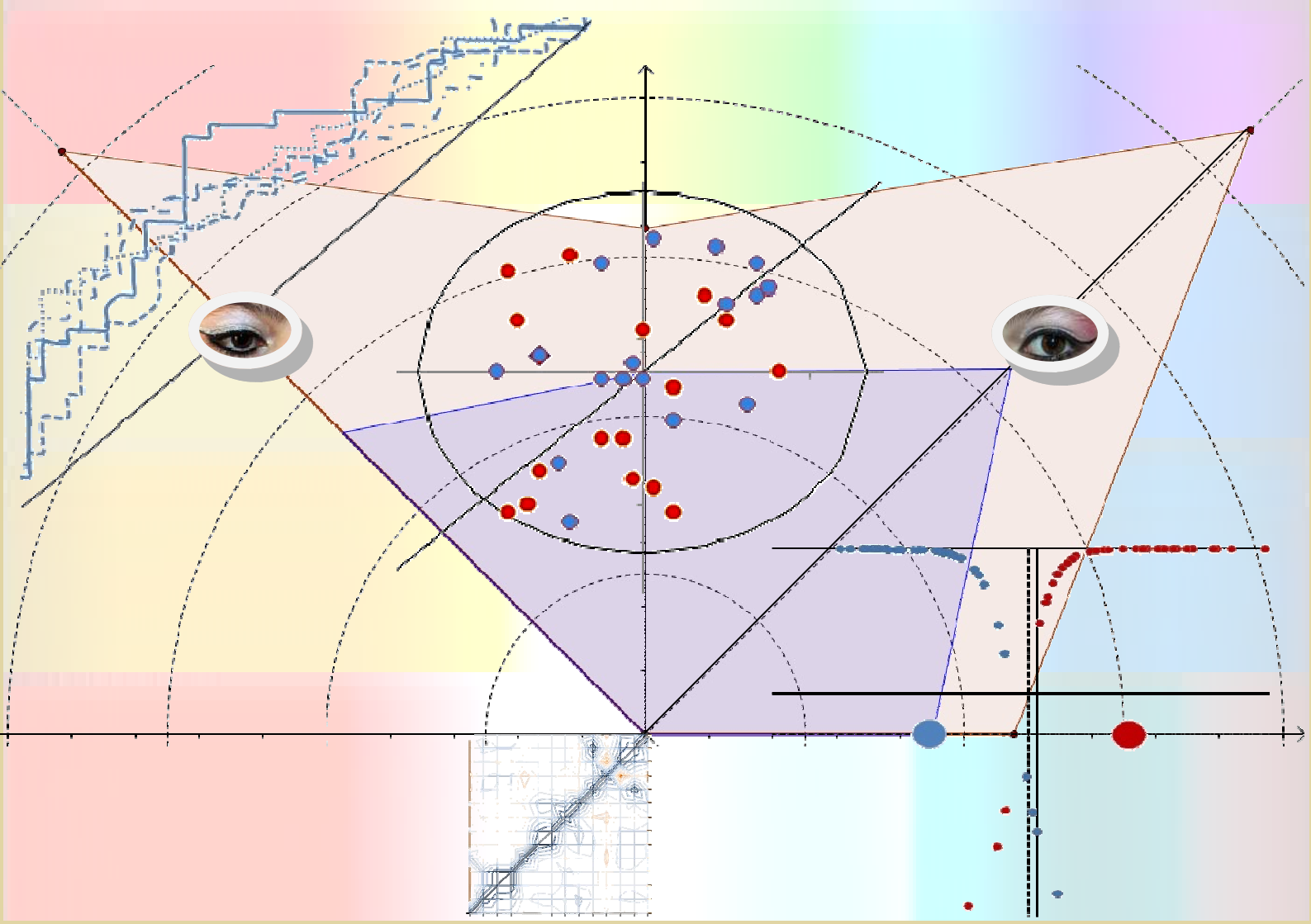

Bárbara Cecília Bessa dos Santos Oliveiros Paiva

Técnicas de Classificação, Diagnóstico e Avaliação de Risco em Doenças com Compromisso da Visão

Classification, Diagnosis and Risk Assessment Methods in Diseases with Visual Impairment

Tese de doutoramento em Ciências da Saúde, Ramo de Ciências Biomédicas, orientada por Prof. Doutor Miguel Castelo-Branco, Professor Doutor Joaquim Murta e Prof. Doutor Alexandre Silva e apresentada à Faculdade de Medicina da Universidade de Coimbra

Agosto 2014

· U C ·

Universidade de Coimbra

Bárbara Cecília Bessa dos Santos Oliveiros Paiva

Técnicas de Classificação, Diagnóstico e Avaliação de Risco em Doenças com Compromisso da Visão

Classification, Diagnosis and Risk Assessment Methods in Diseases with Visual Impairment

Agosto 2014

UNIVERSIDADE DE COIMBRA

Tese de Doutoramento em Ciências da Saúde, Ramo de Ciências Biomédicas, orientada por Prof. Doutor Miguel Castelo-Branco, Professor Doutor Joaquim Murta e Prof. Doutor Alexandre Silva e apresentada à Faculdade de Medicina da Universidade de Coimbra.

A Faculdade de Medicina da Universidade de Coimbra não aceita qualquer responsabilidade em relação à doutrina e à forma desta dissertação.

(Regimento da Faculdade de Medicina Da Universidade de Coimbra, 1931, Artº 108, § único)

To my children, Marta, Filipa e Bernardo

To my husband, Nuno

ACKNOLEGMENTS

In such a work, it is not possible to thank all who have contributed. Over time there were many people who helped in several ways to its achievement. It would be a long list to name all of them. However, I cannot fail to mention the people who, by their direct intervention, played a leading role in the implementation of this thesis.

To Professor Miguel Castelo-Branco, whom I greatly admire given his intellectual and human qualities, I would like to thank for the encouragement, motivation, guidance and friendship that were crucial to this work. My questions and suggestions, as well as the work I developed, always received his attention, with his usual jovial competence. His scientific example was and continues to be fundamental for my training.

To Professor Doctor Joaquim Murta that, from the outset, has always been receptive to collaborate with the established objectives, and for the demonstrated availability.

To Professor Alexandre Silva, I would like to express my gratitude for his indefatigable attention, and for always being available as a co-tutor; his ability to encourage, to read and to motivate the search for new paths and solutions, and for his safe and timely advice. I wish to express my deep gratitude for his encouragement since the beginning of my Master's Thesis, in which he also played a strong role in co-orientation.

To the Faculty of Medicine, University of Coimbra, Portugal, I would like to thank for the support to elaborate the practical work.

To the Institute of Biomedical Imaging and Life Sciences, to the Institute of Nuclear Sciences Applied to Health, to the Ophthalmology, Endocrinology, Cardiology and Imaging Services of the Coimbra Hospital and University Centre, I would like to thank the cooperation provided for obtaining data.

There were also three key people to the realization of this thesis, with a very strong and active contribution. To each one I would like to show my most sincere appreciation.
To my great friend since my youth, Mrs. Catarina Viegas, and to my husband, companion since the days of High School, Engineer Nuno Paiva, thank you for your precious help with the Platform DeGóis. Both were tireless inserting the data of my Curriculum Vitæ.
To my friend and colleague, Researcher Ana Cristina Santos, for the exhaustive reading of all the chapters of the thesis, prior to any tutor, having given me valuable suggestions and recommended insightful changes.

A very special word to my colleagues in the Laboratory of Biostatistics and Medical Informatics, who accompanied me daily, for the good working environment they provided me, being always helpful and available. In particular, I thank Professor Francisco Caramelo who always listen to my doubts and answered to my questions with particular interest, giving critical responses, as it is his feature.

To Professors Maria Filomena Botelho and Emanuel Ponciano, without whom I would never have started this journey. My sincere thanks for their support.

My great esteem and gratitude goes also go to my Mother who will always be connected to any work developed by me, for having instilled in me her fondness for the Theory of Probability and her experience and practice of Statistical Analysis.

I am also grateful to my husband, Nuno, and my three children Marta, Filipa and Bernardo, for their patience and tireless support. Their complaints were more than fair, always!
Marta e Filipa, I would like to thank you for never have charged me for the lack of support in your studies, especially in mathematics and physics, in the last months. Bernardo, I am sorry I have not accompanied you in your wills, proper of your age. It would be fair from you three to do it!
Nuno, I thank you for replacing me, with our children, in an extraordinary way.

To my family and friends, thank you for everything!

INDEX

FIGURE INDEX

TABLES INDEX

GLOSSARY

General abbreviations

AAO	American Academy of Ophthalmology
BDR	Background Diabetic retinopathy
CEO	Clinical and Experimental Optometry
DCCT	Diabetes Control and Complications Trial
DRS	Diabetic Retinopathy Study
DRVS	Diabetic Retinopathy Vitrectomy Study
EDICT	Epidemiology of Diabetes Interventions and Complications Trial
ETDRS	Early Treatment Diabetic Retinopathy Study
HRC	High Risk Characteristics
ICD	International Classification of Diseases
ICO	International Council of Ophthalmology
NPDR	Non-proliferative diabetic retinopathy
NSC	National Screening Committee
OPO	Ophthalmic and Physiological Optics
OVS	Optometry and Vision Science;
PDR	Proliferative Diabetic Retinopathy
SDRGS	Scottish Diabetic Retinopathy Grading Scheme
SQL	Structured Query Language
UKPDS	United Kingdom Prospective Diabetes Study
WHO	World Health Organization

Clinical and demographic assessment

ALT	alanine transaminase
AST	aspartate transaminase
BMI	Body mass index
C-peptide	Connecting peptide
DBP	Diastolic blood pressure
EVC	Erythrocyte variation coefficient
Gama GT	Gamma-glutamyl transferase
HDL	Hight density Lipoprotein
IFCC	International Federation of Clinical Chemistry
IFCC	International Federation of Clinical Chemistry Working Group
LDL	Low density Lipoprotein
MCH	Mean corpuscular haemoglobin
MCHC	Mean corpuscular haemoglobin concentration
MCV	Mean corpuscular volume
MPV	Mean platelet volume
NGSP	National Glycohemoglobin Standardization Program
NGSP	National Glycohemoglobin Standardization Programme
PVC	Platelet variation coefficient
SBP	Systolic blood pressure
TSH	Hormonology measured Thyroid stimulating hormone

Optical Coherence Tomography abbreviations

CS	Central Subfield from Volume Scan
G	Global measure from RNFL
II	Inner Inferior quadrant Volume Scan
IN	Inner Nasal quadrant from Volume Scan
IS	Inner Superior quadrant Volume Scan
IT	Inner Temporal quadrant Volume Scan
N	Nasal quadrant from RNFL
NI	Nasal-Inferior quadrant from RNFL
NS	Nasal-Superior quadrant from RNFL
OCT	Optical Coherence Tomography
OI	Outer Inferior quadrant Volume Scan
ON	Outer Nasal quadrant from Volume Scan
OS	Outer Superior quadrant Volume Scan
OT	Outer Temporal quadrant Volume Scan
RNFL	Retinal Nerve Fibre Layer from OCT
T	Temporal quadrant from RNFL
TI	Temporal-Inferior quadrant from RNFL
TS	Temporal-Superior quadrant from RNFL
VS	Volume Scan from OCT

Decision Tree Algorithms abbreviations

AID	Automatic Interaction Detector
CART	Classification And Regression Tree
CHAID	Chi-squared Automatic Interaction Detection
CLS	Concept Learning Systems
ELISEE	Exploration of Links and Interactions through Segmentation of an Experimental Ensemble
IDEA	Interactive Data Exploration and Analysis
MAID	Multivariate Automatic Interaction Detector
QUEST	Quick, unbiased, efficient, statistical tree[
THAID	THeta Automatic Interaction Detector

Statistical abbreviations

95% CI	95% Confidence Interval
AUC	Area under the ROC curve
CCC	Concordance correlation coefficient
DF	Degrees of Freedom
DFA	Discriminant Function Analysis
GLM	General Linear Model
GRM	General Regression Model
ICC	Intra-class correlation coefficient
LBCI	Lower bound of the 95% Confidence Interval
NLR	Negative Likelihood Ratio
NPV	Negative Predictive Value
P25	Percentile 25 or 1st quartile
P50	Percentile 50 or 2nd quartile or Median
P75	Percentile 75 or 3rd quartile
pCCC	Pseudo-Concordance correlation coefficient
PLR	Positive Likelihood Ratio
PPV	Positive Predictive Value
ROC	Receiver Operating Characteristic curve
SD	Standard Deviation
SEM	Standard Error of the mean
Sens	Sensitivity
Spec	Specificity
UBCI	Upper bound of the 95% Confidence Interval

RESUMO

As doenças da visão incluem a cegueira e a baixa visão, e afetam cerca de 4,25% da população mundial. Cerca de 80% destas podem ser prevenidas ou curadas. Estas estimativas, da Organização Mundial de Saúde, referem que 82% das pessoas com cegueira têm 50 ou mais anos. A sua prevalência está relacionada com o envelhecimento da população, emergindo neste contexto as doenças do segmento posterior. Nestas, inclui-se a retinopatia diabética, uma manifestação clínica da diabetes *mellitus*. Esta doença sistémica é a principal causa de novos casos de cegueira em todo o mundo, entre os 20 e os 74 anos de idade, sendo a complicação referida causada por danos acumulados ao longo do tempo sobretudo nos pequenos vasos sanguíneos na retina.

A diabetes, especialmente do tipo 2, está entre as principais causas de morte e de invalidez, apresentando um elevado peso económico em todo o mundo. Teme-se que esta doença se torne epidémica, dado o aumento da sua incidência e prevalência devido ao crescimento e ao envelhecimento das populações, e ainda a alterações no estilo de vida tais como a redução da atividade física e o aumento da obesidade. Assim, a retinopatia diabética foi adicionada à lista de prioridades no que diz respeito a doenças da visão evitáveis. As últimas estimativas de prevalência de diabetes na população portuguesa entre os 20 e os 79 anos datam de 2012, e referem uma prevalência de 12,9%, representando um aumento de 1,2% desde 2009. Neste ano, a retinopatia diabética foi referida como a principal causa de cegueira na população portuguesa em idade ativa. A necessidade de diagnosticar precocemente ambas as doenças é fundamental em todos os contextos socioeconómicos, a fim de reduzir os seus custos diretos e, principalmente, os custos indiretos e intangíveis, quer para os diabéticos e seus familiares, quer para os Serviços Nacionais de Saúde. Apesar de os métodos para diagnóstico destas doenças estarem claramente definidos, a necessidade de encontrar novos marcadores e classificadores não invasivos, utilizados para rastreio noutros contextos médicos, tornou-se de extrema importância.

Para construir um modelo que identificasse marcadores da diabetes tipo 2, utilizou-se uma amostra de treino constituída por 96 casos, dos quais 49 eram diabéticos tipo 2, com idade compreendida entre os 40 e os 75 anos. O grupo de diabéticos foi usado para o desenvolvimento de um classificador de retinopatia diabética em diabéticos tipo 2, na mesma faixa etária, sendo a amostra constituída por 40 sujeitos, dos quais 20 tinham retinopatia diabética não-proliferativa.

Foi avaliada a correlação e concordância entre as medidas obtidas para os olhos direito e esquerdo, obtidas por Tomografia de Coerência Óptica, concluindo-se que um olho era suficiente para a análise. Foi seleccionado o olho dominante, já que os testes visuais psicofísicos foram realizados apenas neste olho. Foi construída uma medida global do desempenho para cada teste psicofísico (velocidade, visão acromática e visão cromática nos eixos Protan, Deutan e Tritan) com base nos valores obtidos para os meridianos 0º, 45º, 90º e 135º, em cada sujeito.

