupervised method used for multivariate categorical data. Different numbers of clusters from one to eleven have been tested and the BIC, AIC and CAIC criteria, combined with the entropy calculation and the R-squared criterion, helped at determining the optimal number of clusters. For each discovered cluster, the distribution of the variable allows a better analysis of the data set and help at finding the variables influencing the most the injury severity of the pedestrian. The authors preferred the k-means algorithm, an other unsupervised method, on the Montreal data set because the latent class algorithm regrouped 90% of the data in the first two clusters, regardless of the selected clusters number. Both methods are conclusive and provide information on the link between the pedestrian injury severity and the factors that caused the collision (traffic, lighting conditions, environmental conditions). Throughout this publication, the authors putted forward the importance of the data set segmentation and the analysis of contributing factors.

Furthermore, Geurts et al. [28] analysed crash data from the National Institute of Statistics in Belgium to discover patterns. They divided the traffic accident data into clusters by using a latent class clustering, which is in this case more effective as the Poisson distribution. Each cluster gives a indication of the accident risk factor. The profile of the clusters of traffic roads is conducted with the association algorithm in order to discover the most interesting patterns. Wherewith, the authors defined the roads circumstances which mainly involve an accident on the Belgian roads and their frequency through the clusters. The comparison between the clusters helped at finding the frequent patterns and at improving the knowledge on the accident characteristics.

Magidson et al. [22] have also studied particularly this type of clustering and compared it to the k-means method, by demonstrating the key and weak points of the latent class clustering. They used a simulated data favourable to k-means method. They demonstrated

that when the true group membership is known, it outperforms the k-means technique. Additionally, the latent class clustering presents more advantages in comparison with the k-means methods: it allows a better determination of the number of clusters by using diverse diagnostics (such as the BIC statistic) and does not need to standardise the variables. Without the variable standardization, the variance between the clusters can differ and the scale difference produces clusters with diverse variation. The clusters having the most variation will automatically dominate the others.

Finally, Babuska et al. [24] focused their study on the Gustafson-Kessel (GK) clustering algorithm and on its fuzzy covariance matrix (see Chapter 2.2 Section 2.2.2.3). They described two techniques for improving the calculation of the fuzzy covariance matrix of the GK algorithm by using a small number of data. The authors focused their theory on the singularity of the covariance matrix and overfitting problem. If the covariance matrix is singular (i.e. an eigenvalue which defines the shape of the cluster is equal to zero or the ratio between the minimal and maximal eigenvalue is very large), the inverse of the covariance matrix F_i defined in Eq. 2.30 cannot be calculated. In this case, the algorithm is sometimes not able to compute a correct partitioning of the data. The first proposed technique which should avoid this problem is focused on the limitation of the ratio between the maximal and minimal eigenvalues of the covariance matrix. The second technique aims at avoiding the problem of overfitting which occurs when the number of data in a cluster is too low. Additionally to the limitations of the ratio between the maximal and minimal eigenvalues, another method is proposed by the authors: the add of a scaled identity matrix to the covariance matrix. Both modifications have been included in the GK algorithm and applied on chemist data (enzymes) with of the aim of improving GK performance for small data sets. For the crash data, the GK algorithm presents an advantage in comparison with other clustering methods: the adaptation of the distance between a point and the clusters centre for the creation of the clusters does not restrict the clusters to a circular shape. The implementation of the two above mentioned modifications should not be necessary for the collected data for this doctoral thesis.

Independently of the chosen clustering method, the clusters have to be validated. Many validity indexes have been developed during the last years. Xie et al. [9] benchmarked some current validity indexes used with fuzzy algorithms, such as the Dunn and Davies-Bouldin validity criteria. The last one presents the particularity that the distance used for calculating the compactness and the separation of the clusters can be adapted to an Euclidean or Mahalanobis distance. The performance of four validity indices have been evaluated by Maulik et al. [23]: the Davies-Bouldin, the Dunn's, the Calinski Harabasz Indices and the Index I . For that, three clustering methods (k-means, single link and simulated annealing) have been applied on artificial data sets and for each method, the four validity indices have been tested. This experimentation demonstrated that the Index I indicated the correct number of clusters, independently of the clustering technique but one of its disadvantage is the use of the Euclidean distance to form the clusters, what influences the clusters shape. The DBI provided also good results even if the number of clusters was not always optimal. Its advantage is its adaptability to the clustering method through the distance's calculation (Euclidean or Mahalanobis distance).

The clustering can be applied on the dummy biomechanical values, especially the GK algorithm which is able to detect different cluster's shapes. After the output's clustering, a classification method could help find relationships between the restraint system param-

eters (inputs) and the occupant damage (outputs).

2.1.1.2/ CLASSIFICATION METHODS

In the recent years, the classification methods became one of the most commonly supervised learning methods applied on traffic accident data. They allow the extraction of knowledge on factors responsible of severe injuries.

In point of fact, Chong, Abraham and Paprzycki [30] compared two classification methods, the artificial neural networks and the decision trees, concerning their ability to predict the drivers' injury severity in head-on front impact point collisions. The data set contains reports of 6,4 million police accidents in the U.S.A. and includes information about the accidents happened between 1995 and 200 (e.g. the year, the vehicle model and the travel speed). The experiments showed the most important factors in fatal injury: driver's seat belt usage, light conditions on the road and alcohol consumption of the driver. They also demonstrated that in this very severe case, the decision tree outperforms the artificial neural networks in accuracy, independently of the injury class. The difference between the neural network and the decision trees can reach until 14% in the case of fatal injury. The performance of the decision trees is confirmed in another publication [34].

In fact, in this study [34], they presented some models using four machine-learning approaches: neural network, decision trees, hybrid model involving decision trees and neural network and support vector machines. The aim of this comparison is to find the best model that classifies the injury severity (separated in 5 categories) more accurately. This study is focused on the same data set as in their precedent publication [30]. After trying all 4 methods, the authors concluded that the decision trees outperforms the neural network for the 5 injury classes and shows similar results as the hybrid model in accuracy, what confirms the performance of the decision trees with crash data.

As the decision trees, the rough set theory is also frequently used in research to identify factors and to predict the injury severity during a crash. Tian, Yang and Zhang [53] applied two data mining methods with the aim of analysing road traffic accidents on crash data containing the factors responsible of an accident as physiological factors (e.g. fatigue, alcohol, etc.) or mental factors (e.g. driving habits, mental, etc.). The first method is the rough set theory which allows an approximation of a set through an upper and a lower approximation sets. The second one is the theory of association rules which identifies relationships between variables. By using both methods together, they define a standard model set for analysing road traffic accidents. This study shows the advantages of the data mining when the amount of data is relatively important and explained a simple way to identify the key factors in order to prevent and reduce the road traffic accidents.

Another data mining used with crash data is the Principle Component Analysis (PCA) method. It helps find relationships between variables. This technique consists of transforming a number of correlated variables into a number of uncorrelated variables called principal components in order to reduce the number of variables. Hilmann [46] used this method in his thesis on the structural optimization in the automobile passive safety. After applying the PCA to reduce the number of variables and to find some relationships between the variables, he used the decision trees, especially the CART, to identify the design variables having an influence on the crash car performance. The authors present this method as containing various advantages over the others, such as the little data preparation and the handling of both continuous and categorical data. The decision trees

are also simple to understand and interpret.

More recently, in 2013, Chang and Chien [62] were focused on the non-parametric regression models in order to discover the factors that influence the driver injury severity during a crash. The study explores the truck accidents data collected in the Taiwan area for 2005 and 2006: the data base regroups 705 cases of truck collisions. The data contains information about the injury severity level of each occupant, the occupant himself (age, gender, degree of sobriety), the time of the accident and the type of vehicle involved in the collision. All these data produced a database of 1620 observations (after the deletion of missing data). Two types of non-parametric tree-based methods are practicable: the classification and the regression trees (depending of the target value). In this publication, the driver injury severity is separated into 3 categories (no injury, injury and fatal), and for this reason the classification tree has been developed (categorical target value). The Gini Index criterion has been chosen to evaluate the node impurity and to build the decision tree. The analysis of the classification tree showed that the main factors influencing the driver injury severity are the drinking-driving, the seatbelt use and the vehicle type.

Few years ago, Chang and Wang [38] already applied non-parametric classification tree techniques on traffic injury severity data. The data was supplied by the Taiwan Ministry of Transportation and Communications and regrouped the same information type as in their publication in 2013 [62]. They concluded that the vehicle type is the most important variable associated with an accident.

The next publication confirms the popularity and the potential of the Classification and Regression Trees (CART) in the road safety field. According to Lopez et al. [66], the CART combined with the Gini Index is the most commonly applied algorithm with traffic accident data. They first preferred to test three decision trees methods (CART, ID3 and C4.5) to compare their performance. The database contains information about the level of occupant injury severity, the factors that occur (factors based on the original dataset but also on others studies) such as the context of the collision (lighting) and the driver characteristics (age and gender). The data has been collected during 7 years (2003-2009) in Spain and is focused on the accidents happened on two-lane rural highways. Even if the CART showed higher values in the precision and specified parameters, they did not want to privilege a method to the other because the C4.5 algorithm also produced very good results (only the ID3 presented worst results). The C4.5 algorithm seems to be more comfortable for studying the influence of a specific category on the injury severity because of its non-binary structure. The CART becomes more familiar for an overall study.

In a similar publication, Abellan, Lopez and al. [60] studied the decision trees, especially the extraction of their rules, in order to extract more knowledge from data. The selected data is the same as in the precedent publication of Lopez [66]. The standard decision tree is based on a selected split criterion which allows the separation of a non terminal node into branches. The structure is formed with the help of decision rules, defined as "*IfAthenB*". *A* represents one of the variables (here the factors that occur, the driver characteristics) and *B* gives the consequences of the accident (here the occupant injury severity). Each rule begins at the root node of the tree, and each variable that intervenes in the tree division engenders an "*If*". The rule ends in terminal nodes with "*Then*" and presents the results. Independent of the number of variables, only one decision tree is formed. The authors preferred to compile a method which develops a decision tree for each variable in order to extract more knowledge. Two split criteria have been applied: the G_{lx} based on the Gini Index (see Eq. 2.41) and the Information Gain Ratio based on

Shannon's entropy. Both criteria enable to select the attributes (or variables) which are used for building the decision tree and which are placed in a node and branching. They developed a method, called "Information Root Node Variation", which compiles different trees obtained by varying the root node. Each variable is taken as root and produces one decision tree, from which a rules set is extracted. Only rules with the most severe consequences have been considered in order to extract much more knowledge about the factors causing an accident. They extracted more than 70 significant validated rules and concluded thanks to them, the importance of the environment conditions (e.g. safety barriers, visibility, etc.) in the occupant injury severity.

Moreover, Iragavarapu et al. [76] and Kashani et al. [51] also suggested the same approach as the above authors (CART method combined with the Gini Index applied on accident data). The originality of the publication of Kashani, Shariat-Mohaymany and Ranjbari [51] is the application of the Variable Importance Measure (VIM) to discover the most important variable causing injury severity. They showed that on the rural roads in Iran, the non wearing of the seat belt causes severe injuries (VIM equal to 0,8214), very far away from the type of roads, improper overtaking or speeding (VIM equal to 0,1484).

As the literature review shows, lots of authors analysed the relationships between injury severity and risks factors for automobiles accidents by using non-parametric modelling techniques, especially the CART. But no one was focused on the relationships between occupant injury severity and passive safety systems. The classification tree combined with the Gini Index can be considered in this thesis to discover the main parameters of the restraints systems that influence the occupant injury severity (regrouped into clusters, i.e. the outputs are categorical values).

2.1.2/ SYNTHESIS OF THE RESEARCH BIBLIOGRAPHY

As stated in the literature review, many authors focused their study on the injury severity during an automobile crash. With the aim of creating an overview of the used methods, Table 2.1 classifies the common approaches.

In the automotive field, both supervised and unsupervised learning methods have been used. The majority of the publications are focused on factors influencing the severity of the occupant injuries. It aims at identifying these key factors in order to understand, reduce and predict the injury severity.

According to Table 2.1, the decision trees method seems to have a good reputation for the analysis of crash data, not only because of its performance but also because of its ease of use. This method can be applied on the crash data collected for this doctoral thesis with the restraint systems parameters as inputs and the dummy biomechanical values as outputs. Because of the variation between two similar sled tests (due to foreign parameters), it seems to be more suitable to cluster the results before applying the CART method. It will help discover the influence of each input on the dummy injury severity.

2.2/ THE KDD PROCESS AND DATA MINING METHODS

After resuming the literature review for data analysis, especially automobile crash data, a description of the KDD process and data mining methods including classification and

Table 2.1: Synthesis of the supervised and unsupervised methods used in the literature

Methods	Authors' references
Supervised learning methods	
Decision trees (included CART)	[30, 34, 46, 51, 62, 38, 66, 60, 76]
Artificial neural networks	[30, 34]
Hybrid model (Artificial neural networks and decision trees)	[34]
Rough set theory	[53]
Theory of association rules	[53]
Principle component analysis	[46]
Gini index validity	[62, 66, 60, 76, 51]
Shannon's entropy	[60]
Unsupervised learning methods	
K-means methods	[22, 67]
Latent class clustering	[28, 67, 22]
Gustafson-Kessel algorithm	[24]
Dunn index validity	[9, 23]
Davies-Bouldin index validity	[9, 23]
Calinski Harabasz indices validity	[23]
Index I	[23]

clustering is conducted in this chapter.

2.2.1/ DESCRIPTION OF THE KDD PROCESS

"Knowledge Discovery in Databases" (KDD) is the process of discovering useful and understandable patterns and knowledge from a collection of data. The process's goal is the extraction of knowledge from data in large databases, using data mining methods (algorithms). The KDD process is commonly defined with the following steps (Fig. 2.1):

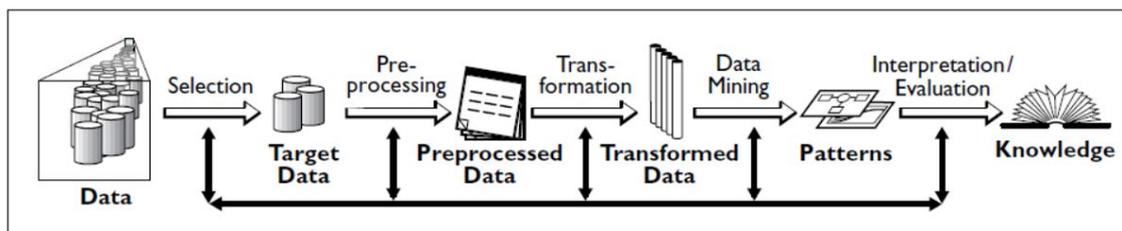


Figure 2.1: Overview of the KDD process's steps [14]

1. Selection: this first step aims at creating a data set using for detecting pattern;
2. Pre-Processing: the data is cleaned (noise removal or missing data replacement) and completed from different sources. It is also defined how many samples are needed;
3. Transformation: if necessary, the attributes are discretised and the data mining algorithms are chosen;

4. Data mining: once the method is selected, the search for patterns through classification and clustering can begin;
5. Interpretation / Evaluation : this last step consists of the interpretation of mined patterns, of the consolidation of discovered knowledge and of the evaluation of the model's limits.

Data mining, one of the most difficult steps of the KDD process, is generally focused on large databases but works also very well with a reduced amount of data. It uses algorithms to segment the data, to extract patterns or trends and to evaluate the probability of future events from data. Leung et al. [58] defined the data mining as

"[...] the process of extracting hidden and useful patterns and information from data"

Data mining involves six common tasks:

- ◆ Summarisation of the data: it helps have a smaller data set which gives a general overview of the data through the abstraction of the whole data set;
- ◆ Classification: this task aims at determining the class of a new data based on its attributes. The initial data set is given as training set which is divided into two or more classes according to the data attributes;
- ◆ Association: it helps discover relationships among data which make them statistically dependent, and establish so-called "association rules" defining the togetherness between data;
- ◆ Clustering: it identifies classes, also called clusters, for a data set whose classes are, unlike the classification, unknown. The data in a class presents similarities in some ways with other data from this class. This task aims at discover sensitivities, in particular which parameters have a significant influence on the outputs;
- ◆ Regression: as the association, it aims at estimating relationships among data. Here, it helps discover how a dependent variable is related to independent variables and predict future output based on given inputs;
- ◆ Anomaly detection: identification of unusual data records that might be interesting or data errors that require further investigation.

The data mining is a complete method that need to be executed for finding relationships between variables.

In the field of machine learning, the automatic learning can be summarised in three steps:

1. Observations of a phenomenon;
2. Construction of a model based on this phenomenon;
3. Predictions and analysing of this phenomenon with the help of the model.

The machine learning regroups two types of methods: supervised and supervised learning methods. In supervised learning, the dataset used by the machine learning algorithm to get the desired outputs is known whereas unsupervised learning systems do not provide any datasets for predicting the outputs. It conduces to a clustering of the data.

2.2.2/ UNSUPERVISED LEARNING THROUGH THE CLUSTERING

In unsupervised learning, there are no explicit outputs values associated with each input. The unsupervised methods have to work with the observed inputs x_i , which are often supposed in this manuscript to be independent samples, i.e. they determine the values of other variables. One aim of the unsupervised learning is to regroup the data which present some similarities into clusters. The unsupervised classification, also called clustering, is the most common task for the unsupervised learning: the algorithm regroups the data with similar attributes according to a split criterion. For the clusters' validation, some algorithms enables their evaluation, such as the Davies-Bouldin Index method.

Generalities on the clustering

The clustering consists of the grouping of objects which present some similarities. This task is a main part of the data mining analysis and is used in many fields such as machine learning or pattern recognition. The first step of this thesis consists of the clustering of the outputs data (dummy biomechanical values) in order to analyse the parameters' influence of them. There are two ways to cluster the data:

- ◆ Hard clustering: an object belongs definitively to only one cluster;
- ◆ Soft clustering: an object can belong to different clusters and has a probability distribution over its cluster.

The term of *cluster* is not precisely defined [19], what engenders the availability of many clustering algorithms. Jain, Flynn and Murty [18] proposed to classify the clustering methods into four groups:

◆ **Hierarchical clustering methods:**

This unsupervised learning method aims at building a hierarchy of clusters. It generates clusters based on the distance between the objects and the cluster's centres. A cluster is defined by the maximum distance needed to connect its parts. It does not provide a single partitioning of the data set but a hierarchy of clusters. The two types of the hierarchical clustering (bottom-up and top-down approaches) are represented in Fig. 2.2.

◆ **Density-based clustering :**

The key point of these algorithms is to create clusters based on density functions. Their main advantage is the arbitrary creation of shaped clusters.

◆ **Grid-based clustering :**

These types of algorithms are mainly proposed for spatial data mining. They quantise the search space into finite number of cells.

◆ **Partitioning clustering methods:**

Also called flat methods, the partitioning methods decompose a set of data into disjoint clusters. According to Adriano Cruz [64], the most used partitioning methods are:

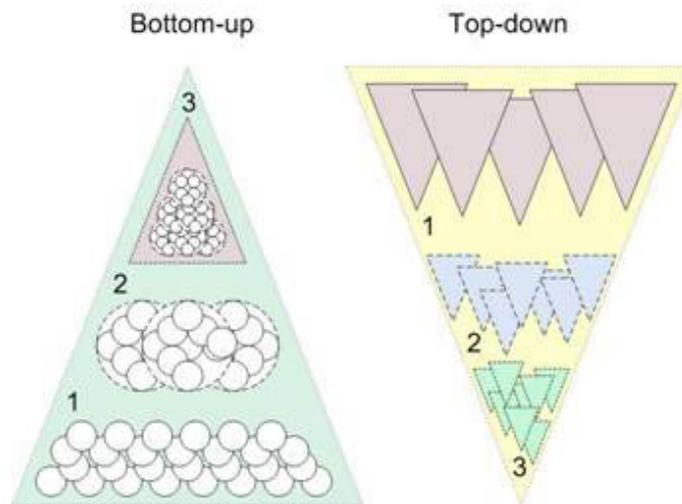


Figure 2.2: Graphical representation of the hierarchical clustering

- ★ K-means: this method is based on the Euclidean distance among objects of the cluster. Each object belongs to a single cluster which its center is the mean value of the objects included in the cluster;
- ★ Fuzzy c-means: this method is a fuzzy version of the k-means. The term fuzzy refers to something not clear, not distinct. In the field of Artificial Intelligence, the term fuzzy-set is introduced on a domain and defined by a membership function which mapped the domain to the unit interval $[0; 1]$. One object can be classed into more than one cluster. Its degree of membership gives an information about its belonging to the cluster relative to the other clusters. This method also uses the Euclidean distance in order to calculate the distance between the i^{th} data and j^{th} cluster's centre;
- ★ Possibility clustering: this method works with the help of the data membership degrees in the clusters;
- ★ Gustafson-Kessel algorithm: this method is similar to the fuzzy c-means but uses the Mahalanobis distance, which is the general case of the Euclidean distance. It allows the formation of ellipsoidal clusters;
- ★ Gath-Geva algorithm: this method is also similar to the fuzzy c-means but uses the Gauss distance;
- ★ K-medoids: this algorithm uses the most centrally object (medoid) as the reference point instead the mean value.

The Gustafson-Kessel algorithm seems to be the most appropriated algorithm for the biomechanical values' clustering because, after a first analysis, it could be possible that for some dummy parts, the spherical clusters are inappropriate, i.e. the algorithm has to be able to recognize ellipsoidal forms. The clustering of the dummy's biomechanical values should help study the influence of the restraint systems parameters (airbag, kneebag, seat belt and dummy position) on it. Sled tests included in the database have been studied, i.e. all the parameters of the restraint systems (see section 3.1.1) and the dummy biomechanical values (outputs) are known. After the creation of the clusters for each dummy part, each output can be gathered into one of the discovered clusters. The study of the sled tests parameters will enable to uncover if their modifications have an

influence on the output's cluster. If the output stays in the same cluster, it means that the modification affects the injury values in the corresponding dummy region only minor.

2.2.2.1/ THE GUSTAFSON-KESSEL ALGORITHM AS UNSUPERVISED LEARNING

The Gustafson-Kessel (GK) algorithm is a powerful clustering technique with a large number of applications in various domains (e.g. image processing, classification, system identification). Its main feature is the local adaptation of the distance metric to the cluster's shape by estimating the cluster covariance matrix and by using the Mahalanobis distance. It is applied as clustering method on the dummy biomechanical values.

It associates a point and a matrix to each cluster, which represent the cluster centre and its covariance respectively. Whereas the original fuzzy c-means method makes the implicit hypothesis that clusters are spherical (through the Euclidean distance [47]), the GK algorithm uses an adaptive distance norm, the Mahalanobis distance, to compute the distances and thus does not impose any conditions on the cluster's form. The Mahalanobis distance is the general case of the Euclidean distance.

2.2.2.2/ FUZZY CLUSTERING

Fuzzy c-means (FCM) is one of the most known clustering algorithms [3]. This method, developed by Dunn in 1973 and improved by Bezdek in 1981, is frequently used in pattern recognition.

The fuzzy clustering aims at subdividing a data set Y into c subsets (clusters) which are pairwise disjoint, all non empty and reproduce Y via union. It produces fuzzy c -partitions of Y which characterize the membership of each point in every cluster through a membership function. Memberships close to 1 correspond to a high degree of similarity between the sample and a cluster while memberships close to zero imply little similarity between the point and this cluster [1]. Obviously, the sum of the memberships for each point must be equal to 1.

The algorithm works by assigning membership to each data point to a cluster through the distance calculation between the point and the cluster's centre. More the data is near to the cluster's centre, more is its membership high to this cluster.

This membership calculation can be mathematically modelled as follows [20]:

Let $Y = \{y_1, y_2, \dots, y_N\}$ be a sample of N observations in \mathbb{R}^n (n -dimensional Euclidean space), y_k is the k -feature vector and y_{kj} the j^{th} feature of y_k . The set of N observations can be represented by a $(n \times N)$ matrix:

$$Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1N} \\ y_{21} & y_{22} & \dots & y_{2N} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ y_{n1} & y_{n2} & \dots & y_{nN} \end{bmatrix} \quad (2.1)$$

In the pattern-recognition terminology, the columns of this matrix are called *patterns* or *objects*, the rows are called *features* or *attributes*, and Y is called *pattern* or *data matrix*.

If $c \in \mathbb{N}$, $2 \leq c < n$, a conventional c -partition is defined by (Y_1, Y_2, \dots, Y_c) subsets (or clusters) of Y which verify:

$$Y_i \neq \emptyset \quad 1 \leq i \leq c \quad (2.2)$$

$$Y_i \cap Y_j = \emptyset \quad i \neq j \quad (2.3)$$

$$\bigcup_{i=1}^c Y_i = Y \quad (2.4)$$

\emptyset represents the empty stand. The membership function of the sample can be yet modelled as follows:

Considering U a real $(c \times N)$ matrix with $U = [u_{ik}]_{c \times N}$. U is the matrix representation of the partition Y_i in Eq. 2.2 with:

$$u_i(y_k) = u_{ik} = \begin{cases} 1 & y_k \in Y_i \\ 0 & \text{otherwise} \end{cases} \quad (2.5)$$

$$\sum_{i=1}^c u_{ik} = 1 \quad \text{for } 1 \leq k \leq N \quad (2.6)$$

$$\sum_{k=1}^N u_{ik} > 0 \quad \text{for } 1 \leq i \leq c \quad (2.7)$$

The space of possible hard partition matrices for Y , called *hard partitioning space* [3], is thus defined as following:

$$M_{hc} = \left\{ U \in \mathbb{R}^{c \times N} \mid u_{ik} \in \{0, 1\} \forall i, k \quad \sum_{i=1}^c u_{ik} = 1, \forall k \quad 0 < \sum_{k=1}^N u_{ik}, \forall i \right\} \quad (2.8)$$

U is called a *hard c -partition* of Y . Bezdek et al. [3] refer to U as a fuzzy c -partition of Y when the elements of U are numbers in the interval $[0;1]$ which continue to satisfy Eq. 2.6 and Eq. 2.7. In Eq. 2.8, Eq. 2.5 is replaced by:

$$u_{ik} \in [0, 1], \quad 1 \leq i \leq c, \quad 1 \leq k \leq N \quad (2.9)$$

and thus the corresponding fuzzy partitioning space for Y is yet defined as:

$$M_{fc} = \left\{ U \in \mathbb{R}^{c \times N} \mid u_{ik} \in [0, 1], \forall i, k \quad \sum_{i=1}^c u_{ik} = 1, \forall k \quad 0 < \sum_{k=1}^N u_{ik}, \forall i \right\} \quad (2.10)$$

It can be concluded that $M_{hc} \subset M_{fc}$ (hc means hard clustering and fc fuzzy clustering), i.e. the fuzzy clustering algorithms can obtain hard c -partitions.

Several clustering criteria have been proposed for identifying the optimal fuzzy c -partitions in Y . The most popular and well studied method is associated with the generalized least-square errors functional, also known as the objective function of the fuzzy c -means:

$$J_m(Y; U, v) = \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m \cdot d_{ik}^2 \quad (2.11)$$

$$= \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m \cdot \|y_k - v_i\|_A^2 \quad (2.12)$$

$$= \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m \cdot (y_k - v_i)^T \cdot A \cdot (y_k - v_i) \quad (2.13)$$

with:

- ◆ $Y = \{y_1, y_2, \dots, y_N\} \subset \mathbb{R}^n$ = the data;
- ◆ c number of clusters in Y ; $2 \leq c < n$;
- ◆ $m \in [1, \infty[$: weighting exponent;
- ◆ $U = [u_{ik}]_{c \times N} \in M_{fc}$: fuzzy c -partition of Y ;
- ◆ $v = [v_1, v_2, \dots, v_c]$: vector of centres, $v_i \in \mathbb{R}^n$;
- ◆ $v_i = (v_{i1}, v_{i2}, \dots, v_{in})^T$: centre of the cluster i ;
- ◆ $\|\dots\|_A$: induced A -norm on \mathbb{R}^n with A , called norm-inducing matrix, a positive-definite ($n \times n$) matrix.
- ◆ $d_{ik}^2 = \|y_k - v_i\|_A^2$: squared distance between an observation y_k and the cluster's centre v_i computed in the A -norm (or matrix-norm)

In particularly, each term of Eq. 2.11 can be defined as follows:

- ◆ d_{ik}^2 = squared A -distance from point y_k to centre of mass v_i ;
- ◆ $(u_{ik})^m \cdot d_{ik}^2$ = squared A -error incurred by representing y_k by v_i weighted by the membership of y_k in cluster i ;
- ◆ $\sum_{i=1}^c (u_{ik})^m \cdot d_{ik}^2$ = squared A -error due to y_k replacement by all centres $\{v_i\}$;
- ◆ $\sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m \cdot d_{ik}^2$ = overall weighted sum of generalized A -errors due to the replacement of Y by v ;

The parameter m controls the relative weights placed on each of the squared errors d_{ik}^2 . If $m = 1$, J_m minimizes only for hard c -partitions. Thus v_i corresponds to the geometric centroid of the corresponding Y_i cluster. According to authors, no theoretical or computational evidence distinguishes an optimal m . Each choice for m defines one FCM algorithm when all other parameters are fixed. The range of useful values for m seems to be between 1 and 30 and for most studies, $1, 5 \leq m \leq 3$ gives good results.

The parameter A of J_m controls the shape that optimal cluster can assume in \mathbb{R}^n . Because every norm on \mathbb{R}^n is defined as:

$$\langle x, y \rangle_A = x^T \cdot A \cdot y \quad (2.14)$$

there are infinitely many A -norms available in Eq. 2.11, however in practice only a few of them are used. The FCM allows a choice of three norms [4]:

$$\begin{aligned} A = I &\Rightarrow d_{ik} \sim \text{Euclidean norm} \\ A = D_y^{-1} &\Rightarrow d_{ik} \sim \text{Diagonal norm} \\ A = C_y^{-1} &\Rightarrow d_{ik} \sim \text{Mahalanobis norm} \end{aligned}$$

Each norm is induced by a specific weight matrix.

Let's define:

$$c_y = \frac{1}{N} \sum_{k=1}^N y_k \quad (2.15)$$

$$C_y = \sum_{k=1}^N (y_k - c_y) \cdot (y_k - c_y)^T \quad (2.16)$$

respectively the sample mean (Eq. 2.15) and the sample covariance matrix (Eq. 2.16) of the data set Y .

When $A = I$, J_m identifies spherical clusters (fuzzy c-means algorithm) and the distance d_{ik}^2 corresponds to the Euclidean distance. When $A \neq I$, the clusters are essentially ellipsoidal (Mahalanobis distance), as represented in Fig. 2.3:

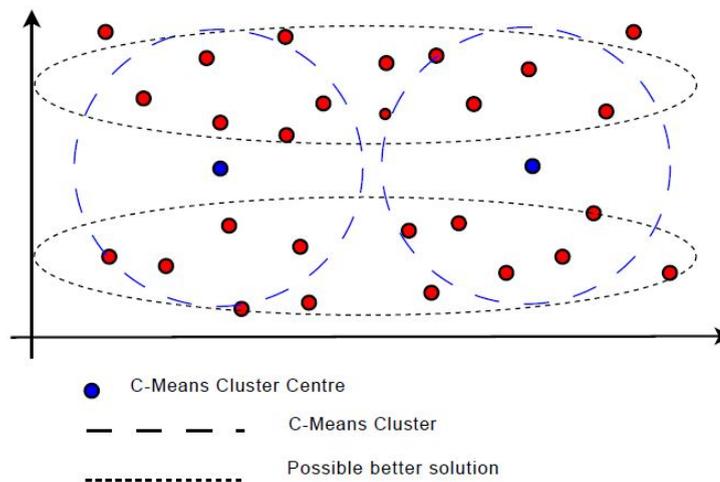


Figure 2.3: Representation of c-means clustering through Euclidean and Mahalanobis distances

Optimal fuzzy clusterings of Y are defined as pairs (U, v) that minimize J_m locally. For $m > 1$, if $y_k \neq v_i$ for all i and k , (U, v) may be locally optimal for J_m if :

$$v_i = \frac{\sum_{k=1}^N (u_{ik})^m \cdot y_k}{\sum_{k=1}^N (u_{ik})^m} \quad 1 \leq i \leq c \quad (2.17)$$

$$u_{ik} = \left(\sum_{j=1}^c \left(\frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}} \right)^{-1} \quad 1 \leq k \leq N \quad 1 \leq i \leq c \quad (2.18)$$

with $d_{ik} = \|y_k - v_i\|_A$

The FCM algorithm is based on the assumption that the clusters are spherical shaped. In real problems, the clusters are of different shapes and with different orientations in the data space. Therefore, Gustafson and Kessel (1979) extended the FCM algorithm by employing an adaptive distance norm for each cluster in order to detect different geometrical shapes in data sets.

2.2.2.3/ THE GK ALGORITHM

The target applications of the Gustafson-Kessel algorithm are similar to those of the fuzzy c-means algorithm. Because of the adaptation of the distance function (Mahalanobis) to the clusters, the results for non-spherical clusters correspond better to a intuitive partition (Fig. 2.3).

Each cluster has its own norm-inducing matrix A_i , which yields the following norm:

$$d_{ikA_i}^2 = \|y_k - v_i\|_{A_i}^2 = (y_k - v_i)^T \cdot A_i \cdot (y_k - v_i) \quad \text{with} \quad 1 \leq k \leq N \quad 1 \leq i \leq c \quad (2.19)$$

The matrices A_i are used as optimization variables in the c-means functional, allowing each cluster to adapt the distance norm to the local topological structure of the data. Let A denote a c-tuple of the norm-inducing matrices corresponding to the clusters $A = (A_1, A_2, \dots, A_c)$. The objective functional J (Eq. 2.11) of the GK algorithm is defined by:

$$J(Y; U, v, A_i) = \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m \cdot d_{ikA_i}^2 \quad (2.20)$$

To minimize this objective function J with respect to A_i , A_i must be constrained. It means that the determinant of A_i has to be constrained, so that:

$$|A_i| = \det(A_i) = \varrho_i \quad \varrho_i > 0, \quad \forall i \quad (2.21)$$

The expression of A_i is obtained by using the Lagrange multiplier method. and the modification of the parameters ϱ_i allows the modification of the cluster's shape while its volume remains constraint.

In order to take the constraints of constant cluster volumes, c Lagrange multipliers have to be introduced, such that the objective function (Eq. 2.20) results in :

$$J = \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m \cdot d_{ikA_i}^2 - \sum_{i=1}^c \lambda_i \cdot (\det(A_i)) \quad \lambda_i, i \in \mathbb{N}_{\leq c} \quad (2.22)$$

Taking the derivative of J with respect to the matrix A_i , we go beyond the restriction to symmetric, positive definite matrices with determinant equal to 1, and instead we consider all regular matrices of $\mathbb{R}^{p \times p}$. The partial derivatives with respect to all matrix elements can be then calculated. With the assumption that the set of irreversible matrices is differentiable in all directions under the continuous mapping, we have:

$$\nabla y_k^T \cdot A_i \cdot y_k = y_k \cdot y_k^T \quad \text{and} \quad \nabla \det(A_i) = \det(A_i) \cdot A_i^{-1} \quad (2.23)$$

For minimizing the objective function, the gradient should be equal to 0:

$$0 = \nabla J = \left(\sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m \cdot (y_k - v_i) \cdot (y_k - v_i)^T \right) - \lambda_i \cdot \det(A_i) \cdot A_i^{-1} \quad (2.24)$$

The derivatives with respect to the Lagrange multipliers lead to the constraint $\det(A_i) = 1$ for $i \in \mathbb{N}_{\leq c}$. With the notation $S_i = \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m \cdot (y_k - v_i) \cdot (y_k - v_i)^T$, Eq. 2.24 can be yet written:

$$0 = S_i - \lambda_i \cdot A_i^{-1} \quad (2.25)$$

$$\Leftrightarrow S_i \cdot A_i = \lambda_i \cdot I \quad (\text{because of the reversibility of } A_i) \quad (2.26)$$

$$\Leftrightarrow \det(S_i \cdot A_i) = \lambda_i^p \quad (2.27)$$

$$\Leftrightarrow \lambda_i = \sqrt[p]{\det(S_i)} \quad \text{with } \det(A_i) = 1 \quad (2.28)$$

By replacing the Lagrange multiplier λ_i in Eq. 2.25, we obtain :

$$A_i = \sqrt[p]{\det(S_i)} \cdot S_i^{-1} \quad (2.29)$$

It remains to be shown that A_i also satisfies the constraint of a positive definite and symmetric matrix. For that, we assume that there are p linearly independent vectors $\xi \in \mathbb{R}^p$ in the data set. Then, the matrices $\xi \xi^T$ are symmetric and positive semi-definite and also their weighted sum and consequently A_i is symmetric and positive definite. The irreversibility of the matrix A_i is also satisfied.

Instead of the matrices S_i , Gustafson and Kessel use the so-called *fuzzy covariance matrices* F_i defined as:

$$F_i = \frac{\sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m \cdot (y_k - v_i) \cdot (y_k - v_i)^T}{\sum_{k=1}^N (u_{ik})^m} \quad (2.30)$$

However, the factor $\frac{1}{\sum_{k=1}^N (u_{ik})^m}$ is not relevant for the result, because the matrices are scaled to the unit determinant.

Finally the matrix A_i can be resumed as :

$$A_i = \sqrt[p]{Q_i \cdot \det(F_i)} \cdot F_i^{-1} \quad \text{with } Q_i^1 \text{ the volume of the } i^{\text{th}} \text{ cluster} \quad (2.31)$$

The substitution of Eq. 2.31 in Eq. 2.19 gives a generalized squared Mahalanobis distance norm, where the covariance is weighted by the membership degrees in U .

Finally, the steps for finding the best clustering with the GK method can be resumed as follows:

Given the data set X , a number of clusters comprised between 2 and n ($2 \leq c < n$), the weighting exponent $m > 1$, the termination tolerance $\varepsilon > 0$ (usually $\varepsilon = 0,001$ is taken), an initial matrix $U^{(0)} \in M_{fc}$ mostly randomly initialized, a step l with $l \in [0, \dots, L_{max}]$ and the cluster volumes q_i (without a prior knowledge, q_i is simply fixed at 1 for each cluster, i.e. the GK algorithm only can find clusters of approximately equal volumes). First the iteration step l is setted at 0. The process follows the steps:

For $l = 1, 2, \dots, L_{max}$

- ◆ Step 1: Calculation of the clusters centres v_i :

$$v_i^{(l)} = \frac{\sum_{k=1}^N (u_{ik}^{(l-1)})^m \cdot y_k}{\sum_{k=1}^N (u_{ik}^{(l-1)})^m} \quad 1 \leq i \leq c \quad (2.32)$$

- ◆ Step 2: Determination of the cluster covariance matrix F_i :

$$F_i = \frac{\sum_{k=1}^N \sum_{i=1}^c (u_{ik}^{(l-1)})^m \cdot (y_k - v_i^{(l)}) \cdot (y_k - v_i^{(l)})^T}{\sum_{k=1}^N (u_{ik}^{(l-1)})^m} \quad (2.33)$$

- ◆ Step 3: Calculation of the distance between an observation Y_k and a clusters centre v_i :

$$d_{ikA_i}^2 = \left\| y_k - v_i^{(l)} \right\|_{A_i}^2 = (y_k - v_i^{(l)})^T \cdot A_i \cdot (y_k - v_i^{(l)}) \quad \text{with } 1 \leq k \leq N \quad \text{and } 1 \leq i \leq c \quad (2.34)$$

- ◆ Step 4: Actualisation of the partition matrix by comparing $U^{(l+1)}$ to U^l :

$$u_{ikA}^{(l)} = \left(\sum_{j=1}^c \left(\frac{d_{ikA_i}}{d_{jkA_i}} \right)^{\frac{2}{m-1}} \right)^{-1} \quad 1 \leq k \leq N \quad 1 \leq i \leq c \quad \text{until } \|U^{(l)} - U^{(l-1)}\| < \varepsilon \quad (2.35)$$

- ◆ Step 5: When $\|U^{(l)} - U^{(l-1)}\| < \varepsilon$, the algorithm stops. Otherwise, $U^{(l-1)} = U^{(l)}$ is considered and the algorithm returns to Step 1.

Practically through experiments, Bezdek et al [4] reveal that numerical convergence is usually achieved in 10-25 operations. The GK algorithm is computationally long because the inverse and the determinant of the cluster covariance matrix must be calculated for each iteration.

Whether local minima of J_m are good clusterings of Y is another matter. In order to determine the validity of the clusters, several types of validity measures have to be applied. One of the most popular is the DBI method.

2.2.2.4/ THE DAVIES BOULDIN INDEX

2.2.2.5/ GENERALITIES

Clustering is an unsupervised learning process, that is why the evaluation of the clustering algorithms is very important. Before using the GK algorithm, the choice of the number of

clusters is the most important parameter.

Two main approaches to determining the appropriate number of clusters in data can be distinguished:

- ◆ **Validity measures:** they are scalar indices that assess the goodness of the obtained partition. Clustering algorithms generally aim at locating well separated and compact clusters. When the number of clusters is chosen equal to the number of groups that actually exist in the data, it can be expected that the clustering algorithm will identify them correctly. If it is not the case, misclassification appears, and the clusters are not likely to be well separated and compact. In fact, most cluster validity measures are designed to quantify the separation and the compactness of the clusters. However according to Bezdek [3], the concept of cluster validity is open to interpretation and can be formulated in different ways. Consequently, many validity measures have been introduced in the literature [12, 8, 3].
- ◆ **Iterative merging or insertion of clusters:** the basic idea of cluster merging is to start with a large number of clusters, and successively reduce this number by merging clusters that are similar or compatible with respect to some defined criteria [10, 11]. The opposite approach is also possible, i.e. start with a small number of clusters and iteratively insert clusters in the regions where the data points have low degree of membership in the existing clusters [8].

The aim of the cluster validity is to find the partitioning that best fits the underlying data. The process of evaluating the results of a clustering algorithm is called *cluster validity assessment*. Two measurement criteria have been proposed [15] for evaluating and selecting an optimal clustering:

- ◆ Compactness: the member of each cluster should be as close to each other as possible. A common measure of compactness is the variance;
- ◆ Separation: the clusters themselves should be widely separated. There are three common approaches measuring the distance between two different clusters:
 - ★ the distance between the closest members of the clusters;
 - ★ the distance between the most distant members of a cluster;
 - ★ the distance between the centres of the clusters.

There are different techniques for the evaluation of the result of the clustering algorithms [40]:

- ◆ External Criteria
- ◆ Internal Criteria
- ◆ Relative Criteria

The evaluation of the clusters' pertinence is a complicated task. Because of the computational complexity of external and internal criteria techniques, Kovacs et al. [35] focused their study on the relative criteria technique which compares different clustering schema. This technique aims at defining the best clustering schema with the help of the validity index. The most used measurement techniques for the cluster validation are ([35], [21]):

- ◆ Dunn index: ratio of the maximal distance which separates two objects from the same cluster over the minimum distance which separates two objects from different clusters;
- ◆ Davies-Bouldin index: ratio of the within cluster spread for the cluster i over the separation between the i^{th} and the j^{th} cluster;
- ◆ Root-Mean-Square Standard Deviation (RMSSD) and R-Squared (RS) validity indices: usable only if the clustering is hierarchical. The RMSSD measures the homogeneity of the clusters for each step of the hierarchical algorithm [21]. The RS validity index is the ratio of the sum of squares between the clusters over the total sum of squares of the entire data set;
- ◆ SD validity index: it measures the compactness and the homogeneity of the discovered clusters through their variance and the variance of the entire data set;
- ◆ S-Dbw validity index: additionally to the properties of the SD index (compactness and homogeneity), it also takes the density of the clusters into account.

These validity indices work principally with separated hyper-sphere shaped clusters. Because of its adaptability to the Mahalanobis distance, the DBI method can be combined with the GK algorithm to form non spherical clusters for the dummy biomechanical values. It enables to keep a certain coherence between the clustering and its validity method.

2.2.2.6/ DAVIES-BOULDIN INDEX

This criterion evaluates the well-separated and compact clusters between the i^{th} cluster and the cluster's centroid through the similarity measure of clusters R_{ij} , which corresponds to:

$$R_{ij} = \frac{\bar{d}_i + \bar{d}_j}{d_{i,j}} \quad (2.36)$$

with

- ◆ \bar{d}_i the average distance between each point in the i^{th} cluster and his centroid
- ◆ \bar{d}_j the average distance between each point in the i^{th} cluster and the centroid of the j^{th} cluster
- ◆ $d_{i,j}$ the Euclidean distance between the centroids of the i^{th} and j^{th} clusters

The similarity measure R_{ij} must satisfy the following conditions [2]:

- ◆ $0 \leq R_{ij}$
- ◆ $R_{ij} = R_{ji}$
- ◆ if $\bar{d}_i = 0$ and $\bar{d}_j = 0$ then $R_{ij} = 0$
- ◆ if $\bar{d}_j > \bar{d}_c$ and $d_{ij} = d_{ik}$ then $R_{ij} > R_{ik}$

- ◆ if $\bar{d}_j = \bar{d}_c$ and $d_{ij} < d_{ik}$ then $R_{ij} > R_{ik}$

The DBI is thus defined as:

$$DBI = \frac{1}{k} \cdot \sum_{i=1}^k \max_{j \neq i} R_{ij} \quad (2.37)$$

with k the number of clusters

$\max_{j \neq i} R_{ij}$ refers to the worst-case scenario and corresponds to R_{ij} for the most similar cluster j to cluster i . That is why the optimal clustering has the lowest DBI value (compact and well-separated clusters, i.e. no similarities between them). For a better correlation with the GK algorithm and in order to detect non-spherical clusters, the average distances \bar{d}_i and \bar{d}_j can be replaced by the Mahalanobis distance.

For the FMVSS NR.208 legislation, the DBI values are calculated from 4 up to 10 clusters. To be used in the next study of relevant inputs parameters, the clusters have to be exploitable. A clustering of the output's values in 1 or 2 clusters seems to be not appropriated because the parameters' influence cannot be evaluated if the number of clusters is too small. The US NCAP rating is a challenge for the engineers because this rating is based on the FMVSS NR.208 and the restraint systems have to work with the HIII 50th dummy and the HIII 5th dummy, with and without belt. That is why it is preferable for this rating to cluster the outputs into many clusters. Therefore the DBI values are calculated from 6 up to 15 clusters.

However, the very good results obtained in EURO NCAP rating do not involve the necessity of a high number of clusters. The DBI values are estimated for 3 until 10 clusters, but the expert knowledge will often help regroup some clusters together and thus improve the quality of the EURO NCAP clustering for the study of the input's influence.

2.2.3/ SUPERVISED LEARNING THROUGH THE CLASSIFICATION

Unlike the unsupervised method, this learning method is based on the knowledge of examples coming from the database. The objective is to model the relation between an input value $X = x_n$ and an output value $Y = y_n$ in order to predict a new output Y_{new} . It produces an inferred function $y_n = f(x_n) + \epsilon$, where ϵ represents the noise.

Two types of problems can be solved with the help of the supervised learning:

1. the regression if $Y \in \mathbb{R}$, i.e. if the output value belongs to the continuous group. The most famous regression methods are the Bayesian statistics and the Gaussian process regression;
2. the classification if $Y = \{1 \dots I\}$, i.e. if Y is a finite set of outputs. The most famous classification methods are the decision tree learning, the naive Bayes classifier, the neural network and the nearest neighbour algorithm (used also for regression).

The performance of the model stemming from a method of learning can be estimated by its capacity of prediction. The generalisation error E represents the number of learning

examples for which the prediction calculated with the function f is false. This generalisation error of the hypothesis function f , given X , is defined by:

$$E(f, \tau_N) = \frac{1}{N} \sum_{i=1}^N \psi(f(x_i), y_i) \quad (2.38)$$

with $\tau_N = (x_i, y_i)_{i=1}^N$

ψ is the error criterion on Y (as small as possible) and defined by:

$$\psi = \begin{cases} 1 & \text{if } f(x_i) \neq y_i \\ 0 & \text{if } f(x_i) = y_i \end{cases}$$

A confusion matrix is a tool helping at summarizing the classification performance of a classifier (or rule) [31]. It is a two-dimensional matrix; each column of the matrix represents the instances in the predicted class, while each row represents the instances in the real class. Table 2.2 presents an example of confusion matrix for a three-class classification task (A, B and C classes). The first row of the matrix indicates that eleven objects belong to the class A but just eight objects are correctly classified as belonging to A. The three other objects are misclassified. Note that the exact proportion of successes p_i for the row and for the column equals 1.

Table 2.2: An example of confusion matrix

		Predicted class			Total
		A	B	C	p_i
Real Class	A	8 $p_{AA} = 0,2222$	2 $p_{AB} = 0,0556$	1 $p_{AC} = 0,0278$	0,3056
	B	0 $p_{BA} = 0$	5 $p_{BB} = 0,1389$	7 $p_{BC} = 0,1945$	0,3334
	C	10 $p_{CA} = 0,2778$	3 $p_{CB} = 0,0833$	0 $p_{CC} = 0$	0,3611
Total	p_j	0,5	0,2778	0,2223	1

2.2.3.1/ BACKGROUND OF THE BASIC TREE MODEL

Decision trees are predictive modelling approaches used in machine learning and data mining and regroup classifiers and regression models. They are represented graphically as hierarchical structures, which make them easy to interpret. The decision tree algorithm creates a model (tree) that predicts the target value Y based on several input variables X_i . For building the tree, a learning sample composed of historical data enables the classification of new data. This sample contains information on input variables and on class variables. Within a decision tree, each node corresponds to an input variable. Each branch represents one or many states of this variable and connects to the next or terminal nodes, which specify the expected value of the class variable Y . Decision trees are built recursively, starting with the data set by the root node. Then, with the help of a split criterion, the learning sample is split into smaller and smaller parts. The process is repeat

until all nodes are pure, i.e the conditions in each subset belong to the same class. This method enables also an evaluation of the interactions and influences between variables.