Posteriormente, foi necessário proceder a uma redução de variáveis, tendo-se comparado os grupos através do teste t-Student para amostras independentes ou do teste de Mann-Whitney, de acordo com a distribuição amostral. Apenas prosseguiram em análise as variáveis que apresentaram diferença estatisticamente significativa entre os grupos, ao nível de significância de 5%. Subsequentemente, foi usada a análise *Receiver Operating Characteristic* (ROC), com o mesmo nível de significância, e identificou-se o conjunto das variáveis que, individualmente, podiam separar os grupos.

Tornou-se assim possível a aplicação de métodos de classificação estatística, tais como a análise discriminante, a regressão logística e a utilização de algoritmos de árvore de decisão, ao conjunto de variáveis remanescentes. O desempenho dos classificadores estatísticos obtidos para a diabetes tipo 2 foi comparado, quer na amostra de treino, quer num conjunto de novos indivíduos participantes. O desempenho dos classificadores para a retinopatia diabética não proliferativa foi avaliado apenas na amostra de treino, mas tenciona-se também testá-lo, futuramente, num conjunto de novos sujeitos. O desempenho dos classificadores foi avaliado através da avaliação da sua acuidade, determinada pela área sob a curva ROC obtida para as probabilidades *a posteriori* de cada um dos modelos, e pela sensibilidade e razão de verossimilhança positiva determinada para as classificações nos grupos.

Um classificador final é apresentado, quer para diabéticos tipo 2 com idades entre 40 e 75 anos de idade, quer para a retinopatia diabética não-proliferativa em diabéticos tipo 2, na

mesma faixa etária, assim como os seus valores preditivos positivos ajustados para os dados mais recentes da prevalência de cada doença na população portuguesa.
A visão cromática relativa ao eixo dos cones Tritan parece desempenhar um papel dominante para a classificação de ambas as doenças.

Palavras-chave: Classificadores estatísticos; Diabetes tipo 2; Retinopatia Diabética; Análise Discriminante; Regressão Logística; Árvores de Decisão

ABSTRACT

Visual impairment, which includes blindness and low vision, affects about 4.25% of the world population, and about 80% is avoidable, since it can be prevented or cured. Those estimates, from the World Health Organization, refer that 82% of blind people are aged 50 or more. The largest proportion of visual impairment is necessarily related to the increase of the ageing of populations, and where posterior segment (retinal) diseases dominate. Among these diseases, there is diabetic retinopathy, an ocular manifestation of diabetes mellitus. This systemic disease, is the leading cause of new cases of blindness around the world in persons aged between 20 and 74 years old, and occurs as a result of long-term accumulated damage to the small blood vessels in the retina. Furthermore, the eye is considered to play an important role in the diagnostic of systemic diseases due to its composition. Every part of the eye is able to give important clues for diagnosis.

Diabetes mellitus, especially type 2, is among the leading causes of death, disability and economic loss throughout the world. It is feared to become an epidemic disease, since its incidence and prevalence are increasing, mainly due to population growth and ageing, as well as a result of alterations in lifestyle, which are leading to the reduction of physical activity and to the increase of obesity. With its increase, diabetic retinopathy was gained a prominent role in the list of preventable visual impairment. The latest prevalence estimates for diabetes in the Portuguese population aged between 20 and 79 years date from 2012, and referred a value of 12.9%, which represents an increment of 1.2% since 2009. In fact, in 2009, diabetic retinopathy was referred as the leading cause of blindness for the Portuguese population in active age.

The need for early diagnosis of both the diseases and its ocular complications is crucial in all socioeconomic contexts, in order to reduce its burden due to its direct costs, and mainly due to its indirect and intangible costs, either for diabetics and their families, or for the National Health Services. In spite of the fact that methods for diagnosing those diseases are clearly

defined, the need to find new markers and non-invasive classifiers used for screening in other medical contexts has become of extreme importance.

A training sample for determination of markers for type 2 diabetes was used, comprising 96 cases, of which 49 were type 2 diabetics, aged between 40 and 75 years old. The group of diabetics was used to build a classifier for diabetic retinopathy in type 2 diabetics in the same age group, and the sample comprised 40 subjects from which 20 had non-proliferative diabetic retinopathy.

Correlation and concordance between measures obtained by Optical Coherence Tomography in the left and right eyes of the same subjects was evaluated, leading to the conclusion that only one eye was needed for the analysis. Hence, the dominant eye was selected for analysis since visual psychophysics tests were performed only in that eye. A global measure of the performance, for each subject, in each one of the visual psychophysics tests (speed, achromatic vision and chromatic vision over the Protan, Deutan and Tritan axes) was build, based upon values obtained for the 0º, 45º, 90º and 135º meridians.

Afterwards, a variable reduction was performed applying an independent samples t test or a Mann-Whitney test, according to data distribution, and only the variables that showed statistical significances, at 5% significance level, were selected to remain in the analysis. Subsequently, a Receiver Operating Characteristic curve was applied to each one of the remaining variables, using the same significance level, and the set of variables which were able to separate groups, individually, was identified.

By then, it was possible to apply different statistical classifying methods, such as discriminant analysis, logistic regression and decision tree algorithms. The performance of the classifiers obtained for type 2 diabetes was compared either in the training set, or in a test set of new subjects. Non-proliferative diabetic retinopathy classifiers were only tested on the training sample, at the moment. Hereafter, we intend to test their performance in a set of new cases. The performance of those classifiers was assessed using accuracy measures, determined by the area under the ROC curve for the posterior probabilities of models, and according to its sensitivity and positive likelihood ratio for group classification.

A final classifier is presented, either for type 2 diabetics aged between 40 and 75 years, or for non-proliferative diabetic retinopathy in type 2 diabetics for the same age group, as well as its positive predictive values adjusted for the latest data on the Portuguese prevalence for each disease.

Whichever the clinical category (presence of disease or complications), chromatic vision over the Tritan cone seems to play a main role for the classification of both diseases.

Keywords: Statistical Classifiers; Type 2 Diabetes; Diabetic Retinopathy; Discriminant Analysis; Logistic Regression; Decision Trees

PART I

CONTEXTUALIZATION

CHAPTER 1

INTRODUCTION

Visual impairment, which includes low vision and blindness, affects about 4.25% of the world population, and estimates from the World Health Organization, dated from 2010, refer that low vision corresponds to 86% of visual impaired people around the world. Moreover, those estimates refer that about 80% of vision impairment is avoidable since it can be either prevented, or cured. On the other hand, blindness prevalence is almost constant since 2002, according to the World Health Organization, but visual impairment prevalence is rising since 2002, from 2.59%, due to the increase on low vision prevalence. Higher differences are observed in the South-Eastern and Western Pacific regions, mainly due to the increase of incidence in India and China, where the prevalence of visual impaired people was, in 2010, respectively 21.9 and 26.5%. It is also known, according to the World Health Organization, that about 90% of visual impaired people live in developing countries, and that 82% of blind people are aged 50 or more.

The main causes of visual impairment around the world are uncorrected refractive errors, cataract, glaucoma, age related macular degeneration, diabetic retinopathy, trachoma and corneal opacities. The largest proportion of blindness, as well as low vision, is necessarily related to the ageing of populations, which is increasing, and where posterior segment (retinal) diseases emerge most. Therefore, age related macular degeneration, glaucoma and diabetic retinopathy are becoming a dominant cause of visual impairment. Furthermore, there are many diseases, namely systemic diseases, with relevant ocular manifestations. The case of diabetes, which is the leading cause of new cases of blindness around the world in persons aged between twenty and seventy four years old according to the World Health organization, is one example. The ocular manifestations of diabetes are diabetic retinopathy

and macular oedema, which affects up to 80% of those who have had the disease for 15 years or more, and occurs as a result of long-term accumulated damage to the small blood vessels in the retina.

Moreover, diabetes is related to nephropathy, being the leading cause of kidney failure, cardiac diseases (about half of diabetics die from cardiovascular disease), and other systems failures which are also known to be related with the eye. Whenever diabetes is combined with reduced blood flow, neuropathy on the feet increases the chance of foot ulcers, infection and eventual need for limb amputation.

Furthermore, the eye is considered to play an important role in the diagnostic of systemic diseases, since it is composed by many different types of tissues and every part of the eye is able to give important clues for diagnosing systemic diseases, which signs may be evident on the outer surface of the eye (such as eyelids, conjunctiva and cornea), on the middle of the eye, or at the back on the retina.

In fact, diabetes mellitus is among the leading causes of death, disability and economic loss throughout the world, and feared to become an epidemic disease. Systemic diseases are a major cause of mortality and, according to the World Health Organization, chronic non-communicable diseases are, by far, the leading cause of mortality in the world.

According to that organization, simple lifestyle measures, such as maintaining a normal body weight, being physically active, eating a healthy diet and reducing sugar and saturated fats intake, and also avoiding alcohol and tobacco use are measures that have shown to be effective in preventing or delaying the onset of type 2 diabetes. Adequate treatment of diabetes, as well as the control of blood pressure and some lifestyle factors such as tobacco use, regular exercise and food habits are, thus, important.

Diabetes mellitus is one of the most common chronic diseases in nearly all countries, especially type 2 diabetes, and its incidence is becoming higher, mainly due to population growth and ageing, but also as a result of alterations in lifestyles, which are leading to the reduction of physical activity and to the increase of obesity.

With the increase of diabetes in the population, diabetic retinopathy was added to the priority list of avoidable visual impairment.

In Portugal, the prevalence of diabetes in the population between 20 and 79 years of age was, in 2009, 11.7% and estimates were that one in each three Portuguese (34.9%) was already diabetic without having its knowledge, or that he/she was at risk of becoming diabetic. Likewise, half of diabetics do not know that they have the disease. Data from 2012 refer that diabetes prevalence is 12.9%, which represents an increment of 1.2% in only three years. The diabetes prevalence is higher in men, increasing with age and with body mass index. Gestational diabetes prevalence, which is related to later onset of type 2 diabetes, also has increased from 3.4% in 2005 and 2006 to values of 4.9% in 2011 and 4.8% in 2012, especially in older pregnant.

According to data from 2009 published by the Portuguese Study Group on Vitreo and Retina (PSGVR), diabetic retinopathy is the leading cause of blindness in the active age group. The PSGVR refers that It develops nearly in all persons with type 1 diabetes, and in more than 77% of people that survived for 20 or more years with type 2 diabetes. The threat to sight can be due to the growth of new vessels leading to intraocular haemorrhage and possible retinal detachment with profound global vision loss, and also to localised damage to the macula or fovea of the eye with loss of central visual acuity. However, the World Health Organization referred in 2005 that evidence-based treatment could reduce the risk for severe vision loss and blindness from proliferative diabetic retinopathy by more than 90%.

Therefore, the need to early diagnose diabetes is crucial in all the socioeconomic contexts, since treatment, hospitalization and complications of diabetes represent a burden both for diabetics and for National Health Services. On the other hand, intangible costs either for a diabetic person or for his/her family are considerably high and, in a certain way, higher than direct and indirect costs. Methods for diagnosing diabetes are clearly defined, but the need to find new markers of the disease, as well as simpler non-invasive classifiers that may be used for screening in other medical contexts (which may raise the suspicion of the presence of diabetes) are emerging. On the other hand, simple markers of the presence of diabetic retinopathy are also useful in order to try to delay the progression of the disease.

Particular care has to be taken when considering the use of applied statistics to ophthalmology, among other medical areas that use data obtained from two similar organs. Often, correlated information from both eyes was used in the past as if two independent measures were being analysed, artificially duplicating the sample size in some cases,

whenever the unit of analysis was the subject and not the eye. Other times, eyes were analysed separately, in spite of the fact that the relevant unit of analysis was the subject, or some other criteria was used, such as choosing only one eye for analysis, or the mean of both eyes. These options prevent the problem of statistical dependence but information is lost. In any case, different results may be achieved according to the chosen eye, and perhaps the best choice is to randomly select one eye for analysis, instead of selecting the right or the left eye, or the best or the worse eye, or even to use the mean or median of both eyes. Decision must take into account the context of the problem. However, as we had tests performed in the dominant eye, and tests performed in both eyes, with strong correlation and concordance between eyes, then the chosen eye for analysis should be the dominant eye for all tests, and results and conclusions must report always to measures obtained for the dominant eye. We will also propose a measure for evaluating concordance between eyes whenever data is not normally distributed, and to evaluate the random error between eye measurements, according to the nature of the case in order to evaluate and compare random error between controls and type 2 diabetics.
Having defined procedures for handling, in a given way that we found appropriate for the present context, with duplicate data, then we were able to establish statistical classifiers and to identify possible markers of either type 2 diabetes or diabetic retinopathy.

In statistics, classification problems are the ones that allow the assignment of individuals to a given group, according to the set of characteristics of the subject that must be quantifiable. Any mathematical function or algorithm that implements that procedure is called a classifier. This situation, in machine learning, is considered to be an instance of supervised learning, while for statistics, it is a problem that may be related either to supervised or to unsupervised learning, since involves grouping data into categories according to measures of similarity or measures of dissimilarity. In fact, classification may be included in a more general problem of pattern recognition and profile detection, and then the classical probability theory is necessarily involved in the process. Other forms of classifying are related to artificial intelligence, which often involve machine learning procedures, but that are more distant from the previous methods, and data mining algorithms. Data mining is procedure that reflects a mixture of machine learning and statistical theory, with a goal in mind. It is often applied when massive data are available, namely whenever we have more variables than cases to analyse, which is of course problematic. We applied data mining

techniques to reduce the number of variables since, initially, we had more than 100 variables available measures in less than 100 cases. This reduction of variables was achieved using classical statistical techniques, such as comparison between groups that were being studied, and more recent techniques such as Receiver Operating Characteristic curves applied to the remaining set of variables, which have showed differences between groups, in order to use only variables that were able to discriminate between groups in the exploratory analysis. Hence, both univariate statistical techniques were used just as exploratory statistics in order to conduct a variable reduction for the use of multivariate classification methods.

Classification was, then, performed using classical statistical techniques such as discriminant analysis (which has strong assumption related to data distribution) and regression procedures, namely logistic regression (without assumption on data distribution, which is usually binary, but with assumptions on multicolinearity), and using algorithms that are related to data mining procedures, such as decision tree (without assumptions). These methods were applied to a subset of all available data, named the training sample, which was formed with the available data on the 31th of December of 2013. Data available after this date were used to test developed classifiers, and to evaluate their performance and accuracy in previsions, so that the best model would ideally be used for classification of new subjects, determining its predictive values according to the most up to date prevalence data.

CHAPTER 2

OBJECTIVES

The main purpose of this work is to develop approaches toward a non-invasive simple classifier for type 2 diabetes based upon measures obtained just in the dominant eye, for volume scan density, retinal nerve fibre layer thickness (obtained with Optical Coherence Tomography) and visual psychophysics tests on speed, achromatic and chromatic vision, according to sociodemographic factors, daily habits, and parameters from collected blood samples, which may be used for screening purpose in subjects aged between 40 and 75 years old.

We also intend to build a non-invasive simple classifier for the presence of non-proliferative diabetic retinopathy on type 2 diabetics aged between 40 and 75 years old, based upon the same factors.

For each one of the proposed classifiers, either for type 2 diabetes or for diabetic retinopathy, the posterior probability for the presence of the condition will be defined.

Along with the defined outcome measures, we want to give an emphasis on statistical and probability methods concerning several points, enumerated below.

Concerning measures obtained for the left and the right eye, we intend to evaluate the need to use both eye information based upon correlation and concordance between eyes. In spite of the existence of recent guidelines for this subject, we propose other alternative methods which may be more useful according to data distribution.

Still regarding this subject, we intend to use graphical methods for the evaluation of statistical correlation between both eyes, enabling the evaluation of error in measurement between eyes, namely the random error, in order to ascertain about sample size.

Concerning multiple measurements in some tests, such as psychophysics tests, which are performed at four different meridians (meridians 0º, 45º, 90º and 135º), we intend to build a global measure based on a simple mathematic algorithm, and to evaluate its discriminatory ability compared to individual measures for each test.

We intend to compare classification methods using three different approaches. On one hand, we intend to compare classical statistical classifiers, such as discriminant analysis and logistic regression methods, with decision tree algorithms. On the other hand, we intend to compare models that have strong assumptions on data distribution, such as discriminant analysis, with methods without assumptions, such as logistic regression and decision trees, as well as methods that use quantitative information (discriminant analysis and decision trees) against logistic regression performed in the same variables after being dichotomized.

For each one of the developed classifiers, we will hopefully be able to identify markers of the disease.

For the best classifier obtained, either for type 2 diabetes, or for diabetic retinopathy, a posterior probability function will be defined in order to allow classification of new cases.

PART II

THEORETICAL FRAMEWORK

CHAPTER 3

VISUAL IMPAIRMENT AND SYSTEMIC DISEASES WITH OCULAR MANIFESTATIONS

SECTION A

VISUAL IMPAIRMENT

1. Definition of Visual Impairment

The definition of visual impairment, low vision and blindness used by the International Statistical Classification of Diseases (ICD), injuries and causes of death[1] states that visual impairment includes low vision as well as blindness. According to those norms, low vision is defined as visual acuity within 3/60 inclusive and 6/18 exclusive, or a corresponding visual field loss to less than 20 degrees in the better eye with best possible correction (ICD-10 visual impairment categories 1 and 2); blindness is defined as visual acuity of less than 3/60, or a corresponding visual field loss to less than 10 degrees in the better eye with best possible correction (ICD-10 visual impairment categories 3, 4 and 5).

2. Wold wide epidemiological estimates

Estimates on global blindness and visual impairment are recent. The first known estimate, according to Resnikoff[2] was published in 1995 based on data from the year of 1990. Previsions were made to the year of 1996 and projected onto the year of 2020. These data provided a base for the Global Initiative for Elimination of Avoidable Blindness, known as "VISION 2020: the Right to Sight"[3]. Since the first estimates on blindness and visual impairment were published, almost all regions belonging to the World Health organization have carried out population-based studies. In fact, several countries conducted studies mostly in cataract surgery, which provided general information on the status of visual impairment in adults aged 50 or older. Recently, many studies have specifically targeted older adults as a mean of updating global and regional estimates on visual impairment.

The availability of new data is crucial for the constant update on the estimates, especially on other target populations such as younger adults or children, in order to prevent visual impairment, as well as the identification of its major causes.

The latest estimates refer 285 million people worldwide affected by visual impairment, which corresponds to 4.25% of the world population. From these 285 millions, 39 millions are blind (14%) and the other 86% have low vision, but 80% of the vision impairment is avoidable since it can be prevented or cured. Cataract and uncorrected refractive errors were identified as the main causes of avoidable vision impairment, and un-operated cataract and glaucoma as the main causes of avoidable blindness.

The World Health Organisation (WHO) owns data from its member states, which are divided into six regions (African Region, Regions of Americas, European Region, South-East Asia Region, Eastern Mediterranean Region and Western Pacific Region).

The last published prevalence data on age-specific blindness by the World Health Organisation dates from 2010, but there were previous estimates published related to the years of 1995, as mentioned above, and on 2002 and 2004. Besides this, data published is not comparable since data was published in different subgroups, either according to gender, or to different age groups, or even on different sub-region divisions.

The report on 2002 data was related to population-based studies on seventeen sub-regions of those six regions (Afr D: Bebin, Cameroon, Cape Verde, Equatorial Guinea, Gambia, Ghana, Mali, Mauritania, Niger, Nigeria, Sierra Leone, Sudan, Togo; Afr E: Central African Republic, Congo, Ethiopia, Kenya, South Africa, United Republic of Tanzania; Amr A: United States of America; Amr B: Barbados, Brazil, Paraguay; Amr D: Peru; Emr B: Lebanon, Oman,

Saudi Arabia, Tunisia; Emr D: Morocco; Eur A: Denmark, Finland, Iceland, Ireland, Italy, Netherlands, United Kingdom; Eur B1: Bulgaria, Turkey; Eur B2: Turkmenistan; Eur C: no population-based studies identified; South-East Asia Region: Sear B: Indonesia, Malaysia, Philippines, Thailand; Sear D: Bangladesh, India, Nepal, Pakistan; Wpr A: Australia; Wpr B1: China, Mongolia; Wpr B2: Cambodia, Myanmar, Viet Nam; Wpr B3: Tonga, Vanuatu). It used estimates of population size and structure based on the 2002 demographic assessment of the United Nations Population Division[4], as used by the World health report 2003[1].

For the 2002 estimates, they have considered 55 countries grouped into 17 regions according to the Global Burden of Disease 2000 Project[5,6], and used estimations of population size and structure from the 2002 demographic assessment of the United Nations Populations Division[4] as published on the World Health Report on 2003[1].

Prevalence of visual impairment using the ICD-10 definition of best corrected visual acuity and the 2002 world population, estimated that the number of people with visual impairment exceeded 161 million of people, from which 37 million were blind and 124 million had low vision, which represented a global prevalence of 2.59% on visual impairment (0.57% on blindness and 2.00% on low vision). Median ratio on low vision to blindness was 3.7, but ranged between 2.4 to 5.8 by region considered, which shows big variability by regions, as observed in Figure 1.

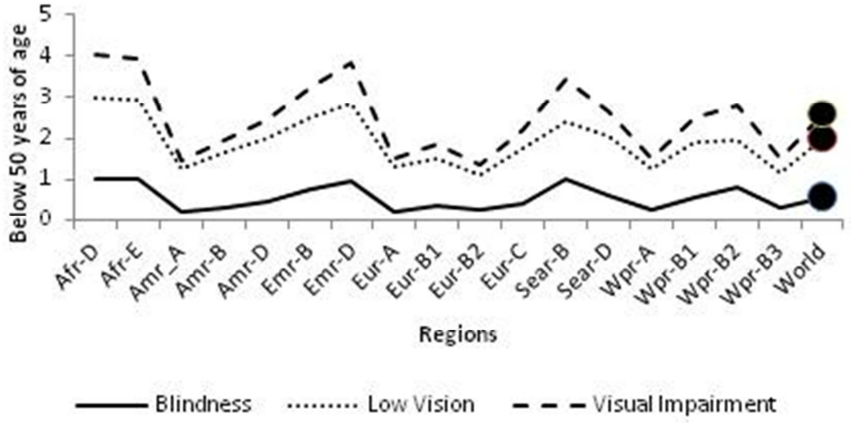

Figure 1 – Visual Impairment prevalence, per region (lines) and in the world (circles), in the year of 2002. Data obtained at WHO's public domain[1].

Childhood blindness (under 15 years of age) is a significant problem, with a global prevalence of 0.07%, but prevalence of blindness more than duplicate on ages between 15 and 49 years (0.16%), and increases significantly in the elderly population (1.68%). Variability between WHO regions, by age group, may be observed in Figure 2.

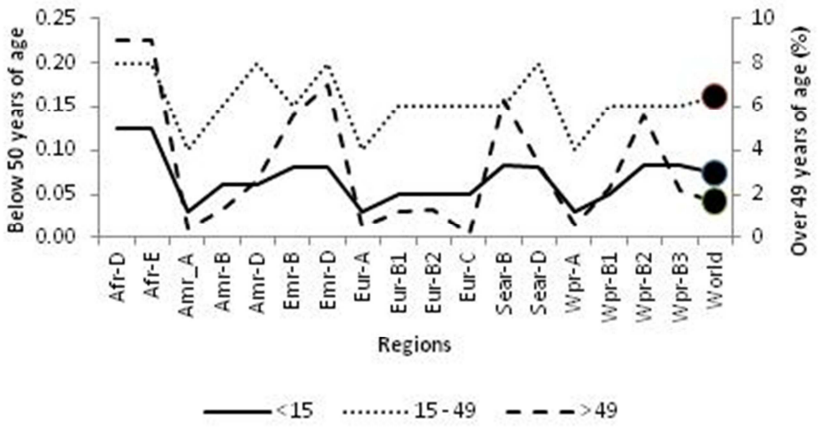

Figure 2 – Blindness prevalence, per region (lines) and in the world (circles), in the year of 2002, according to age group. Data obtained at WHO's public domain[1].

It is also known that the prevalence of visual impairment in females, adjusted for age, is 1.50 to 2.20 times higher than in males. Between the years of 1990 and 2002, the increase number of blind people was estimated in 8.57%, although there was an increase on world population of 18.50%. However, the increase of low vision people between those twelve years was 80%.

The 2010 report used 53 surveys from 39 countries and estimated a number of 285 million of visually impaired people in the world. Methods were not the same but world estimates were precise so they could be compared. It represents an increase of 77% on visual impairment, although the distribution of these increases is much heterogeneous. The actual prevalence of visual impairment is 4.24% (against 2.59% in 2002), with the maintenance of the blindness prevalence since 2002, but an increase from 2.00% in 2002 to 3.65% in 2010 in low vision, which represents an increase of almost 98% in the number of visually impaired people in the world, in spite of an 8.43% of population growth. Higher differences appear in the South-Eastern and Western Pacific regions, mainly due to India and China, where prevalence of visual impairment was, in 2010, of 21.9% and 26.5%, respectively, whereas the prevalence in those regions excluding these countries was of 9.8% and 5.2%, respectively[7]. In the world, the percentage of increase in low vision was almost 100% (Figure 3), with a population growth of 6.80%.

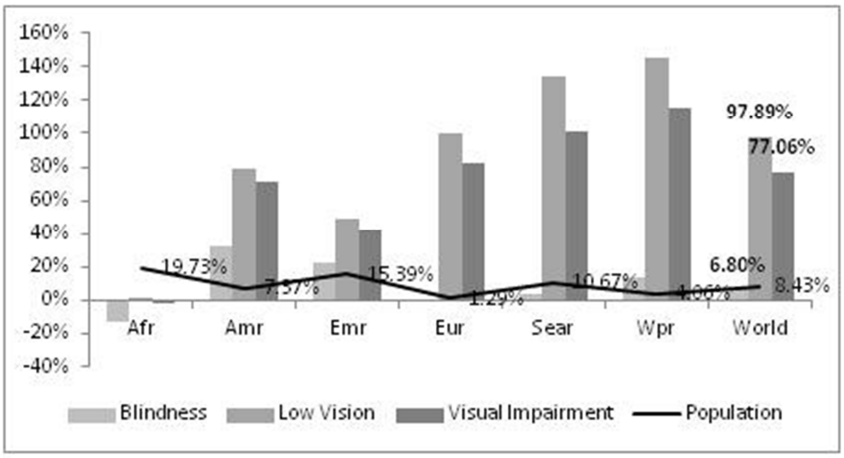

Figure 3 – Percentage of increase in the number of blindness, low vision, and visual impaired people between 2002 and 2010 per region (bars). Population growth is represented by a line. Data obtained at WHO's public domain[7].

The latest update on these statistics refer that about 90% of the world's visually impaired live in developing countries, and that 82% of blind people are aged 50 or more.

3. Main causes of Visual Impairment

According to data published by the World Health Organization relative to the year of 2010[7], the principal causes of visual impairment around the world are uncorrected refractive errors (43%) and cataracts (33%). Other important causes are glaucoma (2%), age related macular degeneration (AMD), diabetic retinopathy, trachoma and corneal opacities, all about 1%, while a large proportion (18%) of causes remain undetermined.

Despite advanced in surgical procedures, cataract remains as the leading cause of blindness in visual impaired populations (51%), followed by glaucoma (8%), an eye disease known for centuries, but which remains on the public health agenda due to difficulties in its early diagnosis and frequent necessity of life-long treatment. Age-related macular degeneration (AMD) ranks third among the global causes of visual impairment with a blindness prevalence of 5%; in fact, it is the primary cause of visual deficiency in industrialized countries. An emerging important cause of visual impairment is uncorrected refractive errors, related to avoidable vision impairment. The increase of diabetes among many population groups has caused diabetic retinopathy to be added to the priority list.

Cataract, onchocerciasis, and trachoma are the principal diseases for which world strategies and programmes have been developed. For glaucoma, diabetic retinopathy, uncorrected refractive errors, and childhood blindness (except for xerophthalmia), the

development of screening and management strategies for use at the primary care level is ongoing at WHO.

The largest proportion of blindness is necessarily related to ageing. Although cataract is not a major cause of blindness in developed countries, globally it is still the leading cause. Cataract is even more significant as a cause of low vision; it is the leading cause of low vision in all sub-regions of WHO.

Estimates published 1995[8], relative to the year of 1990, identified the main causes of blindness and low vision identified as cataract, trachoma, glaucoma, onchocerciasis, and xerophthalmia. However, there was insufficient data on blindness from causes such as diabetic retinopathy and age-related macular degeneration as there were no specific estimative of the global prevalence of diabetes and older populations.

In 2002[9], uncorrected refractive errors such as myopia, hyperopic or astigmatism were identified as the main causes of visual impairment (43%), so they begun to be considered on prevalence and population based studies; cataracts were once more identified, in 2002, as they were in 1995, as the leading cause of blindness (33%) followed by glaucoma (2%). In fact, among blind people, cataract represented (in 2002) almost half of the causes of blindness (47.8%), followed by glaucoma (12.3%), age-related macular degeneration (8.7%) and Corneal opacities and diabetic retinopathy (approximately 5% each).

Estimates from 2004[2] refer once more cataract and uncorrected refractive errors as the main causes of blindness among all causes of blindness. Note that uncorrected refractive errors are an avoidable cause of vision impairment in most of the cases. Oschocerciasis was still considered, at that date.

According to the 2010 update on these statistics, the number of people visually impaired from infectious diseases has greatly reduced in the last 20 years, suggesting progressively reduced incidence. An example is onchocerciasis (Figure 4), which remains endemic in some African regions and few isolated regions of South-America, but was no longer considered in 2010 as a major cause of blindness.

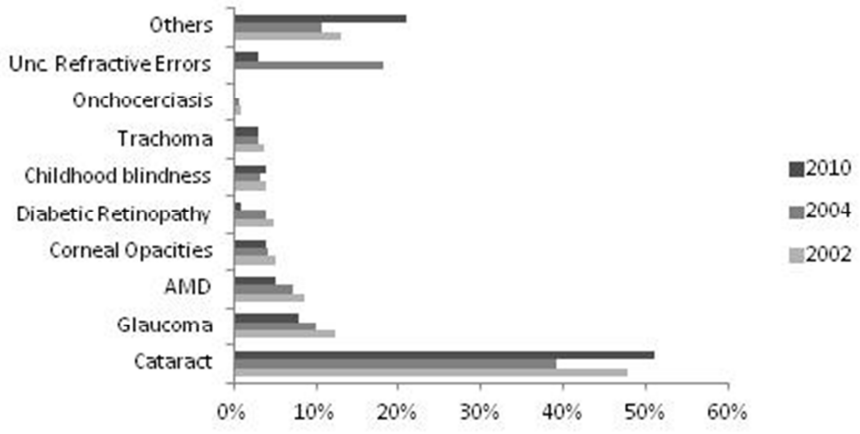

Figure 4 – Main causes of visual impairment in the years of 2002, 2004 and 2010. Data obtained at WHO's public domain[7].

However, posterior segment (retinal) diseases are a major cause of visual impairment worldwide, and are likely to become more important due to the rapid growth of the aging population, and the proportion of the total visual impairment and blindness from age-related macular degeneration, glaucoma and diabetic retinopathy is currently greater than from infective causes such as trachoma and corneal opacities.