The following decision tree (Fig. 2.4) gives as example the repartition of legislations and ratings for the driver side:

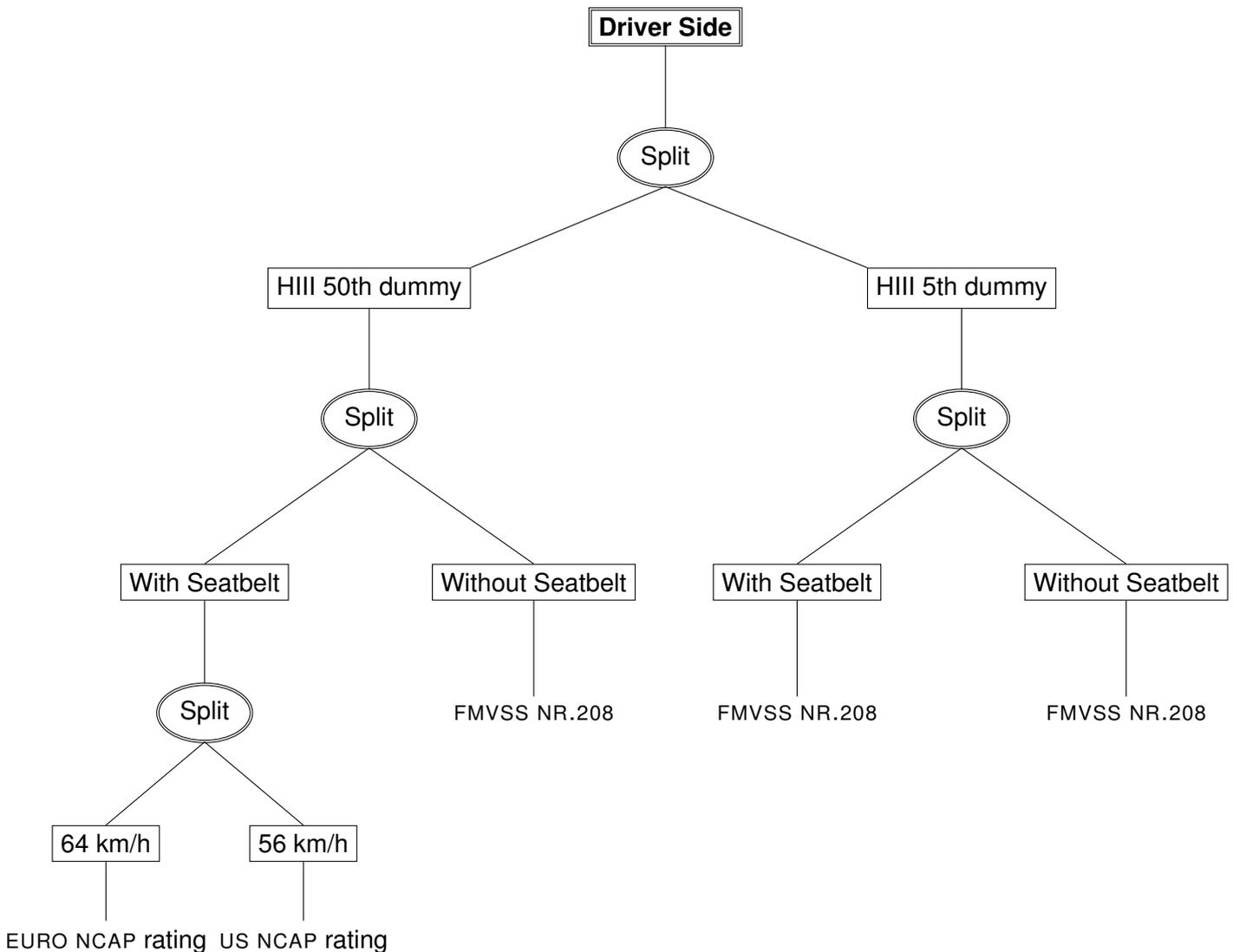


Figure 2.4: Determination of the legislation with a decision tree (driver side)

Notice that the values of the predicted variables (EURO NCAP rating, US NCAP or FMVSS NR.208) are at the bottom of the tree and the predictors (HIII 50th dummy, with seatbelt, without seatbelt, HIII 5th dummy) come into the system at each node of the tree.

The top node contains the entire sample. Each remaining node contains a subset of the sample in the node directly above it. Furthermore, each node contains the informations of the samples in the nodes connected and directly below it. Each node can be considered as a cluster of objects, or cases, that is to be split in further branches in the tree. Tree prediction models add two ingredients: the predictor and predicted variables labelling the nodes and the branches.

Decision trees are composed of two main types:

- ◆ classification tree: the target variables are categorical (finite set of elements) and the class of each observation of the learning sample is known;
- ◆ regression tree: the target variables are continuous values and the observations' classes are unknown.

Classification And Regression Tree (CART) method was developed by Breiman et al. [5] and can handle both numerical and categorical variables. Only binary trees can be produced by splitting a node into two child nodes repeatedly. The root node contains the whole learning sample.

2.2.3.2/ CART AS SUPERVISED LEARNING

CART, a recursive partitioning method, builds classification and regression trees for predicting continuous variables (regression) and categorical variables (classification). In general, the objective of the tree-building algorithms is to determine a set of *if – then* rules (logical split conditions) that enables the prediction or the classification of a new data.

2.2.3.3/ NOTATIONS

This paragraph resumes the variables that are used to define the DBI:

- ◆ $Y = \{y_1, y_2, \dots, y_N\}$: a dependent variable which represents the target variable. It can be categorical or continuous;
- ◆ If Y is categorical with J classes, its class y_k takes values in $c = \{1, \dots, J\}$ with $J \in \mathbb{N}$;
- ◆ $X = \{x_1, x_2, \dots, x_N\}$: a sample of N observations in \mathbb{R}^n (n -dimensional Euclidean space);
- ◆ n : the set of all predictor variables;
- ◆ x_k : k -feature vector;
- ◆ x_{kj} : j^{th} feature of x_k ;
- ◆ $x_k = (x_{k1}, x_{k2}, \dots, x_{kn})^T$.

The set of N observations Y is represented by a $(N \times n)$ matrix:

$$Y = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \cdot & \cdot & \dots & \cdot \\ x_{k1} & x_{k2} & \dots & x_{kn} \\ \cdot & \cdot & \dots & \cdot \\ x_{N1} & y_{N2} & \dots & x_{Nn} \end{bmatrix} \quad (2.39)$$

The predictor can be categorical or continuous.

- ◆ $H = \{(x_1, y_1), \dots, (x_N, y_N)\}$: the whole learning sample;

- ◆ $H(t)$: the learning samples that falls in node t ;
- ◆ $\pi(j)$ with $j = 1, \dots, J$: the prior probability of $Y = j$ when $Y \in C = \{1, \dots, J\}$;
- ◆ $p(j, t)$ with $j = 1, \dots, J$: the probability of a case in class j and node t ;
- ◆ $p(t)$: the probability of a case in node t ;
- ◆ $p(j|t)$ with $j = 1, \dots, J$: the probability of a case in class j given that it falls into node t ;
- ◆ $\Delta i(s, t)$: the splitting criterion with s the best split at node t .

2.2.3.4/ THE GROWING PROCESS

The basic idea of tree growing is to choose a split among all the possible splits at each node, so that the resulting child nodes are the "purest". In the DBI algorithm, only univariate splits are considered. It means that each split depends on the value of only one predictor variable. All possible splits consist of possible splits of each predictor. If X denotes a *nominal categorical variable* of I categories, there are $2^{I-1} - 1$ possible splits for this predictor. If X is a *continuous variable* with K different values, there are $K - 1$ different possible splits on X . A tree is grown starting from the root node by repeatedly using the three following steps on each node:

1. Find each predictor's best split:
 - ◆ For each continuous predictor variable, sort its K values from the smallest to the largest. Then, determine the $K - 1$ possible split points by taking a value between two consecutive values. Mostly, the average value of two consecutive value is taken. Then, let $S = [s_1, s_2, \dots, s_{K-1}]$ be the split sample, each split points s_i is evaluate as following:
If $X \leq s_i$, the case goes to left child node, otherwise, goes to the right. The best split s is obtained for the one that maximizes the splitting criterion;
 - ◆ For each categorical predictor, examine each possible subset of categories (calling it A , if $x \in A$, the case goes to the left child node, otherwise, goes to the right.) to find the best split.
2. Find the node's best split.
Among the splits found in Step 1 (a best split is found for each predictor variable), choose the one that maximizes the splitting criterion;
3. Split the node using the best split chosen in Step 2 if the stopping rules are not satisfied.

2.2.3.5/ SPLITTING CRITERIA AND IMPURITY MEASURES

Split criteria are applied in order to split the learning sample into smaller subsets and thus to build the tree. The most commonly used criteria were listed in the literature [62] [66] [60]:

- ◆ **Gini Index G_{Ix} based on Gini impurity:** this measure of diversity evaluates the node impurity of the classification tree and is defined as:

$$Gini(S) = 1 - \sum_i p_i^2 \quad (2.40)$$

The split criterion G_{Ix} enables the building of the decision tree and is based on the Gini Index. It is defined as follows:

$$G_{Ix}(S, A) = \sum_i \frac{|S_i|}{|S|} \cdot Gini(S_i) \quad (2.41)$$

This criterion should be as low as possible and is often used to build CART algorithms.

- ◆ **Information Gain based on Shannons entropy:** used in ID3 algorithms, this split criterion is based on the entropy from information theory and measures the degree of confusion. The process of pruning consists of simplifying and adjusting the model of the tree more closely to the data set used to build it.

The Information Gain is based on the Shannons entropy $E(S)$ of a data set S , defined as:

$$E(S) = - \sum_{i=1}^m p_i \cdot \log_2(p_i) \quad (2.42)$$

If the entropy is high, it means that the impurity is maximal, i.e. the set S is from a uniform distribution. If the entropy is low, it means that the node is pure.

The Information Gain $IG(S, A)$ for an attribute A is thus defined as:

$$IG(S, A) = E(S) - I(S, A) \quad (2.43)$$

$$= E(S) - \sum_i \frac{|S_i|}{|S|} \cdot E(S_i) \quad (2.44)$$

where S_i are the subsets from the set S and $I(S, A)$ the weighted average over all sets resulting from the split.

Contrary to the Gini Index, $IG(S, A)$ should be maximized, i.e. the average entropy $I(S, A)$ has to be minimized (because the entropy is constant for all attributes). The attribute A that maximizes this criterion reduces the disorderedness.

- ◆ **Information Gain Ratio:** applied by C4.5 algorithms, the Information Gain Ratio $IGR(S, A)$ is an amelioration of the Information Gain by taking the intrinsic information $IntI(S, A)$ of a split into consideration. It takes the number and size of tree branches into consideration when selecting an attribute A . It is defined as:

$$IGR(S, A) = \frac{IG(S, A)}{IntI(S, A)} \quad (2.45)$$

$$= - \frac{IG(S, A)}{\sum_i \frac{|S_i|}{|S|} \cdot \log\left(\frac{|S_i|}{|S|}\right)} \quad (2.46)$$

Categorical target variable

If Y is categorical, the Gini Index measure for impurity is mostly used among splitting criteria.

At node t , the probabilities $p(j, t)$, $p(t)$ and $p(j|t)$ are estimated by:

$$p(j, t) = \frac{\pi(j) \cdot N_{w,j}(t)}{N_{w,j}} \quad (2.47)$$

$$p(t) = \sum_{j=1}^J p(j, t) \quad (2.48)$$

$$p(j|t) = \frac{p(j, t)}{p(t)} \quad (2.49)$$

where

$$N_{w,j} = \sum_{n \in H} I(y_n = j) \quad (2.50)$$

$$N_{w,j}(t) = \sum_{n \in H(t)} I(y_n = j) \quad (2.51)$$

with $I(a = b)$ the indicator function taking the value 1 when $a = b$, 0 otherwise.

This criteria is based on the Gini impurity². As an impurity measure, it reaches a value 0 when only one class is present at a node. With prior estimated from class sizes and equal misclassification costs, the Gini impurity measure is computed as the sum of products of all pairs of class proportions for classes present at the node. It reaches its maximum value when class sizes at the node are equal. The Gini index is equal to 0 if all cases in a node belong to the same class. The Gini impurity measure at a node t is defined as :

$$imp(t) = \sum_{i,j} p(i|t)p(j|t) \quad (2.52)$$

$$= \sum_{i,j} p(i|t) \cdot \left(\sum_{i \neq j} 1 - p(i|t) \right) \quad (2.53)$$

$$= 1 - \sum_i p(i|t)^2 \quad (2.54)$$

The Gini splitting criterion is the decrease of impurity defined as:

$$\Delta imp(s, t) = imp(t) - G_{div} \quad \text{with} \quad G_{div} = p_L \cdot imp(t_L) + p_R \cdot imp(t_R) \quad (2.55)$$

where p_L and p_R are probabilities of sending a case to the left child node t_L and to the right child node t_R respectively, and G_{div} is the combine Gini impurity measure of the child nodes. G_{div} has to be minimized in order to maximize $\Delta imp(s, t)$. The probabilities p_L and p_R are estimated through Eq. 2.56:

² refers to impurity function which measures the extent of purity for a region containing data points from possible different classes

$$p_L = \frac{p(t_L)}{p(t)} \quad \text{and} \quad p_R = \frac{p(t_R)}{p(t)} \quad (2.56)$$

Continuous target variable

When Y is continuous, the splitting criterion defined in Eq. 2.55 is used with the Least Squares Deviation (LSD) impurity measures.

2.2.3.6/ STOPPING RULES AND EVALUATION OF THE TREE

Stopping rules are necessary to control the growing process. It is stopped when [5]:

- ◆ a node is pure, i.e. all cases in the node have identical values of the dependant variable Y ;
- ◆ all cases in a node have identical values for each predictor;
- ◆ the current tree depth recalls the user-specified maximum tree depth limit value;
- ◆ the size of a node is less than the user-specified minimum node size value;
- ◆ the split of a node results in a child node whose node size is less than the user-specified minimum child node size value.

Once the tree is built, its quality of the prediction has to be evaluated. It can also help "prune back" the tree, i.e. select a simpler tree than the original, but containing the same information. There are two main model validations to estimate how the predictive model perform in practice:

- ◆ Cross validation: it involves the partitioning of a sample of data into complementary subsets. One is called training set and the other validation test. The model fits a function with only the training set and validates the model on the validation set;
- ◆ K-fold cross validation: the dataset is separated into K subsets randomly. One fold is used as validation set and the other $K - 1$ folds are used as training set. The classifier fits a function using only the training set and then predicts the output values for a data included in the validation test. This process is repeated K times. The average error rate based on each error (i.e. the difference between the real and the predicted values) enables the evaluation of the model;
- ◆ Leave-one-out cross validation: it corresponds to the K-fold cross validation with $K = N$ (N the number of points in the dataset). This method performs very well but can be expensive to compute.

SLED TESTS CLASSIFICATION THROUGH K-NN METHOD

The Knowledge Discovery in Databases (KDD) and data mining methods are being increasingly developed in industrial applications, especially automobile crash studies. KDD is the process of discovering useful and understandable patterns and aims at extracting knowledge from data, using data mining methods.

In order to extract knowledge from data in databases, the steps of the KDD are applied on the restraint systems parameters (inputs): selection of the data, pre-processing, transformation, application of a data mining method and evaluation of the model. This method is adapted to large databases as well as databases containing a limited amount of data.

3.1/ APPLICATION OF THE KNOWLEDGE DISCOVERY IN DATABASES

3.1.1/ STEP 1: DATA SELECTION

The KDD process is applied on the passive restraint systems, especially the parameters which could have an influence on the biomechanical criteria of the dummy. Three components of the passive safety have been studied: the airbag (driver and passenger), kneebag and seat belt. The dummy position and the quality of the pulse have been also analysed. In the analysis, the potential modification of the environment like the buck, instrumental panel or seat, is not taken into consideration. An expert elicitation has been conducted, it aims at defining the key factors of the components with the help of a questionnaire filling in by passive safety experts. 57 parameters on the driver side and 46 parameters on the passenger side have been selected as influencing the occupant's injuries during a crash. Table 3.1 resumes the selected parameters for each restraint system.

These parameters have been researched and stored for different regulations and ratings. The number of complete available sled tests is indexed in Table 3.2.

3.1.2/ STEP 2: DATA PRE-PROCESSING

The collected data covers the period 2000-2014. The data has been prepared and completed from different sources (dummy instrumentation with electronics devices, measure-

Table 3.1: Input parameters influencing dummy biomechanical values

	Driver airbag	Passenger airbag	Kneebag (both sides)	Seat belt (both sides)
Numerical inputs	<ul style="list-style-type: none"> * Volume * Time to fire * Steering wheel diameter * Steering wheel angle * Steering column breaking distance * Steering column breaking deformation force * Steering column maximum deformation force 	<ul style="list-style-type: none"> * Volume * Time to fire 	<ul style="list-style-type: none"> * Volume * Time to fire 	<ul style="list-style-type: none"> * Load limiter values * Webbing on spool once the dummy is belted
Categorical inputs	<ul style="list-style-type: none"> * Type (round, not round) * Type of fabric (350 dtex Polyester, 470 dtex Nylon, etc.) * Number of tear seams (with their orientation and length) * Inflator name * Number of ventholes (with their diameter) * Active bag vent (with its size and time to fire) * Tethers (with their length and position) * Inflator name * Inflator type (pyrotechnic, hybrid or cold gas) * Number of stages (single or dual stages) * Steering wheel design (2, 3 or 4 spokes) * Steering column type (mechanical or electrical) 	<ul style="list-style-type: none"> * Type (L-shape, Square-shape, 3D-shape, etc.) * Type of fabric (350 dtex Polyester, 470 dtex Nylon, etc.) * Divided panel * Number of ventholes (with their diameter) * Active bag vent (with its size and time to fire) * Tethers (with their length and position) * Inflator name * Inflator type (pyrotechnic, hybrid or cold gas) * Number of stages (single or dual stages) 	<ul style="list-style-type: none"> * Number of ventholes (with their diameter) * Number of tether * Buckle pretensioner (with its time to fire, length of pretension) 	<ul style="list-style-type: none"> * Type (constant, degressive, regressive or switchable load limiter) * Retractor pretensioner (with its time to fire) * Anchor pretensioner (with its time to fire, length of pretension) * Angle between shoulder/seat and D-ring * Height adjustment (with its position) * Dynamic locking tongue

Table 3.2: Database including sled tests with known input parameters

Regulation / rating	Dummies	Sled tests included in the database	
		Driver	Passenger
EURO NCAP rating	50th / 50th	27	140
US NCAP rating	50th / 5th	115	216
FMVSS NR.208 with seat belt	5th / 50th	51	61
FMVSS NR.208 without seat belt	50th / 5th	80	42
FMVSS NR.208 without seat belt	5th / 50th	34	62

ments of the dummy position) and it is possible that some input values are not available and must be estimated.

This step of data replacement is one of the most important challenges of the data pre-processing and is known in literature as "missing data". Tseng et al. [29] were confronted with this problem and they developed a new method based on cluster properties. Batista and Monard [26] also studied this problem by using the 10-NNI method, a k-nearest neighbour imputation algorithm. The term of imputation denotes a procedure that replaces the missing values by plausible values. This replacement is independent of the machine learning algorithm. For this study, the k-NN imputation method is applied as learning algorithm and allows the treatment of the missing values included in the database. The algorithm for this analysis employs some known relationships that already exist between sled tests from the same legislation, sled tests from the same project or theoretical relationships between restraints systems and biomechanical values.

Two types of "missing data" are distinguished:

- ◆ Case 1: the parameter's value is not given but this parameter must be defined, e.g. the dummy position or temperature. This case will be reported in this section;
- ◆ Case 2: the parameter's value is not given because the restraint system is not used or does not have some properties, e.g. if the seat belt system does not have a buckle pretensioner, data such as time to fire or length of pretension cannot be recorded. These values are not directly considered as missing values.

There are three possibilities to solve the problem of missing data (from case 1):

1. Repeat the sled test: this solution seems to be the best one but it is also expensive and sometimes impossible (old projects, trim parts no more available);
2. Average calculation: this method replaces the missing value through the average of the values of other sled tests;
3. Imputation method by using the k-NN: some algorithms can be used in order to define the missing value according to the values included in the database.

The created algorithm takes 3 types of missing data into consideration:

1. The missing data is independent of the sled test, e.g. the dummy temperature. In this case, the average of all dummy temperature values included in the database is calculated;
2. The missing data is dependent on the project, e.g. the dummy position. The dummy position is dependent on the trim parts (seat, instrumental panel) and car body. In this case, the algorithm first selects the sled tests only from the same project and then calculates the average of the values from these sled tests;
3. The missing data is dependent on the project and other parameters, e.g. the "webbing on spool" of the seat belt. It is one of the parameters which influences directly the dummy shoulder force. In this case, the algorithm first selects only the sled tests of the same project, then compares their dummy shoulder force (with a tolerance of +/- 500N) and finally calculates the average of these sled tests.

This algorithm only intervenes if a parameter is missing and should be defined. After running this algorithm, the database is complete and is ready to be exploited for the next KDD step, the data transformation.

3.1.3/ STEP 3: DATA TRANSFORMATION

Because of the different provenance and scale of the data, it has to be transformed before applying a data mining method. Some parameters of restraint systems (e.g. the type of seatbelt) are defined through a nominal description. In order to interpret it in the algorithm, a numerical code has been given for each nominal description (Fig. 3.1):

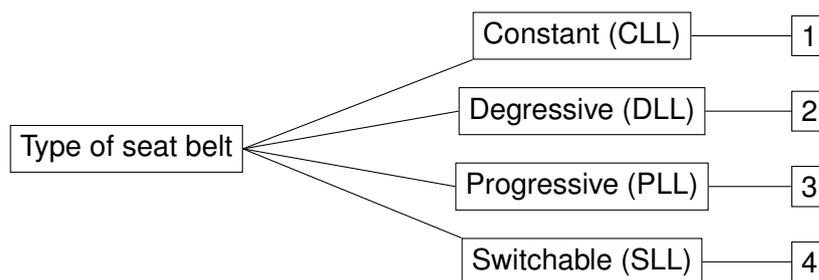


Figure 3.1: Numerical transformation of nominal parameters

Moreover, the parameter's values have a different range. For example, the vent hole diameter of the airbag often does not exceed 60 mm and the length of the tethers can vary between 100 mm and 300 mm. Their use without any changes gives automatically more weight to a data with an important spread and under evaluate the other [33]. The data transformation step thus consists in the normalization of all numerical values into a range (except for the binary values). The range $[-1;1]$ has been chosen arbitrarily.

Table 3.3 presents numerical values of some parameters P_i for 4 tests and Table 3.4 their corresponding normalised value. The "New Test" is not already done, only the inputs parameters are known and the objective is the classification of this "New Test" in the database.

Table 3.3: Example of parameter's values

Parameter	P1	P2	P3	P4	P5	P6	P7	P8
Test 1	13	406	NaN	121	114	0	30	3,5
Test 2	18	NaN	NaN	130	135	0	30	3
Test 3	13	411	300	118	109	15	35	3,5
Test 4	13	411	304	125	118	NaN	35	NaN
New test	18	440	NaN	NaN	133	30	40	2,5

If the missing data is represented as NaN, the parameter is not available in the specific test.

After transforming and normalising the data, the database is yet homogeneous and the classification through a data mining method can be performed.

Table 3.4: Example of normalised parameter's values

Parameter	P1	P2	P3	P4	P5	P6	P7	P8
Test 1	-1	-1	NaN	-0,5	-0,615	-1	-4	1
Test 2	1	NaN	NaN	1	1	-1	-1	0
Test 3	-1	-0,706	-1	-1	-1	0	0	1
Test 4	-1	-0,706	1	0,167	-0,308	NaN	0	NaN
New test	1	1	NaN	NaN	0,846	1	1	-1

3.1.4/ STEP 4: DATA MINING METHOD: K-NEAREST NEIGHBOURS CLASSIFICATION

Wu et al. [45] listed the top 10 algorithms in data mining in the IEEE International Conference on Data Mining 2006. The k-nearest neighbours (k-NN) classification is one of those and uses the Euclidean distance for estimating the similarity between the inputs' parameters of a new test x_{new} and those of the i^{th} training set x_{train}^i . The k-NN is mathematically represented by Eq. 3.1:

$$d(\mathbf{x}_{new}, \mathbf{x}_{train}^i) = \sqrt{\sum_{j=1}^p (x_{new}(j) - x_{train}^i(j))^2} \quad (3.1)$$

The main target of this algorithm is the calculation of the Euclidean distance between a new sled tests and each test included in the database by comparing the inputs parameters (setup, airbags, dummy specification). The parameters taken into account for the determination of the nearest neighbour are indexes in Table 3.1.

In order to collect the information of the sled tests and to calculate the smallest $d(\mathbf{x}_{new}, \mathbf{x}_{train}^i)$ which represents the next nearest neighbour, a graphical user interface (GUI) has been created (Fig. 3.2(a)). The programme includes sled tests data of different customers from 2000-2014 and also the k-NN programme.

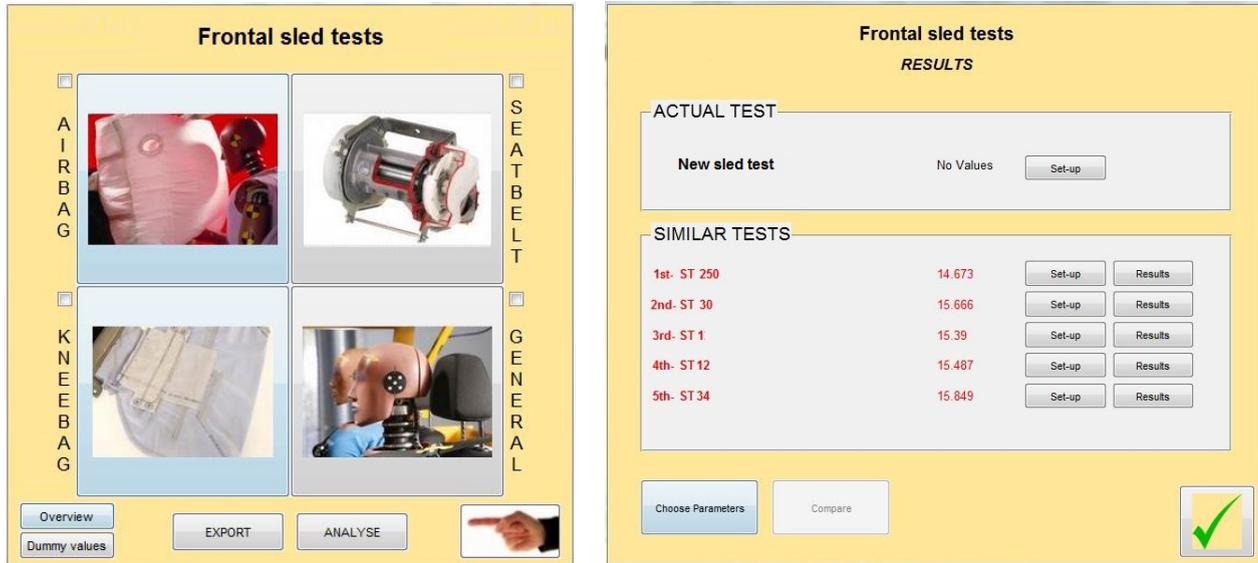
This GUI has been created during the Master internship of Pablo Jueas (from ENSMM engineering school) that I have supervised in 2013. The detailed presentation of the GUI panels can be found in Appendix C.

Thanks to it, the classification of the sled tests can be executed and presented to the user.

3.1.5/ STEP 5: DATA CLASSIFICATION

The last step of the KDD process, the classification, consists in finding the closest training instance by minimizing the Euclidean distance given by Eq. 3.1.

Through the GUI, the parameters of the sled tests included in the database can be compared (Fig. 3.2(b)), and the configuration's difference can be analysed. The sled tests are classified according to their Euclidean distance values (Eq. 3.1). The smaller this value is, the more the input configuration of the new one is similar to its compared test. It gives for the upcoming tests a first indication of the expected results for different regulations. Some additional tools implemented in the GUI allow a personalisation of the comparison.



(a) General panel of the GUI

(b) Example of sled tests classification

Figure 3.2: Graphical User Interface (GUI)

Thanks to this method, expected sled test trends can be estimated and the initial configuration of the sled test can be adapted before running the test. With the knowledge of the relationships between in- and output parameters, it could even become possible to influence future sled test results positively before having run it via the variation of input parameters.

3.2/ VALIDATION OF THE K-NN CLASSIFICATION

Two algorithms have been programmed in order to classify the data: the first one allows the replacement of missing values and the second one the classification of the sled tests through the k-NN method. The "leave-one-out cross validation" (LOOVC) method has been applied for the validation of these algorithms.

The validations have been conducted with the data from EURO NCAP rating on passenger side

3.2.1/ ALGORITHM FOR MISSING DATA

LOOCV is a model validation method of machine learning which uses one value as validation and the remaining values as a training set. This method enables the validation of the algorithm which should be able to replace a missing value. The algorithm has been validated for the 3 types of missing data described in Section 3.1.3:

- ◆ dummy temperature : this missing value is determinate through the average calculation;

- ◆ dummy position through the distance "left knee, shorter distance ahead" (chosen arbitrary): this missing data depends on the sled test and is obtained with the help of the created algorithm;
- ◆ webbing on spool of the seat belt, which depends on the sled test project but also has a influence on other parameters (such as the shoulder force applied on the dummy). The created algorithm enables the estimate value of the missing value.

The data set contains 140 sled tests: 139 values are used as training set and 1 value is the validation value. For each missing value, the algorithm calculates an approximate value and its corresponding relative error E defined as:

$$E = \frac{|\text{exact value} - \text{estimated value}|}{|\text{exact value}|} \quad (3.2)$$

Fig. 3.3, 3.4 and 3.5 present the curves of the exact and approximate values and of the relative error for the different missing parameters.

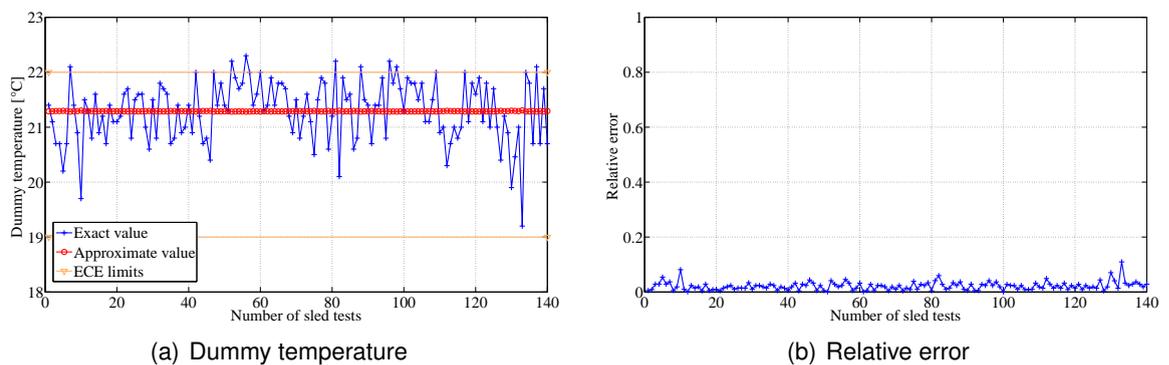


Figure 3.3: Estimation's curves for the dummy temperature

As shown in Fig 3.3, all dummy temperature values estimated by the algorithm are similar. In fact, for parameters which are independent on the sled test, the approximate value is determined through the average of all available values. Lots of dummy temperatures are between 20°C and 22°C. The average temperature is estimated as 21,3°C. The maximal difference is +2,2°C for only one sled test. It is acceptable since, according to the European legislation, the dummy temperature must be between 19°C and 22°C (these limits are also valid for the rating).

The dummy is defined by 6 positions. Only one has been chosen arbitrary to validate the algorithm: the distance between the left dummy knee and the instrumental panel (shorter distance). Fig. 3.4 shows the dummy position values estimated with the help of the algorithm. For such types of parameters which depend on the sled test project, the algorithm takes only the data coming from the same project into consideration. The database contains 140 values at the moment, and per project, it is possible that only one sled test is available. That is why some missing values cannot be determined (rupture in the curve). The maximal difference between the exact and approximate values is 22 mm. Even if the sled tests come from the same project (and thus the same buck), it is difficult to seat the dummy exactly at the same position as the reference. That is why the experts

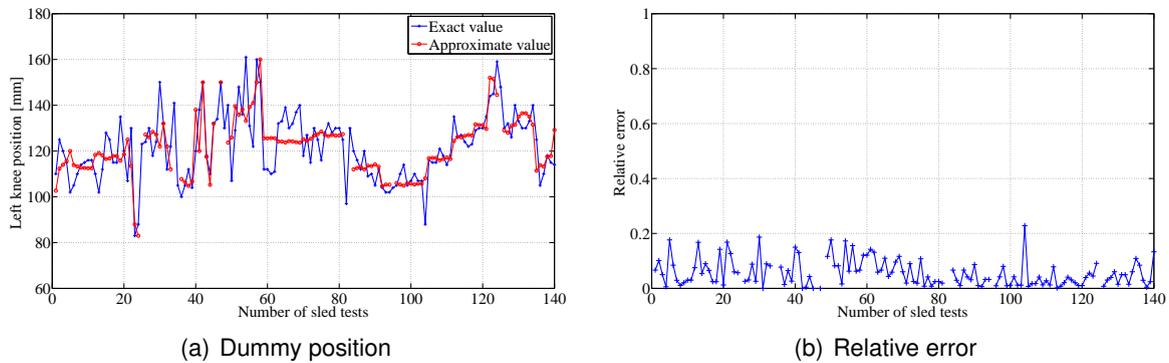


Figure 3.4: Estimation's curves for the dummy position (type 1 of the possible missing values)

estimate that a tolerance of 10 mm is acceptable and should not influence the results. A difference of 22 mm could have an influence on the results.

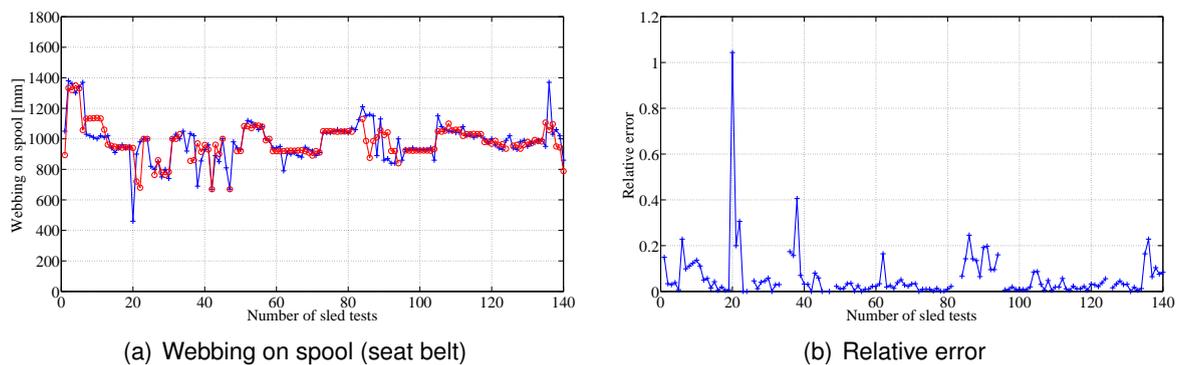


Figure 3.5: Estimation's curves for the webbing on spool of the seat belt

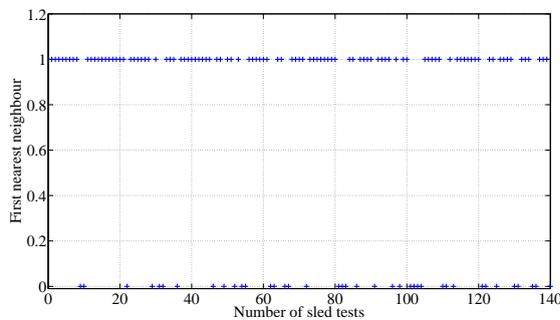
As illustrated in Fig. 3.5, the lengths of the webbing on spool estimated with the help of the algorithm are similar in trend to the real values. Some approximate values are a little bit far away from the real values. The webbing on spool determination is based on the data of the sled tests coming from the same project and on their dummy shoulder force. In fact, the webbing on spool has a direct influence on this criterion. If only a few sled tests from the same project are available, the estimate value can differ from the real one. These approximations can be rectified by adding more sled tests.

These three validations of the algorithm responsible for the replacement of missing values confirm its potential and allows its use for replacing the missing data if desired.

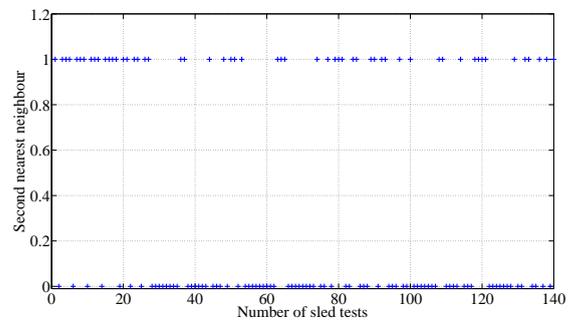
3.2.2/ THE SLED TESTS CLASSIFICATION

One sled test of the database is taken as validation test, the remaining tests as training set (leave-one-out cross validation). This validation compares two classifications: the first one is based on the clear database, i.e. all the inputs parameters are known. For

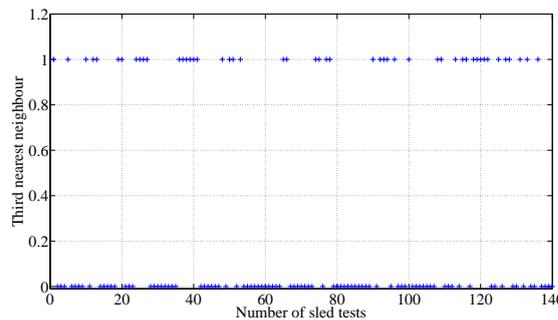
the second classification, three parameters have been deleted and thus replaced by the algorithm itself. The missing values are the three parameters described in Section 3.1.2. There are the most often missing values. The graphs in Fig. 3.6 represent the comparison between the k-NN sled tests, with $k \in \{1, 2, 3\}$, determined with and without missing values. If the value is equal to 1, it means that the k-NN sled test obtained without any missing parameters is the same as the k-NN sled test obtained when three values are missing.



(a) Comparison between the 1-NN sled test determined with and without missing values



(b) Comparison between the 2-NN sled test determined with and without missing values



(c) Comparison between the 3-NN sled test determined with and without missing values

Figure 3.6: Comparison's curves between the k-NN sled test determined with and without missing parameters

As showed in Fig. 3.6, for 72% of cases, the 1-NN sled test estimated with the help of the algorithm is unchanged by replacing the missing values. It means that the replacement of the missing values by the algorithm affects only minor the determination of the 1-NN sled test. In only 40% and 35% of cases, respectively the 2-NN and the 3-NN are identical.

3.2.3/ ESTIMATION OF THE EURO NCAP SCORE WITH ALL KNOWN PARAMETERS

For this validation, all inputs parameters of the validation test are considered as known. The validation aims at estimating with the help of the k-NN method its expected score for the EURO NCAP rating on the passenger side. Each 140 sled test included in the database is taken one by one as validation test. The curves of the exact and estimate scores are represented in Fig. 3.7. The rating scores are the real obtained rating score of the sled

test, the rating score obtained by its nearest neighbor sled test and the average of the ratings scores obtained by its two first nearest neighbours sled tests. The relative error, according to Eq. 3.2, is also calculated to evaluate the precision of the algorithm.

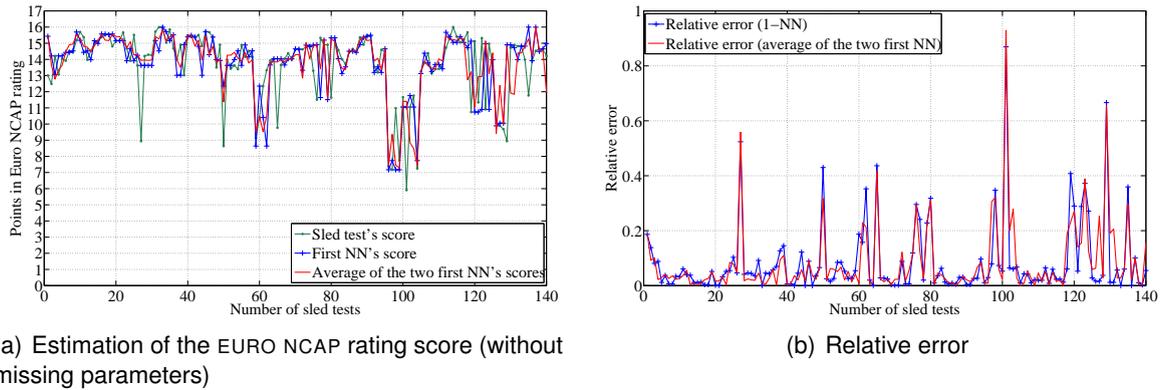


Figure 3.7: Estimation's curves of the EURO NCAP rating score with all known parameters

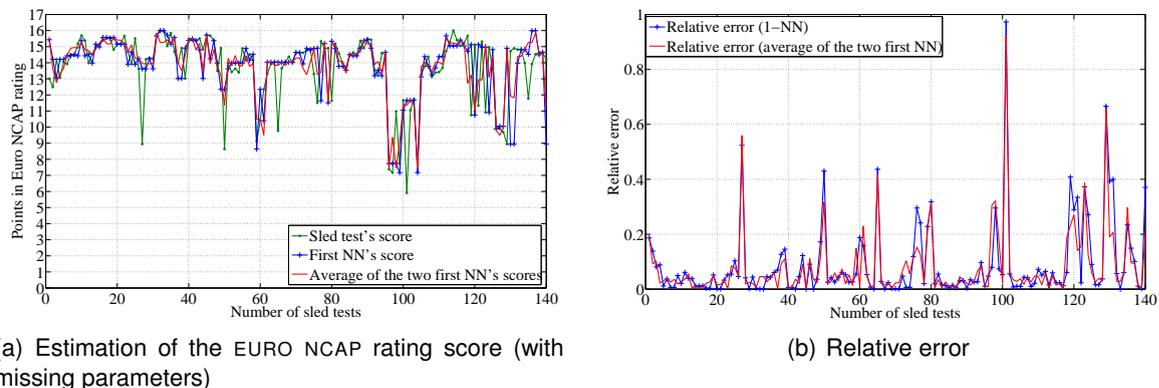
According to the error's curve, the average of the score obtained by the 2 first NN sled tests seems to give a good approximation of the expected EURO NCAP score. The estimation is not successful if the sled test obtained a low score in the rating (because the database contains mostly score above 10 points).

3.2.4/ ESTIMATION OF THE EURO NCAP SCORE WITH UNKNOWN PARAMETERS

This validation is a combination of the two last validations. As usual, the LOOCV is applied. The aim of this validation is the estimation of the score that can be expected in EURO NCAP rating for the passenger side. Three inputs parameters (dummy temperature, dummy position and webbing on spool) have been manually deleted and estimated by the created algorithm. Fig. 3.8 presents diverse rating scores and their corresponding error. The rating scores are the real obtained rating score of the sled test, the rating score obtained by its nearest neighbor sled test and the average of the ratings scores obtained by its two first nearest neighbours sled tests.

The relative error curves between each rating score are similar and do not significantly advantage a method to the other. We can remark that the sled tests for which the score is not satisfactory are the same as in Section 3.1.3 (Fig. 3.7). It confirms that the algorithm has some difficulties to estimate the score if the sled tests present a low one in the EURO NCAP rating.

These two last validations confirm the performance but also show the limits of the algorithm which should learn on the existing data for replacing missing parameters' values (no distinct influence on the results).



(a) Estimation of the EURO NCAP rating score (with missing parameters)

(b) Relative error

Figure 3.8: Estimation's curves of the EURO NCAP rating score with unknown parameters

3.3/ DISCUSSION ON THE K-NN CLASSIFICATION

In order to study and classify the sled tests, the 5 steps of the KDD process have been followed. The first step, the data selection, defines the parameters of each restraint system which could influence the dummy biomechanical criteria. Thanks to an expert elicitation, 57 parameters on the driver side and 46 parameters on the passenger side (driver / passenger airbag, kneebag, seat belt, environment) have been identified by passive safety experts as influencing the dummy biomechanical values. Others external parameters such as the car pulse and dummy position have been also selected. These parameters are independent of the considered legislation / rating but additionally parameters are considered for the driver side (steering column and steering wheel).

After the data selection, the next step consists of the data pre-processing. This is a critical task of the KDD process because of the so-called "missing data". If a parameter is missing, it has to be defined to enable its exploitation by the data mining method. This is the aim of the created algorithm which is able to deal with three types of missing data, according to their dependency to others parameters.

Once the database is complete, the data transformation enables an homogenisation of the data through a numerical transformation for categorical data and a normalisation for the numerical values.

Then the k-nearest neighbour classification is applied as data mining method. The inputs configuration of a new sled test is compared with the configuration of all sled tests included in the database through the calculation of the Euclidean distance.

Finally, according to the Euclidean distance estimated between two sled tests, a rank of the nearest neighbour of a new sled test informs the engineers to the expected results in the chosen regulation.

For a better data processing and practical use for the engineers, a GUI has been created for entering, saving, analysing and classifying the sled tests. This rank responds to an industrial request and enables the analysis of passive safety systems of existing sled tests results before running a new one.

CLUSTERING OF THE DUMMY BIOMECHANICAL VALUES

4.1/ OUTPUTS' CLUSTERING EURO NCAP

In order to regroup dummy biomechanical values that are similar into a same group, an unsupervised learning method, the clustering, is applied. The influence of the input parameters on these clusters will be then analysed. In this chapter, the clustering of the data is focused on the EURO NCAP rating, especially the passenger side. The results of the clustering for the driver side are resumed in Section 4.1.2. As a reminder, for the adult occupant protection, the EURO NCAP score is the sum of the scores obtained by the four dummy parts (head/neck, chest, knee/femur and leg/foot). Each part can reach up to 4 points.

4.1.1/ CLUSTERING FOR THE HIII 50TH DUMMY (PASSENGER SIDE)

4.1.1.1/ GENERALITIES

The considered biomechanical values for the estimation of the EURO NCAP score (adult protection, frontal impact) are referenced in Table 4.1 with the upper and lower values necessary for the score's calculation. The points are computed considering the worst injury criterion (and also applicable modifiers, that are not here taken into account). Between the upper and the lower limits, the score is estimated by linear interpolation. Because 86% of the studied sled tests on the passenger side do not have a knee bag, which should help the reduction of the legs damages, the biomechanical values of the femur compression, knee slider displacement, tibia compression and tibia index are not regrouped in clusters. The eventual damages on these body regions could come from the contact between the legs and the buck environment (glove box), which is not modifiable by the supplier himself. Moreover, the chest damages is represented only by the chest deflection (the viscous criterion is non-critical).

The biomechanical values of the dummy have been clustered according to the body region:

- ◆ outputs' combination for the dummy head: acceleration (3 ms) and criterion HIC 36;
- ◆ outputs' combination for the dummy neck: neck shear, tension and extension;

Table 4.1: Dummy biomechanical values and their limits for the EURO NCAP rating (available for the driver and the passenger)

Injury criteria	Lower limit (4 points in EURO NCAP rating)	Upper limit (0 point in EURO NCAP rating)	Evaluated values (driver side)	Evaluated values (passenger side)
Head Injury Criterion 36 (HIC 36) Acceleration (3 ms)	650	1000	✓	✓
	72 g	88 g	✓	✓
Neck shear	@ 0 ms 1,9 kN @ 25-35 ms 1,2 kN @ 45 ms 1,1 kN	@ 0 ms 3,1 kN @ 25-35 ms 1,5 kN @ 45 ms 1,1 kN	✓	✓
Neck tension	@ 0 ms 2,7 kN @ 35 ms 2,3 kN @ 60 ms 1,1 kN	@ 0 ms 3,3 kN @ 35 ms 2,9 kN @ 60 ms 1,1 kN	✓	✓
Neck extension (Moc-)	42 Nm	57 Nm	✓	✓
Chest deflection	22 mm	50 mm	✓	
Chest Viscous Criterion (VC)	0,5 m/s	1,0 m/s	✓	—
Femurs compression (left and right)	3,8 kN	9,07 kN (@ 10 ms 7,56 kN)	✓	—
Knee slider displacement (left and right)	6 mm	15 mm	✓	—
Tibia compression (left and right)	2 kN	8 kN	✓	—
Tibia index (left and right)	0,4	1,3	✓	—

- ◆ outputs' combination for the dummy chest: chest deflection (criterion VC uncritical for the EURO NCAP rating).

The legs (knee, femur) are not studied on the passenger side (absence of knee bag in EURO NCAP rating).

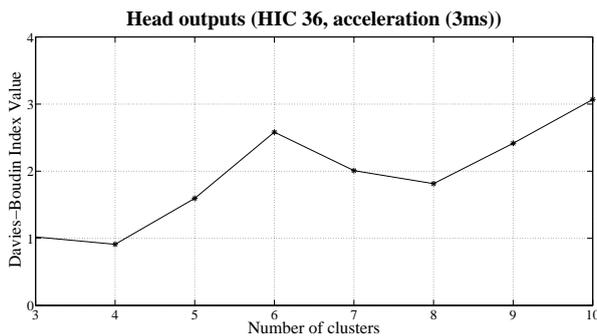
As explained in Section 2.2.2.1, the GK algorithm is one of the clustering methods which can help identify the similarities (clusters) between the biomechanical values. The Davies-Bouldin Index (DBI) method allows the evaluation of the formed clusters. The more the value is low, the more the evaluated cluster is separated and compact. For the passenger side, 431 sled tests coming from the EURO NCAP database are available in order to figure out the clusters (November 2014). All their biomechanical values are correctly measured and plausible. A clustering is conducted for each dummy part.

4.1.1.2/ DUMMY HEAD: CLUSTERS' FORMATION AND EVALUATION

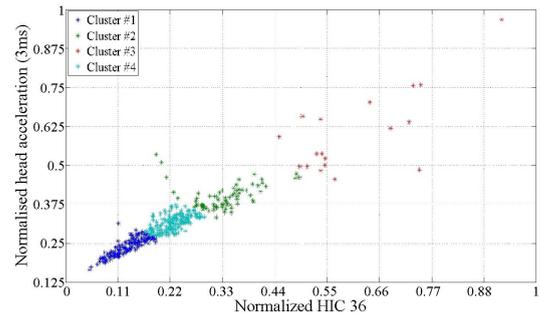
For the dummy head, 431 values of the database are taken into account. The limits for reaching 4 points are listed in Table 4.1.

The GK algorithm is applied on these values, as described in Section 4.1.1.1, with the two available head outputs: the acceleration (3 ms) and the criterion HIC 36. It aims at creating clusters for only the dummy head in order to discover the influence of a pa-

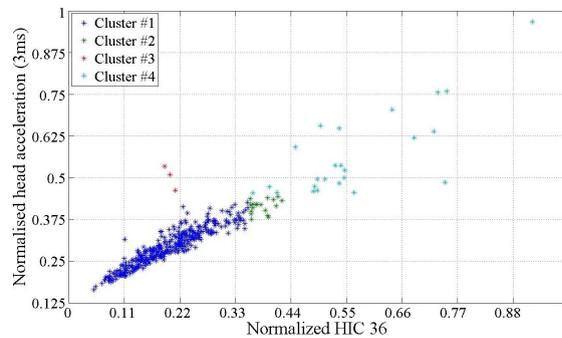
parameter's change only on this dummy part. After running the algorithm, the DBI criterion evaluates the compactness and the separation of the found clusters. Fig. 4.1(a) shows the DBI values from 3 to 10 clusters for the two head outputs.



(a) Representation of the DBI values



(b) Clustering dummy head (4 clusters): graphical representation



(c) Clustering dummy head (4 proposed clusters): graphical representation

Figure 4.1: Head combination (passenger, Europa, HIII 50th dummy)

The best clustering according to the DBI is the clustering with 4 clusters (lowest DBI value), which are represented in Fig. 4.1(b). The properties of the discovered clusters are indexed in Table 4.2.

As shown in Fig. 4.1(b), the clusters are pretty well-separated, in the sense that each value belongs to only one cluster. According to the clusters' repartition (Table 4.2), the clusters #1 and #2 are very similar in size and in results (both clusters include sled tests with 4 points) and the cluster #4 gathers sled tests with a low score for both outputs. The score's range for both outputs in cluster #3 is large and gathers all possible scores. It can be thus difficult to evaluate the influence of a parameter by keeping this clustering. In fact, a change of a parameter that involves a cluster's change must involve a modification of the score's value. That is why the more distinguish the clusters are, the more precise the influence study will be. For this reason, the distribution proposed by the GK algorithm into a cluster is not optimal.