On Figure 5, we may observe the world's distribution of visual impairment, in percentage, on the year of 2010.

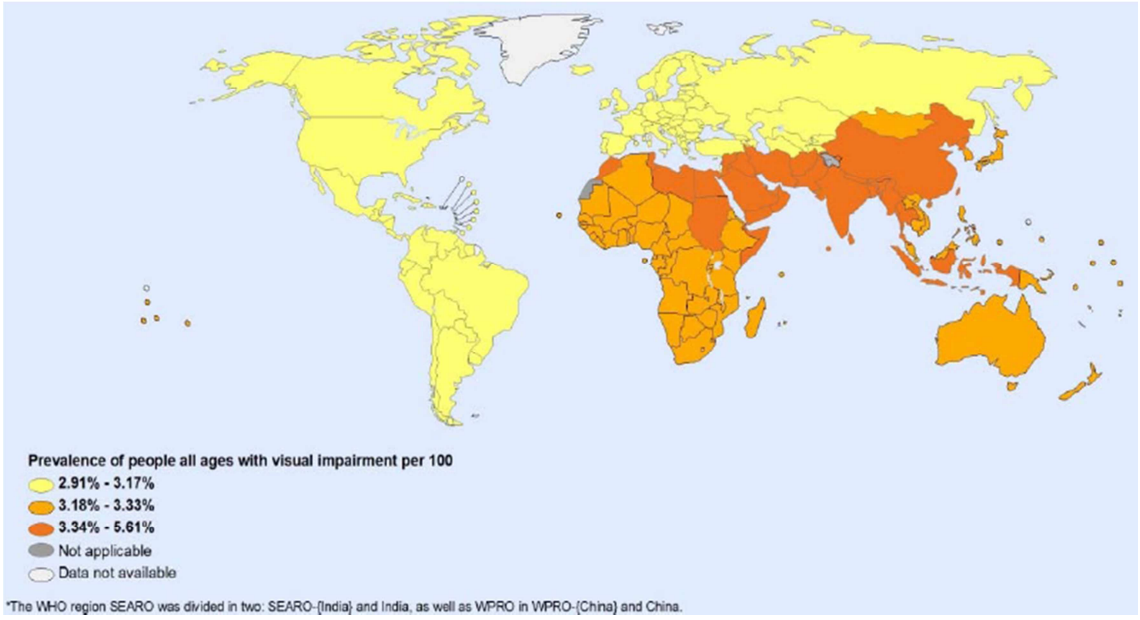

Figure 5 – Prevalence of visual impairment around the world. Figure obtained at WHO's public domain[7].

4. Risk factors for visual impairment

Approximately 65% of visually impaired people are aged 50 or older and about 20% of world population are in these age group. On the other hand, in most of the countries, especially on developed and developing countries, the elder population is increasing so, the number of people at risk of aged-related visual impairment, around the world, is also increasing.

Overall, visual impairment worldwide decreased since 1990, as a result of a reduction in visual impairment due to infectious diseases, despite the aging of populations. Globally, 80% of all visual impairment can be prevented or cured. In 2013, the World Health Assembly approved the "2014-19 Action Plan" for the universal access to eye health, with the aim of achieving a measurable reduction of 25% of avoidable visual impairment by 2019. One of the aims is to eliminate trachoma from the world by the year of 2020 and as a response to the increasing burden of chronic eye disease; the World Health Organization has coordinated the development of research projects and policies for diabetic retinopathy, glaucoma, age-related macular degeneration and refractive errors. These last ones are responsible for about 12 millions of visual impaired children below age 15, on the overall estimated 19 million of visually impaired children below 15, a condition that could easily be diagnosed and corrected.

SECTION B

SYSTEMIC DISEASES WITH OCULAR MANIFESTATIONS

1. Chronic Systemic diseases

A systemic disease is a disease that affects a number of organs and tissues, or affects the body as a whole. Although most medical conditions will eventually involve multiple organs in advanced stage, such as multiple organ dysfunction syndrome, diseases where multiple organ involvement appears in early stages are considered to be systemic diseases.

The eye and nails are considered to play an important role in the diagnostic of systemic diseases. The eye is composed by many different types of tissues and its unique feature makes the eye susceptible to a wide variety of diseases and provides insights into many body systems. Every part of the eye gives important clues for diagnosing systemic diseases, whose signs may be evident on the outer surface of the eye such as eyelids, conjunctiva and cornea, on the middle of the eye, or at the back on the retina.

Fingernails and toenails may also indicate some systemic diseases, since they can cause disruption in the nail growth process. For instance, pitting looks like depressions in the hard part of the nail, and it is usually associated to psoriasis, affecting 10 to 50 per cent of patients with that disorder[10]; it can also be caused by some systemic diseases, including Reiter's syndrome and other connective tissue disorders such as sarcoidosis, pemphigus, alopecia areata and incontinentia pigmenti[11].

A chronic disease is a health condition or disease that is persistent or long-lasting in its effects, and usually lasts for more than three months. The most common chronic diseases generally known by individuals are arthritis, asthma, diabetes, chronic obstructive pulmonary disease (COPD), cancer or acquired immunodeficiency syndrome (AIDS).

Chronic diseases are a major cause of mortality. The World health Organization (WHO) reports that chronic non-communicable diseases are, by far, the leading cause of mortality in the world. In fact, in 2005 chronic non-communicable diseases represented 35 million deaths and exceeded 60% of all deaths.

2. The case of diabetes

There are many diseases known to cause ocular or visual changes. An eye condition that results, directly or indirectly, from a disease process in another part of the body is said to be an ocular manifestation of that disease. It is known that Diabetes is the leading cause of new cases of blindness around the world in persons aged between twenty and seventy four years old, being its ocular manifestations diabetic retinopathy and macular oedema. These ocular manifestations affects up to 80% of those who have had the disease for 15 years or more. Moreover, diabetes is related to nephropathy, cardiac diseases and other systems which are also known to be related to the eye. In fact, diabetes mellitus is among the leading causes of death, disability and economic loss throughout the world.

Diabetes mellitus is a group of metabolic diseases caused either because the pancreas does not produce enough insulin or because cells do not respond to the insulin that is produced (insulin resistance), causing the person to have high levels of blood sugar or glycaemia. In fact, it is considered to be a metabolic disorder of multiple aetiologies characterised by chronic hyperglycaemia with disturbances of carbohydrate, protein and fat metabolism resulting from defects in insulin secretion and/or insulin action.

There are three main types of diabetes mellitus:

- Type 1, which results from the failure insulin production, so it is needed to be daily injected. Therefore, many times it is called insulin-dependent diabetes (IDDM) and usually appears in early or juvenile ages. This type of diabetes involves β-cell destruction.
- Type 2, which results from insulin resistance and/or insulin secretion, where there is a failure on the insulin usage from the cells, sometimes combined with an insulin deficiency. Usually, it is called non-insulin dependent diabetes (NIDDM) or adult-onset diabetes, since it is a disease that is usually diagnosed after thirty years-old.
- Gestational diabetes occurs in pregnant women without a previous diagnose of diabetes, and may precede the development of type 2 diabetes

Other forms of diabetes are less prevalent, and may include congenital diabetes, genetic defects of insulin secretion or steroid diabetes induced by high doses of glucocorticoids, among other causes.

2.1 The Insulin mechanism and consequences of its failure

All forms of diabetes have been treatable since insulin became available..

Human insulin protein is composed by 51 amino acids, has a molecular weight of 5808 Dalton, and is an A-chain and B-chains linked together by disulfide bonds.

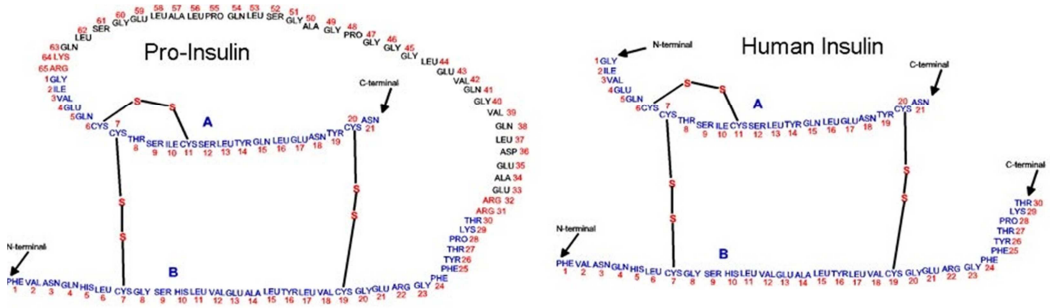

Figure 6 – Pro-Insulin and Human Insulin molecule. Figure obtained at a public domain.

Insulin is a peptide hormone produced by beta cells of the pancreas, and it is released by the same cells whenever glucose levels arise. It is necessary to regulate the carbohydrate and fat metabolism in the body as it induces cells in the liver, skeletal muscles and fat tissue to absorb and consume glucose from the blood, the main source of cell energy, in a process known as glycolysis, where glycogen is synthesised and converted into fatty acids. On the other hand, it inhibits gluconeogenesis and glycogen degradation. Insulin causes cells in the liver, skeletal muscles and fat tissue to absorb glucose from the blood, which is stored by the liver as glycogen, and by the adipocytes (fat cells) as triglycerides. Insulin stops the use of fat as a source of energy as it inhibits the release of glucagon, and it is provided in a constant proportion to remove the excess of glucose from the blood, which would be toxic to the organism. If blood glucose falls below a critical level, body begins to use stored sugar as an energy source through glycogenolysis, which breaks down the glycogen stored in liver and muscles into glucose to be used as an energy source. As it is also a central metabolic control mechanism, its status is also used as a signal control to other body systems, causing anabolic effects throughout the body.

If control of insulin fails, the result is diabetes mellitus, and insulin must be medically taken[12].

Over time, diabetes can damage the heart, blood vessels, eyes, kidneys, and nerves, as diabetes mellitus increases the risk of heart disease and stroke. Half of people with diabetes die of cardiovascular disease. To join this, whenever diabetes is combined with reduced blood flow, neuropathy in the feet increases the chance of foot ulcers, infection and eventual need for limb amputation.

On the other hand, diabetic retinopathy is an important cause of blindness, and occurs as a result of long-term accumulated damage to the small blood vessels in the retina. Diabetes is also among the leading causes of kidney failure[13].

The overall risk of dying among people with diabetes is, at least, the double than the risk of their peers without diabetes[14].

Simple lifestyle measures have been shown to be effective in preventing or delaying the onset of type 2 diabetes. To help prevent type 2 diabetes and its complications, people should achieve and maintain healthy body weight, while being physically active, eat a healthy diet and simultaneously reduce sugar and saturated fats intake, and avoid tobacco use as smoking increases the risk of cardiovascular diseases[13,14].

Untreated diabetes can cause many serious long-term complications, as mentioned before, such as cardiovascular disease, nerve damage, chronic renal failure known as nephropathy, and diabetic retinopathy. Adequate treatment of diabetes is thus important, as well as the control of blood pressure and some lifestyle factors such as tobacco use, regular exercise and food habits. Diabetes mellitus is one of the most common chronic diseases in nearly all countries, and continues to increase in number and significance as changing in lifestyles is leading to the reduction of physical activity and to the increase of obesity[14].

With the increase of diabetes on population, diabetic retinopathy was added to the priority list of avoidable visual impairment[7].

2.2 Estimates on diabetes

The incidence of diabetes is increasing, mainly due to population growth, aging, urbanization and increasing prevalence of obesity and physical inactivity.

According to the world health organization, diabetes prevalence by age and sex extrapolated to all 191 World Health Organization member states, and applied to United Nations' population estimates for 2000 for all age-groups worldwide, was estimated to be 2.8% in 2000 and previsions for 2030 are of 4.4%. The total number of people with diabetes is projected to rise from 171 million existed in 2000 to 366 million in 2030. The

prevalence of diabetes is higher in men than women, but there are more women with diabetes than men. The urban population in developing countries is projected to double between 2000 and 2030. The most important demographic change to diabetes prevalence across the world appears to be the increase in the proportion of people 65 years of age.
These findings indicate that the "diabetes epidemic" will continue even if levels of obesity remain constant. Given the increasing prevalence of obesity, it is likely that these values provide an underestimate of future diabetes prevalence.

Based on the Prevalence Study on Diabetes in Portugal[15], 11.7% of the Portuguese population between 20 and 79 years of age had diabetes mellitus, in 2009. Besides this fact, one in each three Portuguese (34.9%) is already diabetic or is in risk of becoming diabetic and half of diabetics do not know that they are. This study found that gender as a risk factor since prevalence is expected to be between 12.5 and 15.5% in man and between 8.5 and 10.6% in women, with 95% confidence interval. Age is also correlated to diabetes, in Portuguese people, since prevalence increases as people grow older. In addition to that, pre-diabetes prevalence, which is related to impaired fasting glucose and/or impaired glucose tolerance, is estimated to be of 23.3% in the Portuguese Population.
More recent data, from the report published in 2013, relative to 2012, by the National Observatory for Diabetes in Portugal, the prevalence of diabetes in Portugal within the ages of 20 to 79 is 12.9%. The impact of the ageing of the Portuguese population has increased diabetes prevalence in 1.2% since 2009. Notice that 44% of the population with diabetes did not yet been diagnosed. According to that report, prevalence between 40 and 59 years old is of 12.7%, and between 60 and 79 years old more than duplicates (27,0%), although 10,3% of those 27,0% have not yet been diagnosed.
It is also known that prevalence is higher in man (15.4%) than in women (10.6%), and this latest report estimates diabetes prevalence for men in 17.6% or 30.3% if they are in the age group of 40-59 or of 60-79 years, respectively, being the data in women of 8.2% and 24.3%, respectively.
Prevalence data was reported also according to Body mass index, which is known to be related to diabetes, as 5.8% if BMI is below 25 kg/m^2, as 12,7% if BMI is between 25 kg/m^2 inclusive and 30 kg/m^2 exclusively, and as 20.3% if BMI is, at least, 30 kg/m^2.
On the other hand, alterations on fasting blood glucose (AFBG) and low tolerance to glucose (LTG) reach 26.8% of the Portuguese population aged between 20 and 79 years old,

distributed as 10.1% of population with AFBG, 14.0% of population with LTG, and 2.6% with both, a condition that may lead to the development of type 2 diabetes so the global prevalence of diabetes and hyperglycemias is of 39.6%.

Every year, diabetes incidence is estimated to increase between 0.5 to 0.9%. Gestational diabetes prevalence, a condition that is related to type 2 diabetes development after the age of 50, has increased from 3.4% in 2005 and 2006 to 4.9% in 2011 and 4.8% in 2012, being this prevalence related to the age of the mother during pregnancy (13.5% if mother is aged 40 or more, 5.9% between 30 and 39 years, 2.9% between 20.29 years and 1.4% below 20).

The total number of diabetic medical consultations in the National Health Service grew from 1 877 259 in 2011 to 2 202 224 in 2012, and the mean number of diabetes medical consultations per patient remained the same between those years.

On the other hand, almost half of the diabetic patients have or will have diabetic retinopathy. The prevalence of diabetic retinopathy among all types of diabetes, in Portugal, is 44.1%, and in 25.1% of the diabetic subjects, it is classified as proliferative.

By all this reasons, and the chance in lifestyles that leads to less activity, the increase of obesity and the ageing of the population, there is some consensus that so diabetes my become epidemic.

3. Impact on the eye - Diabetic retinopathy

Diabetic retinopathy is a chronic progressive disease of the retinal microvasculature associated with the prolonged hyperglycaemia and other conditions linked to diabetes mellitus such as hypertension. This potentially sight-threatening disease is a micro vascular complication that may occur either on type 1 or type 2 diabetes. It is a disease of the retinal microvasculature associated with prolonged hyperglycaemia and other conditions linked to diabetes mellitus, such as hypertension, hypercholesterolemia or dyslipidaemia. In fact, it develops nearly in all persons with type 1 diabetes, and in more than 77% of people that survived to 20 or more years of type 2 diabetes[16].