An analysis of the dummy head values shows that already 89% of the sled tests get the maximal score. It is thus useful to gather all sled tests with 4 points overall in only one cluster, even if this cluster contains the majority of the sled tests. That is why a new clustering based on this observation is proposed (Table 4.3). For each output, two

Table 4.2: Dummy head, 4 clusters: characteristics (passenger, Europa, HIII 50th dummy)

Item	Number of elements in the cluster	Head acceleration (3 ms)	HIC 36
		Range of values	
#1	164	[26 g ; 50 g]	[92 ; 345]
#2	164	[44 g ; 59 g]	[303 ; 532]
#3	83	[53 g ; 85 g]	[344,86 ; 895]
#4	20	[73 g ; 155 g]	[816 ; 1668]
		Number of points obtained for the dummy head (EURO NCAP rating)	
#1	164	4 points	4 points
#2	164	4 points	4 points
#3	83	0,64 - 4 points	1,2 - 4 points
#4	20	0 - 3,76 points	0 - 2,10 points

clusters are formed by taking the score into consideration, one with results of 4 points and another one with results under 4 points. It aims at helping the engineers to understand which head output value has to be improved. Fig. 4.1(c) represents the new proposed clusters graphically.

Table 4.3: Dummy head, 4 clusters: proposed clustering (passenger, Europa, HIII 50th dummy)

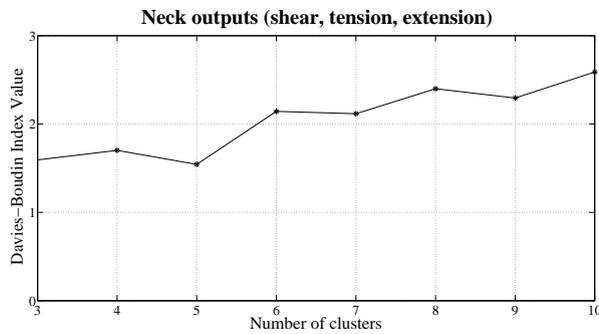
Item	Number of elements in the cluster	Head acceleration (3 ms)	HIC 36
		Range of values	
#1	382	< 72 g	< 650
#2	20	< 72 g	> 650
#3	3	> 72 g	< 650
#4	26	> 72 g	> 650
		Number of points obtained for the dummy head (EURO NCAP rating)	
#1	382	4 points	4 points
#2	20	4 points	< 4 points
#3	3	< 4 points	4 points
#4	26	< 4 points	< 4 points

These clusters will be used as classes for the study of the parameters' influence.

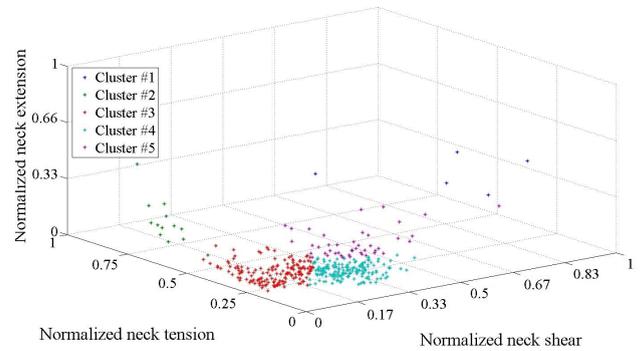
4.1.1.3/ DUMMY NECK: CLUSTERS' FORMATION AND EVALUATION

The procedure for the dummy neck is the same as for the head, even if three outputs have been used to form the clusters: the neck shear (F_x), the neck tension (F_z) and the neck extension (M_{Oyc}). Fig. 4.2(a) shows the DBI values for 3 until 10 clusters for the neck outputs combination.

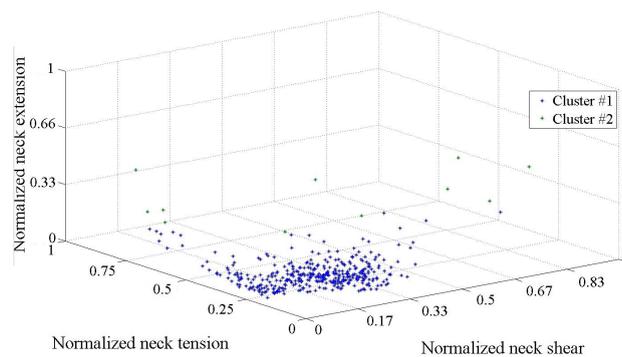
According to the DBI, the best clustering should be with 5 clusters (lowest DBI value),



(a) Representation of the DBI values



(b) Clustering dummy neck (5 clusters): graphical representation



(c) Clustering dummy neck (2 proposed clusters): graphical representation

Figure 4.2: Neck combination (passenger, Europa, HIII 50th dummy)

as represented in Fig. 4.2(b). Table 4.4 resumes the values' range and scores of the sled tests for each cluster. The overall score corresponds to the minimal score obtained between the 3 neck outputs.

As before for the head, this clustering does not present the propriety of uniqueness. It means that sled tests with 4 points overall can be gathered either in cluster #3, #4 or #5. The influence of a parameter's change cannot thus be correctly evaluated.

However, we can remark that 97% of the sled tests have already reached the 4 points overall for the neck. Therefore, it does not make sense to separate them into several clusters. That is why sled tests obtaining 4 points overall can be regrouped into a same cluster. Because of the low number of sled tests with an overall score under 4 points (11 sled tests out of 431), their separation into several clusters according to the critical neck outputs is not necessary. The clustering based on this observation is represented in Fig. 4.2(c) and resumed in Table 4.5.

The two clusters will be used as classes for the study of the parameters' influence.

Table 4.4: Dummy neck, 5 clusters: characteristics (passenger, Europa, HIII 50th dummy)

Item	Number of elements in the cluster	Neck shear	Neck tension	Neck extension	Neck overall
Range of values					
#1	5	[0.158; 0.781]	[0.150; 0.226]	[0.563; 1]	—
#2	13	[0.002; 0.078]	[0.642; 1]	[0.097; 0.699]	—
#3	170	[0.002; 0.203]	[0.160; 0.561]	[0.039; 0.330]	—
#4	196	[0.112; 0.402]	[0.091; 0.385]	[0.291; 0.223]	—
#5	47	[0.162; 0.762]	[0.129; 0.556]	[0.146; 0.563]	—
Number of points obtained for the dummy neck (EURO NCAP rating)					
#1	5	0 - 4 points	4 points	0 point	0 point
#2	13	4 points	0 point	0 - 4 points	0 point
#3	170	4 points	0 - 4 points	4 points	0 - 4 points
#4	196	0 - 4 points	0 - 4 points	4 points	0 - 4 points
#5	47	0 - 4 points	0 - 4 points	0 - 4 points	0 - 4 points

Table 4.5: Dummy neck, 2 proposed clusters : characteristics (passenger, Europa, HIII 50th dummy)

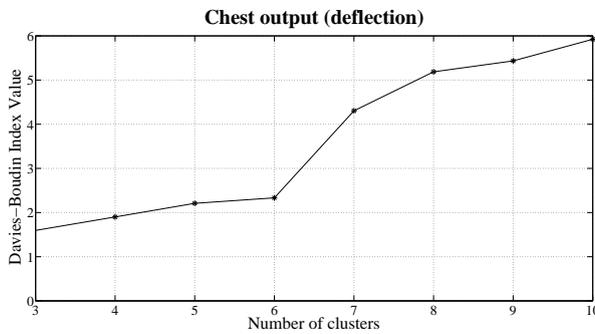
Item	Number of elements in the cluster	Neck score overall (minimal score between the three neck scores)
#1	420	4 points
#2	11	0 - 3,99 points

4.1.1.4/ DUMMY CHEST: CLUSTERS' FORMATION AND EVALUATION

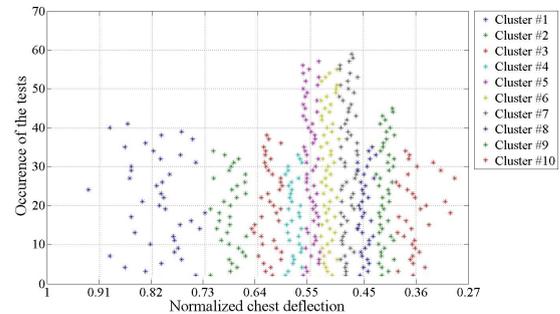
The procedure to form the clusters is the same as for the other dummy parts. For the chest, only one output is considered for forming the clusters: the chest deflection. The chest viscous criterion, which characterises the thoracic injury risk through the velocity of deformation and the compression of the chest, is uncritical for this rating. Fig. 4.3(a) shows the DBI values from 3 until 10 clusters.

According to the DBI, the best clustering should be with 3 clusters (lowest DBI value). But the chest deflection is a sensitive output for which it is often difficult to reach 4 points. Moreover the range for the chest deflection is pretty small (from 22 mm to 50 mm for the HIII 50th dummy) and a modification of a few millimetres can cause the loss of points. This extreme sensitivity is not taken into account by the DBI method, which only evaluates the separation and the compactness of the clusters.

In order to consider this particularity, the optimum number of clusters (10 clusters) has been chosen by the expert knowledge, independently of the DBI values, but according to the separation of the clusters proposed by the GK algorithm. The 10 clusters are formed by the algorithm itself without any modifications and the clusters' characteristics, especially the values' range, represent well the output's sensitivity. Fig. 4.3(b) illustrates the 10 clusters formed by the GK algorithm and Table 4.6 resumes the characteristics of each cluster. This is a very good example to justify why the Mahalanobis distance was



(a) Representation of the DBI values



(b) Clustering dummy chest (10 clusters): graphical representation

Figure 4.3: Chest combination (passenger, Europa, HIII 50th dummy)

preferred to the Euclidean distance (ellipsoidal clusters).

Table 4.6: Dummy chest, 10 clusters: characteristics (passenger, Europa, HIII 50th dummy)

Item	Number of elements in the cluster	Chest deflection
		Range of values
#1	14	[45,65 mm ; 51,07 mm]
#2	26	[40,49 mm ; 45,18 mm]
#3	35	[36,08 mm ; 39,99 mm]
#4	38	[32,67 mm ; 35,56 mm]
#5	61	[30,04 mm ; 32,38 mm]
#6	76	[27,91 mm ; 29,92 mm]
#7	56	[26,10 mm ; 27,87 mm]
#8	49	[23,96 mm ; 25,96 mm]
#9	51	[21,56 mm ; 23,79 mm]
#10	29	[16,3 mm ; 21,11 mm]
		Number of points obtained for the dummy chest (EURO NCAP rating)
#1	14	0 - 0,62 points
#2	26	0,69 - 1,36 points
#3	35	1,43 - 1,99 points
#4	38	2,06 - 2,48 points
#5	61	2,52 - 2,85 points
#6	76	2,87 - 3,16 points
#7	56	3,16 - 3,41 points
#8	49	3,43 - 3,72 points
#9	51	3,74 - 4 points
#10	29	4 points

This clustering is interesting because the chest's sensitivity is expressed by the clusters' limits. Clusters #9 and #10 can both gather sled tests which obtained 4 points but the value's range of the chest deflection is different (the minimum for the cluster #9 is 21,56 mm and the highest value for the cluster #10 is 21,11 mm). 22 mm for the chest deflection

is the limit to reach 4 points and the minimal value in the cluster #9 is 21,56 mm. It means that only sled tests which reached the 4 points narrowly are included in this cluster. In cluster #10 are present only sled tests with a safety margin of almost 1 mm for the chest deflection. The clustering into smaller clusters did not regroup together the sled tests which reached 4 points, that is why the 10 clusters proposed by the GK algorithm have been kept such as it for studying the influence of the inputs' parameters on the biomechanical values.

4.1.2/ CLUSTERING FOR THE 50TH DUMMY (DRIVER SIDE)

The same procedure has been conducted on the driver side. Due to the use of the same dummy (HIII 50th) on the driver side, the clustering procedure is similar to the passenger side. The outputs and their limits are the same as for the passenger side (Table 4.1). For these reasons, only the results are resumed in Table 4.7. The selected number of the clusters for the dummy head, neck and chest are the same as for the passenger side (because of the same reasons detailed for the passenger). Only the limits of the clusters can differ to those for the passenger side. Because of the common availability of a knee bag, a clustering has also been done for the dummy femurs. The graphical representations of the clusters and their values' repartition are indexed in Appendix D.

4.1.3/ CONCLUSION FOR THE EURO NCAP RATING CLUSTERING

For each dummy part (head, neck, chest and femurs for the driver side), the GK algorithm and the DBI method have been applied on the data in order to find similarities between the values and to evaluate it. The database contains 395 results for the driver side and 431 for the passenger side. Independently of the occupant side, several clusters have been modified taking the ground truth (based on the points reached in the rating) into account.

In fact, for the dummy head, lots of sled tests already reached the maximum of points, and a separation of these sled tests into 2 or 3 clusters does not present any advantages for the study of the parameters' influence. For the dummy neck, the proposed clustering separated sled tests which obtained the same points' number in the rating. The clusters have thus been modified according to the expert knowledge such that all the sled tests with 4 points overall are gathered in the same cluster. The GK algorithm is indeed based on the numerical values, independently of the rating score. For the chest, the GK algorithm showed its potential. Despite of the non consideration of the DBI values, the values' repartition into the clusters was very interesting and unmodified.

Table 4.7 and Table 4.8 resume the clusters on the driver and passenger sides respectively that will be considered for studying the parameters' influence. These conclusions show that the DBI method was not be adapted to this crash test data for selecting the clusters' number.

4.2/ OUTPUTS' CLUSTERING FMVSS NR.208 SLED TESTS

Because of the complexity of the U.S. legislation, a separate chapter is proposed. The FMVSS NR.208 legislation gathers four configurations for the frontal tests. Each of these

Table 4.7: Resume for the EURO NCAP clustering (driver, Europa, HIII 50th dummy)

Dummy part	Outputs	Number of clusters
Head	Acceleration (3 ms) and HIC 36	4 (expert knowledge)
Neck	Shear, tension and extension	2 (expert knowledge)
Chest	Deflection	10
Femurs	Forces (left and right)	4 (expert knowledge)

Table 4.8: Resume for the EURO NCAP clustering (passenger side)

Dummy part	Outputs	Number of clusters
Head	Acceleration (3 ms) and HIC 36	4 (expert knowledge)
Neck	Shear, tension and extension	2 (expert knowledge)
Chest	Deflection	10

configurations has to be tested on the driver and the passenger sides:

- ◆ full width, velocity of 35 mph with a Hybrid III 50th Percentile dummy, belted (US NCAP rating for the driver side);
- ◆ full width, velocity of 35 mph with a Hybrid III 5th Percentile dummy, belted (US NCAP rating for the passenger side);
- ◆ full width, velocity of 25 mph with a Hybrid III 50th Percentile dummy, unbelted;
- ◆ full width, velocity of 25 mph with a Hybrid III 5th Percentile dummy, unbelted.

In this chapter, only the passenger side is detailed. The results for the driver side are resumed in Section 4.2.6. Because of the future application of the CART algorithm (each terminal node is assigned with a class), the clusters are chosen according to their possibility to be considered as class, additionally to the legislation's requirements. It is thus important to have clusters that can be used as class.

4.2.1/ CLUSTERING FOR THE FMVSS NR.208 LEGISLATION

The approach to form clusters with FMVSS NR.208 data is the same as for the EURO NCAP rating. The GK algorithm enables the formation of the clusters (and determinesthe values' repartition into it), which are evaluated afterwards using the DBI method. It gives an indication on the best clustering but is not decisive. The number of clusters is chosen according to its DBI value but also according to the expert knowledge and legislation's requirements. It means that the customer's and legal values' limits play an important role in the choice of the clusters' number.

The biomechanical limit values for each biomechanical criterion naturally depend on the dummy type and a comparison of the parameters between sled tests with and without seat belts could produce wrong results. That is why four clustering have been produced for this legislation:

- ◆ a clustering for the Hybrid III 50th Percentile dummy belted;

- ◆ a clustering for the Hybrid III 50th Percentile dummy unbelted;
- ◆ a clustering for the Hybrid III 5th Percentile dummy belted;
- ◆ a clustering for the Hybrid III 5th Percentile dummy unbelted.

638 sled tests, including 265 with seat belt, have been selected with the Hybrid III 50th Percentile dummy as passenger (March 2015) and 714 sled tests, including 514 with seat belt, with the Hybrid III 5th Percentile dummy as passenger (April 2015). The data has been collected from 2001 until 2015.

The upper limits for the biomechanical values are referenced in Table 4.9 for the Hybrid III 50th Percentile dummy and in Table 4.10 for the Hybrid III 5th Percentile dummy. For all the cars going on the market, the biomechanical values of the occupants do not have to exceed the maximal legal values (for belted and unbelted configuration). However most of the customers prefer to have a safety margin and impose the biomechanical values under 80% of these legal biomechanical values.

Table 4.9: Legal and customer's values for the Hybrid III 50th Percentile dummy (FMVSS NR.208 legislation)

HIII 50th percentile dummy		Maximal legal values FMVSS NR.208	Customer's limits 80% of FMVSS NR.208
Head	HIC 15	700	560
Neck	Nte max	1	0,80
	Ntf max	1	0,80
	Nce max	1	0,80
	Ncf max	1	0,80
	Compression	4 kN	3,20 kN
	Tension	4,17 kN	3,34 kN
Chest	a_{3ms}	60 g	48 g
	Deflection	63 mm	50,40 mm
Femurs	Force left	10 kN	8 kN
	Force right	10 kN	8 kN

4.2.2/ CLUSTERING FOR THE HIII 50TH DUMMY BELTED (PASSENGER)

As for the EURO NCAP rating, clusters are formed for each dummy part in order to study the influence of relevant parameters on it. The applied method is the same as in Chapter 4.1. For this configuration (HIII 50th dummy belted), the sled test is not evaluated into a rating. The limits are thus the legal and the customer's limits described in Table 4.9. The database contains 265 sled tests for this configuration. All sled test results are used for creating the clusters. The considered dummy parts are the head, the neck, the chest and the femurs (presence of a knee bag possible).

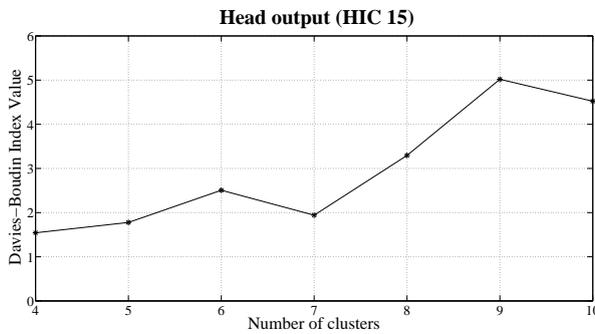
4.2.2.1/ DUMMY HEAD: CLUSTER'S FORMATION AND EVALUATION

The GK algorithm is applied on 265 head values (HIC 15 criterion). The DBI evaluates the compactness and the separation of the clusters. For the FMVSS NR.208, the legislation's

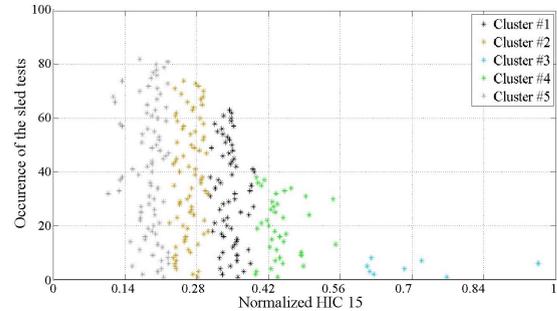
Table 4.10: Legal and customer's values for the Hybrid III 5th Percentile dummy (FMVSS NR.208 legislation)

HIII 5th percentile dummy		Maximal legal values FMVSS NR.208	Customer's limits 80% of FMVSS NR.208
Head	HIC 15	700	560
Neck	Nte max	1	0,80
	Ntf max	1	0,80
	Nce max	1	0,80
	Ncf max	1	0,80
	Compression	2,52 kN	2,02 kN
	Tension	2,62 kN	2,10 kN
Chest	a_{3ms}	60 g	48 g
	Deflection	52 mm	41,60 mm
Femurs	Force left	6,8 kN	5,44 kN
	Force right	6,8 kN	5,44 kN

requirements specify the values' limits for each output. They must be first under the maximal values imposed by the legislation (to be authorized to sell the car in the U.S.A.) and then under the customer's limits (safety margin). These limits are resumed in Table 4.9 and are considered before choosing the most appropriate clusters' number. Fig.4.4(a) shows the DBI values from 4 until 10 clusters for the head output.



(a) Representation of the DBI values



(b) Clustering (5 clusters): graphical representation

Figure 4.4: Head combination (passenger, U.S.A., HIII 50th dummy belted)

According to Fig.4.4(a), an acceptable clustering for the 265 dummy values is between 4 and 7 clusters. The clustering with 4 clusters obtained the lowest DBI value. After an analysis of all possible clustering, the values' separation into 5 clusters enables an interesting grouping of the values (Fig. 4.4(b)).

In fact, according to Table 4.11, the clusters #1, #2 and #5 regroup only values under the customer's requirements (80% of the legal values). Their separation into 3 clusters enables the engineers to evaluate their development's margin. In fact, if the head value is located in cluster #1, any modification in the sled test could produce adverse effects, i.e. the value can be above the 80% limit. The cluster #4 regroups the values comprised between 80% and 112% of the legal values. However only two head values out of 37 that are above 100% are located in this cluster, that is why this cluster is kept such as it is. All the sled tests included in cluster #3 satisfy either customer's or legal requirements. Their results have to be definitely improved.

Table 4.11: Dummy head, 5 clusters: characteristics (passenger, U.S.A., HIII 50th dummy belted)

Item	Number of elements in the cluster	HIC 15
		Range of values
#1	62	[442;563]
#2	75	[335;438]
#3	8	[875 ; 1351]
#4	37	[567;787]
#5	83	[154;325]
		Percentage of the legal values FMVSS NR.208
#1	62	63% - 80%
#2	75	48% - 63%
#3	8	125% - 193%
#4	37	81% - 112%
#5	83	22% - 46%

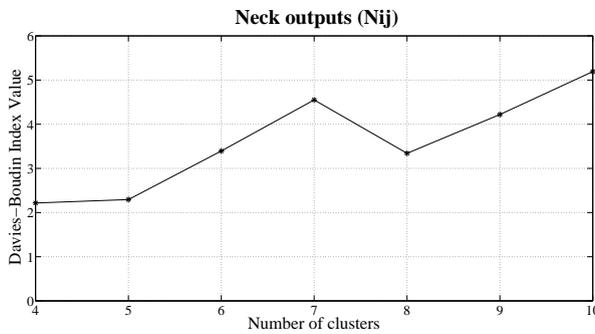
This clustering presents 5 interesting separated clusters. The limit of the legal value ($HIC_{15} = 700$, which corresponds to 100% of the legal value) and the limit imposed by the customer ($HIC_{15} = 560$, which corresponds to 80% of the legal value) are very well represented into distinct clusters. These clusters can be defined as classes for the study on the parameters' influence.

4.2.2.2/ DUMMY NECK: CLUSTER'S FORMATION AND EVALUATION

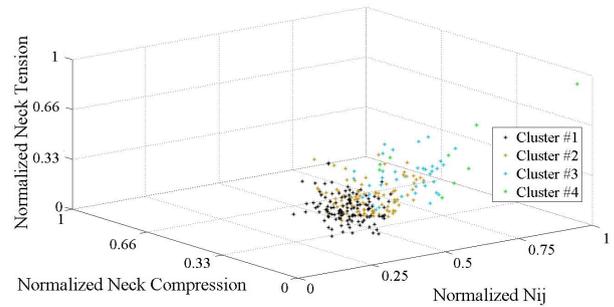
The GK algorithm is applied on the 265 neck results, composed of three outputs: criterion N_{ij} (see Section 1.1.3.3 for more details), neck compression and neck tension. The DBI evaluates the compactness and the separation of the clusters. As for the dummy head, the values must not exceed the legislation's and customer's limits (Table 4.9). These two conditions are taken into account by evaluating the pertinence of the clusters.

For a better clustering of the neck, only the criterion N_{ij} is used to form the clusters because of its predominance in neck score due to its extreme sensitivity. Fig.4.5(a) shows the DBI values from 4 until 10 clusters for the neck output N_{ij} . The values of the neck tension and compression are added after the clustering. According to Fig.4.5(a), the best clustering for the 265 values of the dummy neck is a clustering with 4 or 5 clusters. The 4 clusters are represented in Fig. 4.5(b) and the values' range per cluster is resumed in Table 4.12.

This clustering presents a well separation for the N_{ij} values. We can remark that all the values are under the customer's limits (80% of the legal values). Only 1 value out of 9 (cluster #4) is comprised between 70% and 77% for the N_{ij} . For all 264 values, the margin to customer's, especially legislation's limits, is quite enough. The neck tension and compression have been added after the clustering and have not been considered for the formation of the clusters, that is why the values of the neck compression and tension are not specially well separated. This clustering with 4 clusters can be used such as defined by the algorithm for the input classification through the decision trees method.



(a) Representation of the DBI values

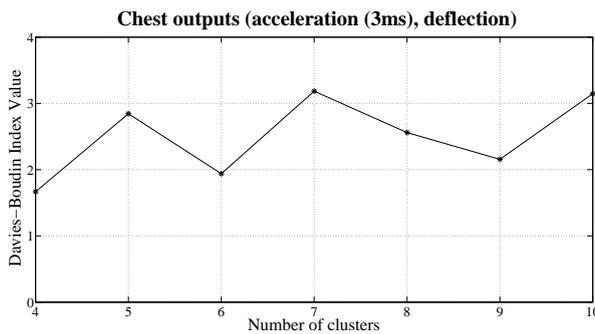


(b) Clustering (4 clusters): graphical representation

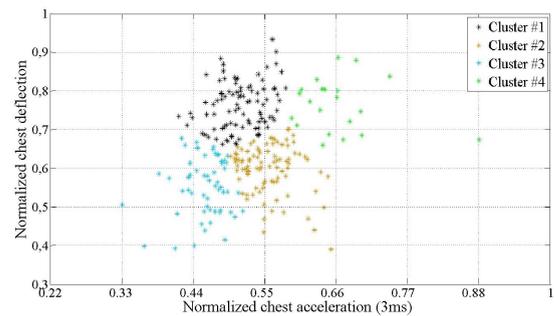
Figure 4.5: Neck combination (passenger, U.S.A., HIII 50th dummy belted)

4.2.2.3/ DUMMY CHEST: CLUSTER'S FORMATION AND EVALUATION

The GK algorithm is applied on the chest values: the acceleration (3 ms) and the deflection. Fig. 4.6(a) illustrates the DBI values from 4 until 10 clusters for the chest outputs.



(a) Representation of the DBI values



(b) Clustering (4 clusters): graphical representation

Figure 4.6: Chest combination (passenger, U.S.A., HIII 50th dummy belted)

According to Fig. 4.6(a), the best clustering for the 265 dummy values is the values' separation in 4 clusters, eventual also in 6 clusters. After analysing both possibilities, the clustering with 4 clusters presents a better separation with respect to the requirements. They are represented graphically in Fig. 4.6(b) and the values' range of each cluster is resumed in Table 4.13.

According to Table 4.13, we can first remark that all the values of the chest deflection stand under the customer's limits (80% of the legal values) and therefore under the legal values. The separation of the acceleration's values do not exactly define the customer's and legal requirements. Cluster #1 can gather values under and above this 80%-limit. But they cannot exceed 88% of the legal values and thus respond in any case to the legal requirements. Cluster #2 is particularly interesting because all these values are comprised between 75% and 99%, what corresponds approximately to the customer's and legal limits (80% and 100% respectively). Some sled tests are included in this cluster even if their value is under 48 g (80% of the legal value). But it represents the practical advantage to prevent the engineers that the margin to the customer's requirements is

Table 4.12: Dummy neck, 4 clusters: characteristics (passenger, U.S.A., HIII 50th dummy belted)

Item	Number of elements in the cluster	N_{ij}	Neck compression	Neck tension
		Range of values		
#1	121	[0,16 ; 0,26]	[31 kN ; 1086 kN]	[404 kN ; 1595 kN]
#2	95	[0,27 ; 0,35]	[30 kN ; 1250 kN]	[412 kN ; 1751 kN]
#3	40	[0,36 ; 0,47]	[59 kN ; 1243 kN]	[389 kN ; 2080 kN]
#4	9	[0,49 ; 0,77]	[53 kN ; 2191 kN]	[280 kN ; 2846 kN]
		Percentage of the legal values (FMVSS NR.208)		
#1	121	16% - 26%	1% - 27%	10% - 38%
#2	95	27% - 35%	1% - 31%	10% - 42%
#3	40	36% - 47%	1% - 31%	9% - 50%
#4	9	49% - 77%	1% - 55%	7% - 68%

very low. As for the cluster #4 which gathers sled tests with a value at 90% of the legal values. Even if some tests satisfy the legal requirements, the safety of the most of them must be completely modified to authorise the car to go on the market. Finally, the cluster #3 is the only cluster that gathers values which respond to both customer's and legal requirements.

These 4 clusters can be thus kept such as it is without any modification.

4.2.2.4/ DUMMY FEMURS: CLUSTER'S FORMATION AND EVALUATION

Because the FMVSS NR.208 regulation imposes the validation of the passive systems for four configurations (two types of dummies belted and unbelted), the kneebag is often a non negligible support to improve the femur values for the sled tests without seat belt. That is why the femur forces (left and right) are also considered.

According to Fig. 4.7(a), the clustering in 6 clusters gives the lowest DBI value. Their graphical representation in Fig. 4.7(b) and the values' repartition in Table 4.14 show a successful clustering, even if the cluster #5 contains a value at 99% (right femur) and 96% (left femur). This value is however very far away from the other 264 values and could come from an error of measurement during the test. For this reason a cluster gathering only this value does not make sense. The femurs clustering is kept such as proposed by the GK algorithm. Moreover, a linearity's relationship between the left and right femurs is visible.

4.2.3/ CLUSTERING FOR THE HIII 50TH DUMMY UNBELTED (PASSENGER)

Even if the HIII 50th dummy is unbelted, it does not change the procedure to form the clusters. That is why only the results of the clustering are resumed in this section (Table 4.15). The curves of the DBI values and the graphical representations of the clusters can be found in Appendix E.

*Table 4.15 : for a better clustering of the neck, only the criterion N_{ij} is used to form the

Table 4.13: Dummy chest, 4 clusters: characteristics (passenger, U.S.A., HIII 50th dummy belted)

Item	Number of elements in the cluster	Chest acceleration (3 ms)	Chest deflection
		Range of values	
#1	95	[38 g; 53 g]	[33 mm; 47 mm]
#2	92	[45 g; 59 g]	[19 mm; 35 mm]
#3	57	[30 g; 47 g]	[20 mm; 34 mm]
#4	21	[54 g; 80 g]	[33 mm; 44 mm]
		Percentage of the legal values (FMVSS NR.208)	
#1	95	63% - 88%	53% - 74%
#2	92	75% - 99%	31% - 56%
#3	57	50% - 78%	31% - 54%
#4	21	90% - 134%	52% - 70%

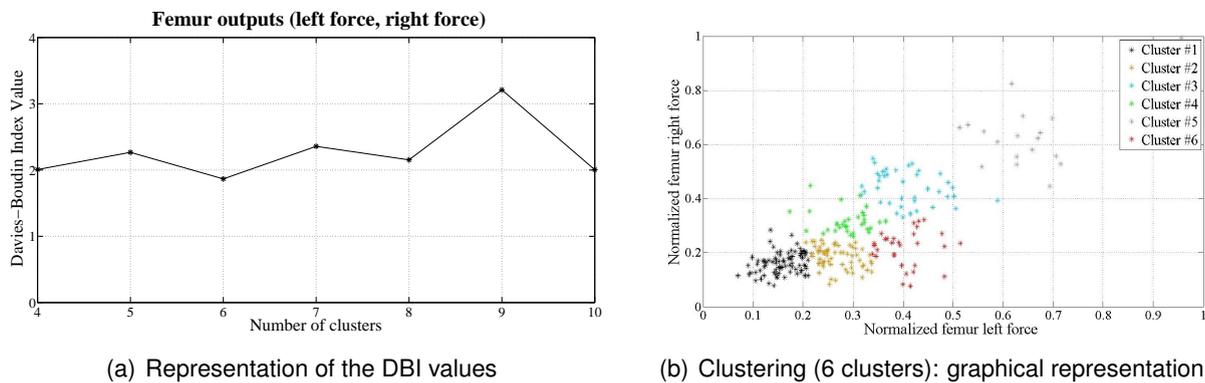


Figure 4.7: Femurs combination (passenger, U.S.A., HIII 50th dummy belted)

clusters because of its predominance in neck score due to its extreme sensitivity. The values of the neck tension and compression are added after the clustering.

For the dummy head, the cluster #1 gathers some values that satisfy the legal requirements but not their of the customer. A value above 58% of the legal values for the head is considered by the passive safety experts already as critical. That is why this cluster has not been modified. The 3 values above 100% have been gathered together by the GK algorithm in the same cluster (cluster #5).

For the dummy neck, one cluster (cluster #3) gathers sled tests that could be not accepted by the customer (values until 86% for the N_{ij}). These sled tests are legally acceptable. On the contrary, the cluster #4, and especially the cluster #5 regroup only sled tests for which the criterion N_{ij} is very critical. The uppermost of them did not satisfy the legal requirements.

For the chest, even if the best DBI value is obtained for 4 clusters, a values' separation in 6 clusters has been privileged. First of all, one cluster (cluster #1) has been created for chest accelerations above 60 g, i.e. legally incorrect, and another one (cluster #3) for which the value of the chest deflection is legally not acceptable (above 63 mm).

Table 4.14: Dummy femurs, 6 clusters: characteristics (passenger, U.S.A., HIII 50th dummy belted)

Item	Number of elements in the cluster	Femurs forces	
		Left	Right
		Range of values	
#1	83	[699 kN - 2102 kN]	[779 kN - 2833 kN]
#2	61	[2061 kN - 3378 kN]	[828 kN - 2473 kN]
#3	39	[3172 kN - 5888 kN]	[3306 kN - 5495 kN]
#4	33	[1731 kN - 3652 kN]	[2600 kN - 4473 kN]
#5	18	[5132 kN - 9565 kN]	[4461 kN - 9949 kN]
#6	31	[3397 kN - 5146 kN]	[774 kN - 3209 kN]
		Percentage of the legal values (FMVSS NR.208)	
#1	83	7% - 21%	8% - 28%
#2	61	21% - 34%	8% - 25%
#3	39	32% - 59%	33% - 55%
#4	33	17% - 37%	26% - 45%
#5	18	51% - 96%	45% - 99%
#6	31	34% - 51%	8% - 32%

The cluster #2 gathers only sled tests that satisfy the legal requirements for the chest acceleration but not their of the customer. Some tests of the cluster #5 present the same results. For only 14 tests out of 61 that are above 100%, it is not necessary to modify the proposed clustering.

Finally, the GK algorithm works also well with the femurs forces. Indeed, the cluster #3 contains sled tests which did not satisfy either the customer or the legal requirements for both femurs. In the cluster #5, only 5 values for the right femur have a value above 80% (but under 90%, what represents a certain safety margin to the legal value). All other clusters contain only sled tests which successful satisfy both demands.

To sum up, the GK algorithm proposed interesting clusters for this configuration even if some of them could be eventually improved. For the chest, a clusters' number near to the lowest DBI value has been preferred (6 clusters instead 4).

4.2.4/ CLUSTERING FOR THE HIII 5TH DUMMY UNBELTED (PASSENGER)

The procedure for the clustering with a HIII 5th dummy unbelted is also the same as for the HIII 50th dummy. Only the limits of the biomechanical values differ due to the use of the HIII 5th dummy. They are referenced in Table 4.10. Only the results of the clustering are resumed in this section (Table 4.16). The curves of the DBI values and the graphical representations of the clusters can be found in Appendix E.

*Table 4.16: for a better clustering of the neck, only the criterion N_{ij} is used to form the clusters because of its predominance in neck score due to its extreme sensitivity. The values of the neck tension and compression are added after the clustering.

The values of the dummy head have been separated in 5 clusters, even if the lowest DBI

value is obtained for 4 clusters. All the values except one are under the customer's limit: the maximal value is 630, what corresponds to 90% of the legal value. The clustering in 5 clusters allows the formation of one cluster that gathers this high value. Because of the general low values of the database (the maximal values do not exceed 469), it has been preferred to gather the values into 5 clusters to separate the high value to the others. All the values included in the database are under the legal limit of 700.

The values of the dummy neck have been gathered in 6 clusters, according to the N_{ij} because of the predominance of this criterion in the neck evaluation. The GK algorithm has been particularly powerful because both customer's and legal limits have been represented. In fact, the clusters #1 to #3 only regroup the N_{ij} values that respond to the customer's requirements, i.e. that are under the 80%-limit. The range of each cluster is approximately 20%. The cluster #4 regroups only values that do not satisfy the customer's but the legal requirements, i.e. all values comprised between 80% and 100% of the legal values. The 14 remaining sled tests have been gathered into 2 different clusters. The first one, the cluster #5 regroups values that are until 123% of the legal values while the cluster #6 regroups only values that are widely above the legal requirements. These last two clusters could be combined together but it was preferred to keep the clustering with the lowest DBI value. This clustering gives also an indication to the engineers about the improvements of the restraint systems that have to be achieved. The number of clusters has been also chosen according to the DBI values for the chest outputs. First of all, we can remark that the deflection values do not exceed 44% of the legal values, what satisfies widely the customer's requirements. Clusters #1, #3, #4, #5, #7 and #8 fan out almost equitably out 162 of 199 values. All these values satisfy the customer's limits for both outputs (acceleration and deflection). The cluster #2 contains acceleration values between 69% and 98%. It means that some values that satisfy the customer's requirements are regrouped together with values that do not satisfy it. Because only 25 values are under the 80%-limit, this cluster is kept such as it. Finally, the cluster #6 is very similar as the cluster #2 but differs with the values' range of the chest deflection.

As for the head, the clustering in 5 clusters for the femurs outputs was preferred to 4 clusters, despite the DBI value. It help at forming a cluster, the cluster #3 that only regroups values above the customer's limit for the left femur and above the legal requirements for the right femur. The cluster #1 also contains values that are above the customer's limit but all these values for left and right femurs stand under the legal values (maximal value : right femur at 6460 N). The remaining 146 values are apportioned between the clusters #1, #2 and #4.

To sum up, the clustering proposed by the GK algorithm does not have been modified. For the dummy head and femurs, a number of clusters different of the lowest DBI value has been chosen in order to have a cluster that regroups the values above the customer's limits (80% of the legal values).

4.2.5/ CLUSTERING FOR THE HIII 5TH DUMMY BELTED (US NCAP RATING, PASSENGER)

The test configuration with a HIII 5th dummy belted on the passenger side is also used for the US NCAP rating according to the FMVSS NR.208 legislation. It means that the sled test does not have to exceed the customer's and legal values (Table 4.10) but additionally the

dummy biomechanical values must be as low as possible to improve the number of stars (see Section 1.1.2.2). That is why it is important in this case to produce many clusters for each output's combination. For this reason, the DBI values have been estimated from 6 to 15 clusters. The procedure for the clustering performed on 169 sled tests stays unchanged that is why only the results are presented in this section (Table 4.17). The curves of the DBI values and the graphical representations of the clusters can be found in Appendix E.

*Table 4.17: for a better clustering of the neck, only the criterion N_{ij} is used to form the clusters because of its predominance in neck score due to its extreme sensitivity. The values of the neck tension and compression are added after the clustering.

**Table 4.17 : only the chest deflection is considered for the calculation of the US NCAP rating score. However, as the chest deflection, the acceleration values have to satisfy the customer's and legal limits (80% and 100% respectively). That is why the clusters have been first created by taking only the values of the chest deflection into consideration and then their of the acceleration have directly been incorporated in the clusters.

The values of the dummy head have been separated in 7 clusters by the GK algorithm, according to its DBI value. The cluster #3 and #4 contain sled tests which mainly do not satisfy the legal requirements. Only 15 sled tests included in cluster #4 respond to the legal conditions. The other 5 clusters contain sled tests which respond to both customer's and legal condition, even if the cluster #1 contains 2 values above this 80%-limit (but under 84%, what stays acceptable). A clustering into more clusters did not gather the values above the 80%-limit in the same group.

The procedure for the clustering of the neck values is the same as for the other FMVSS NR.208 configurations. It means that the clusters have been formed with only the criterion N_{ij} due to its influence on the neck score results. 6 clusters have been formed. The cluster #6 does satisfy either the customer's or the legal requirements. The cluster #5 could be eventual improved because it gathers values between 69% and 102%. It means that this cluster regroups sled tests that respond to both conditions, sled tests that respond to the customer's requirements but also sled tests that do not satisfy both conditions (maximal value at 102% of the legal value). Sohr [49] prescribes that for reaching the 5 stars on the passenger side, the values for the criterion N_{ij} should be under 0,3, i.e. under 30% of the legal value. The engineers are thus more focused on values under 50% of the legal values, what is confirmed here: 433 values out of 514 stand under 50%. Thus it is not essential to separate the cluster #5 into two distinct groups. The clusters #1 to #4 gather the remaining values, from 19% to 67% with the mainly results contained in the clusters #1 to #3. The proposed clustering responds well to the exigences of this critical load criterion.

Concerning the chest outputs, the clustering has been done only with the values of the chest deflection. As explained in Section 1.1.2.2, the chest acceleration is not taken into account for the rating calculation. However, this criterion must respond to the customer's and legal requirements. That is why for the chest, only the value of the chest deflection has to be improved. The lowest DBI value is obtained for 9 clusters. First of all, we can remark that all the values for the chest deflection are under the 80%-limit. The values are separated in 9 clusters and the maximal range for a cluster is 4% of the legal value, what corresponds to 2,1 mm. This restricted range allows the engineers to evaluate the modification of the rating score by changing an input parameter more precisely. The

values of the chest acceleration have been added after the clustering, that is why these values are not well separated. In any case, only 44% of the values are under the limit of 80% and 52% of the values are comprised between 80 and 100% of the legal values. The rating is one of the challenges for the engineers and for this reason, the clustering on the chest deflection in 9 clusters is privileged and kept such as it is. Finally, the GK algorithm provided an interesting clustering of the right and left femur values into 6 clusters. Only a value for the left femur has been measured above 5,44 kN, that is why an additional cluster for this value is not necessary.

To sum up, the clusters have been formed according to the lowest DBI value for each dummy part. Because of their influence on the rating in comparison with other dummy part criteria, the clustering for the neck and for the chest have been created with only one output, the criterion N_{ij} and the deflection respectively. The choice of a higher number of clusters (6 to 15 clusters) in comparison with the other FMVSS NR.208 configurations enables to have clusters with a limited range of values and thus have a better precision for the rating score.

4.2.6/ CLUSTERING ON THE DRIVER SIDE

The method is similar for the driver side: the GK algorithm allows the formation of the clusters that are evaluated through the DBI method. Only the results of the clustering are thus presented in this section. Moreover, we will see in Chapter 5 that not enough sled tests on the driver side have been conducted to enable the study of the input parameters' influence.

769 sled tests, including 438 with seat belt, have been selected with the Hybrid III 50th Percentile dummy as driver (March 2015) and 475 sled tests, including 282 with seat belt, with the Hybrid III 5th Percentile dummy as driver (April 2015). The data has been collected from 2001 until 2015.

The upper limits for the biomechanical values are the same as the passenger and are referenced in Table 4.9 for the Hybrid III 50th Percentile dummy and in Table 4.10 for the Hybrid III 5th Percentile dummy.

The DBI curves and graphical representations of the clusters can be found in Appendix F. As for the passenger side, only the sensible criterion N_{ij} has been taken into account for a better clustering of the neck outputs (N_{ij} , tension and compression). The values of the neck tension and compression have been added after the clustering and must also respond to the customer's and legal conditions.

However, the US NCAP rating imposes the Hybrid III 50th Percentile belted dummy on the driver side. Therefore the clustering has to be performed with a large number of clusters (improvement of the biomechanical values). It means that for the chest, the clustering is based on the values of the deflection because only this criterion is considered in the rating score. However the chest acceleration must respond to the customer's and legal requirements.

HIII 5th dummy, belted and unbelted

Table 4.18 resumes the clustering for the HIII 5th dummy belted and Table 4.19 for the HIII 5th dummy unbelted. For both clustering, the DBI values have been calculated from 4 to 10 clusters. For the dummy head and chest (HIII 5th dummy belted), it has been

decided to form an additional cluster in order to separate better the values.

*Table 4.18 and Table 4.19: for a better clustering of the neck, only the criterion N_{ij} is used to form the clusters because of its predominance in neck score due to its extreme sensitivity. The values of the neck tension and compression are added after the clustering.

HIII 50th dummy, belted and unbelted

Table 4.20 resumes the clustering for the HIII 50th dummy belted. For the driver side, it corresponds to the rating. It means that the number of clusters has to be higher as for the legislation in order to improve the rating score. That is why the DBI values have been calculated from 6 to 15 clusters.

Table 4.21 resumes the clustering for the HIII 50th unbelted dummy. The DBI values have been calculated from 4 to 10 clusters.

*Table 4.20 and Table 4.21: for a better clustering of the neck, only the criterion N_{ij} is used to form the clusters because of its predominance in neck score due to its extreme sensitivity. The values of the neck tension and compression are added after the clustering.

4.3/ OVERVIEW OF THE CLUSTERING RESULTS

The GK algorithm and the DBI method have been applied on data coming from EURO NCAP rating and FMVSS NR.208 sled tests (including US NCAP rating), respectively to form the clusters and to evaluate its compactness and separation. In some cases the clusters needed to be adapted in order to take the expert knowledge (rating requirements, sensibility of some outputs) into account. Table 4.22 resumes the cases where the clusters have been kept such as given by the GK algorithm or not.

As we can see in Table 4.22, the GK algorithm is not well adapted for the collected EURO NCAP data. Only one dummy is evaluated for both sides, the HIII 50th dummy, that is a significant advantage in comparison with the FMVSS NR.208 legislation. As explained in Section 1.1.2.1, each dummy part can receive until 4 points and the rating score is the sum of each dummy region's score. For the dummy head, neck and femurs (only on the driver side), more than 80% of the sled tests already became 4 points, that is why a clustering based on the expert knowledge and rating results is preferred to them proposed by the GK algorithm. However, because of the difficulty for the engineers to reach 4 points for the chest, the panel of the available chest values is more extended and the clusters formed by the GK algorithm (by choosing the clusters' number independently of the lowest DBI value) are already well separated and usable. Many clusters' numbers have been tested and the number of 10 clusters proposed the best separation for both occupant sides.

Concerning the results obtained for the U.S. legislation, as described in Section 1.1.2.2, the restraint systems of a vehicle have to work with two types of dummies (HIII 50th and HIII 5th dummies), with and without seat belt. For these four configurations, the biomechanical values have to be under the limits defined by the legislation but also under the customer's requirements, that impose that the biomechanical values do not have to excess 80% of the legal values (margin safety). The configuration HIII 50th dummy on the driver side and HIII 5th dummy on the passenger side is the official configuration for

the U.S. rating, that is why, the dummy biomechanical values for this configuration have to be additionally to the requirements, as low as possible to reach the best possible score in the US NCAP rating. The score calculation for this rating is also a little bit more complex as for the EURO NCAP rating (estimation of the probability for each dummy part). For these reasons, the DBI values have been calculated from 4 to 10 clusters for the legislation and from 6 to 15 clusters for the rating. Because of the influence of the criterion N_{ij} in the neck score calculation and of the chest deflection in the chest score calculation (only for the rating), the clusters have been first formed with only these values. Finally, according to Table 4.22, the GK algorithm works very well for the FMVSS NR.208 legislation and rating, presumably because of a larger scattering of the data. The DBI values have not been taken into account for some cases, but once the clusters' number has been defined, the GK proved its potential (the proposed clusters have not been modified).

To sum up, the GK algorithm is useful for data with a high scattering, as for the U.S. database. Due to the good and similar results of the EURO NCAP sled tests included in the database (restricted range of values), it is then more complicated for the algorithm to form its clusters. The clusters for this rating had to be often reformed taking the obtained score into account.

The Variational Bayesian methods could be also an alternative to the clustering and would enable add some conditions (or *a priori*), such as the limits of the biomechanical values which depend on the legislation. The publications of Sutton-Charani et al. [69] [74] on decision trees with uncertain data could be very useful to develop new algorithms with the actual database.