According to the Portuguese Study Group on Vitreo and Retina[17], Diabetic Retinopathy is the leading cause of blindness in active age.

The threat to sight can be due to the growth of new vessels leading to intraocular haemorrhage and possible retinal detachment with profound global sight loss, and also to localised damage to the macula or fovea of the eye with loss of central visual acuity.

However, evidence-based treatment can reduce the risk for severe vision loss and blindness from proliferative diabetic retinopathy by more than 90%, as shown by five large multicentre clinical trials conducted in the United Kingdom and United States[18].

3.1 Major earlier studies on Diabetic Retinopathy

The Diabetic Retinopathy Study (DRS), conducted between 1971 and 1975, demonstrated that scatter pan-retinal laser photocoagulation reduces the risk for severe vision loss due to proliferative diabetic retinopathy by 60%[19,20], and provided the first and still most widely used classification system for grading the severity of diabetic retinopathy.

Later, between 1979 and 1990, the Early Treatment Diabetic Retinopathy Study (ETDRS) confirmed previously DRS results on scatter (pan-retinal) photocoagulation, demonstrating that it can reduce the risk for severe vision loss to less than 2% and that focal laser photocoagulation can reduce the risk for moderate vision loss due to diabetic macular oedema by 50%, without having adverse events on the progression of diabetic retinopathy or increasing risk for vitreous haemorrhage[21,22].

The Diabetic Retinopathy Victrectomy Study (DRVS, 1977 – 1987) provided useful information about the timing of vitrectomy surgery to restore useful vision in eyes with non-resolving vitreous haemorrhage, drawing attention to the poor prognosis of eyes which have experienced vitreous haemorrhage, a late complication of diabetic retinopathy[23,24].

During 1983 and 1993, the Diabetes Control and Complications Trial (DCCT) compared conventional blood glucose control with intensive blood glucose control in patients with type 1 diabetes mellitus without or with little diabetic retinopathy. For these patients, it was demonstrated that intensive blood glucose control as reflect of measurements in glycosylated haemoglobin reduced the risk for progression of diabetic retinopathy. Seven years after DCCT was finished, the Epidemiology of Diabetes Interventions and Complications Trial (EDICT) showed that diabetics in intensive control group continued to have a substantial lower risk for progression of diabetic retinopathy than the conventional control group, despite the convergence of glycosylated haemoglobin levels[25].
These two studies became notable for the following findings, which were not the primary outcome measures: They have shown that

- There is no threshold on glycosylated haemoglobin for diabetic retinopathy to occur, that is, there is not a cut point on glycosylated haemoglobin below of which there is no retinopathy; however, there is a linear relationship between achieved glycosylated haemoglobin level and the risk for visual complications of diabetes.
- People receiving intensive blood glucose control had a significant rate of hypoglycaemic reactions so such aggressive control is not benefice.

The United Kingdom Prospective Diabetes Study (UKPDS), performed during 1977 to 1999, had similar findings to the DCCT study, but it was performed on type 2 diabetes[26,27]. Furthermore, it highlighted the role of systemic hypertension as a potential risk factor for the development and progression of diabetic retinopathy, and has also demonstrated the negative effects of high cholesterol and serum lipid concentrations, increasing the risk of retinal complications in patients with diabetes mellitus.

3.2 Classification of Diabetic Retinopathy

The classification and severity grading of diabetic retinopathy have been based, since ever, on ophthalmoscopic visible signs of increasing severity, from no retinopathy through various stages of non-proliferative or pre-proliferative disease, to proliferative disease, but this grading scale may not reflect with accuracy the severity of the disease as maculopathy with severe visual loss may occur in the present of none or moderate ophthalmoscopic signs. Due to this, two different approaches have emerged, depending on the population to be applied: one designed to cover the full range of retinopathy based on the original Airlie House / ETDRS classification, used by most of the ophthalmologists, and another that intends to be applied on population screening.

Modified and simplified versions have been developed from the original Airlie House classification, modified by the DRS developed for the ETDRS, to use in the context of overall severity of ophthalmoscopic signs.

A simplified version was developed for the first version of the guidelines for diabetic retinopathy, in 1977[28]. Later, in 2003, the American Academy of Ophthalmology Guidelines Committee endorsed a reduced version of the ETDRS classification to be used in countries without systematic screening programmes. However, it was updated in order to have a clinical grading system that reflects the vision threatening risk of diabetic retinopathy, and considers five main stages of diabetic retinopathy. The first three stages consider low risk

non-proliferative retinopathy, a fourth stage of severe non-proliferative retinopathy and the fifth grade as proliferative retinopathy; additionally, macular oedema is determined as present or absent and is sub classified based on the involvement of the centre of the macula. There is a considerable overlap between classification systems. All of them recognize two mechanisms that conduce to the loss of vision: retinopathy (risk of new vessels) and maculopathy (risk of damage to the central fovea). Differences between classifications are mainly due to terminology or levels of retinopathy

The World Health Organization[18] suggests the use of the following rating scale for Diabetic Retinopathy (Table 1) and Macular Oedema (Table 2):

Table 1 - The International Clinical Diabetic Retinopathy Severity Scales (Adapted from WHO's public domain[18]).

Proposed disease severity level		**Findings observable (Dilated Ophthalmoscopy)**
No apparent retinopathy		No abnormalities
Nonproliferative Diabetic Retinopathy	Mild	Micro aneurysms only
	Moderate	More than the presence of micro aneurysms but less than severe non-proliferative diabetic retinopathy
	Severe	Any of the following: - More than 20 intra-retinal haemorrhage in each of the four quadrants - Define venous beading in two or more quadrants - Prominent intra-retinal microvascular abnormalities in one or more quadrants
Proliferative diabetic retinopathy		One or more of the following: - Neovascularization - Vitreous or pre-retinal haemorrhage

Table 2 - The International Clinical Macular Oedema Disease Severity Scales (Adapted from WHO's public domain[18]).

Proposed disease severity level		Findings observable (Dilated Ophthalmoscopy)
Apparently absent		No apparent retinal thickening or hard exudates in posterior pole
Apparently present		Some apparent retinal thickening or hard exudates in posterior pole
Diabetic macular oedema present	Mild	Some retinal thickening or hard exudates in posterior pole but distant from the centre of the macula
	Moderate	Retinal thickening or hard exudates approaching the centre of the macula without involving the centre
	Severe	Retinal thickening or hard exudates involving the centre of the macula

The Royal College of Ophthalmologists[29] proposed, in 2012, the following conversion table for classification of diabetic retinopathy on their Guidelines for Diabetic Retinopathy (Table 3).

Table 3 - Conversion table for classification of diabetic retinopathy (Adapted from the Royal College of Ophthalmologists[29]).

ETDRS	NSC	SDRGS	AAO	RCOpht
None (10)	None (R0)	None (R0)	No apparent retinopathy	None
Micro aneurisms only (20)	Background (R1)	Mild Background (R1)	Mild NPDR	Low Risk
Mild NPDR (35)			Moderate NPDR	
Moderate NPDR (43)	Pre-proliferative (R2)	Moderate BDR (R2)		High Risk
Moderately severe NPDR (47)				
A-D Severe NPDR (53)		Severe BDR (R3)		High Risk
E very severe NPDR (53)		Severe BDR (R3)	Severe NPDR	
Mild PDR (61)	Proliferative (R3)	PDR (R4)	PDR	PDR
Moderate PDR (65)				
High Risk PDR (71, 75)				
Advanced PDR (81, 85)				

ETDRS – early treatment Diabetic Retinopathy Study; AAO – American Academy of Ophthalmology; NSC – National Screening Commitee; SDRGS – Scottish Diabetic Retinopathy Grading Scheme; NPDR – Non.proliferative diabetic retinopathy; BDR – Background Diabetic retinopathy; PDR – Proliferative Diabetic Retinopathy; HRC – High Risk Characteristics

Before this, in 2009, the Portuguese Group on Vitreo and Retina Studies[17] had proposed the Guidelines for evaluating diabetic retinopathy, based on the orientations made by the International Council of Ophthalmology (ICO) and by the American Academy of Ophthalmology (AAO) and their "Preferred Practice Patterns". This group proposed an international classification for Diabetic Retinopathy and Macular Oedema based on the ETDRS classification (Table 3), in an effort to facilitate communication between the ophthalmic community and the internists and endocrinologists, based on the ocular fundus observation and on the retinography.

The presence of hard exudates are a sign of a recent macular Oedema or in phase of reabsorption, being the Diabetic Macular Oedema defined as a retinal thickening, which may be observed by the Optical Coherence Tomography (OCT) or Fluorescein Angiography. Although this last method is useful in several situations, it has become less applied since the emergence of the Optical Coherence Tomography as this offers a lower cost, greater safety and likelihood of obtaining useable information. The OCT is an effective mean of evaluation of the retina, either qualitatively or quantitatively, especially in the early detection of retinal thickening and in the follow-up of the macular Oedema[30].

3.3 Prevention of diabetic retinopathy

Despite clinical standards for evaluating and treating diabetic retinopathy cost-effectively, are clearly defined, some treatments that have shown to be effective, such as laser surgery, are still underused. On the other hand, it has been reported that, in patients with type 1 diabetes mellitus, about 26% of patients have never had their eyes examined, being the correspondent proportion of 36% in type 2 diabetes mellitus patients. The pattern for the patients is that they are older, less educated, have a more recent diagnosis than the ones that receive eye care regularly, live in rural areas and receive health care from a family or general practitioner. When these patients are examined, almost 61% are found to have diabetic retinopathy, cataract, glaucoma or another ocular manifestation of diabetes mellitus[18].

It is known that population prevalence of diabetes mellitus is over 6% in most of the high and middle income countries.

According to the Guidelines for Diabetic Retinopathy published by the Portuguese Study Group on Retina[17], it is generally accepted that there are some factors that should be controlled on diabetic, and that could help prevent diabetes complications, such as the metabolic control concerning the reduction as much as possible of the glycosylate haemoglobin to values below 7% and/or fasting blood glucose below 110 mg/dl. In addition, systolic and diastolic pressure should be controlled, with maximum values of 130 mmHg and 80 mmHg, respectively, as also cholesterol and triglyceride levels. Furthermore, obesity should be controlled with an adequate diet and daily physical activity is required, with the maintenance of the renal function.

CHAPTER 4

ROADMAP TO STATISTICAL CLASSIFICATION

SECTION A

THE PROBLEM OF WORKING WITH CORRELATED DATA

1. Introduction

In biomedical research, namely in clinical research, whenever we have a pair of identical organs, there is a challenging problem: data usually involves examining both organs, and measures are often correlated. Ophthalmic research is one of those cases, where we have a pair of eyes to analyse. So, it is extremely important to define who the subjects for analysis are, that is, we should define if we use all the data or if we look at individuals. It depends mostly on the question that is being asked, on the data collected and on the nature of the condition that is being studied. So, a question of major importance is the definition of the Unit of the Analysis.

2. Mathematical Issues on dependent or correlated data

Improper analysis of repeated measurements on the same person not taking in account for correlation between observations, or even the analysis of dependant measurements on the same person as if they were independent is a common error in medical studies. Correlated data arise when pairs or clusters of observations are related and thus are more similar to each other than to other observations in the dataset. There are two different types of dependency:

1) Observations may be related because they come from the same subject, either due to subjects that are measured at multiple time points (repeated measures), or when subjects contribute data on multiple body parts, such as both eyes, hands, arms, legs, or any pair of organs.
2) Observations from different subjects also may be related, such as in the case of the dataset containing siblings, twin pairs, husband-wife pairs, control subjects who have been matched to individual cases, or patients from the same physician practice, clinic, or hospital.

Other type of dependency are cluster randomized trials, which are performed to assign interventions to groups of people rather than to individual subjects (for example, schools, classrooms, cities, clinics, or communities), also are a source of correlated data because subjects within a cluster will likely have more similar outcomes than subjects in other clusters.

Many statistical tests assume that observations are independent and its application to correlated observations will lead to the overestimation of the p-values when we consider within-subject or within-cluster effects and underestimation of the type I error if we intend to analyse between subjects or between cluster effects.

We can state that there is a ***within-subject comparison*** when subjects are compared with themselves under multiple treatments, or at different time points; when they are compared with related subjects, such as twins, these are called ***within-pair*** or ***within cluster comparisons***. Certainly, there is an advantage on doing these, because we can reduce variability. However, analysis that ignores correlation between data will overestimate the variability and artificially increases the p-values, decreasing the chance of observing a significant effect and the correspondent power of the test.

When comparisons are made between unrelated subjects or clusters that have each received just one treatment, these are called ***between-subjects*** or ***between-cluster***

comparisons. In these situations, ignoring correlations in the data will lead to an underestimation of p-values because we cannot assume that results will be independent. For example, if a treatment works in a person's left eye, it is more likely to work in his or her right eye, so we cannot assume that we have two good outcomes from that treatment. If we assume that, we would be artificially increasing the sample size and, by doing that, decreasing the p-values which would lead to significant differences that, in fact, do not exist (type I error).

Correlated data need to be handled by special statistical techniques, which may be challenging to implement and interpret; one solution could be removing correlations by changing the unit of analysis, but it always leads to loss of information.

3. Revision of Literature

Concerning the problem to handle, if it is purely at the ocular level, both eyes should be used; on the other hand, if the problem that is being investigated is related to the individual, then the method of analysis depends on the nature of the condition that is being studied.

If we collect data from just one eye, then we have no problem since each eye represents one individual, but if we collected both eyes information, then we must look at the condition that is being studied before analysis. A major problem is when data has information from one eye only for some individual, and both eyes on others. In this case, it is generally safer to analyse data of only one eye per person.

So, the nature of the condition that is being studied is crucial to define if we use one or both eyes data. However, if sometimes it is obvious that both eyes data should be used, such as in cases of visual disabilities such as squint where the two eye information are needed in order to reflect the disability level of individuals.

Some conditions affect usually only one eye, such as choroidal melanoma (99% of cases) or corneal herpes, or severe ocular trauma (98% of cases)[31]; in these situations we should use the disabled eye information at the level of the individual. Eventually, the "good" eye could be used as a matched control for age and gender, or for other socio-demographic and clinical information. Other extreme conditions such as blepharitis almost always are bilateral (95% of cases)[31]. Due to this, whatever is found in one eye is found in the other, and correlation between eyes is very strong, almost perfect, so there is no use for both eye information and we should use only one eye.

In the middle situation lies the majority of cases, those where we find correlation between both eyes, but not perfect correlation. The simplest and safest statistical way to analyse data is to use only one eye information per subject, but it can lead to a waste of important data. For instance, if we have a sample of 100 individuals, therefore we have 200 eyes; if we use one eye per subject, then there is a waste of information, but if we use the 200 sample we have a bigger sample and then we can have a falsely degree of precision.

In sequence of this, there are some Statistical Guidelines for the analysis of ophthalmological data, published by Richard Armstrong[32].

There are a number of issues raised by the decision of collecting data from one or both eyes. When only one eye is chosen, it must be decided which one should be chosen. On the other hand, if both eyes are chosen, it must be defined how to analyse both eye information. Measurements are usually correlated[33,34], and typical statistical procedures such as t-tests, ANOVA, confidence intervals or other basic nonparametric methods assume that observations are independent when, in fact, they are not. As the variance between eyes (within subject) is usually less than the variance between subjects, the overall variance is likely to be and underestimate of the true variance and the risk of finding falsely significant differences (type I error) increases. It is essential to combine in some way data from both eyes in order to take its correlation into account[35,36].