Table 4.15: Clustering for each dummy part (passenger, U.S.A., HIII 50th dummy unbelted)

Dummy part	Outputs	Number of clusters	Elements in the clusters	Percentage of the legal values (FMVSS NR.208)
Head	HIC 15	5	#1 (25 tests) #2 (144 tests) #3 (133 tests) #4 (63 tests) #5 (3 tests)	58% - 92% 6% - 23% 23% - 38% 39% - 57% 130% - 135%
Neck*	N_{ij} , tension and compression	5	#1 (124 tests) #2 (137 tests) #3 (81 tests) #4 (30 tests) #5 (1 test)	13% - 37%, 3% - 49% and 3% - 52% 38% - 55%, 5% - 60% and 5% - 66% 56% - 86%, 12% - 94% and 3% - 46% 89% - 145%, 12% - 133% and 3% - 66% 214% , 87% and 42%
Chest	Acceleration (3ms) and deflection	6 (lowest DBI: 4)	#1 (1 test) #2 (98 tests) #3 (1 test) #4 (120 tests) #5 (61 tests) #6 (92 tests)	242% and 28% 48% - 95% and 26% - 63% 58% and 130% 61% - 85% and 4% - 25% 83% - 110% and 12% - 33% 38% - 66% and 10% - 30%
Femurs	Forces (left and right)	6	#1 (87 tests) #2 (93 tests) #3 (30 tests) #4 (99 tests) #5 (55 tests) #6 (4 tests)	21% - 44% and 19% - 50% 32% - 58% and 51% - 77% 75% - 113% and 72% - 111% 38% - 67% and 35% - 53% 54% - 79% and 52% - 90% 2% - 6% and 2% - 20%

Table 4.16: Clustering for each dummy part (passenger, U.S.A., HIII 5th dummy unbelted)

Dummy part	Outputs	Number of clusters	Elements in the clusters	Percentage of the legal values (FMVSS NR.208)
Head	HIC 15	5 (lowest DBI: 4)	#1 (74 tests) #2 (332 tests) #3 (53 tests) #4 (9 tests) #5 (1 test)	2% - 16% 16% - 29% 29% - 47% 50% - 67% 90%
Neck*	N_{ij} , tension and compression	6*	#1 (36 tests) #2 (39 tests) #3 (58 tests) #4 (53 tests) #5 (10 tests) #6 (4 tests)	19% - 46%, 1% - 48% and 10% - 53% 48% - 63%, 8% - 47% and 8% - 44% 64% - 79%, 9% - 125% and 6% - 68% 80% - 99%, 28% - 82% and 8% - 86% 101% - 123%, 7% - 85% and 8% - 119% 165% - 219%, 9% - 286% and 13% - 269%
Chest	Acceleration (3 ms) and deflection	8	#1 (2 tests) #2 (29 tests) #3 (47 tests) #4 (15 tests) #5 (18 tests) #6 (8 tests) #7 (34 tests) #8 (46 tests)	27% - 37% and 31% - 39% 69% - 98% and 19% - 33% 48% - 61% and 15% - 28% 54% - 74% and 27% - 44% 65% - 79% and 0% - 12% 78% - 92% and 4% - 17% 42% - 60% and 1% - 18% 59% - 74% and 12% - 25%
Femurs	Forces (left and right)	5 (lowest DBI: 4)	#1 (52 tests) #2 (39 tests) #3 (2 tests) #4 (59 tests) #5 (48 tests)	47% - 88% and 62% - 95% 32% - 55% and 47% - 71% 95% - 101% and 99% - 126% 1% - 52% and 2% - 56% 50% - 80% and 44% - 62%

Table 4.17: Clustering for each dummy part (passenger, U.S.A., HIII 5th dummy belted, US NCAP rating)

Dummy part	Outputs	Number of clusters	Elements in the clusters	Percentage of the legal values (FMVSS NR.208)
Head	HIC 15	7	#1 (54 tests) #2 (81 tests) #3 (7 tests) #4 (27 tests) #5 (83 tests) #6 (142 tests) #7 (120 tests)	61% - 84% 47% - 61% 125% - 172% 87% - 120% 11% - 25% 25% - 36% 36% - 46%
Neck	N_{ij} , tension and compression	6*	#1 (151 tests) #2 (169 tests) #3 (113 tests) #4 (61 tests) #5 (17 tests) #6 (3 tests)	19%- 35%, 1%- 31% and 13%- 68% 36%- 43%, 1%- 37% and 15%- 77% 44%- 52%, 2%- 43% and 14%- 54% 53%- 67%, 3%- 50% and 15%- 78% 69%- 102%, 2%- 65% and 10%- 61% 123%- 154%, 22%- 71% and 19%- 57%
Chest	Acceleration (3 ms) and deflection	9**	#1 (50 tests) #2 (76 tests) #3 (62 tests) #4 (60 tests) #5 (74 tests) #6 (85 tests) #7 (61 tests) #8 (38 tests) #9 (8 tests)	48% - 100% and 26% - 33% 57% - 120% and 34% - 38% 61% - 101% and 38% - 43% 56% - 112% and 43% - 48% 55% - 118% and 48% - 52% 58% - 109% and 52% - 56% 61% - 106% and 57% - 61% 73% - 109% and 62% - 69% 74% - 103% and 72% - 80%
Femurs	Forces (left and right)	6	#1 (64 tests) #2 (16 tests) #3 (43 tests) #4 (69 tests) #5 (156 tests) #6 (166 tests)	35% - 67% and 11% - 44% 46% - 93% and 46% - 73% 20% - 44% and 36% - 66% 25% - 48% and 8% - 26% 17% - 35% and 21% - 36% 0% - 24% and 7% - 32%

Table 4.18: Clustering for each dummy part (driver, U.S.A., HIII 5th dummy belted)

Dummy part	Outputs	Number of clusters	Elements in the clusters	Percentage of the legal values (FMVSS NR.208)
Head	HIC 15	5 (lowest DBI: 4)	#1 (23 tests) #2 (46 tests) #3 (96 tests) #4 (3 tests) #5 (114 tests)	55% - 88% 37% - 53% 10% - 24% 109% - 157% 24% - 36%
Neck	N_{ij} , tension and compression*	4	#1 (130 tests) #2 (99 tests) #3 (50 tests) #4 (3 tests)	23%- 45%, 2%- 27% and 28%- 65% 45%- 65%, 2%- 39% and 34 %- 78% 66%- 105%, 2%- 32% and 43%- 90% 128%- 167%, 8%- 28% and 72%- 96%
Chest	Acceleration (3 ms) and deflection	5 (lowest DBI: 4)	#1 (83 tests) #2 (26 tests) #3 (42 tests) #4 (48 tests) #5 (83 tests)	83% - 103% and 36% - 57% 97% - 131% and 52% - 76% 70% - 94% and 58% - 75% 59% - 82% and 35% - 50% 60% - 88% and 49% - 63%
Femurs	Forces (left and right)	6	#1 (18 tests) #2 (28 tests) #3 (24 tests) #4 (35 tests) #5 (102 tests) #6 (75 tests)	12% - 33% and 16% - 26% 2% - 16% and 10% - 22% 29% - 52% and 1% - 16% 12% - 30% and 8% - 15% 2% - 13% and 1% - 10% 14% - 33% and 1% - 8%

Table 4.19: Clustering for each dummy part (driver, U.S.A., HIII 5th dummy unbelted)

Dummy part	Outputs	Number of clusters	Elements in the clusters	Percentage of the legal values (FMVSS NR.208)
Head	HIC 15	4	#1 (53 tests) #2 (14 tests) #3 (35 tests) #4 (91 tests)	11% - 18% 29% - 46% 19% - 28% 3% - 11%
Neck	N_{ij} , tension and compression*	4	#1 (99 tests) #2 (52 tests) #3 (26 tests) #4 (16 tests)	12%- 37%, 1%- 34% and 12%- 58% 38%- 55%, 2%- 62% and 21%- 58% 60%- 85%, 1%- 78% and 31%- 86% 87%- 142%, 4%- 122% and 30%- 75%
Chest	Acceleration (3 ms) and deflection	7	#1 (16 tests) #2 (47 tests) #3 (30 tests) #4 (7 tests) #5 (32 tests) #6 (50 tests) #7 (11 tests)	47% - 68% and 51% - 90% 52% - 63% and 16% - 48% 69% - 83% and 53% - 83% 89% - 110% and 62% - 93% 62% - 83% and 24% - 51% 28% - 52% and 11% - 37% 66% - 83% and 91% - 117%
Femurs	Forces (left and right)	4	#1 (66 tests) #2 (34 tests) #3 (88 tests) #4 (5 tests)	19% - 69% and 36% - 87% 60% - 100% and 49% - 86% 6% - 56% and 7% - 53% 90% - 125% and 101% - 120%

Table 4.20: Clustering for each dummy part (driver, U.S.A., HIII 50th dummy belted, US NCAP rating)

Dummy part	Outputs	Number of clusters	Elements in the clusters	Percentage of the legal values (FMVSS NR.208)
Head	HIC 15	7	#1 (92 tests) #2 (4 tests) #3 (95 tests) #4 (75 tests) #5 (11 tests) #6 (110 tests) #7 (51 tests)	42% - 55% 130% - 159% 13% - 31% 55% - 71% 96% - 124% 31% - 42% 72% - 94%
Neck	N_{ij} , tension and compression*	6	#1 (59 tests) #2 (150 tests) #3 (135 tests) #4 (67 tests) #5 (19 tests) #6 (8 tests)	15%- 26%, 1%- 19% and 13%- 41% 27%- 32%, 0%- 35% and 12%- 41% 33%- 38%, 1%- 49% and 14%- 50% 38%- 47%, 1%- 30% and 16%- 66% 50%- 65%, 1%- 33% and 46%- 93% 77%- 86%, 2%- 63% and 18%- 74%
Chest	Acceleration (3 ms) and deflection	12	#1 (54 tests) #2 (3 tests) #3 (58 tests) #4 (1 test) #5 (56 tests) #6 (39 tests) #7 (29 tests) #8 (32 tests) #9 (28 tests) #10 (37 tests) #11 (56 tests) #12 (13 tests)	67% - 82% and 27% - 42% 86% - 100% and 56% - 67% 86% - 97% and 44% - 54% 140% and 96% 76% - 86% and 40% - 51% 47% - 75% and 36% - 47% 68% - 80% and 57% - 72% 79% - 87% and 62% - 78% 82% - 95% and 28% - 44% 67% - 78% and 45% - 57% 77% - 87% and 51% - 61% 99% - 122% and 36% - 64%
Femurs	Forces (left and right)	6	#1 (49test) #2 (230 tests) #3 (23 tests) #4 (136 tests) #5 (38 tests) #6 (2 tests)	43% - 68% and 51% - 82% 7% - 23% and 8% - 31% 14% - 30% and 35% - 67% 18% - 37% and 10% - 38% 32% - 55% and 21% - 60% 11% - 17% and 109% - 137%

Table 4.21: Clustering for each dummy part (driver, U.S.A., HIII 50th dummy unbelted)

Dummy part	Outputs	Number of clusters	Elements in the clusters	Percentage of the legal values (FMVSS NR.208)
Head	HIC 15	4	#1 (15 tests) #2 (110 tests) #3 (147 tests) #3 (59 tests)	44% - 80% 16% - 27% 5% - 16% 27% - 41%
Neck	N_{ij} , tension and compression*	4	#1 (153 tests) #2 (112 tests) #3 (47 tests) #4 (19 tests)	17%- 34%, 1%- 47% and 4%- 49% 35%- 49%, 1%- 60% and 5%- 67% 50%- 73%, 2%- 107% and 2%- 81% 76%- 117%, 7%- 130% and 2%- 83%
Chest	Acceleration (3 ms) and deflection	5 (lowest DBI: 4)	#1 (127 tests) #2 (93 tests) #3 (66 tests) #4 (38 tests) #5 (7 tests)	60% - 85% and 15% - 53% 62% - 87% and 53% - 91% 36% - 59% and 9% - 56% 78% - 99% and 25% - 73% 57% - 113% and 49% - 121%
Femurs	Forces (left and right)	8 (lowest DBI: 4)	#1 (69 tests) #2 (83 tests) #3 (12 tests) #4 (36 tests) #5 (44 tests) #6 (38 tests) #7 (22 tests) #8 (27 tests)	33% - 54% and 23% - 43% 7% - 38% and 17% - 41% 39% - 79% and 81% - 102% 51% - 74% and 28% - 55% 26% - 51% and 44% - 70% 49% - 71% and 56% - 77% 84% - 110% and 87% - 113% 78% - 102% and 61% - 88%

Table 4.22: Resume of the use of the clusters given by the GK algorithm according to the dummy part and sled test configuration

Legislation	Configuration	Dummy part	Lowest DBI value taken into account	Clusters given by the GK algorithm taken as proposed
Euro NCAP rating	Driver HIII 50%	Head Neck Chest Femur	× × × ×	× × ✓ ×
	Passenger HIII 50%	Head Neck Chest Femur	× × × —	× × ✓ —
FMVSS Nr.208 legislation	Driver HIII 5% with seat belt	Head Neck Chest Femur	× ✓ (just N_{ij}) × ✓	✓ ✓ (based on N_{ij} clusters) ✓ ✓
	Driver HIII 50% with seat belt (US NCAP rating)	Head Neck Chest Femur	✓ ✓ (just N_{ij}) ✓ ✓	✓ ✓ (based on N_{ij} clusters) ✓ ✓
	Driver HIII 5% without seat belt	Head Neck Chest Femur	✓ ✓ (just N_{ij}) ✓ ✓	✓ ✓ (based on N_{ij} clusters) ✓ ✓
	Driver HIII 50% without seat belt	Head Neck Chest Femur	✓ ✓ (just N_{ij}) × ×	✓ ✓ (based on N_{ij} clusters) ✓ ✓
	Passenger HIII 50% with seat belt	Head Neck Chest Femur	× ✓ (just N_{ij}) ✓ ✓	✓ ✓ (based on N_{ij} clusters) ✓ ✓
	Passenger HIII 5% with seat belt (US NCAP rating)	Head Neck Chest Femur	✓ ✓ (just N_{ij}) ✓ ✓	✓ ✓ (based on N_{ij} clusters) ✓ ✓
	Passenger HIII 50% without seat belt	Head Neck Chest Femur	✓ ✓ (just N_{ij}) × ✓	✓ ✓ (based on N_{ij} clusters) ✓ ✓
	Passenger HIII 5% without seat belt	Head Neck Chest Femur	× ✓ (just N_{ij}) ✓ ×	✓ ✓ (based on N_{ij} clusters) ✓ ✓

PARAMETERS' INFLUENCE THROUGH DECISION TREES

The study of the parameters' influence involves the output's clustering detailed in Chapter 4.1 for the EURO NCAP rating and in Chapter 4.2 for the FMVSS NR.208 legislation and US NCAP rating. The relationships between the input parameters and the clustered dummy biomechanical values are modeled with the help of the CART algorithm. As presented in Chapter 2.2 Section 2.2.3.2, the classification and regression trees are supervised methods that are widely employed in many scientific domains. The fact that the target of the tree is here the clusters, i.e. categorical variables, imposes to build a classification tree.

This chapter aims at presenting a method that enables to understand the relationships between the input parameters and the dummy biomechanical values. Because of a reduced number of available data for the driver side, the study is focused on the passenger side, especially on the EURO NCAP and US NCAP ratings, in which the engineers are mostly interested.

5.1/ DATA DESCRIPTION AND METHODOLOGY

The crash database contains the same data and information as the database created for the k-NN classification, detailed in Chapter 3.1 Section 3.1.1, especially through Table 3.1 that references the selected input parameters and Table 3.2 that sums up the number of available sled tests for each configuration.

For each sled test, 46 attributes describe the restraint systems, the position of the passenger in the vehicle and the car characteristics. The step "pre-processing" of the KDD process is afresh applied on these 46 attributes in order to select those used for building the trees. In fact, even if the database does not contain missing data anymore, some attributes are not necessarily defined for all sled tests. For example, some sled tests do not have a buckle pretensioner and thus this parameter is declared as undefined and cannot be supported by the CART algorithm. By deleting these variables, the new database includes 27 parameters used as the predictors variables (Table 5.1). The data also includes information about the injury severity level for each dummy part by means of the clusters representing the target variable (Table 5.2). Table 5.3 presents the color code that is here used for describing the clusters.

Table 5.1: Data description

Variables	Symbol	Type	Description
Clusters	LEVEL	Qualitative	Target variable
Type of airbag	PAB	Qualitative	1:L-shape; 2:Square-shape; 3: 3D Add-On; 4:3D-shape; 5:Other
Divided panel	DP	Qualitative	0:No;1:Yes
Volume	VOL	Continuous	Volume of the airbag
Number of holes	Num-VH	Qualitative	0:unknown or zero; 1:one; 2:two
Hole diameter 1	H-diam1	Continuous	Diameter of the vent hole
Active bag vent	ABV	Qualitative	0:No; 1:Yes
Tether 1	Tether-1	Qualitative	0:No; 1:Yes
Tether 2	Tether-2	Qualitative	0:No; 1:Yes
Number of stages	Num TTF	Qualitative	0:unknown or zero;1:one;2:two
TTF S1	TTF-S1	Continuous	Time of trigger of the airbag in ms
TTF S2	TTF-S2	Continuous	Time of trigger of the airbag in ms
Type of load limiter	LL	Qualitative	1:constant; 2:degressive; 3:regressive; 4:switchable
Load limiter value 1	LL-V1	Continuous	Value of the load in kN
Retractor pretensioner	RP	Qualitative	0:No; 1:Yes
Buckle pretensioner	BP	Qualitative	0:No; 1:Yes
Anchor point pretensioner	AP	Qualitative	0:No; 1:Yes
Webbing on spool	WB	Continuous	Value webbing on spool in mm
Height adjustment	H-Adj	Qualitative	0:No; 1:Yes
Dynamic locking tongue	DLT	Qualitative	0:No; 1:Yes
Knee bag	KnAB	Qualitative	0:No; 1:Yes
Chin to IP	P1	Continuous	Distance between the chin and the IP
Sternum to IP	P2	Continuous	Distance between the sternum and the IP
Left knee shorter distance ahead	P3	Continuous	Distance between the left knee and the IP
Right knee shorter distance ahead	P4	Continuous	Distance between the right knee and the IP
OLC/OLC++	OLC/OLC++	Continuous	Characteristic of the hardness of the pulse
Time velocity at 0 m/s	Tv0	Continuous	Value of time when velocity = 0 m/s
Dummy temperature	T	Continuous	Temperature of the dummy

Table 5.2: Target variable: reminder of the clusters' number

Rating	Number of clusters per dummy part			
	Head	Neck	Chest	Femur
Euro NCAP	4	2	10	—
US NCAP	7	6	9	6

Table 5.3: Target definition (color code)

Percentage of the legal values									
≤ 80%					[80 ; 100%]			≥ 100%	
No-injury	Injury	Fatal-injury	...

5.2/ EURO NCAP RATING: ANALYSIS OF THE DUMMY CHEST

Analysis of the clusters' material

According to the database, the dummy head and neck are uncritical dummy parts (see Chapter 4.1). The dummy chest is the most sensitive part and for this reason, the engineers are generally only focused on this part. In this section, the case of the chest is discussed and the decision trees for the other dummy parts can be found in Appendix G. Table 5.4 presents the distribution of the input parameters into the 10 clusters. It gives an indication about the parameters that are more present in a specific cluster.

Table 5.4: Sled tests' repartition for the dummy chest (passenger, EURO NCAP)

Clusters / Variables	Cluster #1 45.65-51.07 mm	Cluster #2 40.49-45.18 mm	Cluster #3 36.08-39.99 mm	Cluster #4 32.67-35.56 mm	Cluster #5 30.04-32.38 mm	Cluster #6 27.91-29.92 mm	Cluster #7 26.10-27.87 mm	Cluster #8 23.96-25.96 mm	Cluster #9 21.56-23.79 mm	Cluster #10 16.3-21.11 mm
Number of sled tests	9	2	4	14	11	19	22	17	27	15
PAB										
L-shape	0.0%	4.4%	4.4%	13.3%	6.7%	13.3%	15.6%	15.6%	13.3%	13.3%
Square-shape	11.5%	0.0%	2.6%	1.3%	6.4%	12.8%	19.2%	12.8%	23.1%	10.3%
3D Add-on	0.0%	0.0%	0.0%	25.0%	75.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3D-shape	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	50.0%	50.0%
other	0.0%	0.0%	0.0%	54.5%	0.0%	27.3%	0.0%	0.0%	18.2%	0.0%
Divided Panel										
no	7.1%	1.6%	2.4%	11.0%	7.9%	14.2%	15.7%	13.4%	19.7%	7.1%
yes	0.0%	0.0%	7.7%	0.0%	7.7%	7.7%	15.4%	0.0%	15.4%	46.2%
Volume										
<90 L	0.0%	0.0%	0.0%	25.0%	0.0%	0.0%	0.0%	25.0%	25.0%	25.0%
90-100 L	0.0%	0.0%	2.1%	0.0%	0.0%	8.5%	17.0%	21.3%	34.0%	17.0%
101-110 L	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%
111-120 L	15.8%	1.8%	3.5%	21.1%	10.5%	8.8%	10.5%	7.0%	10.5%	10.5%
> 120 L	0.0%	3.2%	3.2%	3.2%	16.1%	32.3%	25.8%	6.5%	9.7%	0.0%
Number of VH										
no	12.3%	0.0%	3.1%	12.3%	7.7%	10.8%	6.2%	7.7%	21.5%	18.5%
yes	1.3%	2.7%	2.7%	8.0%	8.0%	16.0%	24.0%	16.0%	17.3%	4.0%
ABV										
no	7.4%	1.6%	2.5%	10.7%	6.6%	13.9%	16.4%	13.9%	18.9%	8.2%
yes	0.0%	0.0%	5.6%	5.6%	16.7%	11.1%	11.1%	0.0%	22.2%	27.8%
Tether 1										
no	7.7%	1.7%	2.6%	0.9%	8.5%	15.4%	17.1%	14.5%	22.2%	9.4%
yes	0.0%	0.0%	4.3%	56.5%	4.3%	4.3%	8.7%	0.0%	4.3%	17.4%
Inflator Type										
Pyrotechnic	21.4%	0.0%	0.0%	21.4%	4.8%	9.5%	9.5%	16.7%	14.3%	2.4%
Hybrid	0.0%	2.1%	4.2%	5.2%	9.4%	15.6%	18.8%	10.4%	20.8%	13.5%

20-20.9 °C	0.0%	0.0%	2.7%	18.9%	5.4%	13.5%	13.5%	13.5%	21.6%	10.8%
21-21.9 °C	8.1%	2.3%	3.5%	7.0%	10.5%	16.3%	12.8%	8.1%	19.8%	11.6%
>22 °C	14.3%	0.0%	0.0%	7.1%	0.0%	0.0%	42.9%	28.6%	7.1%	0.0%

According to Table 5.4, some trends about inputs parameters distributed into specific clusters enable to have a first indication about their importance for the dummy chest. Concerning the passenger airbag (PAB), all the studied 140 sled tests with a 3D-shape PAB are located either in cluster #9 or in cluster #10, which correspond to the best clusters according to their values for the chest deflection. Having the new 3D Add-On form seems to be rather a disadvantage for this test configuration because no one of the sled tests equipped with this type of airbag obtained a deflection value under 30,04 mm. In the aim of reaching 4 points in the EURO NCAP rating, the clusters #9 and #10, possibly #8 are the most appropriated. For that, having a PAB with a volume comprised between 101 and 110 L is an advantage because the available sled tests having such a volume are all included in cluster #9. More interestingly, a cold gas inflator, which is not the most popular type of inflator, seems to have a positive effect on the chest deflection because the values of the sled tests having this atypical inflator are located either in cluster #9 or in cluster #10.

Regarding the seat belt, for the sled tests having a DLL that produces a force on the dummy chest comprised between 4.6 kN and 5.5 kN, the value of the chest deflection is more located in the middle of the values' range, i.e. from cluster #4 to cluster #6. Additionally, the analysis of these sled tests shows that a reduced webbing on spool (value <750 mm) produces good results because located in clusters #7 and #8. However if this length increases, it does not mean that the value of the chest deflection is degraded: no conclusions can be drawn from the available data. In theory, the length of the webbing on spool is directly linked with the dummy chest: the longer the webbing on spool is, the lower the shoulder force should be. In fact, the longer the webbing on spool is, the lower force is needed to turn the torsion bar of the seat belt.

Moreover, the dummy position also does not give information. As shown in Table 5.4, many EURO NCAP sled tests already have a similar dummy position, and the restricted values' range does not enable a correct analysis of this parameter. In any case, it is difficult for the engineers to influence or alter this parameter because the dummy position is defined by the legislation.

Finally, the repartition of the Occupant Load Criterion (OLC) shows that if its value is above 28 g, which is already considered as "hard pulse" by the experts, the sled tests cannot pretend to the 4 points for the dummy chest. This parameter is directly linked to the vehicle type and cannot be easily altered by the engineers.

This first analysis of the values' repartition into the clusters enables to discover the trends of decisive parameters. It is supported by the decision tree¹ using the Gini Index as split criterion, represented in Fig. 5.1.

¹Shifted values to avoid representing a real case

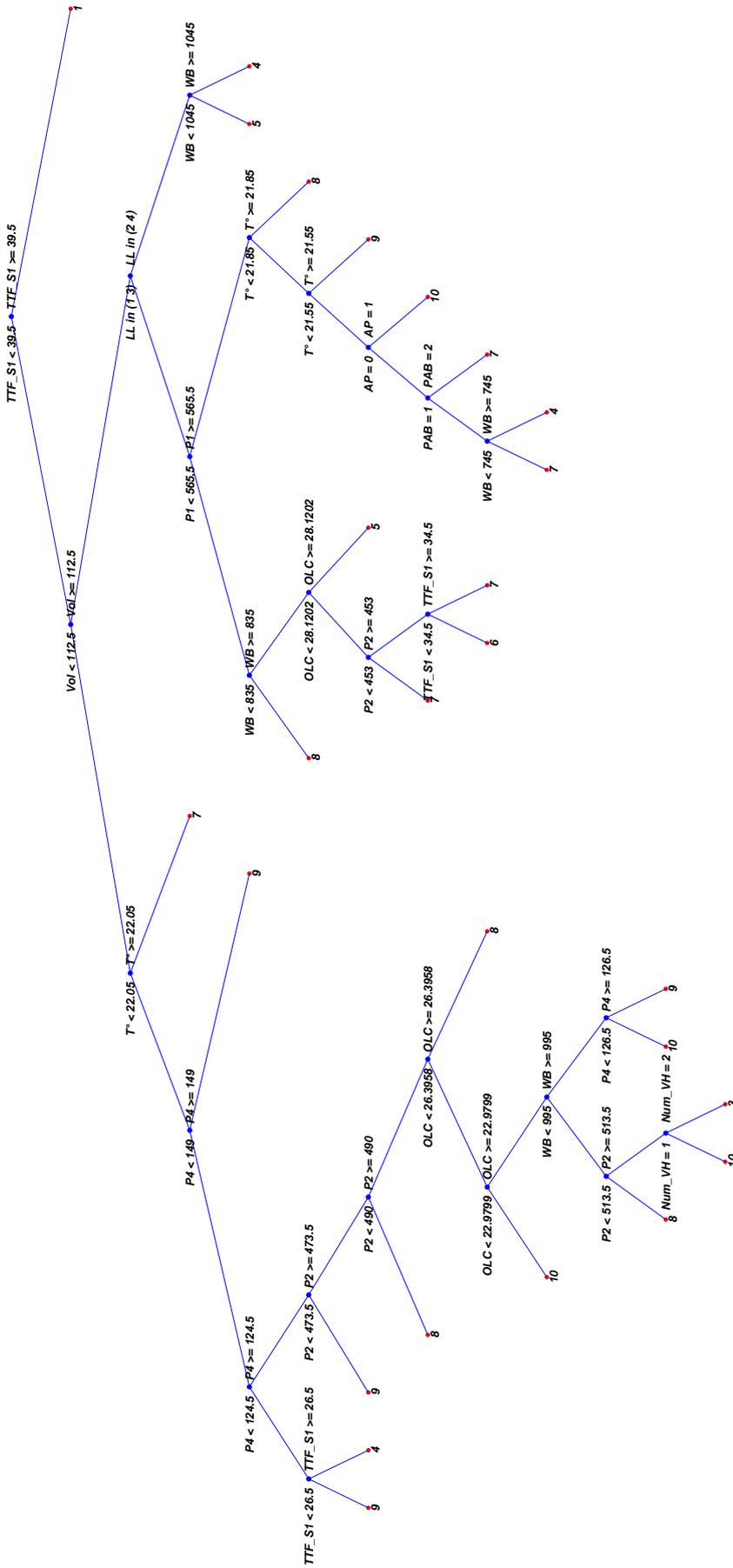


Figure 5.1: Decision tree for the dummy chest (passenger, EURO NCAP)

Analysis of the variables' links via decision tree

An analysis of the tree (Fig 5.1) firstly shows that all the clusters are not represented as terminal nodes and logically the clusters containing the most number of sled tests are the most represented (i.e. clusters #8 to #10). The first conclusion that can be drawn from this decision tree is the influence of the time to fire of the PAB first stage. As a reminder, it corresponds to the time at which the inflator is triggered. The inflator is mainly triggered in 2 times to share out the whole power and thus to avoid the dummy receiving a too high pressure in one time. In fact, a value above 39.5 ms immediately produces a very not satisfying result for the chest deflection : if the inflator is triggered too late, the dummy comes in contact with the airbag when it is deploying. During this deployment, it directly presses on the dummy chest that becomes the whole power proposed by the inflator, which involves a high value for the deflection.

The left branch of the decision tree gathers 26 out of 27 terminal nodes. It means that if the triggering of the first PAB stage is under 39.5 ms, more parameters are necessary to predict the chest value and thus that this parameters is not really relevant in defining the chest deflection (it is not frequent that the first stage is triggered below 39.5 ms). The best results can be obtained according to a combination of different parameters but the dummy temperature and its position mainly have an importance in attributing the cluster. For these two parameters, the engineers do not have a real impact because they are defined by the legislation and must be respected. To reach the two best clusters (#9 and #10), 9 combinations are available. The simple way to reach a very good value for the chest deflection (cluster #9) is an earlier TTF for the first stage (TTF <39.5 ms), a volume for the airbag under 112,5 L, a dummy temperature under 22,05 °C and finally a distance between the right knee and the IP above 149 mm.

To sum up, according to the decision tree, the parameters that mostly influence the value of the chest deflection in a EURO NCAP sled test would be:

- ◆ Time to fire of the first stage (PAB inflator);
- ◆ Volume of the PAB;
- ◆ Dummy position, especially the distance knee / IP;
- ◆ Dummy temperature (its range is defined by the legislation).

The decision tree is supported by the calculation of the Variable Importance Measure (VIM). The higher this measure is, the more is the influence of the corresponding criterion on the tree. This interpretation of the decision tree is thus confirmed by the VIM values (Table 5.5): the parameter "TTF PAB first stage" receives the highest value, followed by the dummy temperature and two dummy positions.

Once the decision tree is built, even if the parameters extracted as influencing the chest deflection are confirmed with the calculation of the VIM, the model has to be validated. As done before, the LOOCV method is applied. This method has been already employed to validate several algorithms (see Chapter 3.2): it takes one sample as validation sample and the (k-1) samples as training set, this operation is repeated k times. The results of the validation are summarised in Table 5.6.

The model is encouraging because 37% of the results have been correctly predicted in the right cluster; it practically means that the model is able to estimate the chest deflection

correctly. It is interesting to point out that each cluster for the chest has a values' range of approximately 2 mm. Also during a repeated sled test, it is possible that the chest value swings between 1 and 3 mm. That is why the percentage of good predicted results has been also calculated if the value has been estimated in the cluster "before" or "after" the right cluster. With this condition, the percentage of right predicted values increases to 68%.

Table 5.5: Calculation of the VIM for the dummy chest (passenger, EURO NCAP)

Predictor Variable	Variable Importance Measure
TTF Stage 1	0.003788111
Dummy T°	0.002771824
Sternum to IP	0.002595893
Right Knee - Shorter Distance Ahead	0.002384063
Webbing on Spool	0.002340877
OLC	0.001631277
Volume	0.00142356
Type of LL	0.00099139
Chin to IP	0.000792477
AP	0.000586081
PAB	0.000494505
Number of VH	0.00040293

Table 5.6: Validation of the decision tree for the dummy chest (passenger, EURO NCAP)

Obtained injury	Predicted injury									
	Cluster #1	Cluster #2	Cluster #3	Cluster #4	Cluster #5	Cluster #6	Cluster #7	Cluster #8	Cluster #9	Cluster #10
Cluster #1	9	0	0	0	0	0	0	0	0	0
Cluster #2	0	0	0	0	0	0	1	0	1	0
Cluster #3	0	0	0	0	0	1	1	2	0	0
Cluster #4	0	0	0	6	3	0	3	1	1	0
Cluster #5	0	0	0	3	0	5	1	2	0	0
Cluster #6	0	0	0	3	2	3	7	1	2	1
Cluster #7	0	0	0	3	0	6	8	2	2	1
Cluster #8	0	0	0	0	0	1	2	10	2	2
Cluster #9	0	0	0	6	0	1	1	2	13	4
Cluster #10	0	0	0	2	0	1	0	4	5	3
Prediction accuracy	37% in the right cluster; 68% in ± 1 cluster									

Finally, the analysis of the decision tree shows that the time to fire of the "PAB first stage" plays an important role in determining the value of the chest deflection. Moreover, the seat belt's parameters are not so decisive as those of the PAB or dummy position. The remaining webbing on spool appears as the most decisive parameter for the seat belt, which is coherent because this parameter is directly linked to the force applied on the chest.

More surprisingly, parameters such as the diameter of the vent hole are not considered for building the tree, although this parameter has an influence on the pressure in the PAB and thus on the force applied on the dummy chest. Unfortunately, the repartition of the sled tests into the clusters are not homogeneous and a better one could produce rules explaining how to avoid having such unexpected results.

Because of the good results obtained by the head and the neck for the EURO NCAP rating, only the chest deflection is studied in this section. The challenge for the engineers in reaching the 5 stars in US NCAP rating is more complex and it is thus more interesting to focus the decision trees analysis on the four dummy parts of the HIII 5th dummy.

5.3/ US NCAP RATING: HIII 5TH BELTED DUMMY (PASSENGER)

This test configuration is one of the challenges for the engineers because of the difficulties to reach 5 stars for the passenger side. The restraint systems have to work with two types of dummies, belted and unbelted, making this rating difficult. For this reason, the analysis of the parameters' influence is conducted on the four dummy parts of the passenger, i.e. the head, neck, chest and femurs. All four parts are considered in this chapter.

5.3.1/ DUMMY HEAD

Analysis of the clusters' material

Table 5.7 provides some information about the injury severity distribution for the dummy head, according to the 27 key factors that characterize the restraint systems and the vehicle, such as the volume of the passenger airbag or the type of seat belt. 103 out of 216 passenger tests received a high level of protection (clusters #5 and #6), 57 a good protection (cluster #7), 53 a protection that should be improved (clusters #1 and #2) and 3 passengers tests are confronted to a protection that probably does not enable to let the car going on the U.S. market.

As shown in Table 5.7, for all the sled tests with a seat belt load limiter value above 5.6 kN and a distance between the dummy sternum and the instrumental panel (IP) above 550 mm, the HIC value is comprised between 251 and 325, what corresponds to 36% and 46% of the legal value respectively. However these two parameters values are very high and unusual, like the triggered of the inflator second stage. In fact, according to Table 5.7, 100% of the sled tests which this value is comprised between 14 ms and 16 ms, had a HIC value that does not satisfy the requirements.

We can also remark that a PAB 3D-shape airbag, with a volume under 110 L and a vent hole diameter comprised between 61 and 70 mm provided good results for the head because a majority of the sled tests with one of these parameters does not have a HIC value that excess 325, which corresponds to 46% of the legal value. Another trend that can be analysed from the sled tests repartition is the dummy temperature. In fact, having

a cold dummy (i.e. a dummy temperature under 20 °C) provides a very good HIC value, what is not surprising. The colder the dummy is, and thus the sensors are, the less sensitive the sensors are and thus the less high the measured values are. That is why the dummy temperature is imposed by the legislation and a dummy temperature under 20 °C is not legal. For the US NCAP rating, it has to be comprised between 20.6 °C and 22.2 °C (i.e. 69°F to 72°F).

Table 5.7: Sled tests' repartition for the dummy head (passenger, US NCAP)

Variables	Cluster 1 430-591	Cluster 2 329-428	Cluster 3	Cluster 4	Cluster 5 74-175	Cluster 6 176-249	Cluster 7 251-325
Number of sled tests	19	34	0	3	43	60	57
PAB							
L-shape	6.2%	16.7%	0.0%	0.6%	16.0%	29.6%	30.9%
Square-shape	23.7%	15.8%	0.0%	5.3%	28.9%	15.8%	10.5%
3D Add-on	0.0%	16.7%	0.0%	0.0%	33.3%	50.0%	0.0%
3D-shape	0.0%	0.0%	0.0%	0.0%	33.3%	66.7%	0.0%
Other	0.0%	0.0%	0.0%	0.0%	42.9%	14.3%	42.9%
Divided panel							
no	12.1%	14.6%	0.0%	1.9%	19.1%	30.6%	21.7%
yes	0.0%	18.6%	0.0%	0.0%	22.0%	20.3%	39.0%
Volume							
<90 L	0.0%	0.0%	0.0%	0.0%	90.0%	10.0%	0.0%
90-100 L	0.0%	0.0%	0.0%	0.0%	33.3%	33.3%	33.3%
101-110 L	0.0%	0.0%	0.0%	0.0%	33.3%	66.7%	0.0%
111-120 L	3.9%	16.1%	0.0%	0.0%	17.8%	31.1%	31.1%
>120 L	60.0%	25.0%	0.0%	15.0%	0.0%	0.0%	0.0%
Number of Vent-holes							
1	0.6%	14.4%	0.0%	0.0%	23.3%	31.7%	30.0%
2	50.0%	22.2%	0.0%	8.3%	2.8%	8.3%	8.3%
Hole diameter 1							
<40 mm	2.6%	17.9%	0.0%	0.0%	15.4%	17.9%	46.2%
40-50 mm	10.7%	17.3%	0.0%	2.7%	29.3%	26.7%	13.3%
51-60 mm	12.2%	17.1%	0.0%	1.2%	8.5%	30.5%	30.5%
61-70 mm	0.0%	0.0%	0.0%	0.0%	40.0%	40.0%	20.0%
>70 mm	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
ABV							
no	41.9%	23.3%	0.0%	7.0%	7.0%	4.7%	16.3%
yes	0.6%	13.9%	0.0%	0.0%	23.1%	33.5%	28.9%
Tether 1							
no	12.6%	15.9%	0.0%	2.0%	15.9%	29.8%	23.8%
yes	0.0%	15.4%	0.0%	0.0%	29.2%	23.1%	32.3%
Tether 2							
no	10.7%	14.7%	0.0%	1.7%	18.6%	31.6%	22.6%
yes	0.0%	20.5%	0.0%	0.0%	25.6%	10.3%	43.6%
Number of Stages							
1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	8.8%	15.7%	0.0%	1.4%	19.9%	27.8%	26.4%
TTF Stage 1							
<14 ms	6.6%	14.3%	0.0%	1.1%	27.5%	24.2%	26.4%
14-16 ms	2.0%	13.3%	0.0%	2.0%	16.3%	36.7%	29.6%
17-20 ms	37.5%	33.3%	0.0%	0.0%	4.2%	8.3%	16.7%
>20 ms	66.7%	0.0%	0.0%	0.0%	33.3%	0.0%	0.0%
TTF Stage 2							
<14 ms	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
14-16 ms	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%
17-20 ms	4.5%	13.5%	0.0%	1.3%	24.5%	32.9%	23.2%

>20 ms	20.0%	20.0%	0.0%	1.7%	8.3%	15.0%	35.0%
Type of LL							
CLL	6.0%	12.0%	0.0%	2.0%	42.0%	14.0%	24.0%
DLL	58.3%	33.3%	0.0%	0.0%	0.0%	0.0%	8.3%
PLL	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
SLL	5.8%	15.6%	0.0%	1.3%	14.3%	34.4%	28.6%
LL Value 1							
<2.5 kN	0.0%	0.0%	0.0%	0.0%	75.0%	25.0%	0.0%
2.5-3.4 kN	7.7%	2.6%	0.0%	2.6%	46.2%	33.3%	7.7%
3.5-4.5 kN	12.9%	17.8%	0.0%	2.0%	15.8%	31.7%	19.8%
4.6-5.5 kN	4.3%	21.4%	0.0%	0.0%	8.6%	18.6%	47.1%
5.6-6.5 kN	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
Retractor Pretensioner							
no	0.0%	0.0%	0.0%	0.0%	0.0%	50.0%	50.0%
yes	8.9%	15.9%	0.0%	1.4%	20.1%	27.6%	26.2%
Buckle Pretensioner							
no	16.5%	15.6%	0.0%	2.8%	17.4%	32.1%	15.6%
yes	0.9%	15.9%	0.0%	0.0%	22.4%	23.4%	37.4%
Anchor Point Pretensioner							
no	11.3%	16.1%	0.0%	1.8%	22.0%	20.8%	28.0%
yes	0.0%	14.9%	0.0%	0.0%	12.8%	51.1%	21.3%
Webbing on Spool							
<750 mm	6.7%	6.7%	0.0%	0.0%	20.0%	60.0%	6.7%
750-800 mm	0.0%	16.1%	0.0%	0.0%	6.5%	51.6%	25.8%
801-850 mm	24.2%	12.1%	0.0%	6.1%	18.2%	18.2%	21.2%
851-900 mm	0.0%	12.9%	0.0%	0.0%	25.8%	32.3%	29.0%
901-950 mm	0.0%	13.6%	0.0%	0.0%	40.9%	31.8%	13.6%
951-1000 mm	0.0%	11.1%	0.0%	0.0%	22.2%	25.9%	40.7%
>1000 mm	17.5%	24.6%	0.0%	1.8%	15.8%	8.8%	31.6%
Height Adjustment							
no	1.3%	11.4%	0.0%	0.0%	30.4%	31.6%	25.3%
yes	13.1%	18.2%	0.0%	2.2%	13.9%	25.5%	27.0%
DLT							
no	9.2%	16.0%	0.0%	1.5%	19.9%	26.7%	26.7%
yes	0.0%	11.1%	0.0%	0.0%	22.2%	55.6%	11.1%
KnAB							
no	0.9%	16.8%	0.0%	0.0%	17.7%	32.7%	31.9%
yes	17.5%	14.6%	0.0%	2.9%	22.3%	22.3%	20.4%
Chin to IP							
<420 mm	15.4%	15.4%	0.0%	2.6%	5.1%	20.5%	41.0%
420-470 mm	14.6%	12.4%	0.0%	2.2%	33.7%	24.7%	12.4%
471-500 mm	0.0%	18.0%	0.0%	0.0%	16.0%	48.0%	18.0%
501-550 mm	0.0%	21.1%	0.0%	0.0%	7.9%	15.8%	55.3%
>550 mm	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sternum to IP							
<420 mm	8.1%	11.1%	0.0%	2.0%	28.3%	28.3%	22.2%
420-470 mm	24.4%	22.0%	0.0%	2.4%	24.4%	17.1%	9.8%
471-500 mm	2.6%	15.4%	0.0%	0.0%	7.7%	48.7%	25.6%
501-550 mm	0.0%	22.2%	0.0%	0.0%	5.6%	16.7%	55.6%
>550 mm	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%
Left Knee - Shorter Distance Ahead							
<40 mm	13.7%	17.6%	0.0%	2.0%	31.4%	15.7%	19.6%
40-49 mm	5.3%	14.0%	0.0%	1.8%	14.0%	21.1%	43.9%
50-59 mm	23.3%	3.3%	0.0%	3.3%	16.7%	40.0%	13.3%
60-69 mm	4.3%	19.6%	0.0%	0.0%	15.2%	37.0%	23.9%
70-79 mm	0.0%	18.2%	0.0%	0.0%	31.8%	22.7%	27.3%

>80 mm	0.0%	25.0%	0.0%	0.0%	0.0%	62.5%	12.5%
Right Knee - Shorter Distance Ahead							
<40 mm	17.2%	24.1%	0.0%	3.4%	27.6%	6.9%	20.7%
40-49 mm	18.5%	7.4%	0.0%	0.0%	29.6%	40.7%	3.7%
50-59 mm	4.9%	13.1%	0.0%	3.3%	11.5%	16.4%	50.8%
60-69 mm	12.2%	16.3%	0.0%	0.0%	12.2%	40.8%	18.4%
70-79 mm	0.0%	17.9%	0.0%	0.0%	28.6%	28.6%	25.0%
>80 mm	0.0%	18.2%	0.0%	0.0%	27.3%	40.9%	13.6%
OLC++							
<26 g	0.0%	8.5%	0.0%	0.0%	38.3%	31.9%	21.3%
26-26.9 g	12.1%	15.2%	0.0%	0.0%	21.2%	21.2%	30.3%
27-27.9 g	20.4%	16.3%	0.0%	6.1%	14.3%	28.6%	14.3%
>28 g	1.9%	22.2%	0.0%	0.0%	7.4%	31.5%	37.0%
Tv0							
<65 ms	4.9%	18.3%	0.0%	0.0%	9.8%	36.6%	30.5%
65-69.9 ms	14.3%	17.1%	0.0%	2.9%	22.9%	20.0%	22.9%
>70 ms	0.0%	3.4%	0.0%	0.0%	37.9%	31.0%	27.6%
Dummy Temperature							
<20 °C	0.0%	0.0%	0.0%	0.0%	50.0%	50.0%	0.0%
20-20.9 °C	5.3%	18.4%	0.0%	1.3%	21.1%	32.9%	21.1%
21-21.9 °C	8.8%	12.4%	0.0%	1.8%	17.7%	27.4%	31.9%
>22 °C	20.0%	24.0%	0.0%	0.0%	24.0%	12.0%	20.0%

Results and discussion (dummy head)

Fig. 5.2 presents the results of the classification for the dummy head² and supports the analysis done with Table 5.7. This tree has 35 terminal nodes, which correspond to the 7 clusters formed by the GK algorithm.

²Shifted values to avoid representing a real case

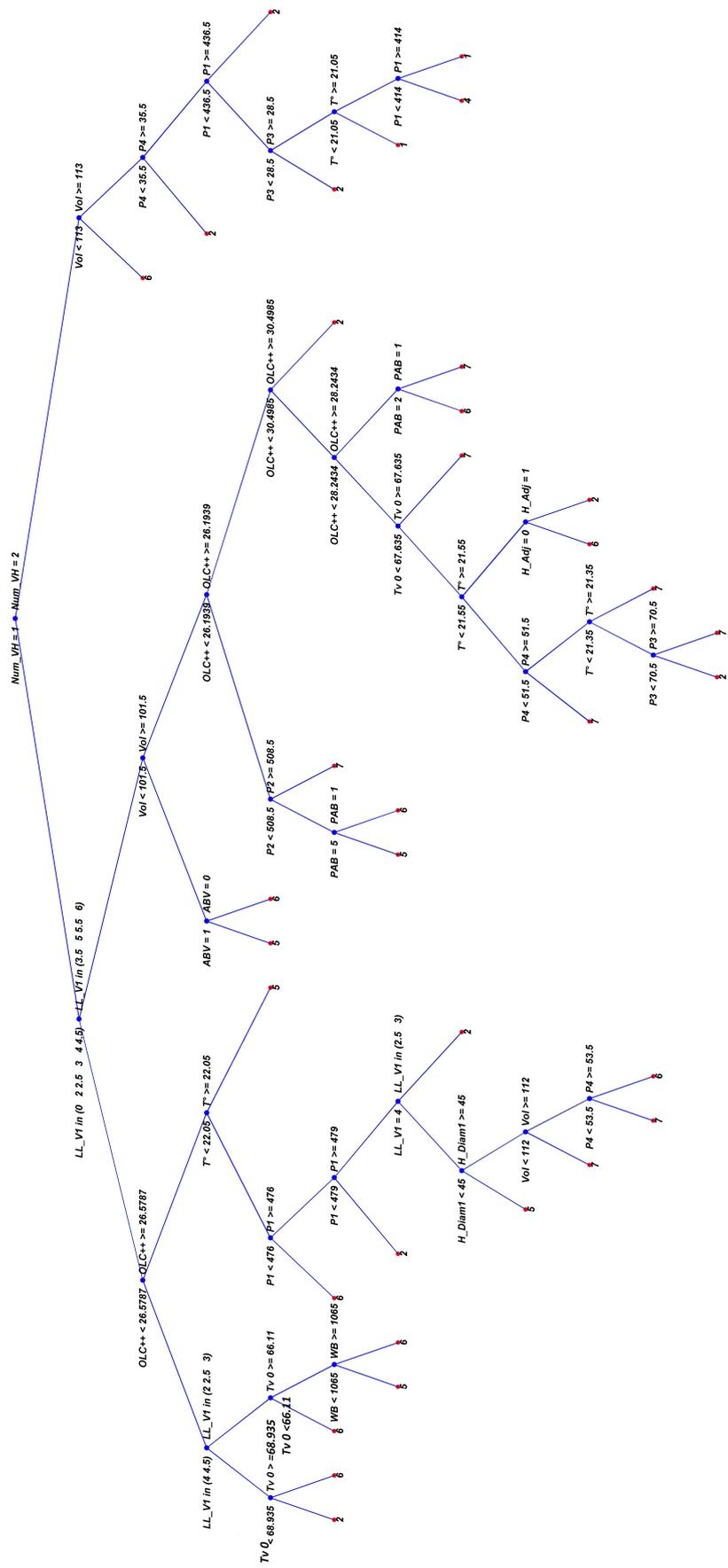


Figure 5.2: Decision tree for the dummy head (passenger, US NCAP)

The first analysis that can be directly drawn from the tree (Fig 5.2) is the influence of the variable "Number of vent holes". Except if the volume of the airbag is under 113 L, having an airbag with 2 vent holes is very harmful for the head. In this case the best possible value for the head is 430 (61% of the legal value), what is insufficient for reaching 5 stars in the rating. This is a little bit surprising because having a second vent hole aims at reducing the pressure inside the airbag that is often too high for a HIII 5th dummy (the airbag has also to work with a HIII 50th dummy for which the pressure in the airbag has to be enough to retain the dummy). The common solution for reducing the airbag pressure for the HIII 5th dummy and maintaining it for the HIII 50th dummy is the use of an Adaptive Bag Vent (ABV) which can be triggered or not. But this variable is not considered as relevant for building the tree.

Also surprisingly, the most impressionable variables for the head are the dummy position. Some seat belt parameters are used as splitters for determining the clusters but a predominance of non restraint systems parameters for estimating the head value is visible. In fact, the value of the pulse through the criterion OLC++ is used as splitter very early in the tree, it means that this criterion is important for the head value.

To sum up, this tree shows that a prediction for the dummy head is complex and involves many non restraint systems parameters, such as the pulse criterion and the dummy position. One of the airbag parameters is nevertheless decisive: the number of vent holes. The second restraint system parameter that is used as splitter for building the tree is the load limiter value of the seat belt. After analysis of the tree, it can be noted that the branches coming from this variable can produce both worst and good results. The head value is finally more sensitive to the crash pulse, dummy position and temperature as to the airbag and seat belt configurations.

Table 5.8 presents the validation of the tree done with the LOOCV method: 53% of the head values have been correctly estimated by the tree.

Table 5.8: Validation of the decision tree for the dummy head (passenger, US NCAP)

Obtained injury	Predicted injury						
	Cluster #1	Cluster #2	Cluster #3	Cluster #4	Cluster #5	Cluster #6	Cluster #7
Cluster #1	13	4	0	0	0	1	1
Cluster #2	5	12	0	0	0	4	13
Cluster #3	0	0	0	0	0	0	0
Cluster #4	3	0	0	0	0	0	0
Cluster #5	0	3	0	0	22	12	6
Cluster #6	0	5	0	0	6	39	10
Cluster #7	0	16	0	0	4	9	28
Prediction accuracy	53%						

The tree can be supported by the calculation of the VIM (Table 5.9). It confirms the importance of the pulse criterion and the number of vent holes of the airbag. The variable "Load limiter value" of the seat belt is considered as influencing the head value, which is in agreement with its position as splitter in the tree. Seven variables, independently of the restraint systems, are ordered in the first 12 variables influencing the head value. This method (Table 5.9) confirms the results obtained by the tree (Fig 5.2).

Table 5.9: Calculation of the VIM for the dummy head (passenger, US NCAP)

Predictor Variable	Variable Importance Measure
OLC++	0.002644228
LL Value 1	0.002366675
Volume	0.001865089
Number of Ventholes	0.001638526
Tv0	0.001467101
Dummy temperature	0.001442678
PAB	0.000799332
Right Knee - Shorter Distance Ahead	0.000797897
Left Knee - Shorter Distance Ahead	0.00076299
Chin to IP	0.000549878
Sternum to IP	0.000498894
H Adjustment	0.000331336
ABV	0.000247574
Webbing on Spool	0.000222404

5.3.2/ DUMMY NECK

Analysis of the clusters' material

Table 5.10 summarises the injury severity distribution for the dummy neck, according to the same 27 attributes as for the dummy head. The target variable, the injury severity level, is separated into 6 distinct clusters, cluster #1 corresponds to "No-injury" and cluster #6 to "Fatal-injury". From the cluster #5, the customer's and thus legal requirements are no more satisfied.

Firstly, for 84 out of 216 sled tests, the measured neck value N_{ij} is under 35 % of the legal value, that signifies that the neck requirement defined by Sohr [49] for reaching the 5 stars is completed (the N_{ij} does not have to exceed 0.3 of the legal value). The restraint systems of the 128 tests included in clusters #2, #3 and #4 have to be definitively improved if the specification sheet imposes 5 stars for the US NCAP rating.

Then, according to Table 5.10, cluster #3 principally contains sled tests for which the airbag form is not standard or unknown. On the contrary, the standard forms "3D Add on" and "3D shape" seems to have a positive influence on the neck criterion, because approximately 67% of the sled tests using these types of airbag are located in cluster #1 ($N_{ij} < 0,35$). Moreover, if its vent hole has a diameter comprised between 40 and 50 mm or between 61 and 70 mm, the probability to be in cluster #1 is improved. The inflator of the PAB also plays an important role through the triggering of its second stage (TTF-2). In fact, for 100% of the sled tests for which the second stage is triggered between 14 and 16 ms, the N_{ij} value has been measured between 0.36 and 0.43 (cluster #2). This is a good indication that is easy to take into consideration because it does not require any

expensive modification.