Armstrong suggests that investigators should consider whether it is advantageous to collect data from both eyes or not, and if one eye is studied and both are eligible, it should be randomly chosen; otherwise, two-eye data can be analysed using eyes as a within subjects if they are not used as case-control.

Armstrong reviewed referenced articles in three optometry journals (OPO – Ophthalmic and Physiological Optics; OVS – Optometry and Vision Science; CEO – Clinical and Experimental Optometry) between 2009 and 2013, eliminating those that involved animal or laboratory studies.

Of the 230 articles remaining and reviewed, 148 (64%) used data from one eye only, and 82 (36%) used data from both eyes. Whenever one eye only was used (148 papers) for analysis, different approaches were used, such as choosing either the right eye or the left eye, or even randomly selected eye, dominant eye, eye with best visual acuity, and the worst or diseased eye, as presented on Figure 7.

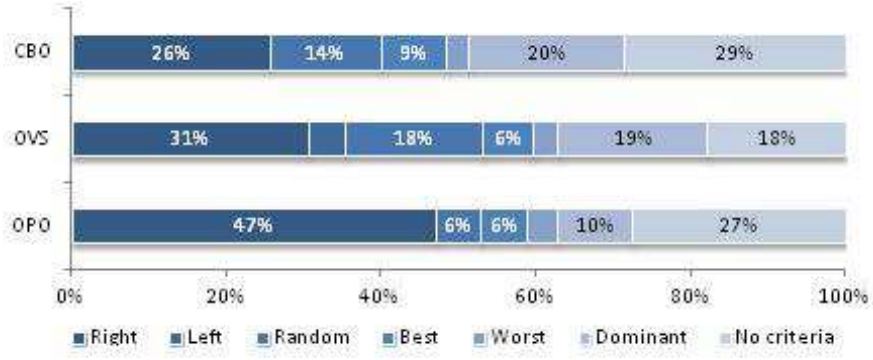

Figure 7 – Distribution of chosen eye to analyse within papers which used data from one eye only, per journal (Adapted from data available at reference 32).

Papers with two-eye information (82) analysed it in several different ways, also (Figure 8): some of them rejected data from the adjacent eye, other used both eyes separately, some used both eye taking into account correlation between eyes or, in other cases, using one eye as a control and the other as a disease or treated eye.

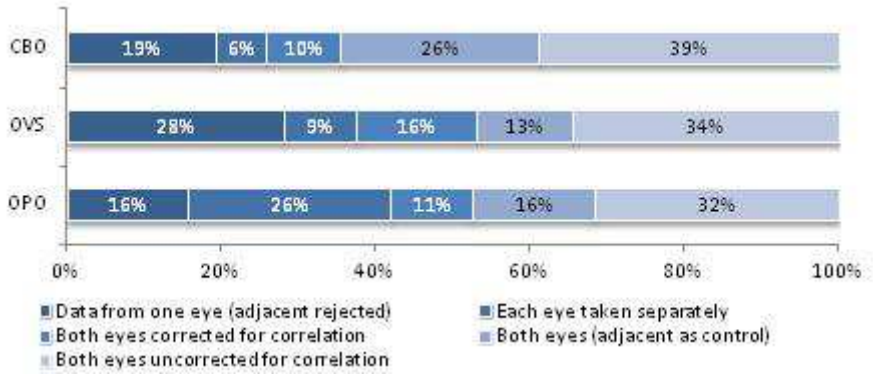

Figure 8 – Distribution of methodologies on number of eyes to analyse, per journal (Adapted from data available reference 32).

On that revision, Armstrong did not find any association between different methods of selecting the eye intended to study when methodology involved only one eye in articles published in those three journals (independence chi-square(12) = 14.48; p = 0.310), or between different methods of analysis of data from both eyes and those three journals (independence chi-square(8) = 7.44; p = 0.510). However, as the number of expected cells in those contingency tables is too high, it is useful to correct those statistics applying a Monte Carlo simulation. By doing this, we can perform the independence chi-square test in a random sample of 10000 cases, for instance, generated by that method. Conclusion is similar to the one that Armstrong[32] stated, but more reliable (Table 4).

Table 4 – Association between eye methods selection and journals.

Data collected	Chi-square (df)	p-value	Monte Carlo simulation	
			p-value	95% CI
One eye only	14.48 (12)	0.271	0.272	0.263 – 0.280
Both Eyes	7.44 (8)	0.490	0.510	0.501 – 0.520

This means that methodologies applied were identically distributed between journals, but apparently there is as predominance of the choice of the right eye when only one eye is chosen for analysis in all the journals (Figure 9.A), and the use of both eyes information uncorrected for correlation when data from both eyes is used in all the journals considered together (Figure 9.B). The left eye is rarely used (in fact, it is only considered at OVS). This shows great heterogeneity in the methodology chosen between studies.

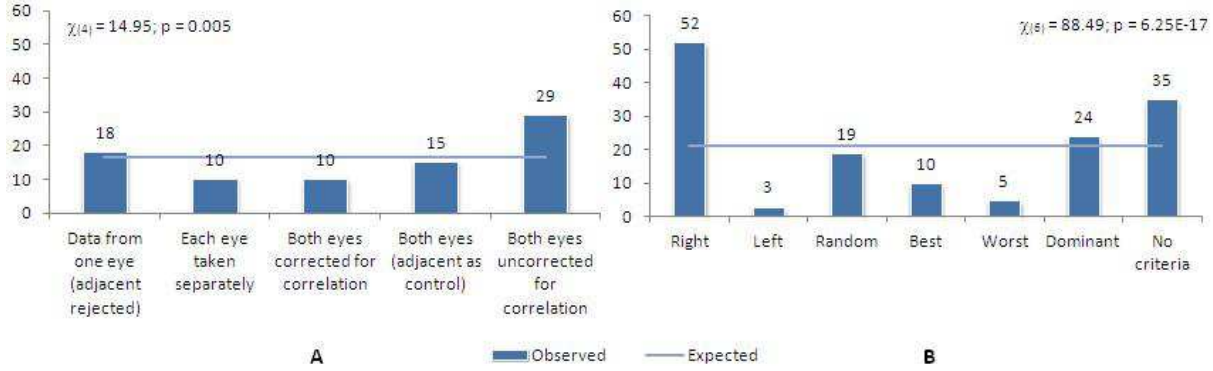

Figure 9 – Overall distribution of eye chosen for analysis, when only one eye was used (A) or when both eye information was used (B) (Adapted from data available reference 32).

Actually, an adjustment chi-square test performed within each journal in papers that considered just one eye shows that there is a prevalence of studies using the Right Eye at OVO (chi-square(6) = 45.35; p = 1.23E-08), of the Right, Random or Dominant Eye in OVS (chi-square(6) = 25.61; p = 2.63E-4), while in CEO there is a only a tendency for the use of the Right Eye or Dominant Eye, although all journals have larger observed papers relative to expected without a given criteria for the choice of the eye (Figure 10.A). When both eye data were collected, the prevalence of two eyes used separately or both used uncorrected for correlation is significantly higher than other methodologies. This pattern is present at OVS papers (Figure 10.B). Analysis of both eyes uncorrected for correlation appear to be the method most frequently used in these journal, but no differences were found in methodologies used in the two other journals.

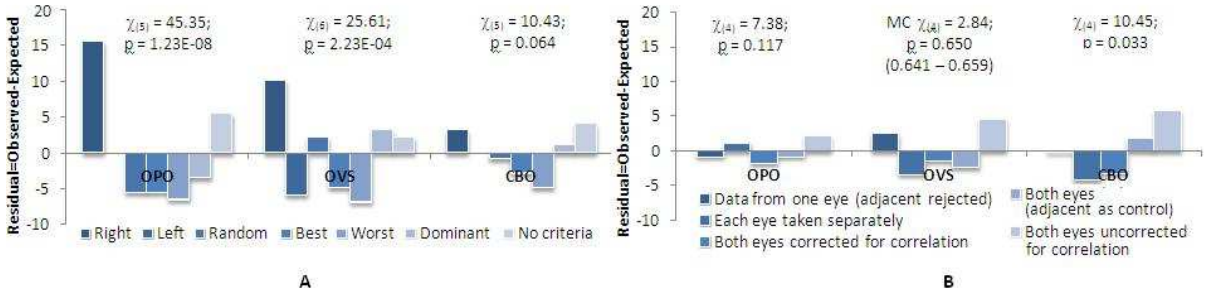

Figure 10 – Distribution of methodology for analysis, when only one eye information or both eye information were used. (Adapted from data available at reference 32). Values between brackets indicate 95% confidence interval for the p-values, if Monte Carlo simulation was applied

Armstrong[32] suggests the use of the following statistical methodology whenever data from both eyes is collected:

Table 5 – Statistical methodology to apply when both eye data is collected, suggested by Armstrong[32].

Objective	Procedure	References
Mean, SD of a sample of right and left eyes	ANOVA nested design (variance components determination)	Armstrong, Eperjesi, Gilmartin[37]
Comparing two groups (correlated data)	Modified Wilcoxon test	Rosner, Glynn, Lee[36] Rosner, Glynn, Lee[39]
Comparing a proportion of eyes with a feature (two samples)	Adjust variance of the difference proportions by calculating asymptotic normal distribution	Fleiss, Levin, Paik[40]
Measure correlation between eyes (no systematic differences between eyes)	Intra-class correlation coefficient	Bland, Altman[41]
Measure correlation between eyes (systematic differences between eyes)	Intra-class correlation coefficient	Rosner, Glynn, Lee[36]
Linear regression	Regression models	Glynn, Rosner[42] Glyyn, Rosner[43]
Level of agreement between eyes	Bland and Altman test of agreement	Bland, Altman[44] McAlinden[45]
Treated eye, other as control (two-way)	Paired t-test	Armstrong[46]
Treated eye, other as control (factorial design)	ANOVA split-plot	Armstrong, Eperjesi, Gilmartin[37]

He also recommends the use of the following flow chart, presented on Figure 11, to plan the statistical analysis:

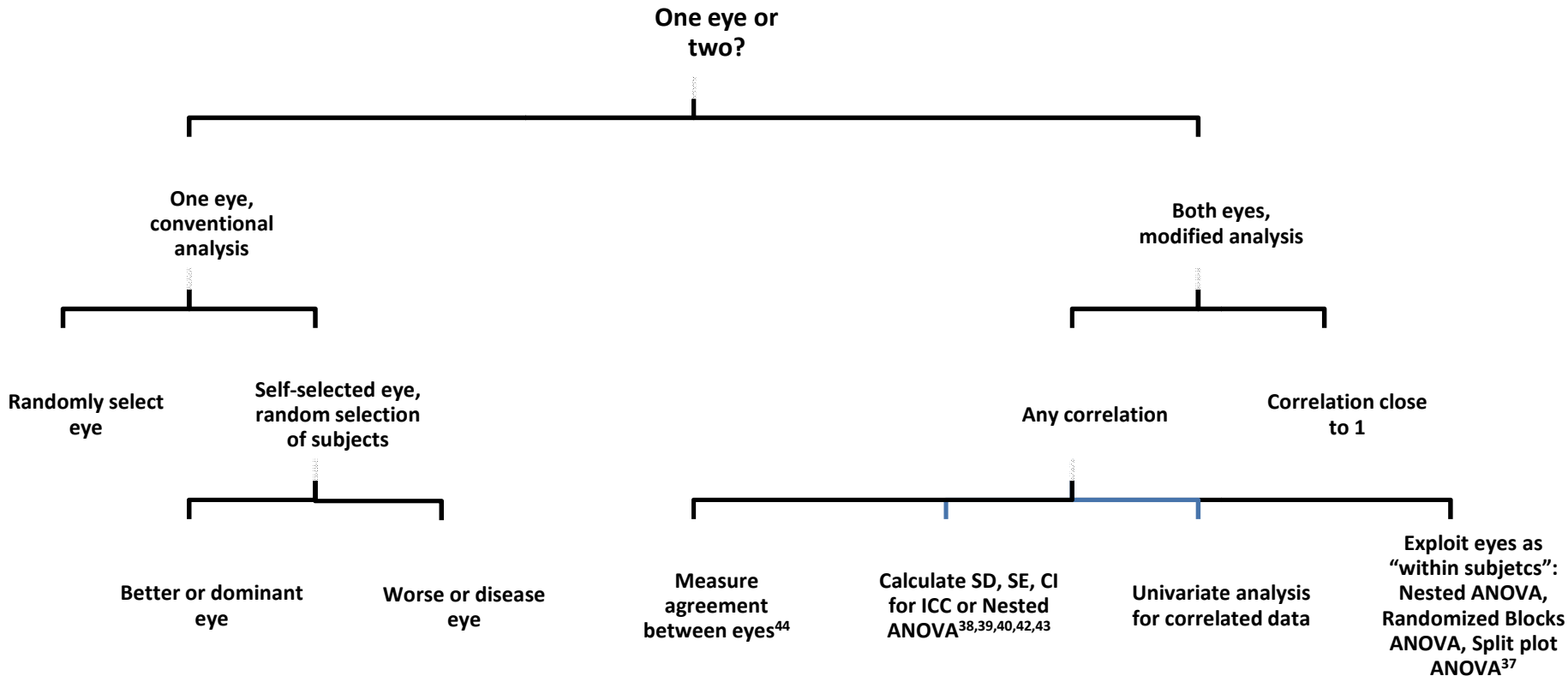

Figure 11 – Flow chart for planning statistical analysis suggested by Armstrong[32], and adapted.

SECTION B

STATISTICAL CLASSIFIERS

1. Introduction

As seen above, there is a concurrent problem in defining or classifying diabetic retinopathy and, moreover, in identifying risk factors for the development of a characteristic that, in no doubt, classifies the individual to a group or traces the profile of a determined group of individuals.

In statistics, classification is the problem of identifying to which groups a new individual belongs, given a set of characteristics. This usually involves having a training set of data containing observations whose group membership is known, given this explanatory variables or features. However, measurement levels of variables must be taken into account since all variables must be quantifiable. Any algorithm that implements classification or any mathematical function that maps input data into a category is known to be a classifier.

In the terminology of machine learning, the term classification is considered to be an instance of supervised learning (while in statistics it is related to unsupervised learning) which means that it is necessary to have available a set of correctly identified observations, while the corresponding unsupervised procedure is known as clustering or cluster analysis, which involves grouping data into categories based on some measure of inherent similarity or, in other cases, dissimilarities[47].

Terminology is quite varied across fields and a class of methods which are often used for classification is regression. Here, independent variables, or regressors, are the explanatory variables and the categories to be predicted are the outcomes, which represent the set of possible values for the dependent variable. Conversely, in machine learning, the observations are called instances while the explanatory variables are the features, grouped into a feature vector, and the possible categories to be predicted are called classes.

Classification and clustering are included in a more general problem related to pattern recognition or profile detection, which often includes probabilistic models. Algorithms of this nature apply statistical inference to find out the bets class for a given instance, which is chosen as the one with the highest probability membership. This kind of algorithms have some advantages on clustering and non-probabilistic algorithms since they offer a confidence interval for the probability of group membership and often they can abstain for classifying if that probability is too low, avoiding the problem of error propagation.

There is a considerable overlap among concepts related to classifiers, related to statistical classifiers, machine learning, data mining and Artificial Intelligence algorithms. In a certain way, these four fields intend to solve the same kinds of problems but with different approaches.

Artificial Intelligence is fairly distinct from the others since it is related to programming a computer or electronic device in order for it to behave as if it had intelligence. However, most experiences to induce intelligence require machine learning algorithms, since it is intended to induce new knowledge from previous learned experiences so a large area of artificial intelligence is machine learning.

This area involves the study of algorithms developed for the automatic extraction of information, without human involvement, with some ideas inspired or directly derived from classical statistics.

Data mining can be said to have born from machine learning, in some aspects and though, from statistics but it is carried out by a person and not by a computer, in a specific situation or data set, with a goal in mind. We can say that this person wants the influence of different pattern recognition techniques developed in machine learning and, quite often, the data set is massive, complex, and frequently with more variables than observations. Data mining problems can be unsupervised, that is, when we do not know the answer (discovery) or supervised when they are used as predictions since we know the answer. Data mining techniques often involves cluster analysis, classification and/or regression trees, or neural networks.

Classical statistics are often related to frequentist or Bayesian methods, and clearly there is an intersection between this topic and optimization theory in order to achieve fields above mentioned. The election prediction statistical procedure for classification is regression, but others are available.

2. Classification Methods

Classification methods were born with the work of Sir Ronald Fisher[49,49], and his work as a co-founder of the actual population genetics. He worked in the context of two group problems, leading to Fisher's Linear Discriminant Function, which is very similar to a multiple regression function, since the intention is to assign a group to a new observation when the dependent variable is qualitative and independent variables are quantitative or binary qualitative.

For the effect, discriminant functions are created as linear combinations of the initial variables in order to maximize the differences between group averages and, at the same time, minimize the probability of incorrect classification of the cases within groups. If we code the two groups for analysis as 0 and 1 and use them as the dependent variable, in general, groups can fit a linear equation[50] of type

$$F(x_1, x_1, \dots x_n) = b_0 + \sum_{i=1}^{n} b_i \times x_i$$

Where b_0 is constant and $b_i, i = \overline{1,n}$ are regression coefficients so interpretation is straight forward and closely follows the logic of multiple regressions, where variables with largest standardized regression coefficients (Equation 1.2) are the ones that most contribute to the prediction of group membership.

$$F^*(x_1, x_1, \dots x_n) = \sum_{i=1}^{n} \beta_i \times x_i$$

2.1 Discriminant Function Analysis

Discriminant function analysis is an extension of Fisher's Liner Discriminant Function as it can be applied to discriminate two or more than two groups. When there are more than two groups, then more than one discriminant function as the one presented above can be estimated. For instance, if there are three groups, two discriminant functions will be estimated: one that discriminated between group 1 and groups 2 and 3 combined, and another one which discriminates between groups 2 and 3. These functions are created in order to optimize the combination of variables that provide the best overall discrimination between groups, sequentially, that is, the first function provides the most overall discrimination between groups, and the second one provides the second most, and so on. Moreover, these functions will be independent and orthogonal so that their contributions to discrimination between groups do not overlap. This can be obtained after a canonical

correlation analysis is performed, so that successive functions and canonical roots (canonical eigenvalues that generate functions) are determined, in a maximum number of the number of group minus one (k-1), as long as the number of variables is at least k-1; otherwise, the number of discriminant functions determined equals the number of variables. This is done thought the Wilk's Lambda Test which represents a generalization of the F distribution, and is performed in a stepwise method that includes, in each step, the variable with the higher Wilk's Lambda, in spite of the importance of eigenvalues, eta-squared and canonical correlation. Moreover, it permits the identification of new cases or the choice of an alternative data set of variables with a similar dimension of the initial model which can discriminate groups. It is also possible to identify similar groups using similarities of centroids, that is, vectors of group means.

Discriminant analysis is often used as a confirmatory analysis for clustering or factorial analysis, a method of reducing variables, but it can also be used as a classification method as a predictive classification of subjects: once discriminant functions are derived, they can be tested in a new set of data to cross-validate the utility of discriminant functions, using classification functions, which cannot be confused with discriminant functions. Classification functions are in the same number as the number of groups, and are used to determine to which group each case most likely belongs, after computing a classification score (S_i) for a new observation[50], defined as:

$$S_i = c_i + \sum_{j=1}^{m} w_{ij} \times x_j \, , i = \overline{1,k}$$

Where k is the number of groups, m is the number of variables, c_i is constant for group i, w_{ij} is the weight for the j'th variable in the computation of the classification score for the i'th group and x_i is is the observed value in variable j for the new case.

By doing this, we can classify the new case to the group where it has the higher classification score, unless prior probability membership are quite disperse. If here are much more observations in one group than in others, *a priori* probability that the case belongs to that group is much higher. So, *a priori* probabilities should be adjusted to represent the proportion observed in the sample if only that represents the true distribution proportion in the population; otherwise, if proportions observed in the sample reflect only the random result of sampling, then *a* priori probabilities should be set to be equal, since *a* priori probabilities have a great impact on the accuracy of the prediction, that is, on *posterior* probabilities. These posterior probabilities are almost proportional to the Mahalanobis

distance, that is, the distance between the point and each group centroid, so classification is performed by choosing the smallest Mahalanobis distance.

The Mahalanobis distance[51] between a multivariate vector $x = (x_1, x_2, \dots, x_p)^T$ and its mean $= (\mu_1, \mu_2, \dots, \mu_p)^T$, with a covariance matrix S is given by

$$d(x) = \sqrt{(x-\mu)S^{-1}(x-\mu)^T}$$

Note that if the covariance matrix is the identity, than the Mahalanobis distance is the Euclidian distance[51]. This is highly improbable with real data, but if we have a diagonal matrix, the Mahalanobis distance is the usually called normalized Euclidian distance, and the distance from a given point in multidimensional space each group centroid is given by

$$d(x,\mu) = \sqrt{\sum_{i=1}^{p} \frac{(x_i-\mu_i)^2}{s_i^2}}$$

In the two group problem, this formula may be used to estimate the probability of a given point in space belongs to a data set, or a group, and classify it according to the higher posterior probability.

For each subject i, the distance between the score obtained in the discriminant function (f_i) and the centroid for each group j (d_j), considering the variance for scores obtained for the discriminant function in group j ($s_{f_j}^2$) defined[51] as

$$d_i = d(f_i, \overline{d_j}) = \sqrt{\frac{(f_i-\overline{d_j})^2}{s_{f_j}^2}}$$

Another way to classify subjects is to use these distances, and classify the subject according to its closeness to each group centroid, after determining the frontier line that separates groups, which is given by[50]

$$f = \frac{n_1\overline{d_1}+n_2\overline{d_2}}{n_1+n_2}$$

where n_1 and n_2 are the number of cases in each group and $\overline{d_1}$ and $\overline{d_2}$ are the centroids for each group.

Based on the distances to the frontier line, or to each group centroid, it is possible to determine posterior probabilities for a given subject belongs to a specified group.

The distribution of the squared distances, D^2, follows a chi-square distribution with one degree of freedom (since there is only one discriminant function) and the right probability of significance for the chi-square obtained is the conditional probability of obtaining that distance, given the group, $P(D|G_j)$, for each one of the groups. Applying the Bayes theorem, it is possible to determine posterior probabilities, given by[50]

$$P(G_j|D) = \frac{P(G_j)\times P(D|G_j)}{\sum_{j=1}^{2} P(G_j)\times P(D|G_j)}, j = 1,2$$

The subject is classified according to the group for he has higher probability of belonging.

However some assumptions should be evaluated, such as:

- Each group is a random sample of a multivariate normal population. The violation of this assumption can lead to incorrect decisions, especially in present of small samples. However, incorrect decisions are in terms of the type II error and rates of classifications, but not in the type I error, that is, violation of multivariate normal distribution will not increase type I error, but it can reduce the power of the test and lead to a higher rate of misclassifications, unless the lack of normality is only due to the lack of symmetry and not due to a non-mesokurtic distribution, according to Sharma[51]. In the presence of leptokurtic or platikurtic distributions, it is common to use logistic, ordinal or multinomial regression as an alternative to discriminant function analysis.
- Homogeneity of covariance matrices which means identical variance within groups, which can be tested using the Box M Tests. The violation of this procedure increases the number of cases classified in the group with higher dispersion and affects almost only the type I error, especially if groups have different dimension. In case of identical dimensions, Sharma[51] considers that the violation of this assumption is no problematic in terms of increase of type I error, even because classification rates are not influenced by this.

2.2 Regression procedures

Also based on multiple regression, some methods have become very popular in classification problems, such as logistic, probit, multinomial or ordinal regression. Perhaps logistic regression is the most popular of those, since it is the simplest to interpret but other may be useful.

Logistic regression is a type of probabilistic statistical classification model used to predict a binary response from a binary predictor or set of predictors; however, it is possible to use ordinal or scale predictors.

Logistic regression is an extension of linear regression and measures the relationship between a binary categorical dependent variable $Y \in \{0; 1\}$ and one or more independent variables ($X_i, i = \overline{1,k}$) by using probability scores as the predictive value on the dependent variable.

In multiple linear regression, the estimated model is obtained by[52]

$$\hat{y} = b_0 + \sum_{i=1}^{k} b_i \times x_i$$

Since $\hat{y}$ can assume only to possible values, let's say, 0 and 1, it would be unrealistic to use that condition. On the other hand, theoretically, the right-hand side of the previous equation can take any value between minus infinity ($-\infty$) and plus infinity ($+\infty$) unless we restrict the values of the regression coefficients ($b_i, i = \overline{0,k}$) and $\hat{y}$ is assumed to be the expected value of a normal distribution. Well, we are dealing with binary variables so that model does not apply to these variables.

Thus, it is more reasonable to consider a regression model which involves the probability of Y being 1 (probability that the event occurs) instead of using Y by itself since $p = P(Y = 1|x) \in [0; 1]$. However, it is still too narrow since $b_0 + \sum_{i=1}^{k} b_i \times x_i \in (-\infty; +\infty)$ but $\frac{p}{1-p} \in (0; +\infty)$ and, if we take the logarithm of this expression, we have a real number $ln\left(\frac{p}{1-p}\right)$ belonging to the interval $(-\infty; +\infty)$. So, it is possible to write the following condition[52]:

$$ln\left(\frac{P(Y=1|x)}{1-P(Y=1|x)}\right) = b_0 + \sum_{i=1}^{k} b_i \times x_i$$

Which is equivalent to

$$\frac{P(Y=1|x)}{1-P(Y=1|x)} = e^{b_0+\sum_{i=1}^{k} b_i \times x_i}$$

And to

$$P(Y = 1|x) = \frac{e^{b_0+\sum_{i=1}^{k} b_i \times x_i}}{1+e^{b_0+\sum_{i=1}^{k} b_i \times x_i}}$$

Or

$$P(Y = 1|x) = \frac{e^{b_0} \times \prod_{i=1}^{k} (e^{b_i})^{x_i}}{1 + e^{b_0} \times \prod_{i=1}^{k} (e^{b_i})^{x_i}}$$

Which reveals that the model links the linear expression $b_0 + \sum_{i=1}^{k} b_i \times x_i$ to the probability $P(Y = 1|x)$, where b_i denote regression coefficients and if a $b_i = 0$ then the correspondent X_i as no association with the dependent variable Y; each one of the e^{b_i} can be interpreted as an odds ratio.

Multinomial logistic regression is a classification method that generalizes logistic regression to a multiclass problem[53], with more than two possible discrete nominal outcomes, used to predict the probabilities of a nominal distributed dependent variable, given a set of independent variables.

Ordinal regression is a classification method applied to predict ordinal dependent variables. The two most common types of ordinal regression models are ordered logit and ordered probit.

In ordered logit, the model applies to data that meets the proportional odds assumption, that is: suppose that the dependent variable Y is ordinal, and has m ordered categories ($C_i, i = \overline{1,m}$) so that the proportion of a statistical population who would answer C_i is denoted by $p_i, i = \overline{1,m}$.

The logarithms of the odds[53] (not the probabilities) of answering in a certain way are given, to the fist m-1 categories, by:

$$\begin{cases} C_1 \ coded \ as \ 0: ln \frac{p_1}{\sum_{i=2}^{m} p_i} \\ C_2 \ coded \ as \ 1: ln \frac{p_1 + p_2}{\sum_{i=3}^{m} p_i} \\ \dots \\ C_k \ coded \ as \ k-1: ln \frac{\sum_{i=1}^{k} p_i}{\sum_{i=k-1}^{m} p_i} \\ \dots \\ C_{m-2} \ coded \ as \ m-3: ln \frac{\sum_{i=1}^{m-2} p_i}{p_{m-1} + p_m} \\ C_{m-1} \ coded \ as \ m-2: ln \frac{\sum_{i=1}^{m-1} p_i}{p_m} \end{cases}$$

The proportional odds assumption is that the number added to each one of these logarithms to get the next is the same in every case so that we obtain an arithmetic sequence and

The probit function (probability unit function) is the quantile function associated with the standard normal distribution, and it has applications in exploratory statistical graphics and

also in specialized regression models, especially with binary response or ordinal response variable, leading to **probit regression models** and **ordered probit regression**, respectively. Generally, the probit function is the inverse of $\Phi(z)$, the cumulative distribution function of the standardized normal distribution Z, and can be expressed as $probit\big(\Phi(z)\big) = z \; and \; \Phi\big(probit(p)\big) = p$

2.3 Bayesian Classifiers

A Bayes classifier is a simple probabilistic classifier based on the Bayes' Theorem and due to that often called Naïve Bayes Classifier, as it uses strong independence assumptions[54].

Bayes' Theorem shows a very simple relation between a conditional probability and its inverse, and is itself a corollary of the Total probability Theorem, and shows how we can change prior probabilities having in account new evidences into posterior probabilities.

Total Probability Theorem states that if we have n independent events $E_i, i = \overline{1,n}$ in the total sample space, Ω, that satisfy simultaneously the following conditions

$$\begin{cases} \bigcup_{i=1}^{n} E_i = \Omega \\ P(E_i) > 0, \forall i = \overline{1,n} \\ E_i \cap E_j = \emptyset, \forall i,j = \overline{1,n} : i \neq j \end{cases}$$

Then

$$P(B) = \sum_{i=1}^{n} P(B \cap E_i) = \sum_{i=1}^{n} P(E_i) \times P(B|E_i)$$

So, from this, and with the same assumptions, the following corollary can be stated, and it is known as the Bayes Theorem:

$$P(E_i|B) = \frac{P(B|E_i) \times P(E_i)}{\sum_{i=1}^{n} P(B|E_i) \times P(E_i)}$$

The Naïve Bayes Classifier assumes that the presence or absence of a feature of a group or of a class is unrelated to another feature, either it is present or absent. For instance, a naïve Bayes classifier may predict the gender of a given subject just because the height is above 1,80 meters, the weight is above 75 kg and the hair is short, considering that this features contribute independently, regardless the presence or absence of other features. In spite of this, it can be trained very efficiently in a supervised learning setting, and they have a good behaviour in many complex real-world situations.

An advantage of this classifier is that it needs a small amount of training data to estimate parameters necessary for classification and, as it assumes independence of variables, it is only necessary to estimate variances and not the entire covariance matrix.

Nowadays, naïve Bayesian classifiers are outperformed by other approaches, such as boosted trees or random forests, methods that are included in decision trees.

2.4 Decision trees

Decision trees are nowadays widely used either as prediction or simply exploratory tools. A decision tree is a predictive model which maps observations about an item into conclusions about its outcome value, and it is used either in statistics, machine learning and data mining and often decision trees are called classification trees or regression trees, depending on the problem that is being solved, although is common the use of the acronym CART standing for Classification And Regression Tree:

- Classification tree analysis is used whenever the predicted outcome is qualitative (the class to which the data belongs)
- Regression tree analysis is used whenever the predictive outcome is quantitative

The tree structure is composed by leaves that represent class labels and branches represent conjunction of features that lead to those class labels, starting from a root with no incoming edges. This kind of models are useful since they put conclusions into a visually space that explicitly represents decisions. The goal is to find the optimal decision tree by minimizing the generalization error. The classical CART algorithm was popularized by Breiman[55], but there are numerous algorithms for predicting continuous or categorical variables from a set of continuous predictors and/or categorical factor effects, and General Linear Models (GLM) or General Regression Models (GRM) are a an example of it, since the design is constructed in a linear combination of those predictors and factors, with or without interactions, being the predicted value continuous. This is the main difference between these algorithms and Discriminant Function Analysis (DFA).