The seat belt is the second restraint system that can be modified by the engineers. The constant and degressive load limiters represent respectively 58% of the sled tests included in cluster #1 and 58.3% in cluster #2. Moreover, if the force applied on the shoulder is under 2.5 kN (LL value), the chances to obtain good results (cluster #1) are very high, as the probability to be in cluster #2 if the LL value is comprised between 5.6 and 6.5 kN. Finally, the chances to reach 5 stars are upgraded if the webbing on spool has a length between 801 and 850 mm.

As shown in Table 5.10, the dummy position also influences the results for the neck, especially the distances sternum-IP and knee-ahead. This first distance should be above 471 mm to hope being in the 3 first clusters and then under 500 mm to stay in the first one. If this value is above 550 mm, the chances to be in cluster #3 are maximum (100% of the sled tests with a value above 550 mm are in this cluster). Cluster#1 contains 72.7% of the sled tests for which the distance between the left knee and ahead is comprised between 70 and 79 mm. But this parameter is hardly altering by the engineers.

Finally, the dummy temperature is one of the modifiable parameters that also influences the neck criterion. Under 20°C, its value can be expected between 0.36 and 0.52 (clusters #2 and #3) and above 22°C, 50% of the sled tests are located in cluster #1.

Table 5.10: Sled tests' repartition for the dummy neck (passenger, US NCAP)

Variables	Cluster #1	Cluster #2	Cluster #3	Cluster #4	Cluster #5	Cluster #6
	0-0.35	0.36-0.43	0.44-0.52	0.53-0.67		
Number of sled tests	84	75	45	8	4	0
PAB						
L-shape	38.3%	35.2%	20.4%	4.3%	1.9%	0.0%
Square-shape	42.1%	36.8%	15.8%	2.6%	2.6%	0.0%
3D Add-on	66.7%	16.7%	16.7%	0.0%	0.0%	0.0%
3D-shape	66.7%	33.3%	0.0%	0.0%	0.0%	0.0%
Other	0.0%	28.6%	71.4%	0.0%	0.0%	0.0%
Divided panel						
no	42.7%	33.1%	19.1%	3.8%	1.3%	0.0%
yes	28.8%	39.0%	25.4%	3.4%	3.4%	0.0%
Volume						
<90 L	40.0%	20.0%	20.0%	10.0%	10.0%	0.0%
90-100 L	33.3%	66.7%	0.0%	0.0%	0.0%	0.0%
101-110 L	66.7%	33.3%	0.0%	0.0%	0.0%	0.0%
111-120 L	39.4%	34.4%	22.2%	2.8%	1.1%	0.0%
>120 L	30.0%	40.0%	15.0%	10.0%	5.0%	0.0%
Number of Vent-holes						
1	41.1%	31.7%	22.2%	2.8%	2.2%	0.0%
2	27.8%	50.0%	13.9%	8.3%	0.0%	0.0%
Hole diameter 1						
<40 mm	30.8%	30.8%	35.9%	0.0%	2.6%	0.0%
40-50 mm	56.0%	25.3%	14.7%	4.0%	0.0%	0.0%
51-60 mm	23.2%	45.1%	22.0%	6.1%	3.7%	0.0%
61-70 mm	55.0%	35.0%	10.0%	0.0%	0.0%	0.0%
>70 mm	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
ABV						
no	27.9%	46.5%	18.6%	7.0%	0.0%	0.0%
yes	41.6%	31.8%	21.4%	2.9%	2.3%	0.0%
Tether 1						
no	41.1%	33.1%	18.5%	4.6%	2.6%	0.0%

yes	33.8%	38.5%	26.2%	1.5%	0.0%	0.0%
Tether 2						
no	42.9%	32.8%	18.1%	4.0%	2.3%	0.0%
yes	20.5%	43.6%	33.3%	2.6%	0.0%	0.0%
Number of Stages						
1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	38.9%	34.7%	20.8%	3.7%	1.9%	0.0%
TTF 1						
<14 ms	47.3%	28.6%	9.8%	1.1%	3.3%	0.0%
14-16 ms	33.7%	39.8%	20.4%	5.1%	1.0%	0.0%
17-20 ms	29.2%	41.7%	25.0%	4.2%	0.0%	0.0%
>20 ms	33.3%	0.0%	33.3%	33.3%	0.0%	0.0%
TTF 2						
<14 ms	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
14-16 ms	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
17-20 ms	40.6%	35.5%	18.1%	3.2%	2.6%	0.0%
>20 ms	35.0%	31.7%	28.3%	5.0%	0.0%	0.0%
Type of LL						
CLL	58.0%	28.0%	12.0%	2.0%	0.0%	0.0%
DLL	0.0%	58.3%	33.3%	8.3%	0.0%	0.0%
PLL	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
SLL	35.7%	35.1%	22.7%	3.9%	2.6%	0.0%
LL Value 1						
<2.5 kN	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2.5-3.4 kN	76.9%	12.8%	7.7%	2.6%	0.0%	0.0%
3.5-4.5 kN	40.6%	37.6%	15.8%	3.0%	3.0%	0.0%
4.6-5.5 kN	12.9%	42.9%	37.1%	5.7%	1.4%	0.0%
5.6-6.5 kN	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
Retractor pretensioner						
no	50.0%	0.0%	50.0%	0.0%	0.0%	0.0%
yes	38.8%	35.0%	20.6%	3.7%	1.9%	0.0%
Buckle pretensioner						
no	42.2%	36.7%	14.7%	4.6%	1.8%	0.0%
yes	35.5%	32.7%	27.1%	2.8%	1.9%	0.0%
Anchor pretensioner						
no	36.9%	35.7%	22.6%	3.6%	1.2%	0.0%
yes	46.8%	29.8%	14.9%	4.3%	4.3%	0.0%
Webbing on Spool						
<750 mm	53.3%	13.3%	13.3%	6.7%	13.3%	0.0%
750-800 mm	41.9%	41.9%	9.7%	6.5%	0.0%	0.0%
801-850 mm	60.6%	30.3%	9.1%	0.0%	0.0%	0.0%
851-900 mm	41.9%	32.3%	22.6%	3.2%	0.0%	0.0%
901-950 mm	54.5%	18.2%	13.6%	9.1%	4.5%	0.0%
951-1000 mm	29.6%	44.4%	22.2%	0.0%	3.7%	0.0%
>1000 mm	17.5%	42.1%	36.8%	3.5%	0.0%	0.0%
Height Adjustment						
no	58.2%	24.1%	16.5%	0.0%	1.3%	0.0%
yes	27.7%	40.9%	23.4%	5.8%	2.2%	0.0%
DLT						
no	38.8%	34.0%	21.8%	3.9%	1.5%	0.0%
yes	50.0%	33.3%	0.0%	0.0%	16.7%	0.0%
KnAB						
no	31.0%	37.2%	23.9%	4.4%	3.5%	0.0%
yes	47.6%	32.0%	17.5%	2.9%	0.0%	0.0%

Chin to IP						
<420 mm	25.6%	48.7%	25.6%	0.0%	0.0%	0.0%
420-470 mm	53.9%	23.6%	15.7%	4.5%	2.2%	0.0%
471-500 mm	44.0%	32.0%	16.0%	4.0%	4.0%	0.0%
501-550 mm	10.5%	50.0%	34.2%	5.3%	0.0%	0.0%
>550 mm	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sternum to IP						
<420 mm	46.5%	29.3%	19.2%	2.0%	3.0%	0.0%
420-470 mm	34.1%	34.1%	24.4%	4.9%	2.4%	0.0%
471-500 mm	51.3%	33.3%	10.3%	5.1%	0.0%	0.0%
501-550 mm	11.1%	52.8%	30.6%	5.6%	0.0%	0.0%
>550 mm	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%
Left Knee - Shorter Distance Ahead						
<40 mm	21.6%	45.1%	25.5%	3.9%	3.9%	0.0%
40-49 mm	26.3%	40.4%	29.8%	3.5%	0.0%	0.0%
50-59 mm	56.7%	36.7%	0.0%	6.7%	0.0%	0.0%
60-69 mm	45.7%	23.9%	23.9%	4.3%	2.2%	0.0%
70-79 mm	72.7%	18.2%	9.1%	0.0%	0.0%	0.0%
>80 mm	37.5%	37.5%	12.5%	0.0%	12.5%	0.0%
Right Knee - Shorter Distance Ahead						
<40 mm	20.7%	48.3%	27.6%	3.4%	0.0%	0.0%
40-49 mm	29.6%	37.0%	18.5%	7.4%	7.4%	0.0%
50-59 mm	24.6%	42.6%	29.5%	3.3%	0.0%	0.0%
60-69 mm	51.0%	30.6%	14.3%	4.1%	0.0%	0.0%
70-79 mm	53.6%	21.4%	17.9%	3.6%	3.6%	0.0%
>80 mm	68.2%	18.2%	9.1%	0.0%	4.5%	0.0%
OLC++						
<26 g	40.4%	29.8%	23.4%	2.1%	4.3%	0.0%
26-26.9 g	42.4%	28.8%	22.7%	6.1%	0.0%	0.0%
27-27.9 g	34.7%	51.0%	12.2%	2.0%	0.0%	0.0%
>28 g	37.0%	31.5%	24.1%	3.7%	3.7%	0.0%
Tv0						
<65 ms	39.0%	36.6%	19.5%	2.4%	2.4%	0.0%
65-69.9 ms	43.8%	31.4%	20.0%	3.8%	1.0%	0.0%
>70 ms	20.7%	41.4%	27.6%	6.9%	3.4%	0.0%
Dummy Temperature						
<20 °C	0.0%	50.0%	50.0%	0.0%	0.0%	0.0%
20-20.9 °C	38.2%	32.9%	22.4%	5.3%	1.3%	0.0%
21-21.9 °C	37.2%	37.2%	21.2%	2.7%	1.8%	0.0%
>22 °C	52.0%	28.0%	12.0%	4.0%	4.0%	0.0%

Analysis of the variables' links via decision tree

27 predictor variables have been used to identify the pattern that influence the level of severity injury for the neck of the HIII 5th dummy in a US NCAP sled test. The binary tree represented in Fig. 5.3 is composed of 34 terminal nodes which represent 4 out of the 6 clusters formed by the GK algorithm. Clusters #5 and #6 are not represented as terminal nodes, because of the low number of sled tests included in these clusters (respectively 4 and 0 sled tests). Cluster #2 is the most represented terminal node with 44% of the total terminal nodes. The second is cluster #3 but relatively far away, with only 26% of the branches that arrive in this terminal node. The analysis of the tree shows

that 16 attributes are used as splitters to frame the tree, it means that these variables are important for classifying the injury severity for the neck. These key factors contain for example the first value of the seat belt load limiter (LL-V1), the time to fire of the inflator second stage (TTF-S2) and the dummy position (P1, P2, P3 and P4). One advantage of the CART is its straightforward interpretation. According to Fig. 5.3³, the split at node 1 is based on the predictor variable "Load limiter value of the seat belt (LL-V1)", that indicates that this value influences the most the injury severity. This is also confirmed by calculating its variable importance (Table 5.11). If this value is equal to 5.5 kN, the neck criterion has a value comprised between 44% and 52% of the legal value. If this value is under 4 kN, except for an exact value of 3.5 kN, the criterion N_{ij} has a maximal value of 0.67, independently of the other parameters. If moreover the length of the webbing on spool is above 995 mm and the seat belt has a degressive or switchable load limiter, the value of the neck is in any case under 52% of the legal value. The prediction of the injury severity can be known by continuing down each tree branch until a terminal node is reached.

To sum up, the parameters that influence the most the criterion N_{ij} for the HIII 5th dummy are:

- ◆ Load limiter value;
- ◆ Second time to fire of the PAB inflator;
- ◆ Dummy position.

³Shifted values to avoid representing a real case

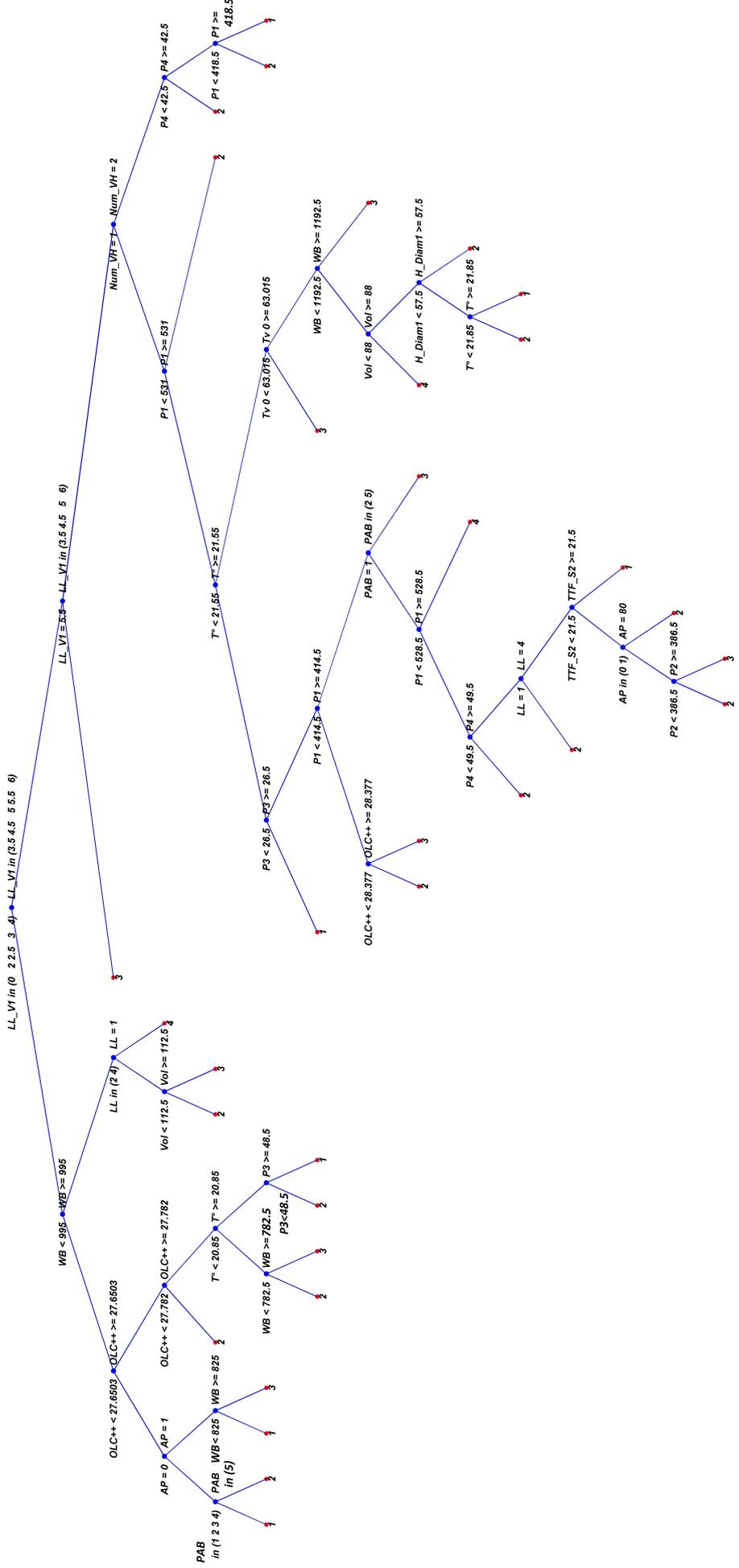


Figure 5.3: Decision tree for the dummy neck (passenger, US NCAP)

This deduction is partially confirmed by the Variable Importance Measure (VIM). According to Table 5.11, the load limiter value and the webbing on spool of the seat belt are the two most important predictor variables (the values of VIM are ordered down). The dummy position also plays a role because three distances out of 4 are classified in the 9 first places.

Table 5.11: Calculation of the VIM for the dummy neck (passenger, US NCAP)

Predictor Variable	Variable Importance Measure
LL Value 1	0.002158827
Webbing on Spool	0.002151536
OLC++	0.001849383
Chin to IP	0.001481697
Dummy temperature	0.000766602
Left Knee - Shorter Distance Ahead	0.00059397
Volume	0.000516073
Right Knee - Shorter Distance Ahead	0.00047086
Anchor Point Pretensioner	0.000455156
PAB	0.000441881
Tv0	0.000400457
Type of LL	0.000344884
Number of Ventholes	0.000290031
Sternum to IP	0.000157179
TTF Stage 2	0.00015229
Hole Diameter 1	0.0000900206

In order to evaluate the performance and reability of this decision tree, a comparison between the predicted and real injury severity has been conducted through the LOOCV method. Table 5.12 resumes the results obtained for the model validation and shows that 44% of the neck values have been predicted in the right cluster by the decision tree. The database contains 84 values for which the value for the N_{ij} criterion is under 0.35 (cluster #1) and 49 of them have been predicted by the tree as belonging to this cluster. It means that 58% of the values have been correctly estimated. Likewise, 34 values out of 75 (i.e. 45%) have been estimated as belonging to cluster #2 and 12 out of 45 (i.e. 27%) to cluster #3.

5.3.3/ DUMMY CHEST

Analysis of the clusters' material

The second sensitive part of the dummy after the neck is the chest. The values for the HIII 5th dummy have been separated into 9 clusters (only formed with the chest deflection) and the same 27 parameters have been used as predictor variables.

According Table 5.13, the form of the PAB is a factor that influences the value of the chest deflection. 66.7% of the sled tests that have a 3D Add-on airbag are located in cluster #2 (chest deflection ≤ 19.74 mm). Moreover, the chances to be in cluster #5 with a 3D shape are maximum (100%). A divided panel also improves the chances to be located

Table 5.12: Validation of the decision tree for the dummy neck (passenger, US NCAP)

		Predicted injury					
		Cluster #1	Cluster #2	Cluster #3	Cluster #4	Cluster #5	Cluster #6
Obtained injury	Cluster #1	49	19	13	2	1	0
	Cluster #2	19	34	18	1	3	0
	Cluster #3	12	17	12	2	2	0
	Cluster #4	2	3	3	0	0	0
	Cluster #5	0	1	3	0	0	0
	Cluster #6	0	0	0	0	0	0
	Prediction accuracy	44%					

in the 2 first clusters. Then a diameter for the vent hole under 40 mm has a negative influence on the chest deflection but a diameter around 70 mm should improve very well the biomechanical value. The diameter regulates the pressure into the bag which is in contact principally in contact with the chest. Higher the pressure is, worst the values for the chest deflection should be. As for the neck, the trigger time of the second stage for the PAB inflator is decisive: all the sled tests for which the second stage has been triggered between 14 and 16 ms are included in cluster #5. The other parameters of the PAB do not influence a lot the chest deflection.

Concerning the seat belt proprieties, the only possible conclusion according to Table 5.13 is the negative influence of a high load limiter value and having a DLT could help improve the chest deflection of the dummy because most of the sled tests equipped with it are located either in cluster #1 or in cluster #2. When the dummy applies a small load on the seat belt webbing, the DLT clamps onto the webbing in order to reduce the amount of webbing that can slip and then to reduce the load on the dummy chest.

Moreover, as for the neck, the dummy position influences the chest deflection, especially the distance dummy sternum - IP. The higher this distance is, the better the criterion value should be because the airbag will have more time to deploy before the dummy comes in contact with. In fact, all the sled tests for which the distance sternum - IP is above 550 mm are located in cluster #2. The distributions of the knees distance show that these parameters are not decisive. More surprisingly, the height adjustment and the webbing on spool are two parameters that are directly in contact with the dummy chest but their influence is more unclear.

Table 5.13: Sled tests' repartition for the dummy chest (passenger, US NCAP)

Variables	Cluster #1	Cluster #2	Cluster #3	Cluster #4	Cluster #5	Cluster #6	Cluster #7	Cluster #8	Cluster #9
[mm]	13.52-17.31	17.32-19.74	19.75-22.39	22.4-24.9	25-27	27.1-29.23	29.24-31.9	32-36.11	36.12-40
Number of sled tests	33	46	39	22	29	25	15	6	1
PAB									
L-shape	18.5%	22.8%	19.1%	8.0%	9.9%	11.7%	6.8%	3.1%	0.0%
Square-shape	5.3%	10.5%	15.8%	18.4%	23.7%	13.2%	10.5%	0.0%	2.6%
3D Add-on	16.7%	66.7%	16.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3D-shape	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
other	0.0%	14.3%	14.3%	28.6%	14.3%	14.3%	0.0%	14.3%	0.0%
Divided panel									
no	8.9%	15.3%	18.5%	10.8%	17.8%	15.9%	8.3%	3.8%	0.6%
yes	32.2%	37.3%	16.9%	8.5%	1.7%	0.0%	3.4%	0.0%	0.0%
Volume									
<90 L	20.0%	60.0%	0.0%	20.0%	0.0%	0.0%	0.0%	0.0%	0.0%
90-100 L	0.0%	0.0%	62.5%	25.0%	0.0%	0.0%	12.5%	0.0%	0.0%
101-110 L	0.0%	0.0%	62.5%	0.0%	37.5%	0.0%	0.0%	0.0%	0.0%
111-120 L	18.3%	24.0%	16.0%	8.6%	9.1%	13.7%	6.9%	2.9%	0.6%
>120 L	0.0%	5.3%	5.3%	21.1%	52.6%	5.3%	5.3%	5.3%	0.0%
Number of Ventholes									
1	18.3%	25.0%	21.1%	9.4%	8.9%	10.6%	5.0%	1.1%	0.6%
2	0.0%	2.8%	2.8%	13.9%	36.1%	16.7%	16.7%	11.1%	0.0%
Hole diameter 1									
<40 mm	0.0%	3.1%	0.0%	6.3%	31.3%	40.6%	12.5%	3.1%	3.1%
40-50 mm	8.5%	11.3%	32.4%	9.9%	12.7%	11.3%	9.9%	4.2%	0.0%
51-60 mm	31.2%	15.6%	14.3%	14.3%	13.0%	5.2%	3.9%	2.6%	0.0%
61-70 mm	8.6%	68.6%	14.3%	5.7%	0.0%	0.0%	2.9%	0.0%	0.0%
>70 mm	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
ABV									
no	0.0%	7.0%	9.3%	14.0%	25.6%	18.6%	14.0%	11.6%	0.0%
yes	19.1%	24.9%	20.2%	9.2%	10.4%	9.8%	5.2%	0.6%	0.6%
Tether 1									
no	9.3%	13.2%	16.6%	11.9%	19.2%	16.6%	8.6%	4.0%	0.7%

yes		29.2%	40.0%	21.5%	6.2%	0.0%	0.0%	3.1%	0.0%	0.0%
Tether 2										
no		9.6%	19.2%	18.1%	10.7%	16.4%	14.1%	7.9%	3.4%	0.6%
yes		41.0%	30.8%	17.9%	7.7%	0.0%	0.0%	2.6%	0.0%	0.0%
Number of Stages										
1		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2		15.3%	21.3%	18.1%	10.2%	13.4%	11.6%	6.9%	2.8%	0.5%
TTF 1										
<14 ms		1.4%	9.7%	0.0%	15.3%	30.6%	26.4%	13.9%	1.4%	1.4%
14-16 ms		27.4%	27.4%	31.6%	8.5%	3.4%	0.9%	0.9%	0.0%	0.0%
17-20 ms		0.0%	29.2%	4.2%	4.2%	12.5%	16.7%	12.5%	20.8%	0.0%
>20 ms		0.0%	0.0%	33.3%	0.0%	0.0%	33.3%	33.3%	0.0%	0.0%
TTF 2										
<14 ms		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
14-16 ms		0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
17-20 ms		32.7%	5.1%	8.2%	19.4%	17.3%	8.2%	7.1%	1.0%	1.0%
>20 ms		0.9%	35.0%	26.5%	2.6%	9.4%	14.5%	6.8%	4.3%	0.0%
Type of LL										
CLL		8.0%	20.0%	34.0%	10.0%	6.0%	12.0%	4.0%	6.0%	0.0%
DLL		0.0%	0.0%	0.0%	0.0%	25.0%	16.7%	41.7%	16.7%	0.0%
PLL		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
SLL		18.8%	23.4%	14.3%	11.0%	14.9%	11.0%	5.2%	0.6%	0.6%
LL Value 1										
<2.5 kN		25.0%	25.0%	0.0%	0.0%	0.0%	25.0%	25.0%	0.0%	0.0%
2.5-3.4 kN		10.3%	25.6%	38.5%	10.3%	12.8%	2.6%	0.0%	0.0%	0.0%
3.5-4.5 kN		13.7%	26.5%	13.7%	11.8%	12.7%	9.8%	9.8%	2.0%	0.0%
4.6-5.5 kN		20.0%	11.4%	14.3%	8.6%	15.7%	18.6%	4.3%	5.7%	1.4%
5.6-6.5 kN		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%
Retractor Pretensioner										
no		0.0%	50.0%	0.0%	0.0%	0.0%	0.0%	50.0%	0.0%	0.0%
yes		15.4%	21.0%	18.2%	10.3%	13.6%	11.7%	6.5%	2.8%	0.5%
Buckle Pretensioner										
no		13.8%	26.6%	14.7%	11.0%	14.7%	8.3%	6.4%	4.6%	0.0%
yes		16.8%	15.9%	21.5%	9.3%	12.1%	15.0%	7.5%	0.9%	0.9%

Left Knee - Shorter Distance Ahead <40 mm 40-49 mm 50-59 mm 60-69 mm 70-79 mm >80 mm	13.7%	11.8%	13.7%	11.8%	15.7%	15.7%	9.8%	7.8%	0.0%
	21.1%	29.8%	21.1%	5.3%	7.0%	10.5%	3.5%	1.8%	0.0%
	10.0%	23.3%	20.0%	10.0%	23.3%	6.7%	3.3%	0.0%	3.3%
	17.4%	28.3%	10.9%	13.0%	10.9%	13.0%	6.5%	0.0%	0.0%
	4.5%	9.1%	36.4%	13.6%	18.2%	9.1%	9.1%	0.0%	0.0%
	20.0%	10.0%	10.0%	0.0%	10.0%	10.0%	30.0%	10.0%	0.0%
Right Knee - Shorter Distance Ahead <40 mm 40-49 mm 50-59 mm 60-69 mm 70-79 mm >80 mm	0.0%	13.8%	17.2%	6.9%	13.8%	24.1%	13.8%	10.3%	0.0%
	14.8%	18.5%	14.8%	14.8%	14.8%	7.4%	7.4%	7.4%	0.0%
	29.5%	27.9%	19.7%	4.9%	8.2%	8.2%	1.6%	0.0%	0.0%
	16.3%	34.7%	14.3%	10.2%	14.3%	6.1%	4.1%	0.0%	0.0%
	0.0%	0.0%	21.4%	17.9%	28.6%	25.0%	3.6%	0.0%	3.6%
	13.6%	13.6%	22.7%	13.6%	4.5%	4.5%	22.7%	4.5%	0.0%
OLC++ <26 g 26-26.9 g 27-27.9 g >28 g 14.8%	8.5%	31.9%	29.8%	12.8%	6.4%	8.5%	0.0%	2.1%	0.0%
	21.2%	21.2%	21.2%	6.1%	9.1%	6.1%	9.1%	6.1%	0.0%
	14.3%	10.2%	16.3%	14.3%	24.5%	12.2%	8.2%	0.0%	0.0%
	22.2%	5.6%	9.3%	14.8%	20.4%	9.3%	1.9%	1.9%	0.0%
Tv0 <65 ms 65-69.9 ms >70 ms	15.9%	15.9%	11.0%	8.5%	18.3%	19.5%	8.5%	1.2%	1.2%
	15.2%	23.8%	21.0%	11.4%	11.4%	5.7%	7.6%	3.8%	0.0%
	13.8%	27.6%	27.6%	10.3%	6.9%	10.3%	0.0%	3.4%	0.0%
Dummy Temperature <20 °C 20-20.9 °C 21-21.9 °C >22 °C	50.0%	50.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	23.7%	28.9%	14.5%	14.5%	9.2%	2.6%	3.9%	1.3%	1.3%
	11.5%	15.9%	19.5%	8.8%	15.0%	17.7%	8.0%	3.5%	0.0%
	4.0%	20.0%	24.0%	4.0%	20.0%	12.0%	12.0%	4.0%	0.0%

Analysis of the variables' links via decision tree

After analysing the distribution of the variables into each cluster, the classification tree⁴ (Fig. 5.4) can give a hand in determining the most influential variables concerning the dummy chest. For building it, 14 predictors variables are used as splitters and only cluster #6 is not considered as terminal node. Clusters #3 and #4 are the most represented clusters with respectively 7 and 10 possibilities out of 38 to arrive at these clusters.

First of all, the analysis of the tree shows that the distance Sternum-IP ("Position 2") is the most decisive parameter and if its value is above 473.5 mm, the pretensioners (retractor and anchor) are more involved in the tree and thus in the prediction of the chest value. Cluster #1, that contains the best values of the chest deflection ($\leq 33\%$ of the legal value), is represented at only 2 terminal nodes and imposes conditions on the dummy position, load limiter value and webbing on spool of the seat belt, volume and eventually form of the PAB, depending of the value of the pulse criterion (OLC++). Additionally, the presence of a tether in the airbag also influences a lot the results and can produced both types of results, depend on the combinations with the other parameters. The tether should aim helping the airbag deploy in a defined direction for a better positioning of the bag. By going down the tree, some conditions for reaching specific deflection values can be defined.

To sum up, the parameters that influence the chest deflection are similar to those that influence the value of the neck :

- ◆ Distance Sternum-IP and more generally the dummy position;
- ◆ Volume of the PAB;
- ◆ Tether in the PAB;
- ◆ Value of the OLC++ pulse criterion;
- ◆ Load limiter value.

⁴Shifted values to avoid representing a real case

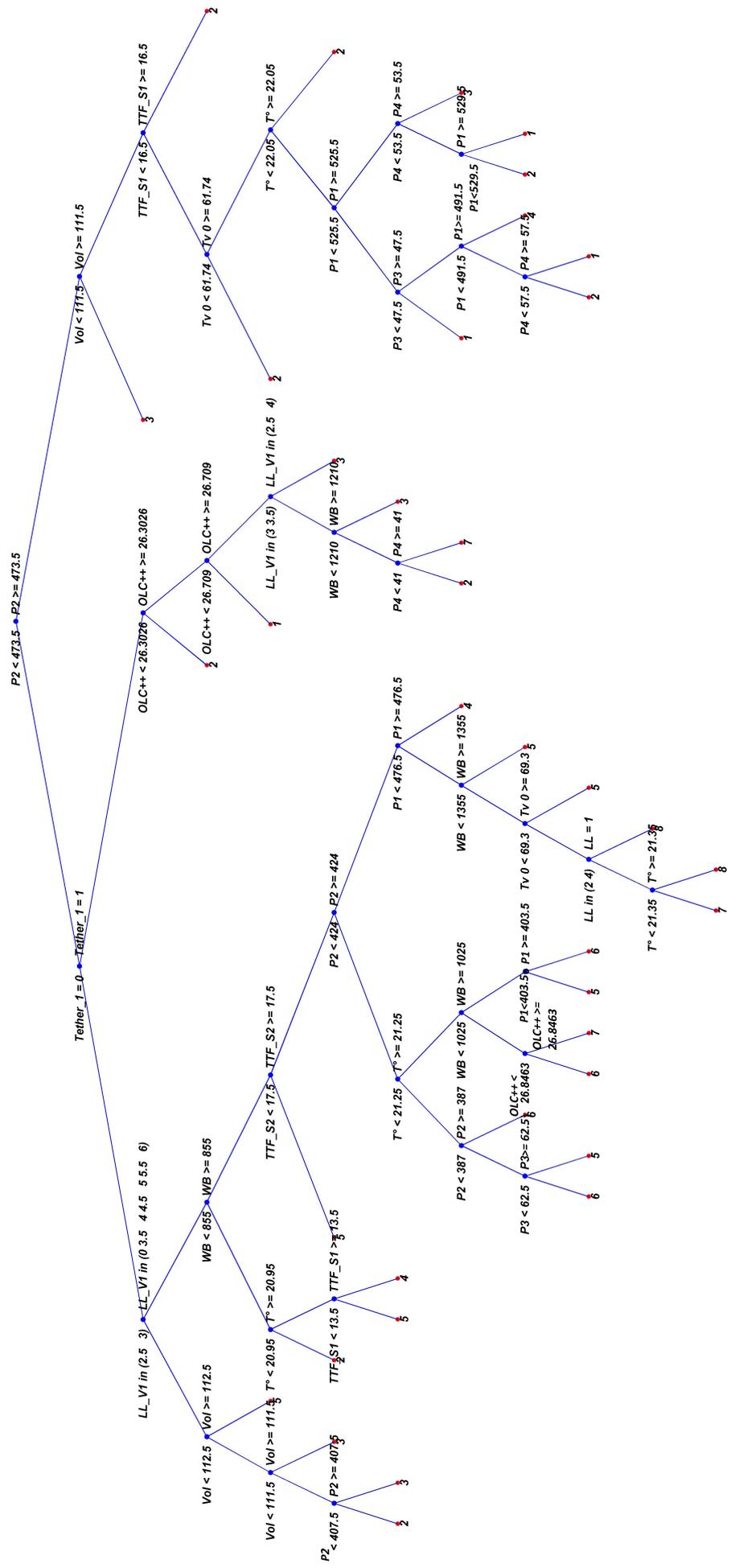


Figure 5.4: Decision tree for the dummy chest (passenger, US NCAP)

Table 5.14 supports the parameters' importance through the calculation of the VIM. It confirms that the dummy position and the airbag volume are two decisive parameters. The further the dummy is from the IP, the later the dummy comes in contact with the airbag. It can thus deploy without any contact with the dummy that could receive too much power from the inflator. If the dummy is too close to the IP, the biomechanical values, especially the chest, could be aggravated because the airbag will deploy in this dummy area. Additionally, the volume of the airbag directly influences the pressure that is applied on the dummy, that is why it is coherent that this parameter is considered as playing a role by attributing the value of the chest deflection.

Table 5.14: Calculation of the VIM for the dummy chest (passenger, US NCAP)

Predictor Variable	Variable Importance Measure
Sternum to IP	0.002094051
Chin to IP	0.001609263
Volume	0.001231309
Dummy temperature	0.001175684
OLC++	0.001069725
Webbing on Spool	0.001002823
LL Value 1	0.000940871
Right Knee. Shorter Distance Ahead	0.000797916
Tether 1	0.000778427
Tv0	0.000582996
TTF Stage 1	0.000563691
Left Knee. Shorter Distance Ahead	0.000543875
TTF Stage 2	0.000383884
Type of LL	0.000196625

Table 5.15 summarises the model validation and shows that 33% of the chest values have been predicted in the right cluster, but also that, if we consider that to be in ± 1 cluster is correct (the range for each cluster is equal to 2 or 3 mm, that is why this condition is plausible), 74% of the sled tests have been correctly predicted by the tree.

Table 5.15: Validation of the decision tree for the dummy chest (passenger, US NCAP)

Obtained injury	Predicted injury								
	Cluster #1	Cluster #2	Cluster #3	Cluster #4	Cluster #5	Cluster #6	Cluster #7	Cluster #8	Cluster #9
Cluster #1	16	12	3	0	2	0	0	0	0
Cluster #2	10	19	12	3	1	0	1	0	0
Cluster #3	5	13	15	0	4	2	0	0	0
Cluster #4	2	4	5	1	5	3	1	1	0
Cluster #5	2	0	5	3	10	5	4	0	0
Cluster #6	0	0	2	1	9	8	5	0	0
Cluster #7	1	0	2	1	3	5	2	1	0
Cluster #8	0	0	0	0	2	0	4	0	0
Cluster #9	0	0	0	0	0	1	0	0	0
Prediction accuracy	33% in the right cluster; 74% in ± 1 cluster								

5.3.4/ DUMMY FEMURS

Analysis of the clusters' material

As for the other dummy parts, an analysis of the content of each cluster has been done. It allows to have a first indication about the parameters and the repartition of the sled tests into the clusters. This distribution is summarised in Table 5.16.

First of all, concerning the PAB form, no conclusions can be drawn from the distribution. This is actually obvious because the PAB airbag aims at protecting the upper body of the human, i.e. the head/neck and the chest. The seat belt and the knee bag should more influence the results of the femur forces.

Concerning the seat belt, 91.7% of the sled tests having a DLL as load limiter received no satisfying results (the range of the obtained values is between 11% and 67% of the legal value). Moreover, a load limiter with a small force (<2.5 kN) or a too high force (>5.6 kN) would not reduce the femur forces and no having a retractor pretensioner is definitively not a good deal. The distribution for the other seat belt parameters do not show any trends.

More surprisingly is the distribution of the sled tests for the knee bag. It does not allow a distinction of the influence of the knee bag. In theory, the knee bag helps reduce the femur forces, especially for the HIII 50th dummy tested without seat belt. It means not be essential for this dummy.

To sum up, the distribution of the sled tests through the different parameters is more complex for the femur forces as for the other dummy parts. No real trends can be drawn from Table 5.16.

Table 5.16: Sled tests' repartition for the dummy femurs (passenger, US NCAP)

Variables	Cluster #1	Cluster #2	Cluster #3	Cluster #4	Cluster #5	Cluster #6
Max. left-right forces	0.7-4.6	3.1-6.3	1.4-3.0	0.5-3.3	1.2-2.4	0-2.2
Number of sled tests	28	32	2	40	16	98
PAB						
L-shape	9.3%	17.3%	0.6%	18.5%	4.9%	49.4%
Square-shape	28.9%	2.6%	2.6%	26.3%	21.1%	18.4%
3D Add-on	0.0%	33.3%	0.0%	0.0%	0.0%	66.7%
3D-shape	0.0%	33.3%	0.0%	0.0%	0.0%	66.7%
other	28.6%	0.0%	0.0%	0.0%	0.0%	71.4%
Divided panel						
no	17.8%	13.4%	1.3%	18.5%	7.0%	42.0%
yes	0.0%	18.6%	0.0%	18.6%	8.5%	54.2%
Volume						
<90 L	30.0%	10.0%	0.0%	20.0%	20.0%	20.0%
90-100 L	0.0%	0.0%	0.0%	33.3%	0.0%	66.7%
101-110 L	0.0%	33.3%	0.0%	0.0%	0.0%	66.7%
111-120 L	7.2%	16.7%	1.1%	17.2%	7.8%	50.0%
>120 L	60.0%	0.0%	0.0%	30.0%	0.0%	10.0%
Number of Ventholes						
1	3.9%	17.2%	0.0%	17.8%	8.9%	52.2%
2	58.3%	2.8%	5.6%	22.2%	0.0%	11.1%
Hole diameter 1						
<40 mm	7.7%	15.4%	0.0%	30.8%	2.6%	43.6%
40-50 mm	18.7%	16.0%	2.7%	12.0%	10.7%	40.0%
51-60 mm	13.4%	15.9%	0.0%	11.0%	6.1%	53.7%
61-70 mm	0.0%	5.0%	0.0%	50.0%	10.0%	35.0%

>70 mm	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
ABV						
no	58.1%	0.0%	4.7%	25.6%	4.7%	7.0%
yes	1.7%	18.5%	0.0%	16.8%	8.1%	54.9%
Tether 1						
no	16.6%	11.9%	1.3%	17.9%	6.0%	46.4%
yes	4.6%	21.5%	0.0%	20.0%	10.8%	43.1%
Tether 2						
no	14.1%	12.4%	1.1%	20.9%	6.2%	45.2%
yes	7.7%	25.6%	0.0%	7.7%	12.8%	46.2%
Number of Stages						
1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	13.0%	14.8%	0.9%	18.5%	7.4%	45.4%
TTF Stage 1						
<14 ms	12.1%	14.3%	1.1%	23.1%	9.9%	39.6%
14-16 ms	3.1%	19.4%	0.0%	17.3%	7.1%	53.1%
17-20 ms	54.2%	0.0%	4.2%	4.2%	0.0%	37.5%
>20 ms	33.3%	0.0%	0.0%	33.3%	0.0%	33.3%
TTF Stage 2						
<14 ms	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
14-16 ms	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%
17-20 ms	7.7%	16.8%	0.6%	14.8%	9.7%	50.3%
>20 ms	26.7%	10.0%	1.7%	26.7%	1.7%	33.3%
Type of LL						
CLL	18.0%	12.0%	4.0%	12.0%	20.0%	34.0%
DLL	91.7%	0.0%	0.0%	8.3%	0.0%	0.0%
PLL	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
SLL	5.2%	16.9%	0.0%	21.4%	3.9%	52.6%
LL Value 1						
<2.5 kN	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2.5-3.4 kN	76.9%	12.8%	7.7%	2.6%	0.0%	0.0%
3.5-4.5 kN	40.6%	37.6%	15.8%	3.0%	3.0%	0.0%
4.6-5.5 kN	12.9%	42.9%	37.1%	5.7%	1.4%	0.0%
5.6-6.5 kN	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%
Retractor Pretensioner						
no	50.0%	0.0%	50.0%	0.0%	0.0%	0.0%
yes	38.8%	35.0%	20.6%	3.7%	1.9%	0.0%
Buckle Pretensioner						
no	42.2%	36.7%	14.7%	4.6%	1.8%	0.0%
yes	35.5%	32.7%	27.1%	2.8%	1.9%	0.0%
Anchor Point Pretensioner						
no	36.9%	35.7%	22.6%	3.6%	1.2%	0.0%
yes	46.8%	29.8%	14.9%	4.3%	4.3%	0.0%
Webbing on Spool						
<750 mm	53.3%	13.3%	13.3%	6.7%	13.3%	0.0%
750-800 mm	41.9%	41.9%	9.7%	6.5%	0.0%	0.0%
801-850 mm	60.6%	30.3%	9.1%	0.0%	0.0%	0.0%
851-900 mm	41.9%	32.3%	22.6%	3.2%	0.0%	0.0%
901-950 mm	54.5%	18.2%	13.6%	9.1%	4.5%	0.0%
951-1000 mm	29.6%	44.4%	22.2%	0.0%	3.7%	0.0%
>1000 mm	17.5%	42.1%	36.8%	3.5%	0.0%	0.0%
Height Adjustment						
no	58.2%	24.1%	16.5%	0.0%	1.3%	0.0%
yes	27.7%	40.9%	23.4%	5.8%	2.2%	0.0%
DLT						
no	38.8%	34.0%	21.8%	3.9%	1.5%	0.0%
yes	50.0%	33.3%	0.0%	0.0%	16.7%	0.0%
KnAB						
no	31.0%	37.2%	23.9%	4.4%	3.5%	0.0%

yes	47.6%	32.0%	17.5%	2.9%	0.0%	0.0%
Chin to IP						
<420 mm	25.6%	48.7%	25.6%	0.0%	0.0%	0.0%
420-470 mm	53.9%	23.6%	15.7%	4.5%	2.2%	0.0%
471-500 mm	44.0%	32.0%	16.0%	4.0%	4.0%	0.0%
501-550 mm	10.5%	50.0%	34.2%	5.3%	0.0%	0.0%
>550 mm	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Sternum to IP						
<420 mm	46.5%	29.3%	19.2%	2.0%	3.0%	0.0%
420-470 mm	34.1%	34.1%	24.4%	4.9%	2.4%	0.0%
471-500 mm	51.3%	33.3%	10.3%	5.1%	0.0%	0.0%
501-550 mm	11.1%	52.8%	30.6%	5.6%	0.0%	0.0%
>550 mm	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%
Left Knee - Shorter Distance Ahead						
<40 mm	21.6%	45.1%	25.5%	3.9%	3.9%	0.0%
40-49 mm	26.3%	40.4%	29.8%	3.5%	0.0%	0.0%
50-59 mm	56.7%	36.7%	0.0%	6.7%	0.0%	0.0%
60-69 mm	45.7%	23.9%	23.9%	4.3%	2.2%	0.0%
70-79 mm	72.7%	18.2%	9.1%	0.0%	0.0%	0.0%
>80 mm	37.5%	37.5%	12.5%	0.0%	12.5%	0.0%
Right Knee - Shorter Distance Ahead						
<40 mm	20.7%	48.3%	27.6%	3.4%	0.0%	0.0%
40-49 mm	29.6%	37.0%	18.5%	7.4%	7.4%	0.0%
50-59 mm	24.6%	42.6%	29.5%	3.3%	0.0%	0.0%
60-69 mm	51.0%	30.6%	14.3%	4.1%	0.0%	0.0%
70-79 mm	53.6%	21.4%	17.9%	3.6%	3.6%	0.0%
>80 mm	68.2%	18.2%	9.1%	0.0%	4.5%	0.0%
OLC++						
<26 g	40.4%	29.8%	23.4%	2.1%	4.3%	0.0%
26-26.9 g	42.4%	28.8%	22.7%	6.1%	0.0%	0.0%
27-27.9 g	34.7%	51.0%	12.2%	2.0%	0.0%	0.0%
>28 g	37.0%	31.5%	24.1%	3.7%	3.7%	0.0%
Tv0						
<65 ms	39.0%	36.6%	19.5%	2.4%	2.4%	0.0%
65-69.9 ms	43.8%	31.4%	20.0%	3.8%	1.0%	0.0%
>70 ms	20.7%	41.4%	27.6%	6.9%	3.4%	0.0%
Dummy Temperature						
<20 °C	0.0%	50.0%	50.0%	0.0%	0.0%	0.0%
20-20.9 °C	38.2%	32.9%	22.4%	5.3%	1.3%	0.0%
21-21.9 °C	37.2%	37.2%	21.2%	2.7%	1.8%	0.0%
>22 °C	52.0%	28.0%	12.0%	4.0%	4.0%	0.0%

Analysis of the variables' links via decision tree

The decision tree⁵ represented in Fig. 5.5 allows some conclusions on the parameters influencing the femur forces of the HIII 5th dummy.

⁵Shifted values to avoid representing a real case

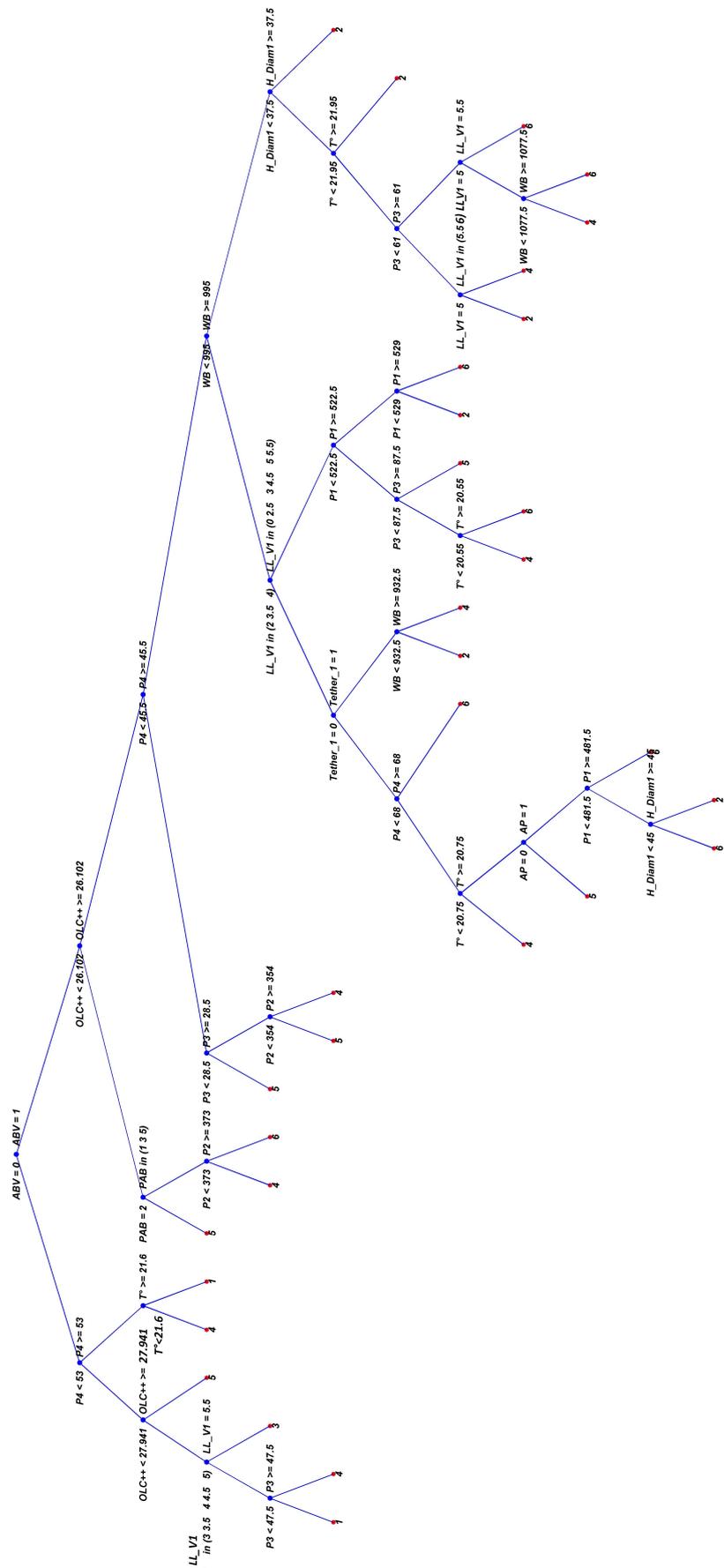


Figure 5.5: Decision tree for the dummy femurs (passenger, US NCAP)

According to the decision tree, an Additional Bag Vent (ABV) that is triggered or not (most often triggered by the HIII 5th dummy in order to reduce the pressure into the airbag) seems to be important for evaluating the femurs forces but not essential because if the airbag does not have an ABV, the results do not exceed anyway 67% of the legal value. This decision tree contradicts the distribution of the sled tests (Table 5.16) that seems to indicate the non importance of the PAB concerning the femurs. Many combinations are then available for reaching the best clusters (i.e. clusters #5 and #6) if an ABV is added on the PAB and the following decisive parameters are mainly parameters concerning the dummy position and the car evaluation. 31 variables have been selected to form this tree and only 13 out of 31 are variables concerning either the passenger airbag (5 variables) or the seat belt (8 variables). It confirms that the PAB globally does not have an influence on the dummy legs. This decision tree is supported by the calculation of the VIM (Table 5.17). This method appoints the ABV as the most important parameter concerning the dummy legs.

It also confirms the importance of the dummy position and external parameters as the OLC++ criterion. The first parameter concerning the PAB is only ordered in the 10th place of the rank.

Table 5.17: Calculation of the VIM for the dummy femurs (passenger, US NCAP)

Predictor Variable	Variable Importance Measure
ABV	0.003051477
OLC++	0.002513954
Right Knee. Shorter Distance Ahead	0.002141061
Webbing on Spool	0.001524438
LL Value 1	0.001483736
Dummy temperature	0.001406577
Left Knee. Shorter Distance Ahead	0.001308783
Chin to IP	0.000996686
Tether 1	0.000697923
Hole Diameter 1	0.000537344
Sternum to IP	0.000506055
KnAB	0.000429752
PAB	0.000280415

To sum up, three different methods - the sled tests' distribution, the decision tree and the VIM - enables to conclude that the passenger airbag does not influence a lot the femurs forces of the HIII 5th dummy. As explained before, this is coherent with the reality that gives as first priority of the PAB the protection of the upper dummy part, i.e. the head/neck and the chest. More surprisingly, the knee bag, which should protect the legs, is not decisive for this dummy. In fact, even if the dummy seat is positioned on the forwards position, the energy built-up by the dummy is surely high but not enough to impose having a knee bag. This restraint system is definitely necessary for FMVSS NR.208 but its influence is more visible on sled tests carried out without seat belt. Finally, the influential parameters for the HIII 5th dummy legs in the US NCAP rating are:

- ◆ Presence of an ABV;
- ◆ Dummy position;
- ◆ Value of the OLC++;

- ◆ Webbing on spool of the seat belt.

Table 5.18: Validation of the decision tree for the dummy femurs (passenger, US NCAP)

		Predicted injury					
		Cluster #1	Cluster #2	Cluster #3	Cluster #4	Cluster #5	Cluster #6
Obtained injury	Cluster #1	49	19	13	2	1	0
	Cluster #2	19	34	18	1	3	0
	Cluster #3	12	17	12	2	2	0
	Cluster #4	2	3	3	0	0	0
	Cluster #5	0	1	3	0	0	0
	Cluster #6	0	0	0	0	0	0
	Prediction accuracy	63%					

5.4/ CONCLUSION

This chapter proposes the application of a famous method, the CART, on data coming from sled tests. The novelty of this chapter is the employment of the decision trees with parameters of restraint systems. In fact, this doctoral research is focused on an industrial demand which consists in finding relationships with the available data between the parameters of seat belt and airbags principally and the dummy biomechanical values. It aims at understand the effects of some parameters on specific dummy parts and helps improve the car safety.

This chapter is focused on the two main ratings: the EURO NCAP and the US NCAP ratings. The legal requirements are in any ways completed before closing the project. These two ratings, especially the US NCAP, are the main challenges for the passive safety engineers.

The EURO NCAP is the easiest one because the European legislation uses the same belted dummy on the passenger side, a HIII 50th dummy. It means that the modifications done on the restraint systems for the rating are only few consequences on the results for the legislation. On the contrary, the FMVSS NR.208 imposes the achievement of four configurations and one of those is used to evaluate the car safety: the HIII 5th dummy belted (passenger side). The eventual modifications proposed by the decision trees have automatically an influence on the results of the other configurations and thus have to be considered carefully.

Table 5.19 resumes the most influential parameters for each rating and dummy parts, proposed by the decision trees and the VIM.

Table 5.19: Resume of the four most influential parameters

Dummy part	US NCAP			
	Euro NCAP	Head	Neck	Chest
Decision tree	Chest	Number of ventholes Dummy position OLC++ LL value	LL value TTF 2nd inflator stage Dummy position Dummy position	Dummy position Volume of the PAB Tether in the PAB OLC++
	Femurs	TTF 1st inflator stage Volume of the PAB Dummy position Dummy temperature	LL value Webbing on spool OLC++ Dummy position	ABV Dummy position OLC++ Webbing on spool
VIM	Head	OLC++ LL Value Volume of the PAB Number of ventholes	LL value Webbing on spool OLC++ Dummy position	Dummy position Volume of the PAB Dummy temperature
	Femurs	TTF 1st inflator stage Dummy temperature Dummy position Dummy position	LL value Webbing on spool OLC++ Dummy position	ABV OLC++ Dummy position Webbing on spool

The two ratings have to be considered separately. For the EURO NCAP rating, the results given by the classification tree are very similar that those proposed by the VIM. The parameter that influences the most the dummy chest according to both methods is the time to fire of the first stage of the PAB inflator. This is not surprisingly because the triggered time imposes the time of deploying the PAB. If this time is too early, the pressure in the bag is already maximum when the dummy comes in contact with the bag, what could involve a high chest deflection value. If the inflator is triggered to late, the dummy has already begun its displacement forwards and comes in contact with the bag when it is deploying. It means that the bag pushes on the dummy chest to have enough place to deploy itself fully. A specific time cannot be estimated through the decision tree. The dummy position and temperature also seem to influence the chest value of the HIII 50th dummy. Finally, the parameter that is "easy" to modify and that can improve the results of the dummy chest is the volume of the PAB.

The second rating in which the engineers have some interests is the US NCAP rating on the passenger side, i.e. with a HIII 5th belted dummy on the passenger side. According to Table 5.19, the results obtained by the decision trees and the VIM are quite similar. Some parameters have been selected as modifying the results of the biomechanical values for each part of the dummy:

- ◆ Dummy position;
- ◆ The value of the OLC++;
- ◆ Value of the load limiter of the seat belt for the dummy head and neck.

The two first parameters are however not parameters of restraint systems and cannot be easily modified because the first one is defined by the legislation and the second depends on the car type. The third is a parameter that can be changed by the engineers but its influence on the head/neck and not on the chest is quite surprising, such as the fact that a presence of a ABV could modify the femurs forces. The main challenges for this rating are the improvement of the values for the neck and the chest and according to the different Tables, some characteristics of the restraint systems (form of the PAB, vent hole diameter, time to fire of the second inflator stage) seem to influence the dummy neck and chest for the HIII 5th dummy in a US NCAP rating sled test.

In conclusion, these methods help identify some parameters as influencing the biomechanical values in a specific test configuration. Those discovered for the HIII 50th dummy in a EURO NCAP test are near to the expert knowledge and their modifications will not affect any legislation results. On the contrary, the modifications proposed by the decision trees for the HIII 5th dummy are quite surprising, except those for the neck and chest. Some modifications have to be considered carefully to avoid a conflict with another test configuration. The dummy position seems to influence each dummy part but is hardly modifiable.

To sum up, the decision trees are very useful to learn on the already carried out tests and try to find some similarities but for the US NCAP rating, it did not really proposed some parameters from the restraint systems influencing the results.

GENERAL CONCLUSION

The traffic volume on the roads is continually growing and therewith the number of traffic accidents. Therefore, the automobile safety is one of the actual major concern. In addition to the country legislation, ratings are carried out by independent assessments to evaluate car safety. Consequently automotive suppliers continually develop new restraint passive systems to reduce the occupant damages during a crash. The injury severity is controlled through high tech dummies during sled tests in order to simulate the injury damages of the car occupant during an impact. This doctoral research is based on an industrial demand of an automobile firm which presented some interests in understanding the relationships between the parameters of restraint systems (frontal airbags, seat belt and knee bag) and the dummy biomechanical values. The thesis is focused on frontal impact crash tests.

The first step of this thesis has been focused on the literature about data mining methods used in automobile domain, especially with crash data. This research enabled to define the state-of-the-art of the actual methods and processes used in this specific domain. However, most of the publications with crash data have been oriented on the circumstances of the accidents, e.g. the gender of the occupant and the weather conditions. Even if some methods proposed by the authors are applicable, the first observation that can be done after the publications' study is the among of available data for this doctoral research. In fact, even if lots of sled tests have been carried out in different test configurations, only for some of them, all necessary information were available. A work on the data had to be thus done before applying any data mining methods, that is why the steps of the Knowledge Discovery in Database have been followed. The first one allows the selection of the parameters: thanks to an expert elicitation, 63 parameters of the restraint systems (driver / passenger airbag, knee bag, seat belt, environment) have been identified as influencing the dummy biomechanical values during a crash. Some data were missing and for that a dedicated simple algorithm has been developed. Once the data has been selected, replacing if necessary and transformed, the data mining method called k-NN enabled us to have a first classification of the crash data through a ranking of their rating scores (EURO NCAP and US NCAP rating tests). It aimed at understanding and optimizing the passive systems before running a test by comparing the configuration of the already carried out sled tests. The step of evaluation and validation of the model has been proceeded by the "leave-one out cross validation" method. The EURO NCAP rating data on the passenger side has been used as database for the model evaluation: four validations have been conducted. The first one evaluates the estimation of a missing value replaced by the algorithm which shows promising results even if some values are estimated far away from real values. These inconsistencies can be solved by adding new tests in the database. The second validation consists of evaluating the sled tests classi-

fication with the aim of comparing the sled tests classification with and without missing data. It has shown that the first-NN is identical in both cases for 75% of the cases. It confirms the potential of the algorithm responsible of the replacement of missing data. The third and fourth validations are focused on the estimation of the EURO NCAP score without and with missing data respectively. It showed that the majority of the estimated scores are close to the reality. Finally, this second task of the doctoral thesis gave a hand in understanding and optimizing the passive safety by comparing, before running the sled test, the systems parameters in order to adapt it to the desired rating score. The proposed method showed its limits when the database does not contain enough sled tests.

After this rating score's estimation, the doctoral researches have been focused on the relationships between the restraint systems parameters and the dummy biomechanical values. Before finding these relationships, a work on the dummy results enabled gather it into clusters. This step has been done with the help of the GK algorithm, combined with experts knowledge. According to the test configuration, the data has been gathered into 2 to 10 clusters: the values of the ratings configurations have been mainly gathered into more clusters as for the legislation tests in order to taken any little changes that could influence the final rating score into account. This step of clustering has been done for both car occupants, the driver and the passenger, and for both legislation and ratings presented in this manuscript. The analysis of the clusters enabled to conclude that the formed groups can be considered as classes, what allowed the use of a classification algorithm to find the relationships and the links between the inputs (parameters of restraint systems) and the outputs (dummy biomechanical values classed into levels of injury severity).

The last step of this thesis was thus focused on the classification methods. A powerful data mining method in the automobile domain is the decision trees, combined with the Gini split criterion. They have been applied on the data, focused only on the passenger side. In fact, the data preparation showed that not enough sled tests (under 100 sled tests per configuration) have been collected and a study on it could not produce solid results. That is why the last task has been focused only on the passenger side and moreover only on the two ratings (main challenges for the engineers). In supporting the decision trees, an additional ranking of the importance of the parameters has been done by calculation their Variable Importance Measure.

The EURO NCAP rating is the easiest one because of the use of only one belted occupant, the HIII 50th dummy. Moreover, the eventual modifications done for this rating will not influence the results of the legislation because of a very similar test configuration. The analysis has been also focused on the dummy chest because of its sensibility. The study of the database showed that the majority of the carried out sled tests obtained without difficulties the 4 points for the head and the neck and because of an absence of a knee bag for the European market, the femurs have not been considered. For the dummy chest, the time to fire of the first PAB inflator stage, that regulates the pressure in the bag during the first contact with the dummy chest, seems to influence the level of severity injury. This pressure can be also modified through a modification of the PAB volume (also defined by both methods as influencing the chest deflection value). Finally, external parameters, the dummy position and temperature, have been also point out as impressionable parameters but their modifications are not so flexible and are controlled by the legislation.

The second rating is the US NCAP rating. This is a part of the FMVSS NR.208 and uses the HIII 5th belted dummy as passenger. Because of their importance in reaching the 5 stars in the US NCAP rating, the neck and the chest have been in particular studied.

It enables to define some characteristics of the restraint systems (form of the PAB, vent hole diameter, time to fire of the second inflator stage) as influencing the dummy neck and chest for the HIII 5th dummy. The parameters selecting as improving the results for the four dummy parts were parameters as the value of the OLC++ criterion or the dummy position that are modifiable only with difficulties because of the legislation and the car vehicle. Moreover, for this rating, the modification of parameters must be considered in the other FMVSS NR.208 configurations.

This study demonstrated that CART models are good alternatives for analysing key factors influencing dummy biomechanical values. It represents an important methodological step in exploring the factors affecting the outputs. The obtained results provided valuable underlying relationships between input factors and their outputs.

To sum up, this doctoral study was based on an industrial demand which was focused on the restraint systems of the car (airbag, knee bag, seat belt), especially on their links with the level of severity injury of the car occupant during a crash. The first results have been produced with the help of the k-NN method which enables to classify a new sled test with the database according to its input configuration. It gives also a first indication on the expected results. The second and third tasks defined the parameters that influence the most the dummy biomechanical values for each rating. This study helped the engineers understand the behaviour of the restraint systems and their role in attributing the score. These methods, implemented in a GUI for a better use, are a considerable help for learning on the already carried out sled tests, analysing the passive systems before running a new one, optimizing the influence of passive restraint systems' parameters on occupant's injury severity and further reducing the number of carried out sled tests.

In term of future works, an upgrading of the database and the removing of parameters which have non-significant influences could lead to a better decision three models. The accuracy of the model deserves also to be proved by applying the constructed tree. Finally other decision trees building algorithms can also be tested to confirm the strengthness of the CART model compared to them or provide other interesting results.

The graphical user interface shall be used by the engineers in order to increase the database and thus the accuracy of the results. These results can be also compared in the future to the results obtained in simulation in order to improve the models. The Variational Bayesian methods could be also an alternative to the clustering and would enable add some conditions (or *a priori*), such as the limits of the biomechanical values which depend on the legislation. The publications of Sutton-Charani et al. [69] [74] on decision trees with uncertain data could be very useful to develop new algorithms with the actual database.

BIBLIOGRAPHY

- [1] ZADEH, L. A. **Fuzzy sets**. *Information and Control* 8 (1965), 338–353.
- [2] DAVIES, D. L., AND BOULDIN, D. W. **A cluster separation measure**. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2 (1979), 224–227.
- [3] BEZDEK, J. C. **Pattern Recognition with Fuzzy Objective Function Algorithms**. Kluwer Academic Publishers, Norwell, MA, USA, 1981.
- [4] BEZDEK, J. C., EHRLICH, R., AND FULL, W. **Fcm: The fuzzy c-means clustering algorithm**. *Computers & Geosciences* 10, 2 (1984), 191–203.
- [5] BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A., AND STONE, C. J. **Classification and Regression Trees**. Wadsworth International Group, Belmont, CA, 1984.
- [6] RICHTER, B., APPEL, H., AND HOEFS, R. **Entwicklung von PKW im Hinblick auf einen volkswirtschaftlich optimalen Insassenschutz: Das Projekt wurde in der Zeit vom 1.8.1980-31.3.1984 durchgeführt**. 1985.
- [7] QUINLAN, J. R. **Induction of decision trees**. *MACH. LEARN* 1 (1986), 81–106.
- [8] GATH, I., AND GEVA, A. B. **Unsupervised optimal fuzzy clustering**. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 7 (1989), 773–780.
- [9] XIE, X. L., AND BENI, G. **A validity measure for fuzzy clustering**. *IEEE Trans. Pattern Anal. Mach. Intell.* 13, 8 (Aug. 1991), 841–847.
- [10] KRISHNAPURAM, R., AND FREG, C. **Fitting an unknown number of lines and planes to image data through compatible cluster merging**. *Pattern Recognition* 25, 4 (1992), 385–400.
- [11] KAYMAK, U., AND BABUŠKA, R. **Compatible cluster merging for fuzzy modelling**. In *Fuzzy Systems, 1995. International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium.*, *Proceedings of 1995 IEEE Int* (1995), vol. 2, IEEE, pp. 897–904.
- [12] PAL, N. R., AND BEZDEK, J. C. **On cluster validity for the fuzzy c-means model**. *Fuzzy Systems, IEEE Transactions on* 3, 3 (1995), 370–379.
- [13] FAYYAD, U., PIATETSKY-SHAPIRO, G., AND SMYTH, P. **From data mining to knowledge discovery in databases**. *AI magazine* 17, 3 (1996), 37–54.
- [14] FAYYAD, U. M., PIATETSKY-SHAPIRO, G., AND SMYTH, P. **The kdd process for extracting useful knowledge from volumes of data**. pp. 27–34.
- [15] BERRY, M. J., AND LINOFF, G. **Data Mining Techniques: For Marketing, Sales, and Customer Support**. John Wiley & Sons, Inc., New York, NY, USA, 1997.

- [16] FRALEY, C., AND RAFTERY, A. E. **How many clusters? which clustering method? answers via model-based cluster analysis.** *THE COMPUTER JOURNAL* 41, 8 (1998), 578–588.
- [17] MOUSSEAU, V., AND SLOWINSKI, R. **Inferring an electre tri model from assignment examples.** *Journal of global optimization* 12, 2 (1998), 157–174.
- [18] JAIN, A. K., MURTY, M. N., AND FLYNN, P. J. **Data clustering: A review.** *ACM Comput. Surv.* 31, 3 (september 1999), 264–323.
- [19] ESTIVILL-CASTRO, V., AND YANG, J. **Fast and robust general purpose clustering algorithms.** In *PRICAI 2000 Topics in Artificial Intelligence*. Springer, 2000, pp. 208–218.
- [20] BABUŠKA, R. **Fuzzy and neural control disc course lecture notes (october 2001).** 55–72.
- [21] HALKIDI, M., BATISTAKIS, Y., AND VAZIRGIANNIS, M. **Clustering validity checking methods: Part ii.** pp. 19–27.
- [22] MAGIDSON, J., AND VERMUNT, J. **Latent class models for clustering: A comparison with k-means.** *Canadian Journal of Marketing Research* 20, 1 (2002), 36–43.
- [23] MAULIK, U., AND BANDYOPADHYAY, S. **Performance evaluation of some clustering algorithms and validity indices.** *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24, 12 (2002), 1650–1654.
- [24] R. BABUSKA, P.J. VAN DER VEEN, U. K. **Improved covariance estimation for gustafson-kessel clustering.** pp. 1081–1085.
- [25] ALEXANDROV, V., DONGARRA, J., JULIANO, B., RENNER, R., AND TAN, C. **Computational Science - ICCS 2001: International Conference, San Francisco, CA, USA, May 28-30, 2001. Proceedings.** No. Teil 2 in Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2003.
- [26] BATISTA, G. E. A. P. A., AND MONARD, M. C. **An analysis of four missing data treatment methods for supervised learning.** pp. 519–533.
- [27] GABER, M. M., KRISHNASWAMY, S., AND ZASLAVSKY, A. **Adaptive mining techniques for data streams using algorithm output granularity mohamed.** In *Workshop (AusDM 2003), Held in conjunction with the 2003 Congress on Evolutionary Computation (CEC 2003)* (2003), Springer Verlag.
- [28] GEURTS, K., BRIJS, T., AND WETS, G. **Clustering and profiling traffic roads by means of accident data.** 1–2.
- [29] TSENG, S., HOWANG, K., AND C.LEE. **A pre-processing method to deal with missing values by integrating clustering and regression techniques.** *Taylor & Francis* (2003).
- [30] CHONG, M. M., ABRAHAM, A., AND PAPRZYCKI, M. **Traffic accident analysis using decision trees and neural networks.** *CoRR cs.AI/0405050* (2004).

- [31] ROSSITER, D. **Technical note: Statistical methods for accuracy assessment of classified thematic maps.** Enschede (NL): International Institute for Geo-information Science & Earth Observation (ITC) URL [http://www. itc. nl/personal/rossiter/teach/R/R.ac. pdf](http://www.itc.nl/personal/rossiter/teach/R/R.ac.pdf) 25, 92 (2004), 107.
- [32] MAIMON, O., AND ROKACH, L., Eds. **The Data Mining and Knowledge Discovery Handbook.** Springer, 2005.
- [33] CHAVENT, M. **Sur la normalisation pour la classification de données intervalles.**
- [34] CHONG, M., ABRAHAM, A., AND PAPRZYCKI, M. **Traffic accident data mining using machine learning paradigms.** *Informatica (Slovenia) - INFORMATICASI* 29, 1 (2005), 89–98.
- [35] KOVÁCS, F., LEGÁNY, C., AND BABOS, A. **Cluster validity measurement techniques.** In *6th International symposium of hungarian researchers on computational intelligence* (2005), Citeseer.
- [36] KRISHNASWAMY, S., LOKE, S. W., RAKOTONIRAINY, A., HOROVITZ, O., AND GABER, M. M. **Towards situation-awareness and ubiquitous data mining for road safety: Rationale and architecture for a compelling application.** In *Proceedings of Conference on Intelligent Vehicles and Road Infrastructure* (2005), pp. 16–17.
- [37] CARUANA, R., AND NICULESCU-MIZIL, A. **An empirical comparison of supervised learning algorithms.** In *Proceedings of the 23rd International Conference on Machine Learning* (New York, NY, USA, 2006), ICML '06, ACM, pp. 161–168.
- [38] CHANG, L.-Y., AND WANG, H.-W. **Analysis of traffic injury severity: An application of non-parametric classification tree techniques.** *Accident Analysis & Prevention* 38, 5 (2006), 1019–1027.
- [39] CICHOS, D., DE VOGEL, D., OTTO, M., SCHAAR, O., AND ZÖLSCH, S. **Crash analysis criteria description.** *Workgroup Data Processing Vehicle Safety* (2006).
- [40] THEODORIDIS, S., AND KOUTROUMBAS, K. **Pattern Recognition (3rd edition).** Academic Press, Inc., Orlando, FL, USA, 2006.
- [41] HADDADIN, S., ALBU-SCHFFER, A., AND HIRZINGER, G. **Dummy crash-tests for the evaluation of rigid human-robot impacts.** In *IARP-IEEE/RAS-EURON Workshop on Technical Challenges for Dependable Robots in Human Environments* (2007).
- [42] DEPARTMENT OF TRANSPORTATION, N. **Consumer information; new car assessment program; notice (part II).** Tech. Rep. DOT HS 810 698, National Highway Traffic Safety Administration, Washington, DC 20590, 2008.
- [43] KRAMER, F., FRANZ, U., GÖRING, T., LORENZ, B., STEFFAN, H., AND SCHÖNEBURG, R. **Passive Sicherheit von Kraftfahrzeugen: Biomechanik - Simulation - Sicherheit im Entwicklungsprozess.** ATZ/MTZ-Fachbuch. Vieweg+Teubner Verlag, 2008.

- [44] NHTSA. **Sign conventions for part 572 subpart e test dummies**. Tech. Rep. TP208-14, NHTSA, Washington, DC 20590, 2008.
- [45] WU, X., KUMAR, V., QUINLAN, J. R., GHOSH, J., YANG, Q., MOTODA, H., MCLACHLAN, G. J., NG, A. F. M., LIU, B., YU, P. S., ZHOU, Z.-H., STEINBACH, M., HAND, D. J., AND STEINBERG, D. **Top 10 algorithms in data mining**. *Knowledge and Information Systems* 14, 1 (2008), 1–37.
- [46] HILMANN, J. **On the development of a process chain for structural optimization in vehicle passive safety**. PhD thesis, Technische Universität Berlin, Berlin, 2009.
- [47] LIU, H.-C., JENG, B.-C., YIH, J.-M., AND YU, Y.-K. **Fuzzy c-means algorithm based on standard mahalanobis distances**. In *Proceedings of the 2009 International Symposium on Information Processing* (2009), pp. 422–427.
- [48] PIRDAVANI, A., BRIJS, T., AND WETS, G. **A multiple criteria decision making approach for prioritizing accident hotspots in developing countries in the absence of crash data**. *Transport reviews* 30, 1 (2009), 97–113.
- [49] SOHR, S., AND HEYM, A. **Benefit of adaptive occupant restraint systems with focus on the new us ncap rating requirements**. In *World Automotive Congress of the Scientific Society for Mechanical Engineering (GTE) and International Federation of Automotive Engineering Societies (FISITA)* (Budapest, 2009).
- [50] HAIDAR, M. **Méthode multicritre d'aide la dcision**. 2010.
- [51] KASHANI, A., SHARIAT-MOHAYMANY, A., AND RANJBARI, A. **A data mining approach to identify key factors of traffic injury severity**. *Promet-Traffic & Transportation* 23 (2010), 11–17.
- [52] NAFI, A., AND WEREY, C. **Aide à la décision multicritère: introduction aux méthodes danalyse multicritère de type electre**. 2010.
- [53] TIAN, R., YANG, Z., AND ZHANG, M. **Method of road traffic accidents causes analysis based on data mining**. In *Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on* (2010), IEEE, pp. 1–4.
- [54] HAGHIGHAT, F. **Application of a multi-criteria approach to road safety evaluation in the bushehr province, iran**. *PROMET-Traffic&Transportation* 23, 5 (2011), 341–352.
- [55] NAYAK, R., EMERSON, D., WELIGAMAGE, J., AND PIYATRAPOOMI, N. **Road crash proneness prediction using data mining**. In *EDBT* (2011), pp. 521–526.
- [56] BEHZADIAN, M., OTAGHSAR, S. K., YAZDANI, M., AND IGNATIUS, J. **A state-of-the-art survey of topsis applications**. pp. 13051–13069.
- [57] KHORASANI, G., YADOLLAHI, A., RAHIMI, M., AND TATARI, A. **Implementation of mcdm methods in road safety management**. In *International Conference on Transport, Civil, Architecture and Environment engineering (ICTCAEE2012) December* (2012), pp. 26–27.

- [58] LEUNG, Y., WONG, M. H., WONG, K.-C., 0048, W. Z., AND LEUNG, K.-S. **A novel web-based system for tropical cyclone analysis and prediction.** *International Journal of Geographical Information Science* 26, 1 (2012), 75–97.
- [59] PROGRAMME, E. N. C. A. **Assessment protocol overall rating.** Tech. Rep. Version 6.0, EUROPEAN NEW CAR ASSESSMENT PROGRAMME, Brussels, 1050, 2012.
- [60] ABELLÁN, J., LÓPEZ, G., AND DE OÑA, J. **Analysis of traffic accident severity using decision rules via decision trees.** *Expert Syst. Appl.* 40, 15 (Nov. 2013), 6047–6054.
- [61] ARAL-AG. **Aral Studie, Trends beim Autokauf.** Aral, 2013.
- [62] CHANG, L.-Y., AND CHIEN, J.-T. **Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model.** *Safety science* 51, 1 (2013), 17–22.
- [63] CHIOU, Y.-C., LAN, L. W., AND CHEN, W.-P. **A two-stage mining framework to explore key risk conditions on one-vehicle crash severity.** *Accident Analysis & Prevention* 50 (2013), 405–415.
- [64] CRUZ, A. **Cluster algorithms.** 2013.
- [65] JOHANNSEN, H. **Unfallmechanik und Unfallrekonstruktion: Grundlagen der Unfallaufklärung.** ATZ/MTZ-Fachbuch. Springer Fachmedien Wiesbaden, 2013.
- [66] LÓPEZ, J. D. O. N., MALDONADO, G. L., AND MULERO, J. A. **Extracting decision rules from police accident reports through decision trees.** 1151–1160.
- [67] MOHAMED, M. G., SAUNIER, N., MIRANDA-MORENO, L., AND UKKUSURI, S. V. **A clustering regression approach: A comprehensive injury severity analysis of pedestrian-vehicle crashes in new york, us and montreal, canada.** *Safety Science* 54 (2013), 27–37.
- [68] ROLLAND, A. **Analyse multicritère pour laide la décision.** 2013.
- [69] SUTTON-CHARANI, N., DESTERCKE, S., AND DENOEU, T. **Learning decision trees from uncertain data with an evidential EM approach.** In *12th International Conference on Machine Learning and Applications, ICMLA 2013, Miami, FL, USA, December 4-7, 2013, Volume 1* (2013), pp. 111–116.
- [70] VERKI, M. R. M., YADOLLAHI, A., KHORASANI, G., TATARI, A., TAHERIAN, H., MOTAMED, H., KHORASANI, M., FAZELPOUR, S., AND MIRMOHAMMADI, F. **Prioritization of suburban road safety plans in iran.** *Journal of American Science* 9, 7s (2013).
- [71] BALAJI, K., AND KUMAR, V. S. **Multicriteria inventory abc classification in an automobile rubber components manufacturing industry.** *Procedia CIRP* 17 (2014), 463–468.
- [72] EURONCAP. **The official site of the european new car assessment programme,** 2014. www.euroncap.com.

- [73] PORTAL, E. U. O. D. **Persons killed in road accidents by sex (care data)**, Januar 2014.
- [74] SUTTON-CHARANI, N., DESTERCKE, S., AND DENOEU, T. **Training and evaluating classifiers from evidential data: Application to E 2 M decision tree pruning**. In *Belief Functions: Theory and Applications - Third International Conference, BELIEF 2014, Oxford, UK, September 26-28, 2014. Proceedings* (2014), pp. 87–94.
- [75] CARHS. **Safety Companion**. CARHS GMGH, Alzenau, Germany, 2015. www.carhs.de.
- [76] IRAGAVARAPU, V., LORD, D., AND FITZPATRICK, K. **Analysis of injury severity in pedestrian crashes using classification regression trees**. In *Transportation Research Board 94th Annual Meeting* (2015), no. 15-2926.
- [77] UETUG, G., KABAKC, N., BEKDIKHAN, O. B., AND AKYUEREK, B. **Multi-criteria decision making-based comparison of power source technologies for utilization in automobiles**. *Journal of Clean Energy Technologies* 3, 3 (2015).
- [78] WIKIPEDIA. **Euro ncap**, Nov. the 5th, 2014.
- [79] ROY, B. **Multicriteria methodology for decision aiding. nonconvex optimization and its applications**. 1996.
- [80] PLANETOSCOPE. **Mortalite - morts d'accidents de la route dans le monde**, September the 16th, 2014. www.planetoscope.com/mortalite/1270-mortalite—morts-d-accidents-de-la-route-dans-le-monde.html.

LIST OF FIGURES

1.1	Reduction of accidents and consideration of the car safety	2
1.2	Preponderance of frontal impact crashes with driver and passenger as most exposed car occupants	3
1.3	Main crash regulations in Europe and USA	3
1.4	European regulation and EURO NCAP rating for frontal impact	5
1.5	The four FMVSS NR.208 test configurations (in-position and full-width test)	7
1.6	Injury criteria and their logarithmic representation using for the U.S. rating score calculation	9
1.7	Description of a sled test facility (HyG)	11
1.8	Dummy sensors	12
1.9	Data measurement and transformation	13
1.10	Injuries frequency for each dummy part	14
1.11	Repartition of the damages' gravity for the head (HIC criterion)	15
1.12	Repartition of the damages' gravity for the chest through its acceleration	17
1.13	Actual and future processes for the development of passive systems	20
2.1	Overview of the KDD process's steps [14]	27
2.2	Graphical representation of the hierarchical clustering	30
2.3	Representation of c-means clustering through Euclidean and Mahalanobis distances	34
2.4	Determination of the legislation with a decision tree (driver side)	42
3.1	Numerical transformation of nominal parameters	52
3.2	Graphical User Interface (GUI)	54
3.3	Estimation's curves for the dummy temperature	55
3.4	Estimation's curves for the dummy position (type 1 of the possible missing values)	56
3.5	Estimation's curves for the webbing on spool of the seat belt	56
3.6	Comparison's curves between the k-NN sled test determined with and without missing parameters	57

3.7	Estimation's curves of the EURO NCAP rating score with all known parameters	58
3.8	Estimation's curves of the EURO NCAP rating score with unknown parameters	59
4.1	Head combination (passenger, Europa, HIII 50th dummy)	63
4.2	Neck combination (passenger, Europa, HIII 50th dummy)	65
4.3	Chest combination (passenger, Europa, HIII 50th dummy)	67
4.4	Head combination (passenger, U.S.A., HIII 50th dummy belted)	71
4.5	Neck combination (passenger, U.S.A., HIII 50th dummy belted)	73
4.6	Chest combination (passenger, U.S.A., HIII 50th dummy belted)	73
4.7	Femurs combination (passenger, U.S.A., HIII 50th dummy belted)	75
5.1	Decision tree for the dummy chest (passenger, EURO NCAP)	98
5.2	Decision tree for the dummy head (passenger, US NCAP)	106
5.3	Decision tree for the dummy neck (passenger, US NCAP)	114
5.4	Decision tree for the dummy chest (passenger, US NCAP)	122
5.5	Decision tree for the dummy femurs (passenger, US NCAP)	128
A.1	Main accident causes and their frequency	157
A.2	Collision's types passenger car / passenger car	158
A.3	History of the major legislation for passive safety	158
A.4	History of the most major ratings for passive safety	158
A.5	Sign conventions for the accelerations, forces and moments [44]	159
A.6	Sign conventions for the accelerations, forces and moments [44]	160
A.7	Localisation of the main sensors in a HIII 50th dummy	161
B.1	Attribution of the points for each dummy part according to the EURO NCAP rating	164
C.1	GUI panel for the sled test's configuration	165
C.2	GUI panel for the four main categories	166
C.3	GUI panel for entering the driver airbag parameters	167
C.4	GUI panel for entering the passenger airbag parameters	167
C.5	GUI panel for entering the kneebag parameters if mounted in the buck	168
C.6	GUI panel for entering the seat belt parameters if available	168
C.7	GUI panel for the dummy position and car specifications	169

C.8	GUI panel for the checking of the entered parameters before running the programme	170
C.9	GUI panel for entering the dummy biomechanical values after the carrying out of the sled test	170
D.1	Head combination (driver, Europa, HIII 50th dummy)	171
D.2	Neck combination (driver, Europa, HIII 50th dummy)	172
D.3	Chest combination (driver, Europa, HIII 50th dummy)	173
D.4	Femurs combination (driver, Europa, HIII 50th dummy)	174
E.1	Head combination(passenger, USA, HIII 50th dummy unbelted)	177
E.2	Neck combination (passenger, USA, HIII 50th dummy unbelted)	178
E.3	Chest combination (passenger, USA, HIII 50th dummy unbelted)	178
E.4	Femur combination (passenger, USA, HIII 50th dummy unbelted)	178
E.5	Head combination (passenger, USA, HIII 5th dummy unbelted)	179
E.6	Neck combination (passenger, USA, HIII 5th dummy unbelted)	179
E.7	Chest combination (passenger, USA, HIII 5th dummy unbelted)	180
E.8	Femurs combination (passenger, USA, HIII 5th dummy unbelted)	180
E.9	Head combination (passenger, USA, HIII 5th dummy belted)	181
E.10	Neck combination (passenger, USA, HIII 5th dummy belted)	181
E.11	Chest combination (passenger, USA, HIII 5th dummy belted)	182
E.12	Femurs combination (passenger, USA, HIII 5th dummy belted)	182
F.1	Head combination (driver, USA, HIII 50th dummy unbelted)	184
F.2	Neck combination (driver, USA, HIII 50th dummy unbelted)	184
F.3	Chest combination (driver, USA, HIII 50th dummy unbelted)	185
F.4	Femurs combination (driver, USA, HIII 50th dummy unbelted)	185
F.5	Head combination (driver, USA, HIII 5th dummy unbelted)	186
F.6	Neck combination (driver, USA, HIII 5th dummy unbelted)	186
F.7	Chest combination (driver, USA, HIII 5th dummy unbelted)	187
F.8	Femurs combination (driver, USA, HIII 5th dummy unbelted)	187
F.9	Head combination (driver, USA, HIII 50th dummy belted)	188
F.10	Neck combination (driver, USA, HIII 50th dummy belted)	188
F.11	Chest combination (driver, USA, HIII 50th dummy belted)	189
F.12	Femurs combination (driver, USA, HIII 50th dummy belted)	189
F.13	Head combination (driver, USA, HIII 5th dummy belted)	190
F.14	Neck combination (driver, USA, HIII 5th dummy belted)	191

F.15 Chest combination (driver, USA, HIII 5th dummy belted)	191
F.16 Femurs combination (driver, USA, HIII 5th dummy belted)	192
G.1 Decision tree for the dummy head (passenger, EURO NCAP) - Shifted values to avoid representing a real case	195
G.2 Decision tree for the dummy neck (passenger, EURO NCAP) - Shifted values to avoid representing a real case	198

LIST OF TABLES

1.1	Definition of the safety performance for the EURO NCAP rating	4
1.2	Weight factors for the EURO NCAP sections	5
1.3	Calculation of the overall EURO NCAP rating score (2014)	6
1.4	EURO NCAP rating score repartition for the driver and passenger	6
1.5	Star rating according to the injury probability and/or the Relative Risk Score (RRS)	8
1.6	Injury criteria and probabilities for the HIII 50th dummy (driver side)	8
1.7	Injury criteria and probabilities for the HIII 5th dummy (passenger side)	8
1.8	Characteristics of the three Hybrid III Percentile dummies (weight and size)	11
1.9	Instrumentation's description of the HIII 5th dummy	12
1.10	Instrumentation's description of the HIII 50th dummy	13
1.11	Description of the four neck movements using for the points calculation in FMVSS NR.208 regulation	16
2.1	Synthesis of the supervised and unsupervised methods used in the literature	27
2.2	An example of confusion matrix	41
3.1	Input parameters influencing dummy biomechanical values	50
3.2	Database including sled tests with known input parameters	50
3.3	Example of parameter's values	52
3.4	Example of normalised parameter's values	53
4.1	Dummy biomechanical values and their limits for the EURO NCAP rating (available for the driver and the passenger)	62
4.2	Dummy head, 4 clusters: characteristics (passenger, Europa, HIII 50th dummy)	64
4.3	Dummy head, 4 clusters: proposed clustering (passenger, Europa, HIII 50th dummy)	64
4.4	Dummy neck, 5 clusters: characteristics (passenger, Europa, HIII 50th dummy)	66
4.5	Dummy neck, 2 proposed clusters : characteristics (passenger, Europa, HIII 50th dummy)	66

4.6	Dummy chest, 10 clusters: characteristics (passenger, Europa, HIII 50th dummy)	67
4.7	Resume for the EURO NCAP clustering (driver, Europa, HIII 50th dummy) . .	69
4.8	Resume for the EURO NCAP clustering (passenger side)	69
4.9	Legal and customer's values for the Hybrid III 50th Percentile dummy (FMVSS NR.208 legislation)	70
4.10	Legal and customer's values for the Hybrid III 5th Percentile dummy (FMVSS NR.208 legislation)	71
4.11	Dummy head, 5 clusters: characteristics (passenger, U.S.A., HIII 50th dummy belted)	72
4.12	Dummy neck, 4 clusters: characteristics (passenger, U.S.A., HIII 50th dummy belted)	74
4.13	Dummy chest, 4 clusters: characteristics (passenger, U.S.A., HIII 50th dummy belted)	75
4.14	Dummy femurs, 6 clusters: characteristics (passenger, U.S.A., HIII 50th dummy belted)	76
4.15	Clustering for each dummy part (passenger, U.S.A., HIII 50th dummy unbelted)	82
4.16	Clustering for each dummy part (passenger, U.S.A., HIII 5th dummy unbelted)	83
4.17	Clustering for each dummy part (passenger, U.S.A., HIII 5th dummy belted, US NCAP rating)	84
4.18	Clustering for each dummy part (driver, U.S.A., HIII 5th dummy belted) . . .	85
4.19	Clustering for each dummy part (driver, U.S.A., HIII 5th dummy unbelted) .	85
4.20	Clustering for each dummy part (driver, U.S.A., HIII 50th dummy belted, US NCAP rating)	86
4.21	Clustering for each dummy part (driver, U.S.A., HIII 50th dummy unbelted) .	87
4.22	Resume of the use of the clusters given by the GK algorithm according to the dummy part and sled test configuration	88
5.1	Data description	90
5.2	Target variable: reminder of the clusters' number	90
5.3	Target definition (color code)	90
5.4	Sled tests' repartition for the dummy chest (passenger, EURO NCAP)	92
5.5	Calculation of the VIM for the dummy chest (passenger, EURO NCAP)	101
5.6	Validation of the decision tree for the dummy chest (passenger, EURO NCAP)	101
5.7	Sled tests' repartition for the dummy head (passenger, US NCAP)	103
5.8	Validation of the decision tree for the dummy head (passenger, US NCAP) .	108
5.9	Calculation of the VIM for the dummy head (passenger, US NCAP)	109

5.10 Sled tests' repartition for the dummy neck (passenger, US NCAP)	110
5.11 Calculation of the VIM for the dummy neck (passenger, US NCAP)	115
5.12 Validation of the decision tree for the dummy neck (passenger, US NCAP) .	116
5.13 Sled tests' repartition for the dummy chest (passenger, US NCAP)	117
5.14 Calculation of the VIM for the dummy chest (passenger, US NCAP)	123
5.15 Validation of the decision tree for the dummy chest (passenger, US NCAP) .	124
5.16 Sled tests' repartition for the dummy femurs (passenger, US NCAP)	125
5.17 Calculation of the VIM for the dummy femurs (passenger, US NCAP)	129
5.18 Validation of the decision tree for the dummy femurs (passenger, US NCAP)	130
5.19 Resume of the four most influential parameters	131
D.1 Dummy head, 4 proposed clusters: characteristics (driver, Europa, HIII 50th dummy)	172
D.2 Dummy neck, 2 proposed clusters: characteristics (driver, Europa, HIII 50th dummy)	173
D.3 Dummy chest, 10 clusters: characteristics (driver, Europa, HIII 50th dummy)	174
D.4 Dummy femurs, 4 clusters: characteristics (driver, Europa, HIII 50th dummy)	175
G.1 Sled tests' repartition for the dummy head (passenger, EURO NCAP)	193
G.2 Calculation of the VIM for the dummy head (passenger, EURO NCAP)	196
G.3 Validation of the decision tree for the dummy head (passenger, EURO NCAP)	196
G.4 Sled tests' repartition for the dummy neck (passenger, EURO NCAP)	196
G.5 Calculation of the VIM for the dummy neck (passenger, EURO NCAP)	199
G.6 Validation of the decision tree for the dummy neck (passenger, EURO NCAP)	199

LIST OF DEFINITIONS

GIDAS	German In-Depth Accident Study
ECE	Economic Commission for Europe (United Nations)
FMVSS	Federal Motor Vehicle Safety Standards
NCAP	New Car Assessment Program
FIA	Fédération Internationale de l'Automobile
NHTSA	National Highway Traffic Safety Administration
US	United States
ADAC	Allgemeine Deutsche Automobil-Club
HIII	Hybrid III
HyG	Hydraulic Generator
MTS	Material Testing Systems
ATD	Anthropomorphic Test Device
ATD	Anthropomorphic Test Device
SID	Side Impact Dummy
BioRID	Biofidelic Rear Impact Dummy
CRABI	Child Restraint Air Bag Interaction
HIC	Head Injury Criterion
HIC15	Head Injury Criterion 15 ms
HIC36	Head Injury Criterion 36 ms
N_{ij}	Normalized Neck Injury Criterion
VC	Viscous Criterion
TI	Tibia Index
KDD	Knowledge Discovery in Databases
SVM	Support Vector Machines
ROC	Receiver Operating Characteristic

NIS	National Institute of Statistics
CART	Classification And Regression Tree
PCA	Principle Component Analysis
MCDM	Multi-Criteria Decision Method
Electre	ELimination Et Choix Traduisant la REalité
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
AHP	Analytic Hierarchy Process
DAU	Data Analysis Unit
TRW	Thompson Ramo Wooldridge
RMSSD	Root-Mean-Square Standard Deviation
RS	R-Squared
RRS	Relative Risk Score
DBI	Davies-Bouldin Index
USA	United States of America
MDB	Moving Deformable Barrier
CRS	Child Restraint System
ESC	Electronic Stability Control
LDW	Lane Departure Warning
LKD	Lane Keeping Device
AEB	Autonomous Emergency Braking
CART	Classification And Regression Tree
VIM	Variable Importance Measure
ID3	Iterative Dichotomiser 3
IG	Information Gain
IGR	Information Gain Ratio
AIS	Abbreviated Injury Scale
k-NN	k-Nearest Neighbour
k-NNI	k-Nearest Neighbour Imputation
IEEE	Institute of Electrical and Electronics Engineers
BIC	Bayesian Information Criterion

AIC	Akaike Information Criterion
CAIC	Consistent Akaike's Information Criterion
FCM	Fuzzy C-Means
DAB	Driver AirBag
PAB	Passenger AirBag
KnAB	Knie AirBag
ABV	Active Bag Vent
TTF	Time To Fire
LL	Load Limiter
RP	Retractor Pretensioner
BP	Buckle Pretensioner
AP	Anchor Pretensioner
DLT	Dynamic Locking Tongue
IP	Instrumental Panel
OLC	Occupant Load Criterion
Tv0	Time with the Velocity equal to 0
IFSTTAR	Institut français des sciences et des technologies des transports, de l'aménagement et des réseaux
HDR	Habilitation à Diriger des Recherches



ANNEXES

COMPLEMENT TO CHAPTER 1

This Annex is a complement of chapter 1 and regroups additional information about passive safety.

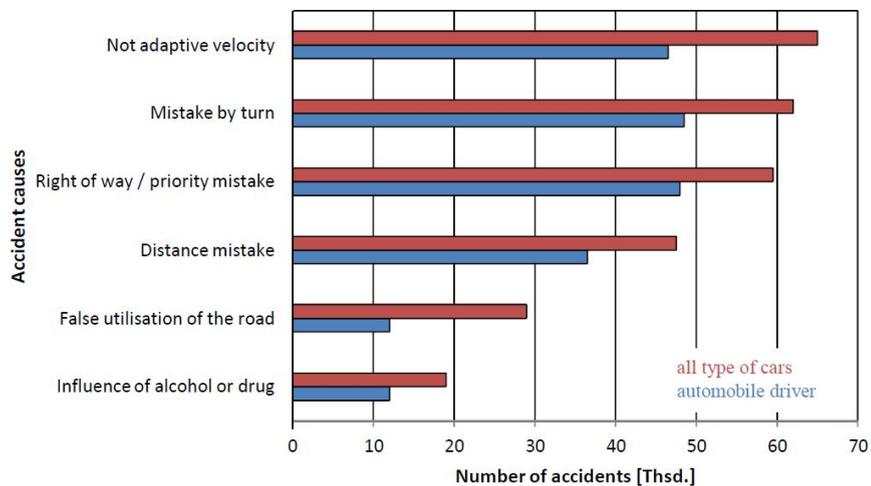


Figure A.1: Main accident causes and their frequency

Many types of collision can be registered (Fig. A.2). The collision's definition between utility vehicle/passenger car, passenger car/passenger car and passenger car/barrier was explained by Richter et al. [6]

Fig. A.3 and Fig. A.4 present the history of the major worldwide legislation and ratings for car safety.

Fig A.6 presents the sign conventions for the accelerations, forces and moments of the dummy.

Fig. A.7 shows the positioning of the sensors in the HIII 50th dummy.

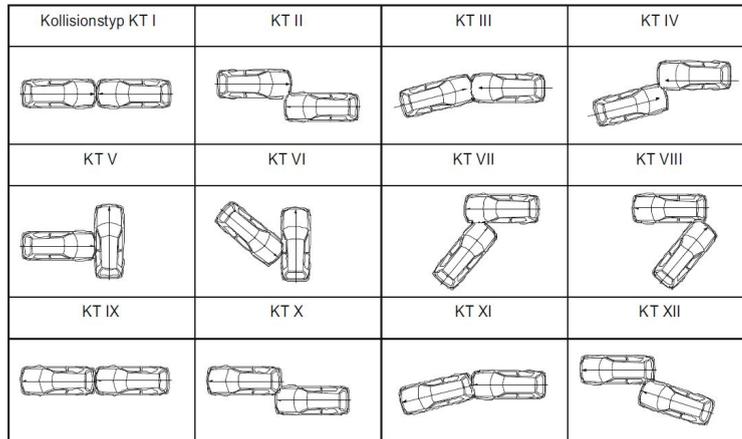


Figure A.2: Collision's types passenger car / passenger car

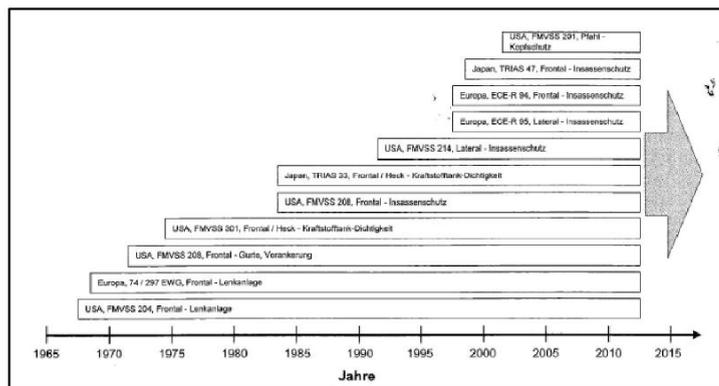


Figure A.3: History of the major legislation for passive safety

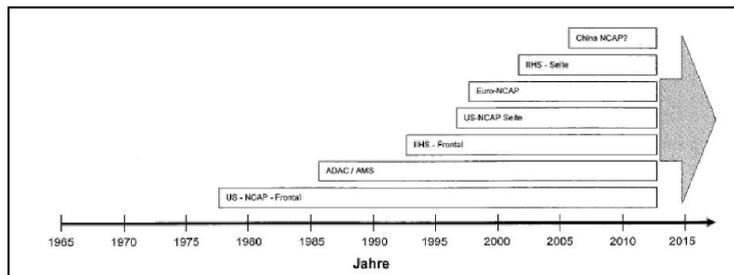


Figure A.4: History of the most major ratings for passive safety

TABLE 1A
SIGN CONVENTIONS FOR
PART 572 SUBPART E TEST DUMMIES

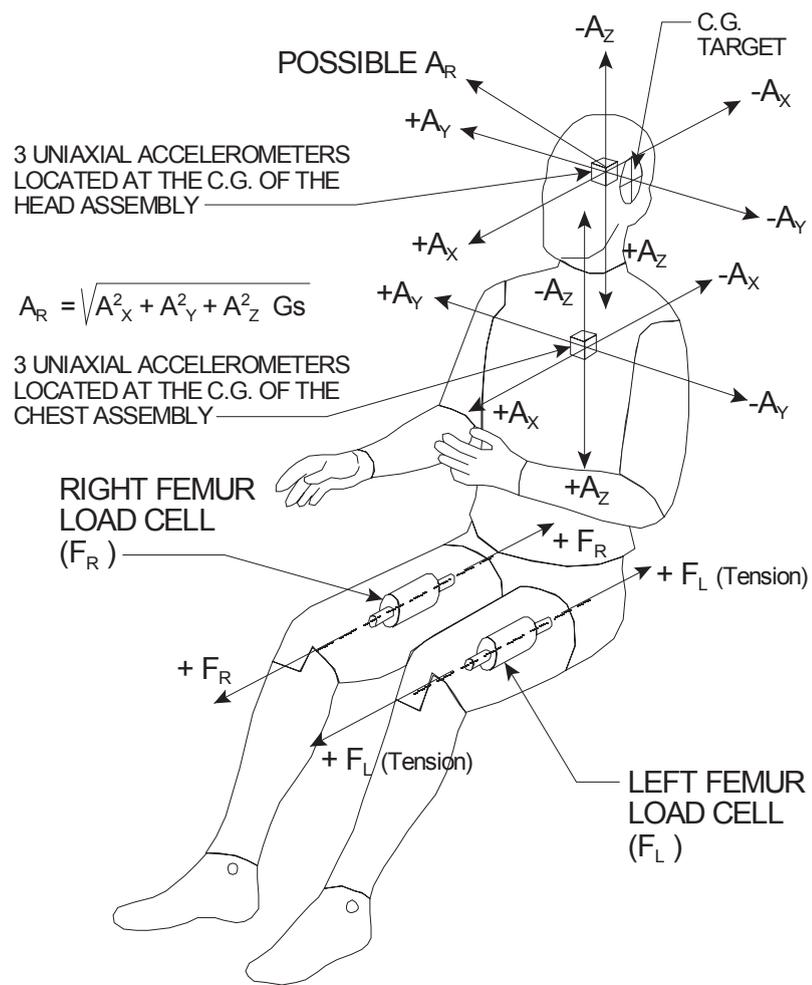


FIGURE 2A

Figure A.5: Sign conventions for the accelerations, forces and moments [44]

TP208-14

A 10

SIGN CONVENTION FOR HYBRID III TRANSDUCER OUTPUTS	
BODY SEGMENT — MEASURED FORCE	POSITIVE OUTPUT DIRECTION
NECK FX SHEAR FY SHEAR FZ AXIAL MX MOMENT (ROLL) MY MOMENT (PITCH) MZ MOMENT (YAW)	HEAD REARWARD OR CHEST FORWARD HEAD LEFTWARD, CHEST RIGHTWARD HEAD UPWARD, CHEST DOWNWARD LEFT EAR TOWARD LEFT SHOULDER CHIN TOWARD STERNUM CHIN TOWARD LEFT SHOULDER
FEMUR FX SHEAR FY SHEAR FZ AXIAL MX MOMENT (ROLL) MY MOMENT (PITCH) MZ MOMENT (YAW)	KNEE UPWARD, UPPER FEMUR DOWNWARD KNEE RIGHTWARD, UPPER FEMUR LEFTWARD KNEE FORWARD (TENSION), PELVIS REARWARD KNEE LEFTWARD, HOLD UPPER FEMUR IN PLACE KNEE UPWARD, HOLD UPPER FEMUR IN PLACE KNEE ROTATED CCW WHEN FACING FRONT OF DUMMY
KNEE CLEVIS - FZ AXIAL	TIBIA DOWNWARD (TENSION), FEMUR UPWARD
UPPER TIBIA MX MOMENT MY MOMENT	ANKLE LEFTWARD, HOLD KNEE IN PLACE ANKLE FORWARD, BOTTOM OF KNEE CLEVIS REARWARD
LOWER TIBIA FX SHEAR FY SHEAR FZ AXIAL MX MOMENT MY MOMENT	ANKLE FORWARD, KNEE REARWARD ANKLE RIGHTWARD, KNEE LEFTWARD ANKLE DOWNWARD (TENSION), KNEE UPWARD ANKLE LEFTWARD, HOLD KNEE IN PLACE ANKLE FORWARD, BOTTOM OF KNEE CLEVIS REARWARD
CHEST DISPLACEMENT	CHEST COMPRESSED - NEGATIVE
KNEE SHEAR DISPLACEMENT	PUSH ON FRONT OF TIBIA - NEGATIVE

NOTE: DIRECTIONS ARE DEFINED IN RELATION TO A SEATED DUMMY

Figure A.6: Sign conventions for the accelerations, forces and moments [44]

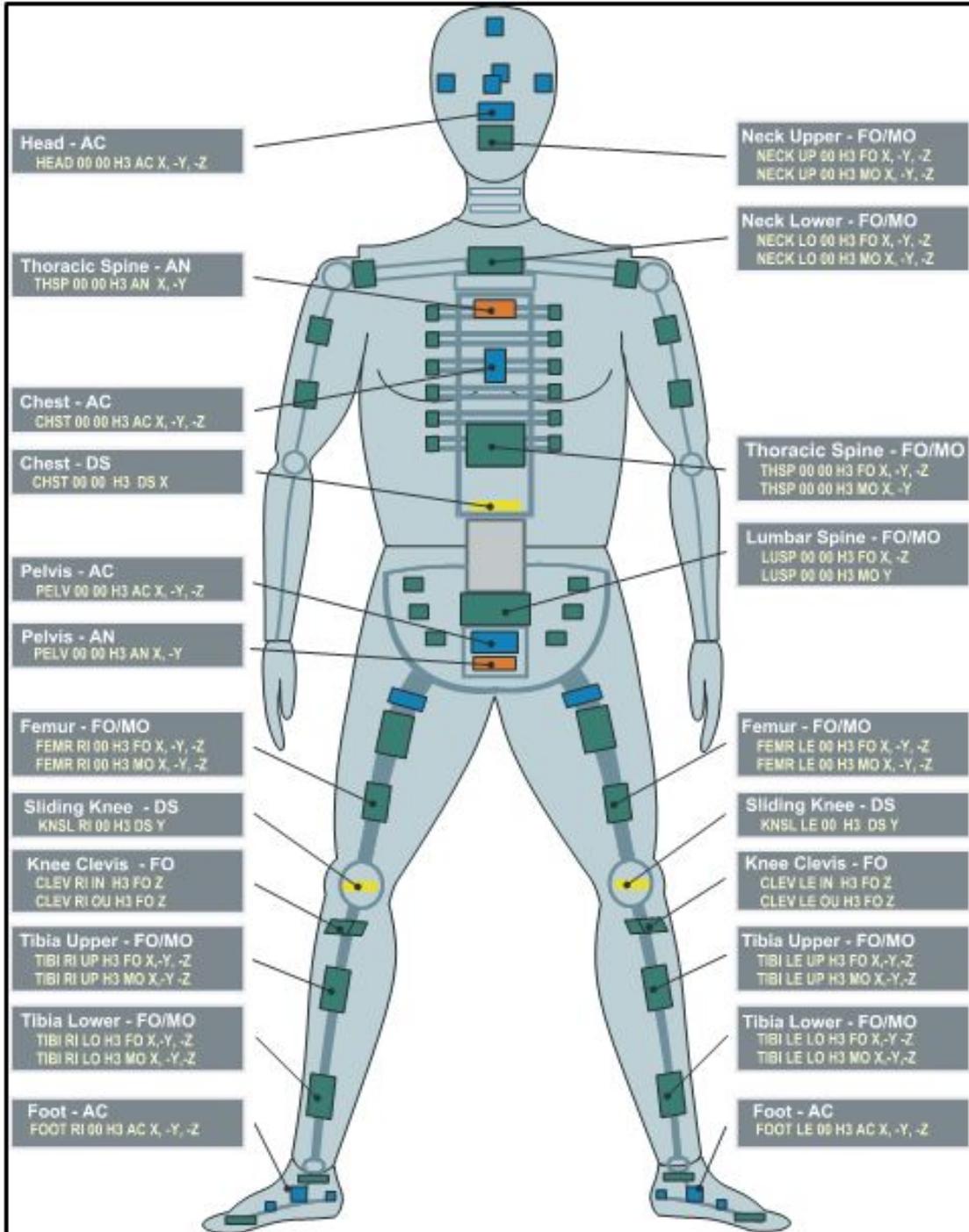


Figure A.7: Localisation of the main sensors in a HIII 50th dummy

B

LEGAL REQUIREMENTS FOR THE ECE LEGISLATION

B.1/ EURO NCAP RATING: ATTRIBUTION OF THE POINTS

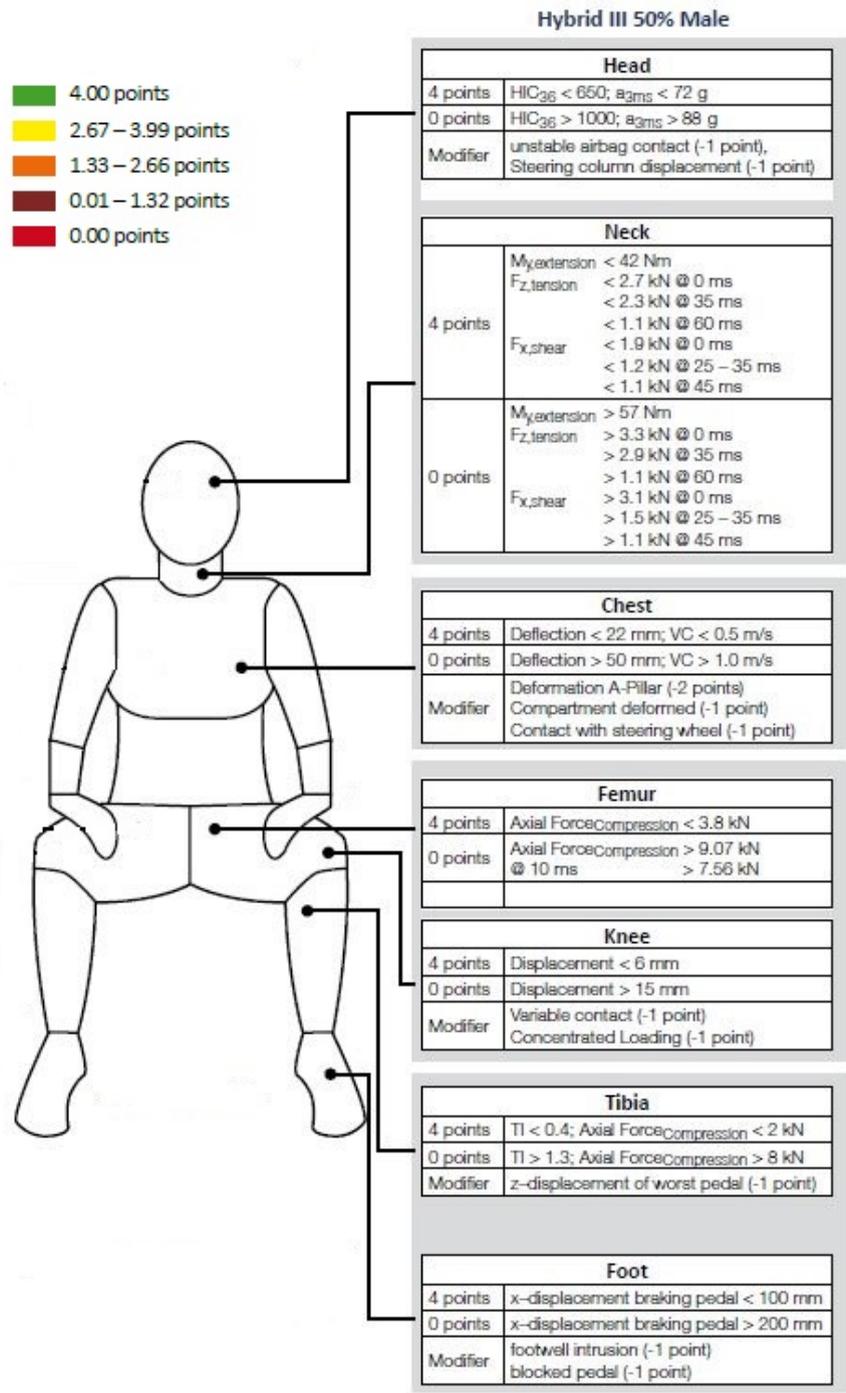


Figure B.1: Attribution of the points for each dummy part according to the EURO NCAP rating

C

PANELS OF THE GRAPHICAL USER INTERFACE

C.1/ CONFIGURATION'S PANEL

The created Graphical User Interface (GUI) has a general panel which the user can choose the configuration of the sled test and enter some general parameters as the test's name, customer's name, type of legislation or rating, dummy, occupant position. This panel is represented in Fig. C.1. The next panels are according to these choices.

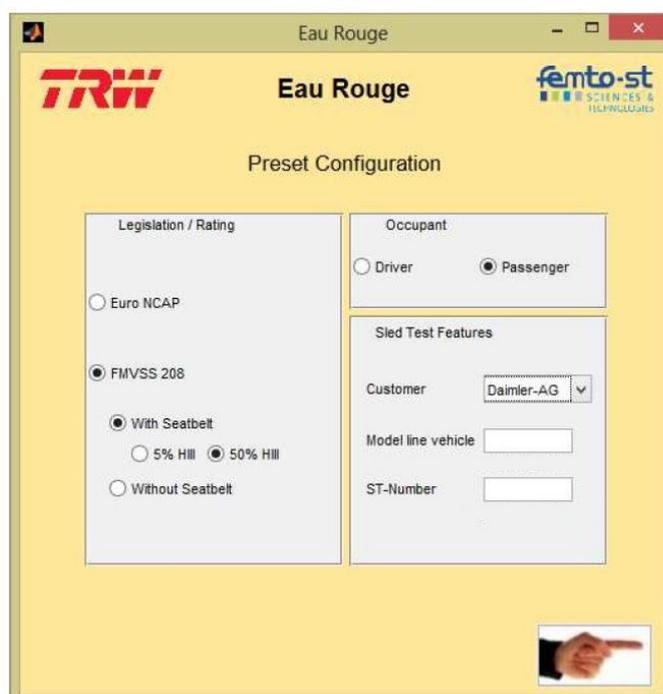


Figure C.1: GUI panel for the sled test's configuration

C.2/ GENERAL PANELS

After choosing the initial configuration of the sled test, the four main categories which can influence the dummy biomechanical values, are available (Fig. C.2):

- ◆ Airbag (driver or passenger, depending on the chosen configuration)
- ◆ Kneebag
- ◆ Seat belt: if the user chooses the FMVSS Nr.208 legislation without seat belt in the configuration's panel, the panel "seat belt" will be unavailable.

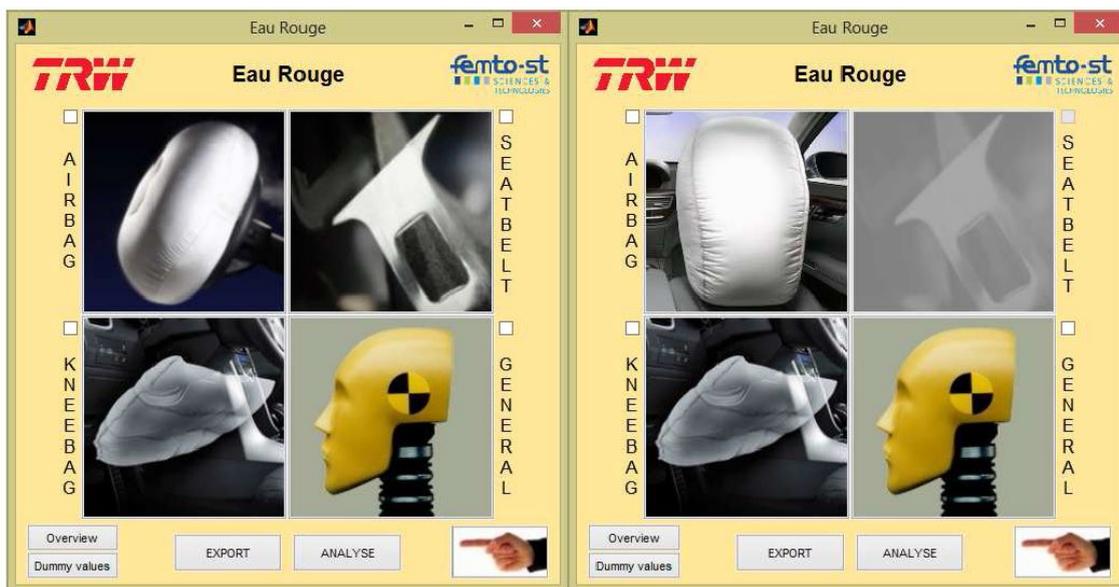


Figure C.2: GUI panel for the four main categories

C.2.1/ PANEL FOR THE DRIVER AND PASSENGER AIRBAG

Because the airbag's parameters are different for the driver and passenger airbag, two panels have been created for the frontal airbag (Fig. C.3 and Fig. C.4). This panel allows the entering of many airbags parameters which can have an influence on the dummy biomechanical values. These parameters have been selected through an expert elicitation. By availability of some parameters (i.e. the active bag vent), additional fields are automatically available in order to "define" this parameters (i.e. through its surface or time to fire). Because the driver airbag are directly connected on the steering system, the extra panel is available for the driver side containing information on the steering wheel and steering column.

C.2.2/ PANEL FOR THE KNEEBAG

This panel is independent of the sled test's configuration, i.e. it is always available. If no kneebag is mounted in the buck, its parameters will not be shown (Fig. C.5).

Figure C.3: GUI panel for entering the driver airbag parameters

Figure C.4: GUI panel for entering the passenger airbag parameters

C.2.3/ PANEL FOR THE SEAT BELT

If the sled test is a EURO NCAP rating or a FMVSS Nr.208 with seat belt, some parameters of the seat belt are asked. As for the other panels, the availability of components causes the appearance of other characteristics (Fig. C.6).

The figure shows two side-by-side screenshots of the 'Kneebag Characteristics' GUI panel. Both panels feature the TRW and femto-st logos at the top. The left panel is titled 'Eau Rouge KNEEBAG' and shows the 'Kneebag available' field with 'Yes' selected. The right panel is also titled 'Eau Rouge KNEEBAG' and shows the same 'Kneebag available' field with 'Yes' selected, plus several other input fields: 'Volume Kneebag' (L), 'Number of ventholes' (1), 'Venthole Diameter' (mm), 'Number of tether', 'Inflator name' (name), and 'Time to fire' (ms). Both panels include a 'Remarks' field at the bottom with the text '*Use points as separators if necessary' and a green checkmark icon.

Figure C.5: GUI panel for entering the kneebag parameters if mounted in the buck

The figure shows two side-by-side screenshots of the 'Seatbelt Characteristics' GUI panel. Both panels feature the TRW and femto-st logos at the top. The left panel is titled 'Eau Rouge SEATBELT' and shows the 'Type of Load Limiter (LL)' set to 'Constant Load Limiter' and 'LL Value 1' in kN. The right panel is also titled 'Eau Rouge SEATBELT' and shows the 'Type of Load Limiter (LL)' set to 'Switchable Load Limiter' with 'LL Value 1' and 'LL Value 2' in kN, and 'TTF Switch' in ms. Both panels include a 'Remarks' field at the bottom with the text '*Use points as separators if necessary' and a green checkmark icon.

Figure C.6: GUI panel for entering the seat belt parameters if available

C.2.4/ PANEL FOR THE SLED TEST'S GENERAL PARAMETERS

This panel resumes first the sled test's configuration by inserting the data entering in the configuration's panel. It allows to verify the configuration before running the programme (Fig. C.7). This panel includes the dummy position: it is defined by the chosen law but, in order to compare it, the distances (and not the coordinate of some points) between the dummy and the instrument panel have been measured. It allows to compare the dummy position between different types of cars in accordance to the position of the instrumental

panel. In fact, theoretic the more nearest from the instrumental panel the dummy is, the more worst the dummy biomechanical values are.

General Characteristics

TRW **Eau Rouge** **femto-st**
GENERAL RESEARCH SCIENCES & TECHNOLOGIES

TEST SET-UP

Rating/Legislation : FMVSS 208 Speed : 56 km/h ODB : 100%

Occupant : Driver Hand drive : Left Dummy Seatbelt : With Seatbelt

Customer : Daimler-AG ST Number:

DUMMY POSITION *All measures in mm

Clearance Chin - Point 1 of Steering wheel Recess of Sternum to Centre of Emblem

Left Knee-Shorter Distance Ahead Right Knee-Shorter Distance Ahead

Left Heel Point Ahead to Floor Angle Right Heel Point Ahead to Floor Angle

Angle of Neck Bracket Clearance between inner Thorax Hole & Webbing

OLC++ Dummy Temperature °C

OLC European Car Classification A mini cars

Tv 0 ms

Remarks *Use points as separators if necessary

Figure C.7: GUI panel for the dummy position and car specifications

C.2.5/ OVERVIEW PANEL

After entering all parameters of each restraint systems, the user can check it before running the programme with the help of the overview panel (Fig. C.8). It is available through the button "Dummy values" available on the general panel (Fig. C.2).

C.3/ SLED TEST'S RESULTS

The majority of the sled tests included in the database are already executed. That is why the dummy biomechanical values are known and can be used in order to predict the results of a new sled test. The panel C.9 allows the entering of the results for EuroNCAP rating and FMVSS Nr.208. This panel is available through the button "Dummy values" available on the general panel (Fig. C.2).

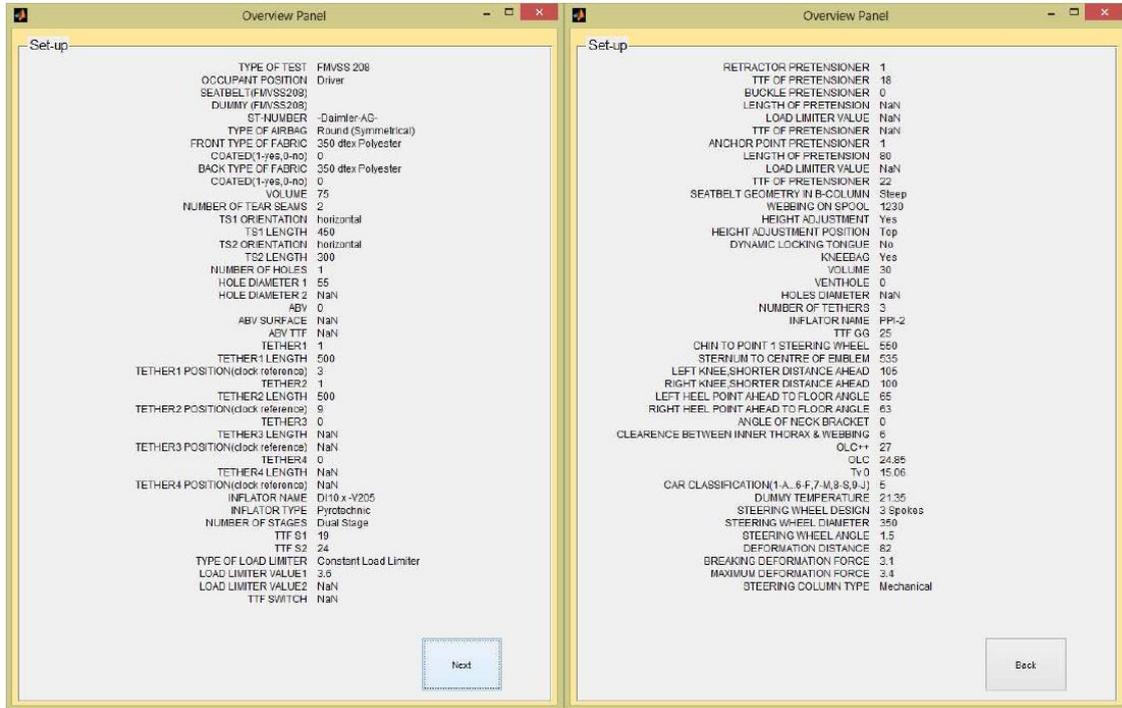


Figure C.8: GUI panel for the checking of the entered parameters before running the programme

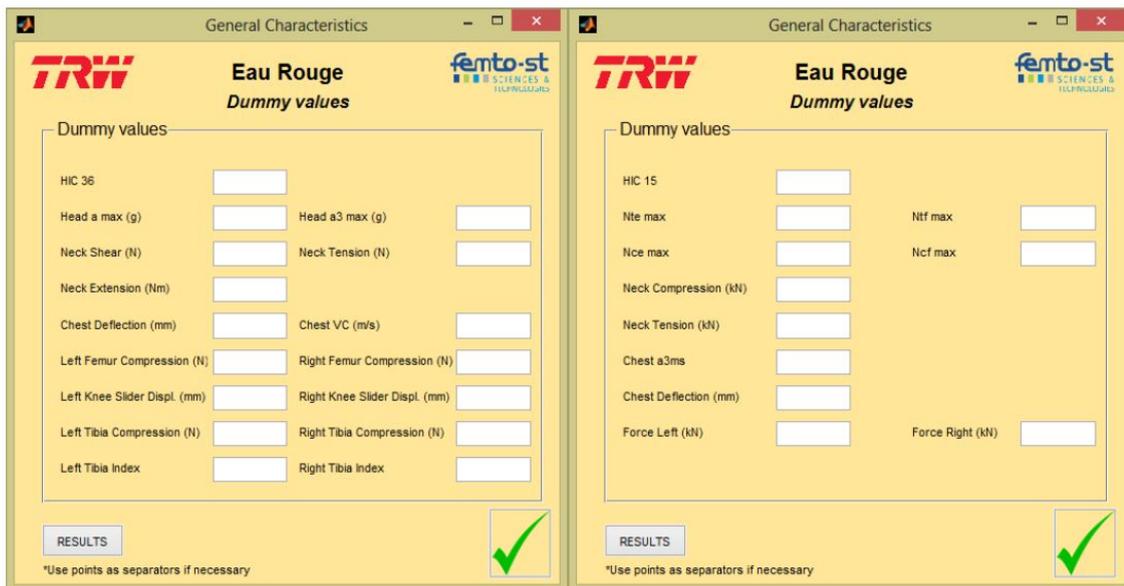


Figure C.9: GUI panel for entering the dummy biomechanical values after the carrying out of the sled test

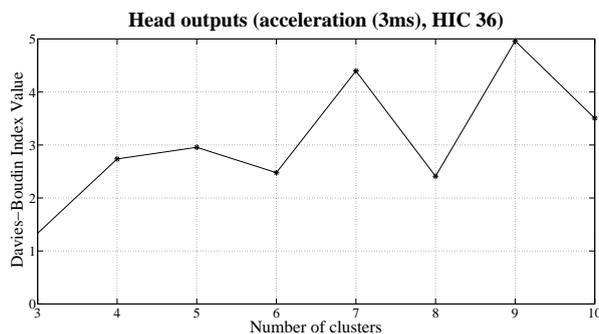
D

COMPLEMENT TO CHAPTER 4.1 (EURO NCAP RATING, DRIVER SIDE)

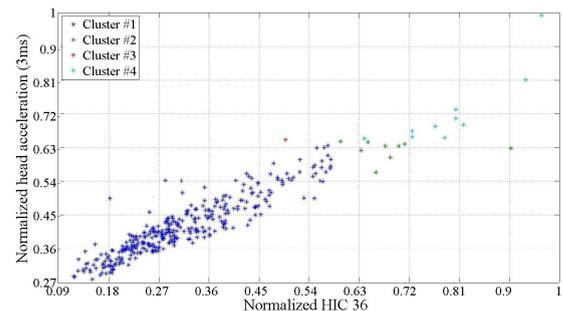
Because the method for the determination of the clusters for the driver side (HIII 50th dummy) in EURO NCAP is the same as for the passenger side, only the DBI curves and the corresponding graphical representation of the clusters will be represented.

D.1/ DUMMY HEAD

For the dummy head, Fig. D.1(a) presents the DBI values. According to it, the best clustering is the clustering with 4 clusters (Fig. D.1(b) and Table D.1).



(a) Representation of the DBI values



(b) Clustering dummy head (4 clusters): graphical representation

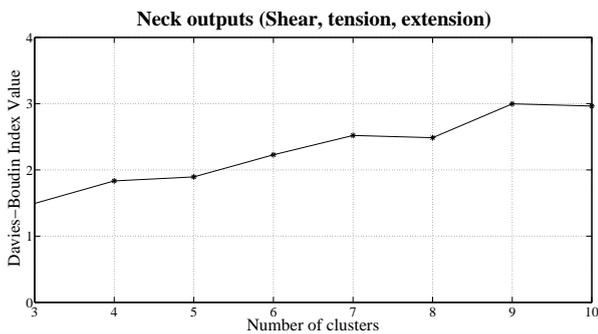
Figure D.1: Head combination (driver, Europa, HIII 50th dummy)

D.2/ DUMMY NECK

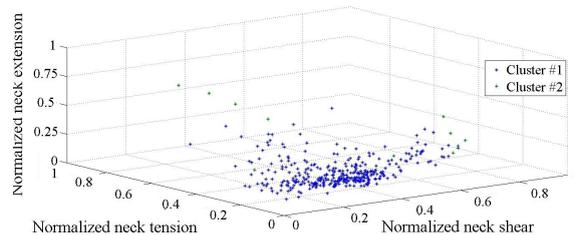
For the dummy neck, Fig. D.2(a) presents the DBI values. According to it, the best clustering is the clustering with 3 clusters but after analysis of the sled tests' results, the clustering with 2 clusters regrouping sled tests with 4 points overall and sled tests which did not reach the 4 points overall seems to be more adapted (Fig. D.2(b) and Table D.2).

Table D.1: Dummy head, 4 proposed clusters: characteristics (driver, Europa, HIII 50th dummy)

Item	Number of elements in the cluster	Head acceleration (3 ms)	HIC 36
		Range of values	
#1	372	[31 g; 71 g]	[130 ; 643]
#2	12	[63 g; 72 g]	[663 ; 1003]
#3	1	72 g	553
#4	10	[73 g; 109 g]	[710 ; 1064]
		Number of points obtained for the dummy head (EURO NCAP rating)	
#1	372	4 points	4 points
#2	12	4 points	< 4 points
#3	1	< 4 points	4 points
#4	10	< 4 points	< 4 points



(a) Representation of the DBI values



(b) Clustering dummy neck (2 proposed clusters): graphical representation

Figure D.2: Neck combination (driver, Europa, HIII 50th dummy)

D.3/ DUMMY CHEST

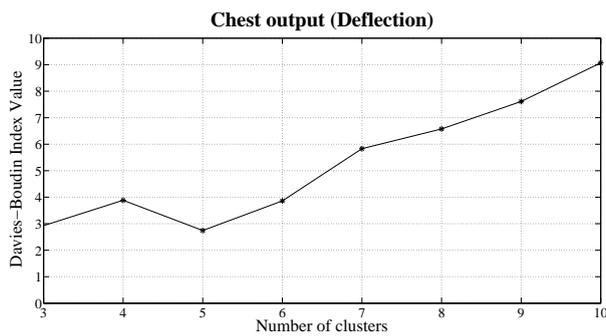
For the dummy chest, Fig. D.3(a) presents the DBI values. According to it, the best clustering is the clustering with 3 clusters. Because of the sensibility of this dummy part, the clustering will be done with 10 clusters, as for the passenger side (Fig. D.3(b) and Table D.3). These clusters have directly been formed by the GK algorithm. Because for one sled test the sensor seems to be defect (chest deflection value by 0), this value has been deleted from the database.

D.4/ DUMMY FEMURS

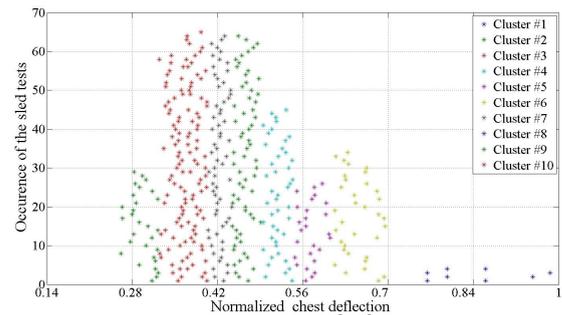
For the majority of the EURO NCAP tests, the driver has a knee bag. That is why the femur forces have to be taken into account. Fig. D.4(a) presents the DBI values. According to

Table D.2: Dummy neck, 2 proposed clusters: characteristics (driver, Europa, HIII 50th dummy)

Item in the cluster	Number of elements	Neck shear	Neck tension	Neck extension
		Range of values		
#1	384	[16 N; 1921 N]	[496 N; 2503 N]	[2 Nm; 46 Nm]
#2	11	[343 N; 2122 N]	[681 N; 4375 N]	[11 Nm; 62 Nm]
		Number of points obtained for the dummy neck (EURO NCAP rating)		
#1	384	4 points	4 points	4 points
#2	11	0 - 4 points	0 - 4 point	0 - 4 points
		Neck score overall (minimal score between the three neck scores)		
#1	384	4 points		
#2	11	0 - 3,99 points		



(a) Representation of the DBI values



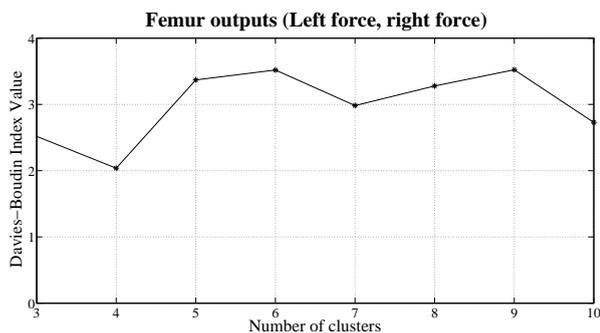
(b) Clustering dummy chest (10 clusters): graphical representation

Figure D.3: Chest combination (driver, Europa, HIII 50th dummy)

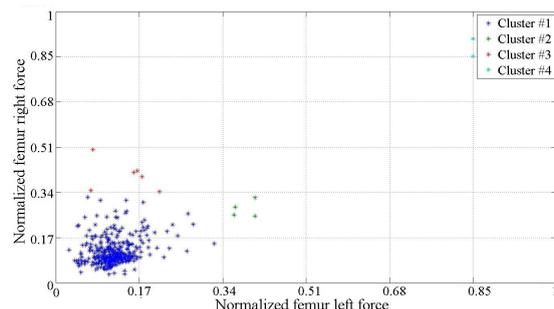
it, the best clustering is the clustering with 4 clusters. But the chosen clusters have been formed according to the obtained score (Fig. D.4(b) and Table D.4).

Table D.3: Dummy chest, 10 clusters: characteristics (driver, Europa, HIII 50th dummy)

Item	Number of elements in the cluster	Chest deflection
		Range of values
#1	4	[54,58 mm ; 56,90 mm]
#2	64	[31,72 mm ; 35,07 mm]
#3	59	[23,19 mm ; 26,20 mm]
#4	45	[35,35 mm ; 38,81 mm]
#5	26	[39,08 mm ; 43,20 mm]
#6	34	[43,78 mm ; 49,66 mm]
#7	64	[28,88 mm ; 31,66 mm]
#8	4	[61,41 mm ; 68,98 mm]
#9	29	[18,66 mm ; 23,03 mm]
#10	65	[26,25 mm ; 28,76 mm]
	Number of points obtained for the dummy chest (EURO NCAP rating)	
#1	4	0 point
#2	64	2,13 - 2,61 points
#3	59	3,40 - 3,83 points
#4	45	1,60 - 2,09 points
#5	26	0,97 - 1,56 points
#6	34	0,05 - 0,89 points
#7	64	2,62 - 3,02 points
#8	4	0 point
#9	29	3,85 - 4 points
#10	65	3,03 - 3,39 points



(a) Representation of the DBI values



(b) Clustering dummy femurs (4 proposed clusters): graphical representation

Figure D.4: Femurs combination (driver, Europa, HIII 50th dummy)

Table D.4: Dummy femurs, 4 clusters: characteristics (driver, Europa, HIII 50th dummy)

Item	Number of elements in the cluster	Femurs forces	
		Left	Right
		Range of values	
#1	379	[315 N ; 3798 N]	[379 N ; 3795 N]
#2	5	[4273 N ; 4774 N]	[2947 N ; 3769 N]
#3	9	[836 N ; 2485 N]	[4031 N ; 5900 N]
#4	2	[9996 N ; 10000 N]	[10018 N ; 10810 N]
		Number of points obtained for the dummy femur (EURO NCAP rating)	
#1	379	4 points	4 points
#2	5	< 4 points	4 points
#3	9	4 points	< 4 points
#4	2	< 4 points	< 4 points

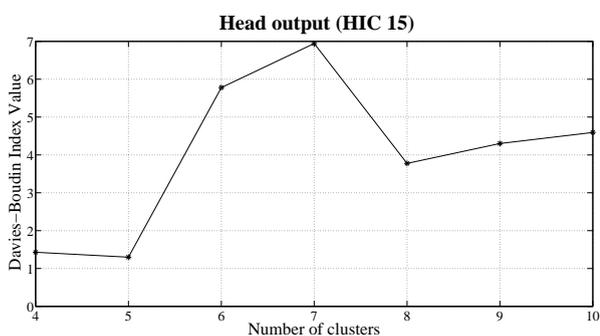
E

COMPLEMENT TO CHAPTER 4.2 (FMVSS NR.208, PASSENGER SIDE)

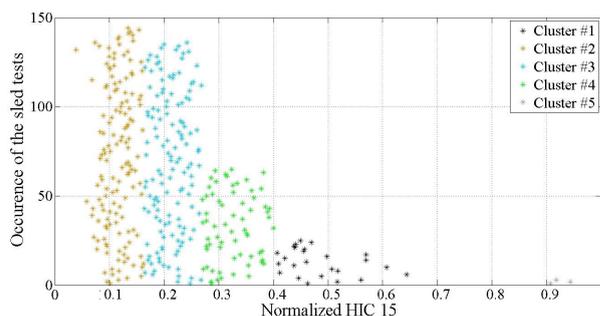
E.1/ PASSENGER SIDE : HIII 50TH DUMMY UNBELTED

This section regroups the DBI curves and the graphical representations of the clusters for the HIII 50th dummy unbelted placed on the passenger side. The head, the neck, the chest and the femurs have been analysed.

E.1.1/ DUMMY HEAD

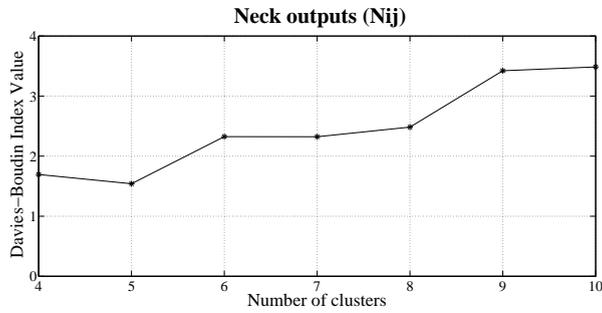


(a) Representation of the DBI values

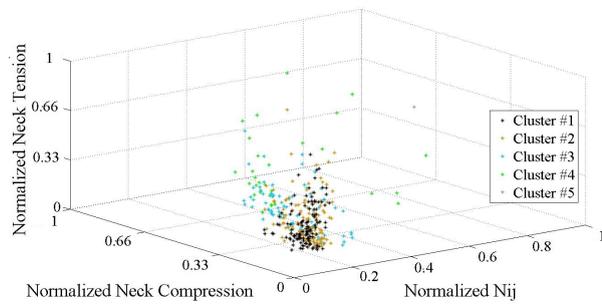


(b) Clustering dummy head (5 Clusters): graphical representation

Figure E.1: Head combination(passenger, USA, HIII 50th dummy unbelted)

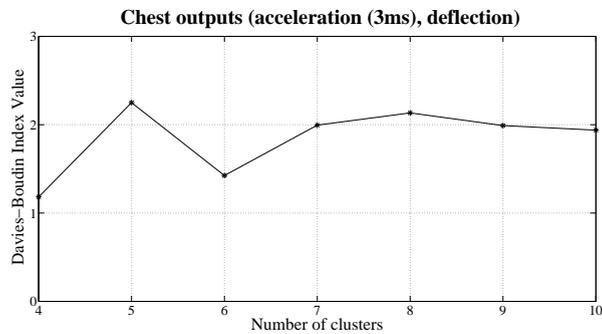


(a) Representation of the DBI values

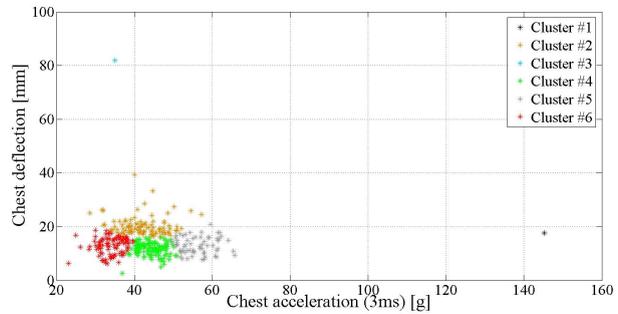


(b) Clustering dummy neck (5 clusters): graphical representation

Figure E.2: Neck combination (passenger, USA, HIII 50th dummy unbelted)

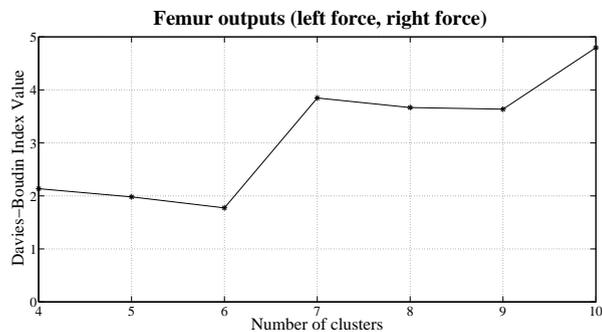


(a) Representation of the DBI values

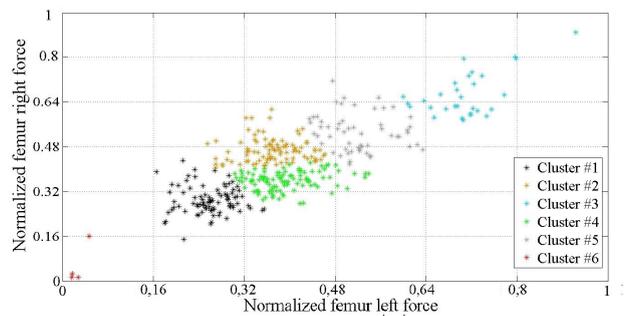


(b) Clustering dummy chest (6 clusters): graphical representation

Figure E.3: Chest combination (passenger, USA, HIII 50th dummy unbelted)



(a) Representation of the DBI values



(b) Clustering dummy femurs (6 clusters): graphical representation

Figure E.4: Femur combination (passenger, USA, HIII 50th dummy unbelted)

E.1.2/ DUMMY NECK

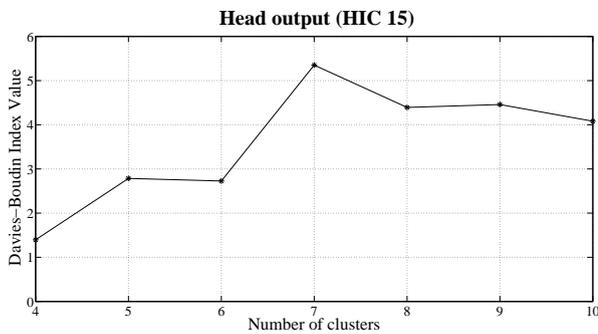
E.1.3/ DUMMY CHEST

E.1.4/ DUMMY FEMURS

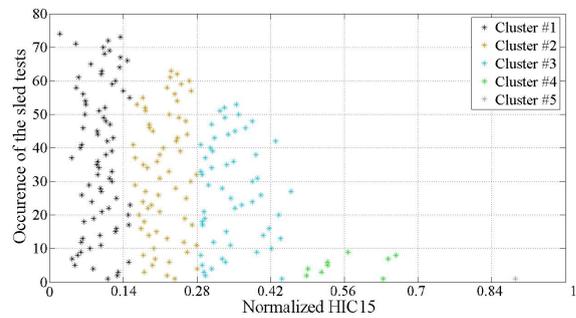
E.2/ PASSENGER SIDE : HIII 5TH DUMMY UNBELTED

This section regroups the DBI curves and the graphical representations of the clusters for the HIII 5th dummy unbelted placed on the passenger side. The head, the neck, the chest and the femurs have been analysed.

E.2.1/ DUMMY HEAD



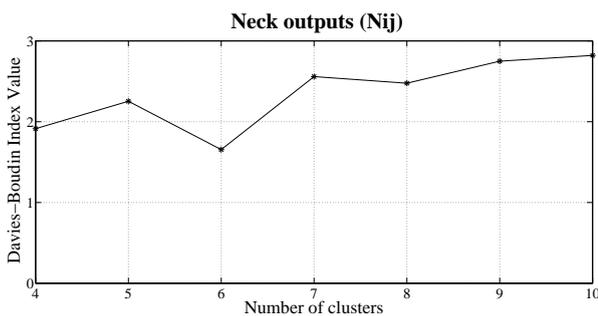
(a) Representation of the DBI values



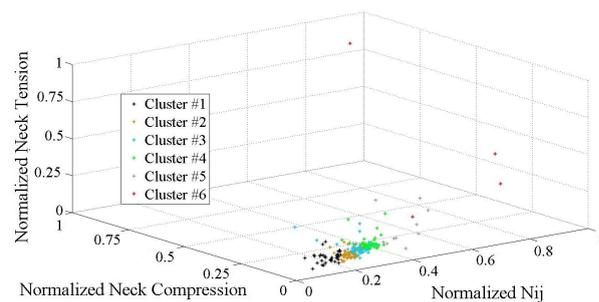
(b) Clustering dummy head (5 clusters): graphical representation

Figure E.5: Head combination (passenger, USA, HIII 5th dummy unbelted)

E.2.2/ DUMMY NECK



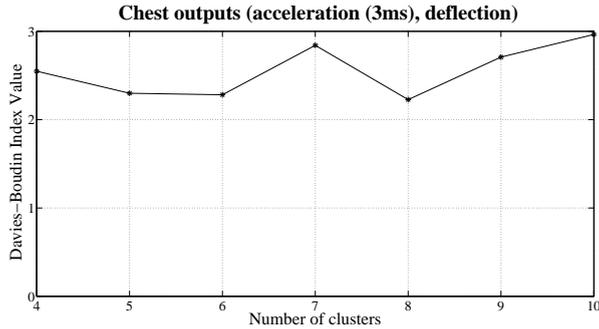
(a) Representation of the DBI values



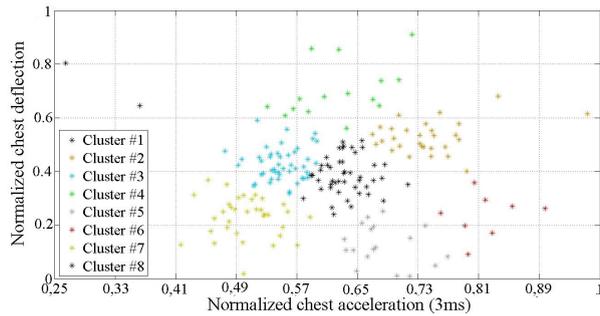
(b) Clustering dummy neck (6 clusters): graphical representation

Figure E.6: Neck combination (passenger, USA, HIII 5th dummy unbelted)

E.2.3/ DUMMY CHEST



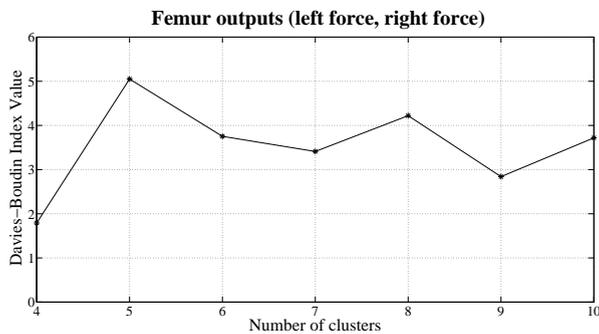
(a) Representation of the DBI values



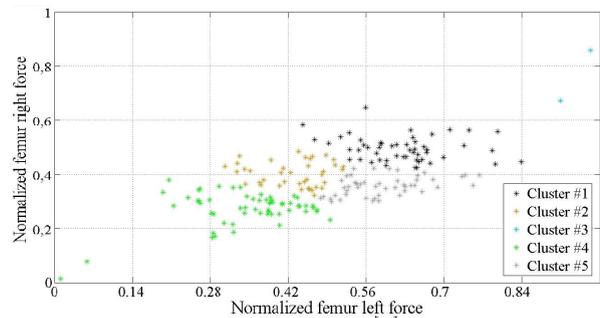
(b) Clustering dummy chest (8 clusters): graphical representation

Figure E.7: Chest combination (passenger, USA, HIII 5th dummy unbelted)

E.2.4/ DUMMY FEMURS



(a) Representation of the DBI values

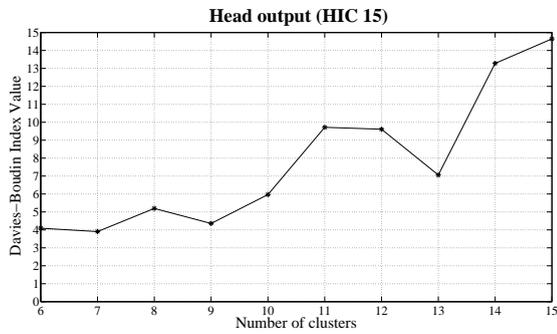


(b) Clustering dummy femurs (5 clusters): graphical representation

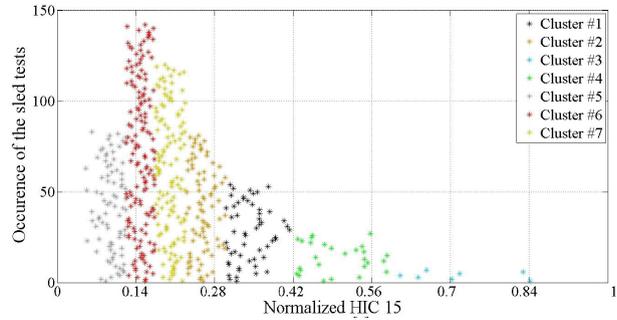
Figure E.8: Femurs combination (passenger, USA, HIII 5th dummy unbelted)

E.3/ PASSENGER SIDE : HIII 5TH DUMMY BELTED (RATING)

This section regroups the DBI curves and the graphical representations of the clusters for the HIII 5th dummy belted placed on the passenger side. It corresponds to the rating. The head, the neck, the chest and the femurs have been analysed.

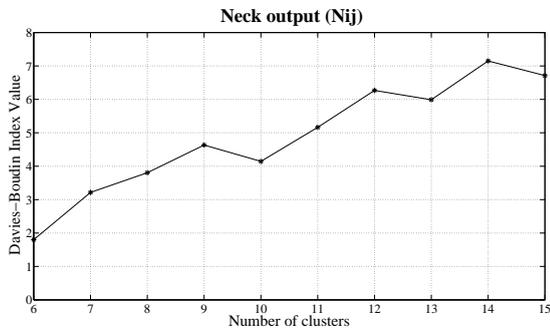


(a) Representation of the DBI values

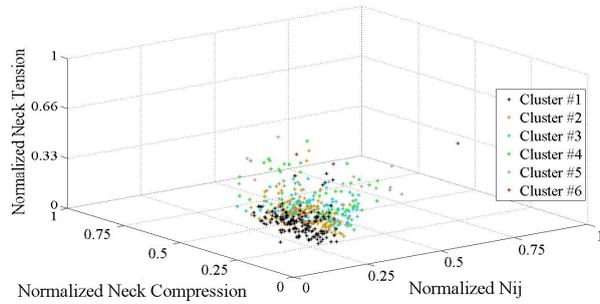


(b) Clustering dummy head (7 clusters): graphical representation

Figure E.9: Head combination (passenger, USA, HIII 5th dummy belted)



(a) Representation of the DBI values



(b) Clustering dummy neck (6 clusters): graphical representation

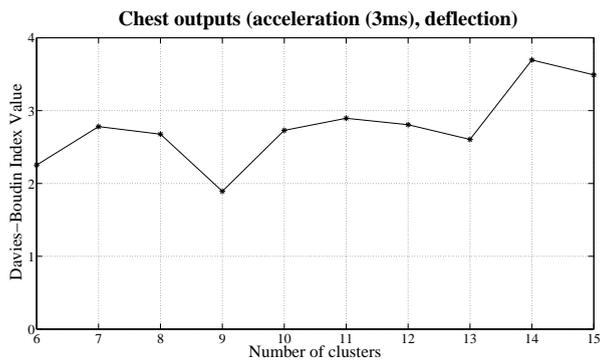
Figure E.10: Neck combination (passenger, USA, HIII 5th dummy belted)

E.3.1/ DUMMY HEAD

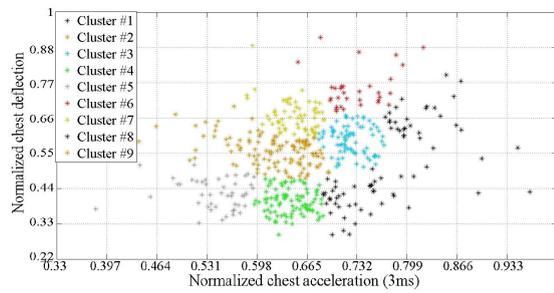
E.3.2/ DUMMY NECK

E.3.3/ DUMMY CHEST

E.3.4/ DUMMY FEMURS

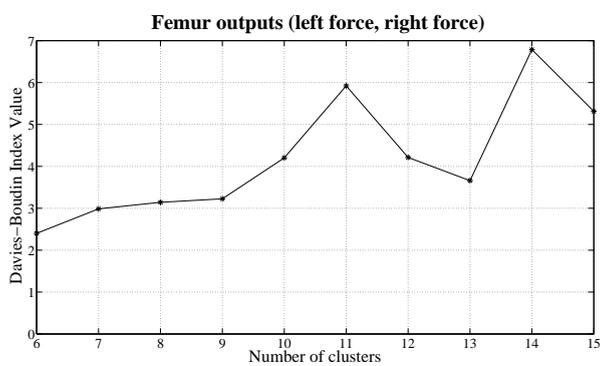


(a) Representation of the DBI values

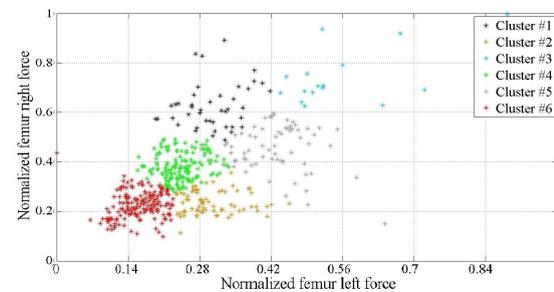


(b) Clustering dummy chest (9 clusters): graphical representation

Figure E.11: Chest combination (passenger, USA, HIII 5th dummy belted)



(a) Representation of the DBI values



(b) Clustering dummy femurs (6 clusters): graphical representation

Figure E.12: Femurs combination (passenger, USA, HIII 5th dummy belted)

F

COMPLEMENT TO CHAPTER 4.2 (FMVSS NR.208, DRIVER SIDE)

Because the method for the determination of the clusters for the driver side in FMVSS NR.208 is the same as for the passenger side for both dummies, only the DBI curves and the corresponding graphical representation of the clusters will be represented.

For the legislation case, the DBI has been calculated from 4 up to 10 clusters. For the rating (HIII 50th dummy belted), the DBI has been calculated from 6 up to 15 clusters.

F.1/ DRIVER SIDE : HIII 50TH DUMMY UNBELTED

This section regroups the DBI curves and the graphical representations of the clusters for the HIII 50th dummy unbelted placed on the driver side. The head, the neck, the chest and the femurs have been analysed.

F.1.1/ DUMMY HEAD

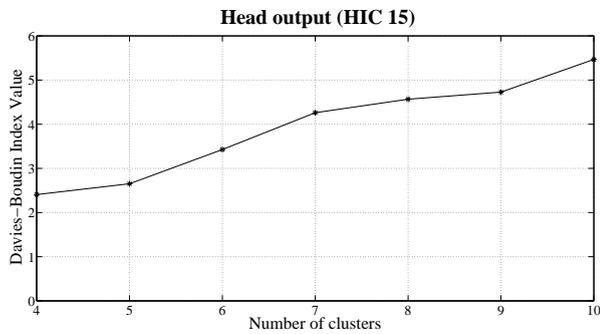
Fig. F.1(a) presents the DBI values for the dummy head. According to it, the best clustering is the clustering with 4 clusters that are represented in Fig. F.1(b). These clusters have directly been formed by the GK algorithm.

These clusters will be used for the study of the parameters' influence on the dummy head for this configuration.

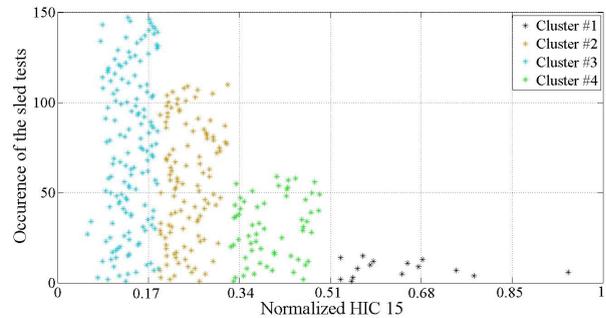
F.1.2/ DUMMY NECK

Fig. F.2(a) presents the DBI values for the dummy neck. According to it, the best clustering is the clustering with 4 clusters that are represented in Fig. F.2(b). These clusters have directly been formed by the GK algorithm.

These clusters will be used for the study of the parameters' influence on the dummy neck for this configuration.

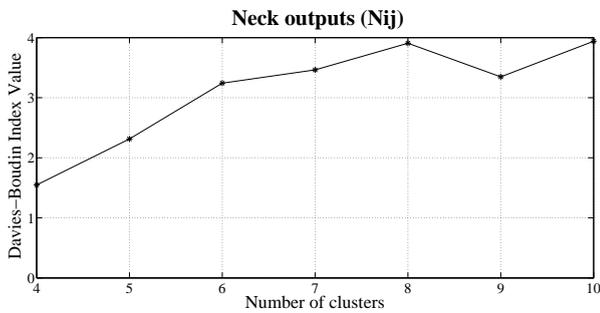


(a) Representation of the DBI values

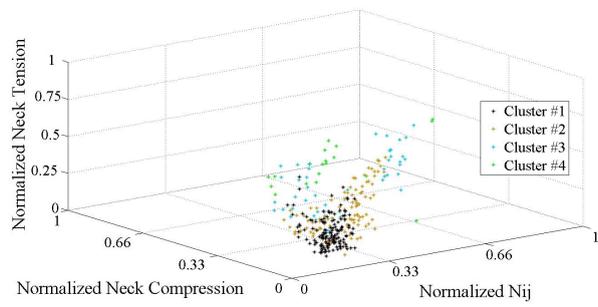


(b) Clustering dummy head (4 clusters): graphical representation

Figure F.1: Head combination (driver, USA, HIII 50th dummy unbelted)



(a) Representation of the DBI values



(b) Clustering dummy neck (4 clusters): graphical representation

Figure F.2: Neck combination (driver, USA, HIII 50th dummy unbelted)

F.1.3/ DUMMY CHEST

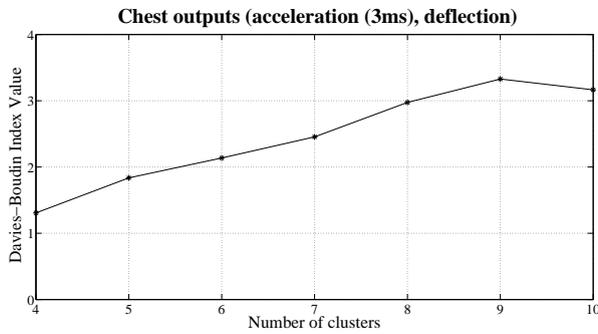
Fig. F.3(a) presents the DBI values for the dummy chest. According to it, the best clustering is the clustering with 4 clusters. A fifth cluster has been manually created in order to regroup the sled tests with values up 80% of the legal values into a same cluster. They are represented in Fig. F.3(b). The four clusters have directly been formed by the GK algorithm and the fifth manually.

These clusters will be used for the study of the parameters' influence on the dummy chest for this configuration.

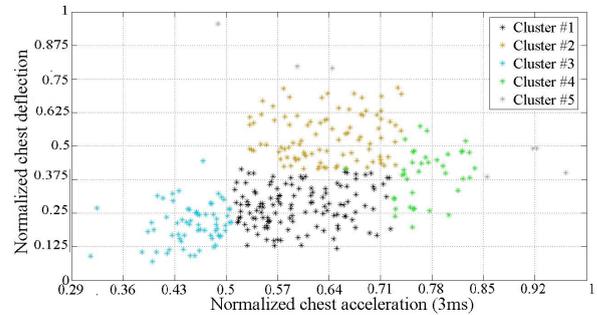
F.1.4/ DUMMY FEMURS

Fig. F.4(a) presents the DBI values for the dummy femurs. According to it, the best clustering is the clustering with 4 clusters. The clustering in 8 clusters presents more advantages because of the separation of the values up 80% and 100% of the legal values. It will be helpful for the future analysis of the input's parameters. The clusters are represented in Fig. F.4(b). They have directly been formed by the GK algorithm.

These clusters will be used for the study of the parameters' influence on the dummy

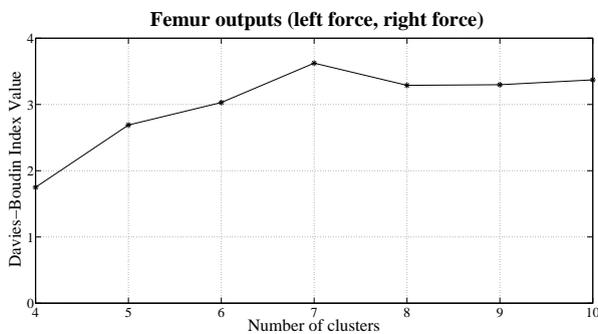


(a) Representation of the DBI values

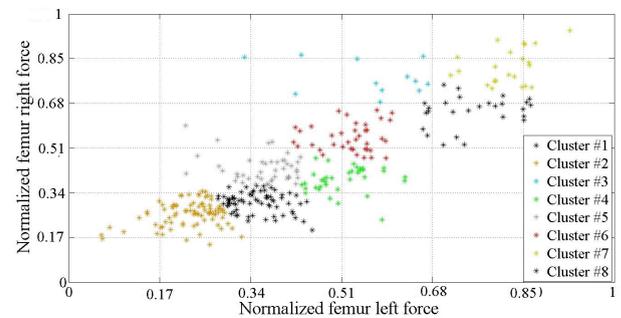


(b) Clustering dummy chest (4 clusters): graphical representation

Figure F.3: Chest combination (driver, USA, HIII 50th dummy unbelted)



(a) Representation of the DBI values



(b) Clustering dummy femurs (4 clusters): graphical representation

Figure F.4: Femurs combination (driver, USA, HIII 50th dummy unbelted)

femurs for this configuration.

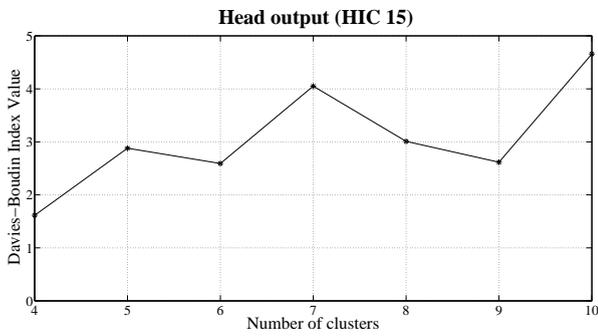
F.2/ DRIVER SIDE : HIII 5TH DUMMY UNBELTED

This section regroups the DBI curves and the graphical representations of the clusters for the HIII 5th dummy unbelted placed on the driver side. The head, the neck, the chest and the femurs have been analysed.

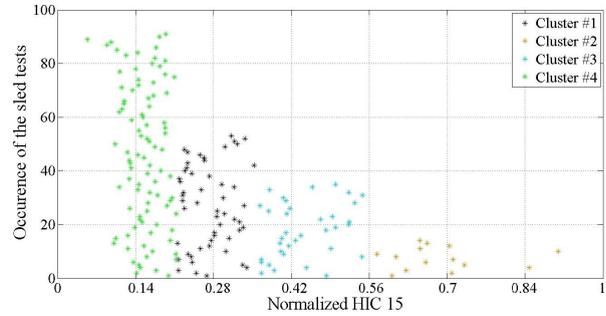
F.2.1/ DUMMY HEAD

Fig. F.5(a) presents the DBI values for the dummy head. According to it, the best clustering is the clustering with 4 clusters that are represented in Fig. F.5(b). These clusters have directly been formed by the GK algorithm.

These clusters will be used for the study of the parameters' influence on the dummy head for this configuration.



(a) Representation of the DBI values

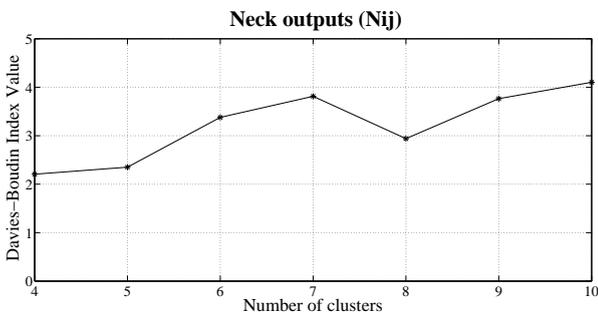


(b) Clustering dummy head (4 clusters): graphical representation

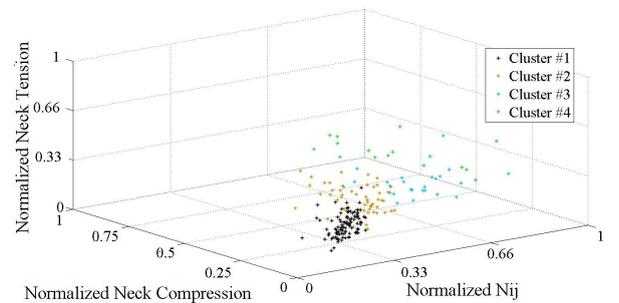
Figure F.5: Head combination (driver, USA, HIII 5th dummy unbelted)

F.2.2/ DUMMY NECK

Fig. F.6(a) presents the DBI values for the dummy neck. According to it, the best clustering is the clustering with 4 clusters that are represented in Fig. F.6(b). These clusters have directly been formed by the GK algorithm.



(a) Representation of the DBI values



(b) Clustering dummy neck (4 clusters): graphical representation

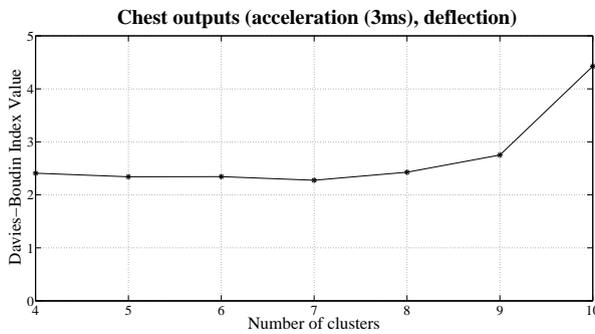
Figure F.6: Neck combination (driver, USA, HIII 5th dummy unbelted)

These clusters will be used for the study of the parameters' influence on the dummy neck for this configuration.

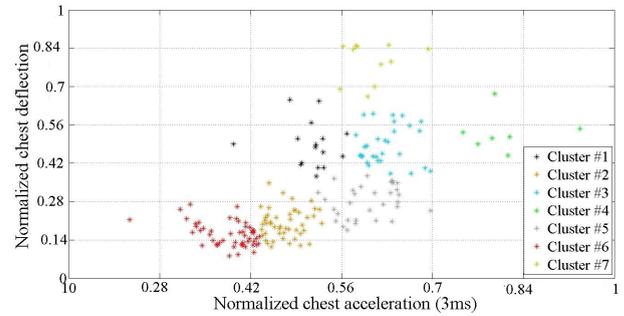
F.2.3/ DUMMY CHEST

Fig. F.7(a) presents the DBI values for the dummy chest. According to it, the best clustering is the clustering with 7 clusters that are represented in Fig. F.7(b). These clusters have directly been formed by the GK algorithm.

These clusters will be used for the study of the parameters' influence on the dummy chest for this configuration.



(a) Representation of the DBI values

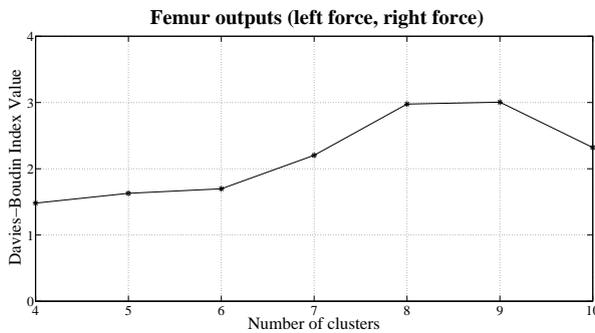


(b) Clustering dummy chest (7 clusters): graphical representation

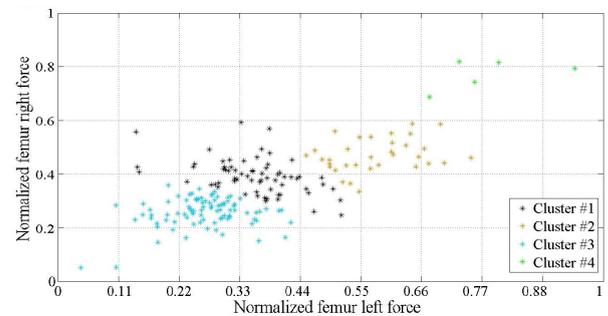
Figure F.7: Chest combination (driver, USA, HIII 5th dummy unbelted)

F.2.4/ DUMMY FEMURS

Fig. F.8(a) presents the DBI values for the dummy femurs. According to it, the best clustering is the clustering with 4 clusters that are represented in Fig. F.8(b). These clusters have directly been formed by the GK algorithm.



(a) Representation of the DBI values



(b) Clustering dummy femurs (4 clusters): graphical representation

Figure F.8: Femurs combination (driver, USA, HIII 5th dummy unbelted)

These clusters will be used for the study of the parameters' influence on the dummy femurs for this configuration.

F.3/ DRIVER SIDE : HIII 50TH DUMMY BELTED (RATING)

This section regroups the DBI curves and the graphical representations of the clusters for the HIII 50th dummy belted placed on the driver side. The head, the neck, the chest and the femurs have been analysed. This case corresponds to the rating, it means that the biomechanical values have to be under the legislation limits but also as low as possible to get the best score. That is why the DBI curves have been done for 6 until 15 clusters.

F.3.1/ DUMMY HEAD

Fig. F.9(a) presents the DBI values for the dummy head. According to it, the best clustering is the clustering with 7 clusters that are represented in Fig. F.9(b). These clusters have directly been formed by the GK algorithm.

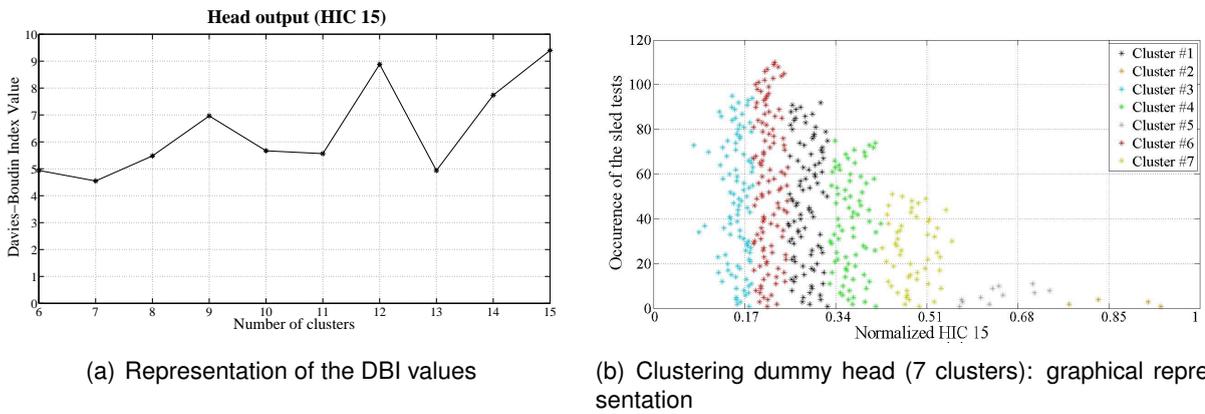


Figure F.9: Head combination (driver, USA, HIII 50th dummy belted)

These clusters will be used for the study of the parameters' influence on the dummy head for this configuration.

F.3.2/ DUMMY NECK

Fig. F.10(a) presents the DBI values for the dummy neck. According to it, the best clustering is the clustering with 6 clusters that are represented in Fig. F.10(b). These clusters have directly been formed by the GK algorithm.

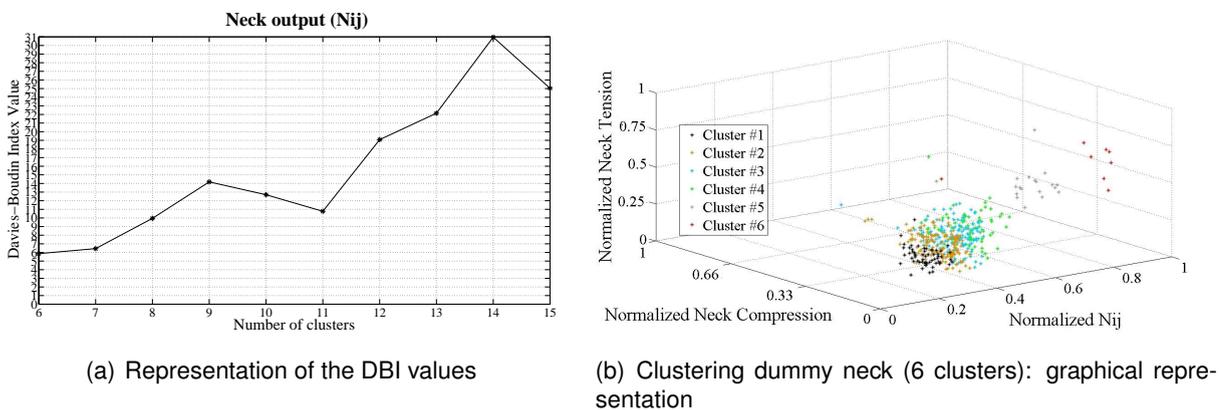
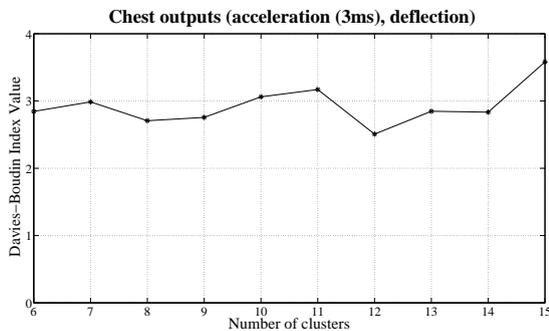


Figure F.10: Neck combination (driver, USA, HIII 50th dummy belted)

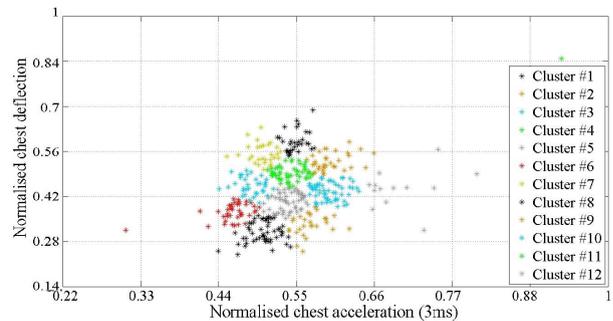
These clusters will be used for the study of the parameters' influence on the dummy neck for this configuration.

F.3.3/ DUMMY CHEST

Fig. F.11(a) presents the DBI values for the dummy chest. According to it, the best clustering is the clustering with 12 clusters that are represented in Fig. F.11(b). These clusters have directly been formed by the GK algorithm. For the US NCAP rating, only the chest deflection is taken under consideration but the chest acceleration has to be under the 80% limit. Because lots of sled tests were upper this limit, the clustering has been done with both outputs.



(a) Representation of the DBI values



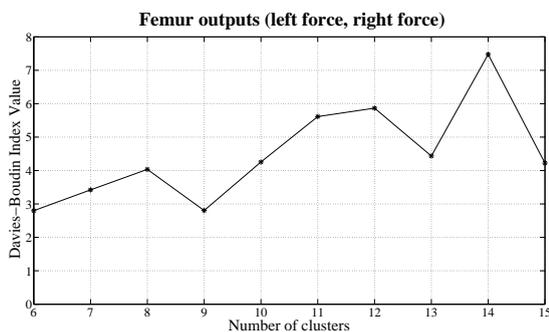
(b) Clustering dummy chest (12 clusters): graphical representation

Figure F.11: Chest combination (driver, USA, HIII 50th dummy belted)

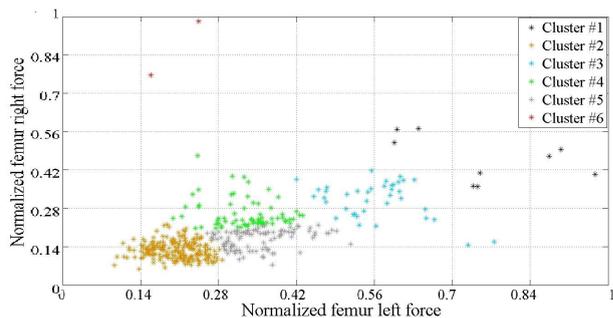
These clusters will be used for the study of the parameters' influence on the dummy chest for this configuration.

F.3.4/ DUMMY FEMURS

Fig. F.12(a) presents the DBI values for the dummy femurs. According to it, the best clustering is the clustering with 6 clusters that are represented in Fig. F.12(b). These clusters have directly been formed by the GK algorithm.



(a) Representation of the DBI values



(b) Clustering dummy femurs (6 clusters): graphical representation

Figure F.12: Femurs combination (driver, USA, HIII 50th dummy belted)

These clusters will be used for the study of the parameters' influence on the dummy

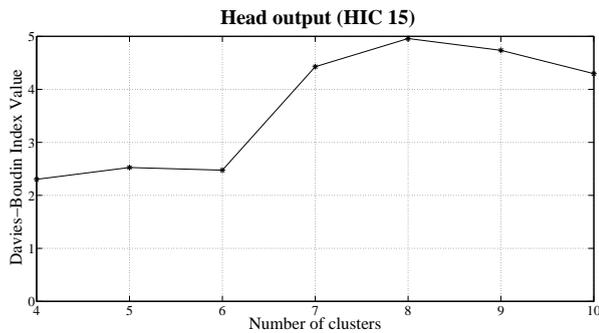
femurs for this configuration.

F.4/ DRIVER SIDE : HIII 5TH DUMMY BELTED

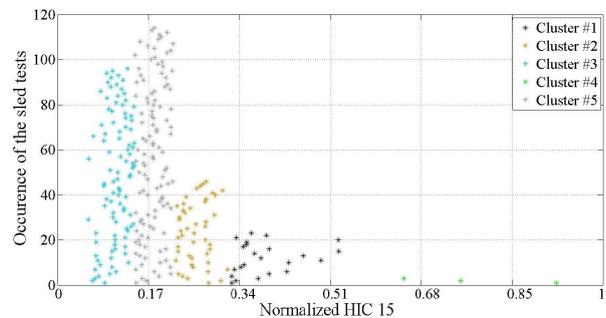
This section regroups the DBI curves and the graphical representations of the clusters for the HIII 5th dummy belted placed on the driver side. The head, the neck, the chest and the femurs have been analysed.

F.4.1/ DUMMY HEAD

Fig. F.13(a) presents the DBI values for the dummy head. According to it, the best clustering is the clustering with 4 clusters. A fifth cluster has been manually created in order to regroup the sled tests with values up 80% of the legal values into a same cluster. They are represented in Fig. F.13(b). The four clusters have directly been formed by the GK algorithm and the fifth manually.



(a) Representation of the DBI values



(b) Clustering dummy head (4 clusters): graphical representation

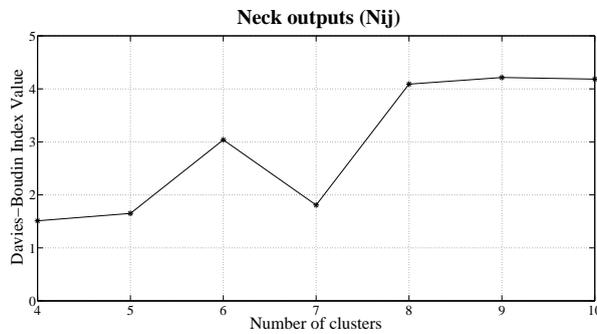
Figure F.13: Head combination (driver, USA, HIII 5th dummy belted)

These clusters will be used for the study of the parameters' influence on the dummy head for this configuration.

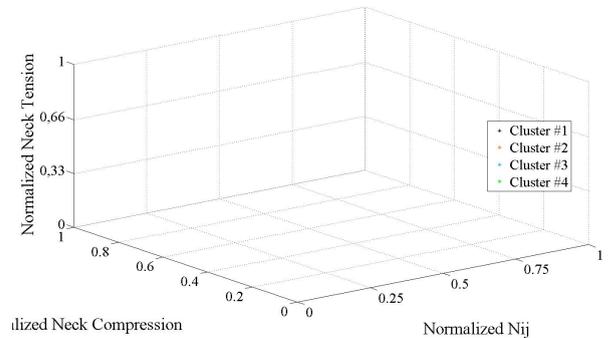
F.4.2/ DUMMY NECK

Fig. F.14(a) presents the DBI values for the dummy neck. According to it, the best clustering is the clustering with 4 clusters that are represented in Fig. F.14(b). These clusters have directly been formed by the GK algorithm.

These clusters will be used for the study of the parameters' influence on the dummy neck for this configuration.



(a) Representation of the DBI values

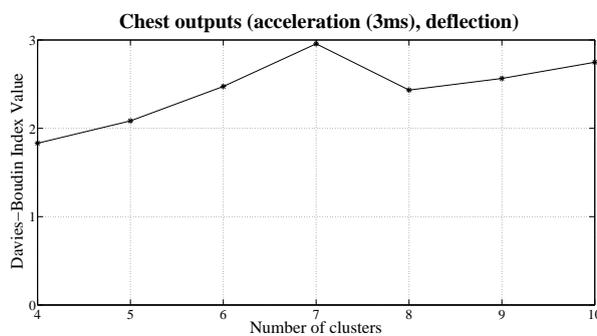


(b) Clustering dummy neck (6 clusters): graphical representation

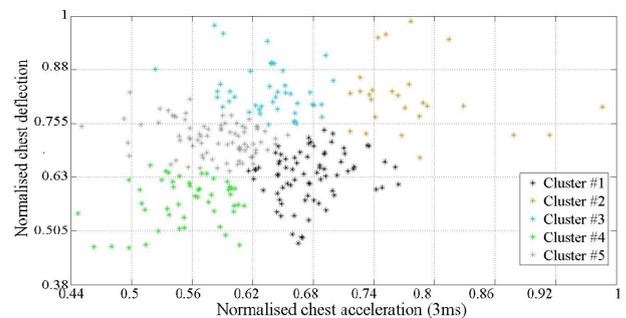
Figure F.14: Neck combination (driver, USA, HIII 5th dummy belted)

F.4.3/ DUMMY CHEST

Fig. F.15(a) presents the DBI values for the dummy chest. According to it, the best clustering is the clustering with 4 clusters. The representation in 5 clusters presents a better separation of the legal values according to the ground truth. They are represented in Fig. F.15(b) and have directly been formed by the GK algorithm.



(a) Representation of the DBI values



(b) Clustering dummy chest (8 clusters): graphical representation

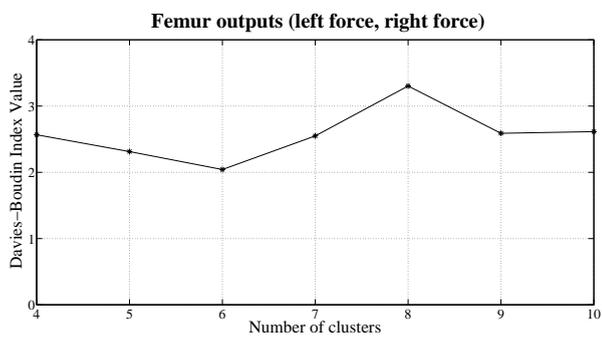
Figure F.15: Chest combination (driver, USA, HIII 5th dummy belted)

These clusters will be used for the study of the parameters' influence on the dummy chest for this configuration.

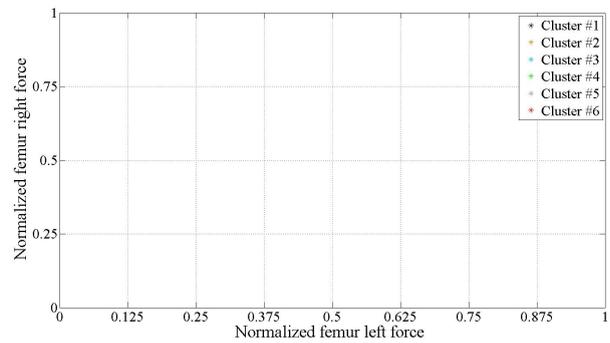
F.4.4/ DUMMY FEMURS

Fig. F.16(a) presents the DBI values for the dummy femurs. According to it, the best clustering is the clustering with 6 clusters that are represented in Fig. F.16(b). These clusters have directly been formed by the GK algorithm.

These clusters will be used for the study of the parameters' influence on the dummy femurs for this configuration.



(a) Representation of the DBI values



(b) Clustering dummy femurs (6 clusters): graphical representation

Figure F.16: Femurs combination (driver, USA, HIII 5th dummy belted)

DECISION TREES EURO NCAP

This Appendix contains the decision trees and validation results that have not been explained in the main report.

G.1/ DUMMY HEAD

Table G.1 resumes the sled tests' repartition according to the head value and the restraint systems parameters. The build of a decision tree¹ (Fig G.1) enables to detect the parameters influencing the dummy head score in a EURO NCAP sled test, supported by the calculation of the VIM (Table G.2). Finally, Table G.3 presents the results obtained by the LOOCV validation and confirms the potential of the model: in fact, 82% of the values have been correctly assigned in the right cluster by the proposed decision tree.

Table G.1: Sled tests' repartition for the dummy head (passenger, EURO NCAP)

Variables	Cluster #1	Cluster #2	Cluster #3	Cluster #4
Number of sled tests	122	10	2	6
PAB				
L-shape	95.6%	2.2%	0.0%	2.2%
Square-shape	79.5%	11.5%	2.6%	6.4%
3D Add-on	100.0%	0.0%	0.0%	0.0%
3D-shape	100.0%	0.0%	0.0%	0.0%
other	100.0%	0.0%	0.0%	0.0%
Divided Panel				
no	86.6%	7.1%	1.6%	4.7%
yes	92.3%	7.7%	0.0%	0.0%
Volume				
<90 L	50.0%	50.0%	0.0%	0.0%
90-100 L	95.7%	0.0%	2.1%	2.1%
101-110 L	0.0%	0.0%	100.0%	0.0%
111-120 L	94.7%	3.5%	0.0%	1.8%
>120 L	67.7%	19.4%	0.0%	12.9%
Number of VH				
no	96.9%	1.5%	0.0%	1.5%
yes	78.7%	12.0%	2.7%	6.7%
ABV				
no	86.1%	7.4%	1.6%	4.9%
yes	94.4%	5.6%	0.0%	0.0%
Tether 1				

¹Shifted values to avoid representing a real case

no	86.8%	7.0%	1.6%	4.7%
yes	90.9%	9.1%	0.0%	0.0%
Inflator Type				
Pyrotechnic	95.2%	4.8%	0.0%	0.0%
Hybrid	83.3%	8.3%	2.1%	6.3%
Cold gas	100.0%	0.0%	0.0%	0.0%
Number of Stages				
1	94.9%	3.4%	0.0%	1.7%
2	81.5%	9.9%	2.5%	6.2%
TTF Stage 1				
<14 ms	91.6%	4.7%	1.9%	1.9%
14-16 ms	91.7%	8.3%	0.0%	0.0%
17-20 ms	46.7%	26.7%	0.0%	26.7%
>20 ms	100.0%	0.0%	0.0%	0.0%
Type of LL				
CLL	91.6%	4.7%	1.9%	1.9%
DLL	91.7%	8.3%	0.0%	0.0%
PLL	46.7%	26.7%	0.0%	26.7%
SLL	100.0%	0.0%	0.0%	0.0%
LL Value 1				
<2.5 kN	50.0%	25.0%	0.0%	25.0%
2.5-3.4 kN	86.7%	6.7%	2.2%	4.4%
3.5-4.5 kN	94.5%	4.1%	1.4%	0.0%
4.6-5.5 kN	100.0%	0.0%	0.0%	0.0%
5.6-6.5 kN	0.0%	0.0%	0.0%	0.0%
Retractor pretensioner				
no	92.4%	3.0%	3.0%	1.5%
yes	82.4%	10.8%	0.0%	6.8%
Bucke pretensioner				
no	81.0%	11.1%	0.0%	7.9%
yes	92.2%	3.9%	2.6%	1.3%
Anchor pretensioner				
no	86.5%	7.1%	1.6%	4.8%
no	92.9%	7.1%	0.0%	0.0%
Seatbelt Position				
Steep	91.2%	2.9%	5.9%	0.0%
Mid	77.8%	15.6%	0.0%	6.7%
Down	91.8%	3.3%	0.0%	4.9%
Webbing on Spool				
<750 mm	100.0%	0.0%	0.0%	0.0%
750-800 mm	100.0%	0.0%	0.0%	0.0%
801-850 mm	83.3%	0.0%	0.0%	6.7%
851-900 mm	69.2%	23.1%	0.0%	7.7%
901-950 mm	75.0%	15.6%	0.0%	9.4%
951-1000 mm	91.7%	4.2%	0.0%	4.2%
>1000 mm	94.6%	1.8%	3.6%	0.0%
Height Adjustment				
no	93.9%	2.4%	2.4%	1.2%
yes	77.6%	13.8%	0.0%	8.6%
DLT				
no	87.4%	6.7%	1.5%	4.4%
yes	80.0%	20.0%	0.0%	0.0%
KnAB				
no	93.4%	3.3%	1.7%	1.7%
yes	47.4%	31.6%	0.0%	21.1%
Chin to IP				
<420 mm	100.0%	0.0%	0.0%	0.0%
420-470 mm	0.0%	0.0%	0.0%	0.0%
471-500 mm	0.0%	0.0%	0.0%	0.0%
501-550 mm	60.0%	20.0%	0.0%	20.0%

>550 mm	91.5%	5.1%	1.7%	1.7%
Sternum to IP				
<420 mm	100.0%	0.0%	0.0%	0.0%
420-470 mm	92.3%	7.7%	0.0%	0.0%
471-500 mm	94.1%	2.9%	0.0%	2.9%
501-550 mm	82.8%	9.4%	1.6%	6.3%
>550 mm	84.6%	7.7%	3.8%	3.8%
Left Knee - Shorter Distance Ahead				
<40 mm	0.0%	0.0%	0.0%	0.0%
40-49 mm	0.0%	0.0%	0.0%	0.0%
50-59 mm	0.0%	0.0%	0.0%	0.0%
60-69 mm	0.0%	0.0%	0.0%	0.0%
70-79 mm	0.0%	0.0%	0.0%	0.0%
>80 mm	87.1%	7.1%	1.4%	4.3%
Right Knee - Shorter Distance Ahead				
<40 mm	0.0%	0.0%	0.0%	0.0%
40-49 mm	0.0%	0.0%	0.0%	0.0%
50-59 mm	0.0%	0.0%	0.0%	0.0%
60-69 mm	0.0%	0.0%	0.0%	0.0%
70-79 mm	0.0%	0.0%	0.0%	0.0%
>80 mm	87.1%	7.1%	1.4%	4.3%
OLC				
<26 g	93.5%	3.7%	1.9%	0.9%
26-26.9 g	79.2%	8.3%	0.0%	12.5%
27-27.9 g	50.0%	33.3%	0.0%	16.7%
>28 g	0.0%	66.7%	0.0%	33.3%
Dummy Temperature				
<20 °C	100.0%	0.0%	0.0%	0.0%
20-20.9 °C	89.2%	5.4%	2.7%	2.7%
21-21.9 °C	84.7%	9.4%	1.2%	4.7%
>22 °C	93.3%	0.0%	0.0%	6.7%

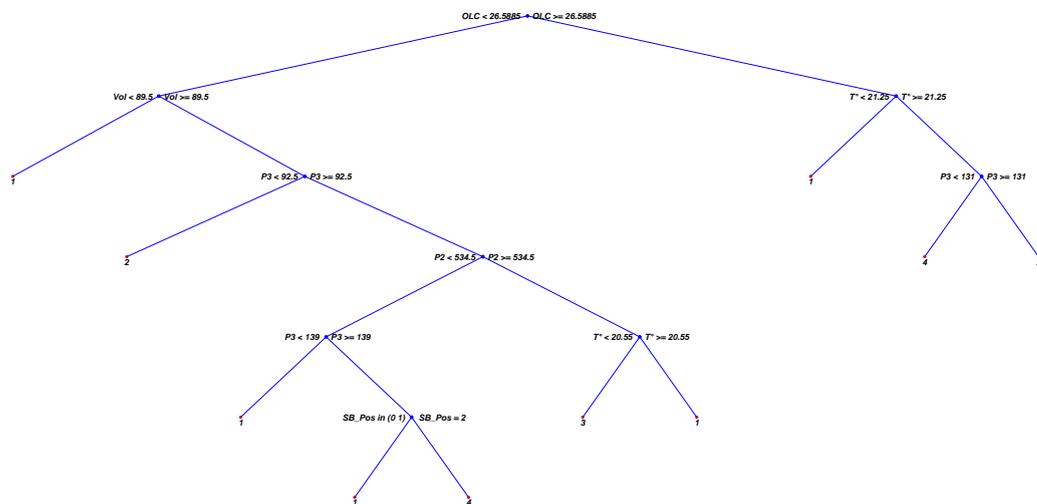


Figure G.1: Decision tree for the dummy head (passenger, EURO NCAP) - Shifted values to avoid representing a real case

Table G.2: Calculation of the VIM for the dummy head (passenger, EURO NCAP)

Predictor Variable	Variable Importance Measure
OLC	0,005472533
Dummy Temperature	0,003769771
Left Knee, Shorter Distance Ahead	0,002906496
Seat belt Position	0,001455026
Volume	0,001341526
Sternum to IP	0,000305456

Table G.3: Validation of the decision tree for the dummy head (passenger, EURO NCAP)

		Predicted injury			
		Cluster #1	Cluster #2	Cluster #3	Cluster #4
Ob. injury*	Cluster #1	112	5	1	4
	Cluster #2	7	3	0	0
	Cluster #3	2	0	0	0
	Cluster #4	5	1	0	0
	Prediction accuracy	82% in the right cluster			
*Obtained injury					

G.2/ DUMMY NECK

Table G.4 resumes the sled tests' repartition according to the neck value and the restraint systems parameters. The build of a decision tree² (Fig G.2) enables to detect the parameters influencing the dummy neck score in a EURO NCAP sled test, supported by the calculation of the VIM (Table G.5). Finally, Table G.6 presents the results obtained by the LOOCV validation.

Table G.4: Sled tests' repartition for the dummy neck (passenger, EURO NCAP)

Variables	Cluster #1	Cluster #2
Number of sled tests	19	121
PAB		
L-shape	2.2%	97.8%
Square-shape	23.1%	76.9%
3D Add-on	0.0%	100.0%
3D-shape	0.0%	100.0%
other	0.0%	100.0%
Divided Panel		
no	15.0%	85.0%
yes	0.0%	100.0%
Volume		
<90 L	50.0%	50.0%
90-100 L	14.9%	85.1%
101-110 L	0.0%	100.0%
111-120 L	10.5%	89.5%
>120 L	12.9%	87.1%
Number of VH		
no	16.9%	83.1%
yes	10.7%	89.3%
ABV		

²Shifted values to avoid representing a real case

no	15.6%	84.4%
yes	0.0%	100.0%
Tether 1		
no	14.7%	85.3%
yes	0.0%	100.0%
Inflator Type		
Pyrotechnic	14.3%	85.7%
Hybrid	13.5%	86.5%
Cold gas	0.0%	100.0%
Number of Stages		
1	22.0%	78.0%
2	7.4%	92.6%
TTF Stage 1		
<14 ms	0.0%	0.0%
14-16 ms	0.0%	0.0%
17-20 ms	10.0%	90.0%
>20 ms	13.8%	86.2%
Type of LL		
CLL	14.0%	86.0%
DLL	0.0%	100.0%
PLL	26.7%	73.3%
SLL	0.0%	100.0%
LL Value 1		
<2.5 kN	25.0%	75.0%
2.5-3.4 kN	17.8%	82.2%
3.5-4.5 kN	9.6%	90.4%
4.6-5.5 kN	0.0%	100.0%
5.6-6.5 kN	0.0%	0.0%
Retractor pretensioner		
no	13.6%	86.4%
yes	13.5%	86.5%
Buckle pretensioner		
no	15.9%	84.1%
yes	11.7%	88.3%
Anchor pretensioner		
no	15.1%	84.9%
no	0.0%	100.0%
Seatbelt Position		
Steep	11.8%	88.2%
Mid	8.9%	91.1%
Down	18.0%	82.0%
Webbing on Spool		
<750 mm	0.0%	100.0%
750-800 mm	0.0%	100.0%
801-850 mm	16.7%	83.3%
851-900 mm	7.7%	92.3%
901-950 mm	37.5%	62.5%
951-1000 mm	16.7%	83.3%
>1000 mm	1.8%	98.2%
Height Adjustment		
no	9.8%	90.2%
yes	19.0%	81.0%
DLT		
no	14.1%	85.9%
yes	0.0%	100.0%
KnAB		
no	12.4%	87.6%
yes	21.1%	78.9%
Chin to IP		
<420 mm	0.0%	100.0%

420-470 mm	0.0%	0.0%
471-500 mm	0.0%	0.0%
501-550 mm	20.0%	80.0%
>550 mm	12.8%	87.2%
Sternum to IP		
<420 mm	0.0%	100.0%
420-470 mm	7.7%	92.3%
471-500 mm	2.9%	97.1%
501-550 mm	23.4%	76.6%
>550 mm	7.7%	92.3%
Left Knee - Shorter Distance Ahead		
<40 mm	0.0%	0.0%
40-49 mm	0.0%	0.0%
50-59 mm	0.0%	0.0%
60-69 mm	0.0%	0.0%
70-79 mm	0.0%	0.0%
>80 mm	13.6%	86.4%
Right Knee - Shorter Distance Ahead		
<40 mm	0.0%	0.0%
40-49 mm	0.0%	0.0%
50-59 mm	0.0%	0.0%
60-69 mm	0.0%	0.0%
70-79 mm	0.0%	0.0%
>80 mm	13.6%	86.4%
OLC		
<26 g	10.3%	89.7%
26-26.9 g	25.0%	75.0%
27-27.9 g	16.7%	83.3%
>28 g	33.3%	66.7%
Dummy Temperature		
<20 °C	0.0%	100.0%
20-20.9 °C	5.4%	94.6%
21-21.9 °C	17.6%	82.4%
>22 °C	13.3%	86.7%

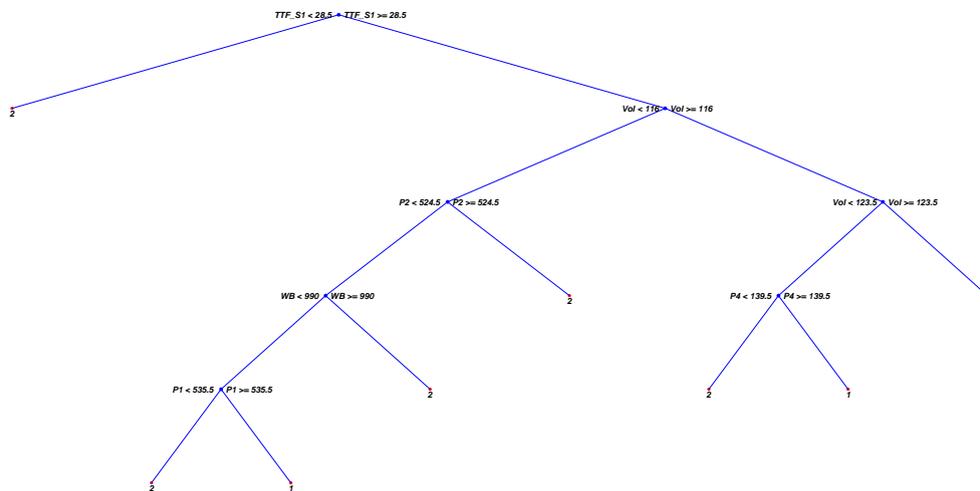


Figure G.2: Decision tree for the dummy neck (passenger, EURO NCAP) - Shifted values to avoid representing a real case

Table G.5: Calculation of the VIM for the dummy neck (passenger, EURO NCAP)

Predictor Variable	Variable Importance Measure
Volume	0,00579165
TTF Stage 1	0,005247813
Sternum to IP	0,005002001
Right Knee, Shorter Distance Ahead	0,001927438
Chin to IP	0,001855288
Webbing on Spool	0,001546073

Table G.6: Validation of the decision tree for the dummy neck (passenger, EURO NCAP)

		Predicted injury	
		Cluster #1	Cluster #2
*O.I.	Cluster #1	8	11
	Cluster #2	9	112
	Prediction accuracy	86% in the right cluster	
*Obtained injury			

Abstract:

Safety is one of the most important considerations when buying a new car. The car has to achieve crash tests defined by the legislation before being selling in a country, what drives to the development of safety systems such as airbags and seat belts. Additionally, ratings like EURO NCAP and US NCAP enable to provide an independent evaluation of the car safety. Frontal sled tests are thus carried out to confirm the protection level of the vehicle and the results are mainly based on injury assessment reference values derived from physical parameters measured in dummies.

This doctoral thesis presents an approach for the treatment of the input data (i.e. parameters of the restraint systems defined by experts) followed by a classification of frontal sled tests according to those parameters. The study is only based on data from the passenger side, the collected data for the driver were not enough completed to produce satisfying results. The main objective is to create a model that evaluates the input parameters' influence on the injury severity and helps the engineers having a prediction of the sled tests results according to the chosen legislation or rating. The dummy biomechanical values (outputs of the model) have been regrouped into clusters in order to define injuries groups. The model and various algorithms have been implemented in a Graphical User Interface for a better practical daily use.

Keywords: Passive safety, restraint systems, frontal sled test, dummy biomechanical values, data mining methods, KDD process, GK algorithm, classification trees

Résumé :

La sécurité automobile est l'une des principales considérations lors de l'achat d'un véhicule. Avant d'être commercialisée, une voiture doit répondre aux normes de sécurité du pays, ce qui conduit au développement de systèmes de retenue tels que les airbags et ceintures de sécurité. De plus, des ratings comme EURO NCAP et US NCAP permettent d'évaluer de manière indépendante la sécurité de la voiture. Des essais catapultes sont entre autres effectués pour confirmer le niveau de protection du véhicule et les résultats sont généralement basés sur des valeurs de référence des dommages corporels dérivés de paramètres physiques mesurés dans les mannequins.

Cette thèse doctorale présente une approche pour le traitement des données d'entrée (c'est-à-dire des paramètres des systèmes de retenue définis par des experts) suivie d'une classification des essais catapultes frontaux selon ces mêmes paramètres. L'étude est uniquement basée sur les données du passager, les données collectées pour le conducteur n'étant pas assez complètes pour produire des résultats satisfaisants. L'objectif principal est de créer un modèle qui définit l'influence des paramètres d'entrées sur la sévérité des dommages et qui aide les ingénieurs à avoir un ordre de grandeur des résultats des essais catapultes selon la législation ou le rating choisi. Les valeurs biomécaniques du mannequin (outputs du modèle) ont été regroupées en clusters dans le but de définir des niveaux de dommages corporels. Le modèle ainsi que les différents algorithmes ont été implémentés dans un programme pour une meilleur utilisation quotidienne.

Mots-clés : Sécurité passive, systèmes de retenue, essai catapulte frontal, valeurs biomécaniques du mannequin, méthodes de data mining, procédé du KDD, algorithmes de GK, arbres de décision

The logo for the SPIM (École doctorale SPIM) is displayed in a large, white, sans-serif font. The letters 'S', 'P', 'I', and 'M' are connected and stylized. A yellow horizontal bar is positioned to the left of the 'S'.