Algorithms for constructing decision trees usually work top-down, choosing a variable that, at each step, best splits the set of items, depending on the metric that is being used. Most popular algorithms are the CART, the CHAID (Chi-squared Automatic Interaction Detection[56]) and the QUEST (Quick, unbiased, efficient, statistical tree[57]) algorithms. The Quest is generally faster than CHAID and CART algorithms, but can only be applied to classification

problems and, in very large data sets, the amount of memory required to compute this algorithm may be impractical.

Perhaps the first published proposal of tree algorithms was done in 1959 by Belson[58], where he addressed a matching issue which was, in fact, a predictive model where the prediction of the second group depend on the outcome observed for a first group. Predictors and outcome are dichotomized, and the tree grows using the difference between the observed count and the number expected under no association assumption, for each one of the two outcome categories.

2.4.1 Evolution of Decision Trees Algorithms

The first algorithms for inducing trees begun to appear from survey data and were mainly developed by statisticians. Perhaps the first tree algorithm proposed was published by Belson[58], in 1959, as a predictive model for the outcome of a second group given the outcome observed for the first group which, in fact, uses a Bayesian procedure. All predictors and outcomes needed to be dichotomized and the growing criteria computed the differences between observed counts and expected counts under the no association assumption which is something similar to a Chi-square test.

In few years, other proposals were made, such as the **AID** (Automatic Interaction Detector) algorithm for growing a binary regression tree for a dependent quantitative outcome, proposed by Morgan and Sonquist[59] in 1963, or the **ELISEE** (Exploration of Links and Interactions through Segmentation of an Experimental Ensemble) algorithm, proposed by Cellard, Labbé and Savisky[60] in 1967, a binary model for categorical dependent variables.

Meanwhile, with the computer development and the use of routines or programmes, the AID tree method became popularized in 1971, by Sonquist, Baker and Morgan[61], and ELISEE somewhere between 1970 and 1972, by Bouroche and Tenenhauss[62,63]. However, AID was always more popular, especially after Sonquist has shown interest in complementing it with a multiple correlation analysis tool. This algorithm was, meanwhile complemented by Messenger and Mandell[64], in 1972, and Morgan and Messenger[65], in 1973, with a tool for categorical outcomes using what was called the Theta Criterion, and resulted in the **THAID** (THeta AID) algorithm.

Others, such as Gillo[66], in 1972 or Guillo and Shelly[67] in 1974, extended the AID algorithm for multivariate quantitative outcome variables, resulting in **MAID** (Multivariate AID).

Before these extensions of the AID algorithm, two other groups worked independently: Hunt's group[68], in 1966, has proposed a series of decision trees induction algorithms known as Concept Learning Systems (**CLS-1** to **CLS-9**), explicitly developed in an Artificial Intelligent perspective for binary (CLS-1 to CLS-8) or multibranching (CLS-9) classification, while Press, Rogers and Shure[69] developed, in 1969, an interactive tree growing tool, the Interactive Data Exploration and Analysis (**IDEA**) that allowed multibranching.

With the exception of the Concept Learning System algorithms, all authors were mainly interested in finding alternatives to the restrictions of linear models, where the effect of explanatory variables are basically addictive, in order to detect important interactions, just to gain better knowledge about how outcomes are linked to explanatory factors, with no concern in improving predictions.

2.4.2 Actual Decision Trees Algorithms

One of the most simple decision tree algorithms is the **ID3**, developed by Quinlan[70] which uses information gain for splitting criteria and stops either because all instances belong to a single value of the target or when the higher information gain is non-positive, without making any pruning. However, this algorithm does not perform any pruning, and it cannot work with numeric information or missing values. Algorithm ID3 evolutes to C4.5[71], using the gain ratio as splitting criteria, stopping when the number of issues to split is below a given threshold and incorporating error-based pruning. Besides this, C4.5 can handle numeric information and missing values.

In 1984, Breiman[55] presesented the **CART** algorithm. The Classification and Regression Tree (CART) Algorithm is characterized by the construction of binary trees as each internal node has exactly two outgoing edges. The splits are selected using the *twoing criteria* and pruning process of the tree is done by cost-complexity. It is possible here to define *a priori* distribution.

On the other hand, this type of procedure can generate regression trees, that is, these models can predict a real number instead of a class and, in this case, splits are generated in order to minimize the least-square deviation, that is, the squared error, and the prediction is each leaf is based on the weighted mean of the node.

The evolution of AID to MAID and THAID has leaded to the development of the **CHAID** algorithm, in 1980, by Kass[56]. In fact, CHAID stands for CHi-square Automatic Interaction Detection so it can be viewed as an evolution of those primary methods, by using the p-

value obtained by a statistical test to find out the pair of values that has least significant difference with respect to the target attribute. Initially, this algorithm was designed for nominal dependant variables so the Pearson Chi-square test was the only one to apply. However, actually it handles all types of dependent variables or target attributes so if its nature is ordinal then a likelihood-ratio test is used and in cases of quantitative attributes the F distribution is the one to use. For each selected pair, CHAID checks whether the adjusted p-value obtained is greater than a threshold and, if so, it merges values and searches for an additional potential pair to be merged, until no significant pairs are found, in order to have each child node made of a group of homogeneous values of the selected attribute. However, CHAID can also stop due to reaching maximum tree depth, or reaching the minimum number of cases in a child node or in a parent node, and it does not perform any pruning, in spite of handling missing values as a single valid category.

The **QUEST** algorithm was developed by Loh and Shih[57] in 1997, and the acronym stands for Quick, Unbiased, Efficient, Statistical Tree. It supports univariate and linear combination splits and, for each split, the association between each input attribute and target attribute is computed using ANOVA F-test or Levene's test if their measurement level is, at least, ordinal, or Pearson's Chi-square if variables are nominal. One advantage of QUEST is that it handles multinomial target attributes and, in this particular case, a two-means clustering is used to create two super-classes and the attribute that obtains the highest association with the target attribute is the one selected for splitting. The optimal splitting point for the input attribute is obtained by Quadratic Discriminant Analysis. This algorithm has a negligible bias, yields binary tree and performs ten-fold cross-validation to prune the tree. Moreover, QUEST is generally faster than CART or CHAID, but for very large data sets, the memory requirements are usually larger and its application may be impractical. Also, it cannot be applied to regression type problems.

There are other algorithms available in literature, but most of them are variations of the previous ones. Decision trees are self-explanatory and an easy to follow procedure and they can easily be converted in a set of procedures, which handles nominal and numeric attributes simultaneously, and work with missing values. Moreover, they are considered a non-parametric method so assumptions about the space distribution and classifier structure are not needed.

However, some decision tree algorithms, like ID3 and C4.5 require the target attribute to have only discrete values. Quinlan[71] points out that decision trees are over-sensitive to the training set in order to irrelevant attributes and noise.

2.4.3 Growing the tree – splitting, stopping and pruning

The aim of these primary methods was mainly segmentation of data into groups with as much difference as possible thus splitting criteria was basically obtained determining measures of association between outcome and split variables. Nowadays, the effort is in order to maximize homogeneity of each group by means of purity measures, and the splitting criteria depend on the nature of the dependent variable.

In most cases, the discrete splitting functions are univariate in the sense that an internal node is split according to the value of a single attribute so that the inducer searches for the best attribute upon to split. These criteria is characterized according to the origin of measure, such as information theory, dependence and distance, related to measures of association, or according to measures of structure, such as impurity based criteria, normalized or not, or Binary criteria.

An **impurity measure** can be defined as follows: given a random variable X with k discrete values $(x_1, x_2, \dots, x_k)_k$), distributed according to the vector$P = (p_1, p_2, \dots, p_k)$, then an impurity measure is a function Φ: $[0,1]^k \to R$ which satisfies simultaneously[72]:

$$\begin{cases} \Phi(P) \geq 0 \\ \Phi(P)\ is\ minimum\ if\ \exists i \in \{1,2,\dots,k\}: p_i = 1 \\ \Phi(P)\ is\ maximum\ if\ \forall i \in \{1,2,\dots,k\}: p_i = \frac{1}{k} \\ \Phi(P)\ is\ symmetric\ with\ respect\ to\ components\ of\ P \\ \Phi(P)\ is\ smooth\ in\ its\ range\ (that\ is, differentiable\ in\ its\ range) \end{cases}$$

Note that if a component of P equals 1, it means that the variable X has only one value, and then the variable is defined as **pure**. On the other hand, if all components are equal, then the level of impurity reaches the maximum.

Given a training set S, the probability vector of the target attribute Y is defined by[72]

$$\mathrm{P_Y}(S) = \left(\frac{|\sigma_{y=c_1} S|}{|S|}, \dots, \frac{\left|\sigma_{y=c_{|dom(y)|}} S\right|}{|S|} \right)$$

And the goodness-of-split due to discrete attribute a_i is defined as a reduction in impurity of the target attribute after partitioning S according to the values $v_{i,j} \in \mathrm{dom}(a_i)$ so that[72]

$$\Delta\Phi(a_i, S) = \Phi\big(P_Y(S)\big) - \sum_{j=1}^{|dom(a_i)|} \frac{\left|\sigma_{y=v_{i,j}} S\right|}{|S|} \times \Phi\left(P_Y\left(\sigma_{y=v_{i,j}} S\right)\right)$$

Some measures of impurity that are often used are presented below:

- Information Gain[73] (IG) is an impurity-based criteria that has its origins on information theory and uses entropy (E) as a measure of impurity:

  $$IG(a_i, S) = E(y, S) - \sum_{v_{i,j} \in dom(a_i)}^{|dom(a_i)|} \frac{\left|\sigma_{y=v_{i,j}} S\right|}{|S|} \times E\left(y, \sigma_{y=v_{i,j}} S\right)$$

  Where

  $$E(y, S) = \sum_{c_j \in dom(y)} - \frac{\left|\sigma_{y=c_j} S\right|}{|S|} \times \log_2\left(\frac{\left|\sigma_{y=c_j} S\right|}{|S|}\right)$$

- Gini Index (GI) is an impurity-based criteria which measures the divergence between probability distributions of the target attribute's value, and is defined by:

  $$GI(y, S) = 1 - \sum_{c_j \in dom(y)} \frac{\left|\sigma_{y=c_j} S\right|}{|S|}$$

  The evaluation criteria for selecting an attribute a_i is defined by the Gini Gain (GG) function as:

  $$GG(a_i, S) = GI(y, S) - \sum_{v_{i,j} \in dom(a_i)} \frac{\left|\sigma_{y=v_{i,j}} S\right|}{|S|} \times GI\left(y, \sigma_{y=v_{i,j}} S\right)$$

- Likelihood-ratio Chi-squared Statistics was defined by Attneave[74], in 1959, as $G^2(a_i, S) = 2 \times \ln(2) \times |S|\, IG(a_i, S)$ and is useful for measuring the statistical significance of the information gain criteria since it is tested under the null hypothesis that input and target attributes are independent and since $G^2(a_i, S) \sim \chi^2_{(dom(a_i)-1)\times(dom(y)-1)}$.
- The DKM criterion was designed by Dietterich, Learns and Mansour[75] in 1996. Later, in 1999, Kearns and Mansour[76] have proven, theoretically, that this criterion requires smaller trees for obtaining the same error than other impurity indexes, such as Information Gain or Gini Index, and is defined by

  $$DKM(y, S) = 2 \times \sqrt{\left(\frac{\left|\sigma_{y=c_1} S\right|}{|S|}\right) \times \frac{\left|\sigma_{y=c_2} S\right|}{|S|}}$$

However, the DKM impurity-based criterion and the other presented above are biased towards attributes with larger domain values. For this reason, it is important to normalize the impurity based measures, which may origin some of the following normalized Impurity based criteria:

- The Gain Ratio (GR) normalizes the Information Gain in order to Entropy, as long as it is not null, by[72]

$$GR(a_i, S) = \frac{IG(a_i,S)}{Entropy(a_i,S)}$$

- The Distance Measure (DM) normalizes the Impurity Measure, such as the Gain Ratio, but in a different way[72]

$$\frac{\Delta\Phi(a_i,S)}{-\sum_{v_{i,j}\in dom(a_i)}\sum_{c_k\in dom(y)}\frac{\left|\sigma_{a_i=v_{i,j}\wedge y=c_k}S\right|}{|S|}\times\log_2\left(\frac{\left|\sigma_{a_i=v_{i,j}\wedge y=c_k}S\right|}{|S|}\right)}$$

Other criterions have been used, such as the Binary criterion, developed for binary trees, is based on the division of the input attribute domain into two sub-domains. For instance, let $\beta(a_i, dom_1(a_i), dom_2(a_i), S)$ denote the binary criterion value for attribute a_i over the sample S when $dom_1(a_i)$ and $dom_2(a_i)$ are its correspondent sub-domains. Then, the value obtained for optimal division of the attribute domain into two mutually exclusive and exhaustive sub-domains is used for comparing attributes. Some examples of binary criteria are the following:

- the Twoing criteria, used in recent CART algorithms, and preferred to the binary criteria when domain of target attributes is relatively wide[55] and the Gini Index may encounter problems. However, if the target attribute is binary, then the Twoing and Gini criteria are equivalent. Note that in multi-class problems, the towing criteria prefer attributes with evenly divided splits.

$$twoing(a_i, dom_1(a_i), dom_2(a_i), S) = 0{,}25 \times \frac{\left|\sigma_{a_i\in dom_1(a_i)}S\right|}{|S|} \times \frac{\left|\sigma_{a_i\in dom_2(a_i)}S\right|}{|S|} \times \times \left(\sum_{c_i\in dom(y)}\left|\frac{\left|\sigma_{a_i\in dom_1(a_i)\wedge y=c_i}S\right|}{\left|\sigma_{a_i\in dom_1(a_i)}S\right|} - \frac{\left|\sigma_{a_i\in dom_2(a_i)\wedge y=c_i}S\right|}{\left|\sigma_{a_i\in dom_2(a_i)}S\right|}\right|\right)^2$$

- The Orthogonal criteria (ORT) is defined through the angle formed by the two vectors $P_{y,1}$ and $P_{y,2}$, which represent the probability distribution of the target attribute in the partitions $\sigma_{a_i \in dom_1(a_i)} S$ and $\sigma_{a_i \in dom_2(a_i)} S$, respectively. This criterion performs better than information gain or the Gini index in some specific problems, and is defined by[72]

  $$ORT(a_i, dom_1(a_i), dom_2(a_i), S) = 1 - cos\theta\left(P_{y,1}, P_{y,2}\right)$$

  Another binary criterion is the one proposed by Friedman, in 1977, and that applies the Kolmogorov-Smirnov distance and is useful to handle target attributes with multiple classes and missing values, and it has been suggested that it outperforms the gain ratio criteria. When assuming a binary target attribute $dom(y) = \{a_1, a_2\}$, the criteria is defined as[72]:

  $$KS(a_i, dom_1(a_i), dom_2(a_i), S) = \frac{\left|\sigma_{a_i \in dom_1(a_i) \wedge y = c_1} S\right|}{|\sigma_{y=c_1} S|} - \frac{\left|\sigma_{a_i \in dom_2(a_i) \wedge y = c_2} S\right|}{|\sigma_{y=c_2} S|}$$

- The AUC-spliting criteria is defined by the selection of the attribute that maximizes the Area Under convex hull the ROC Curve, and this criteria has showed that it outperforms other splitting criteria both with respect to accuracy and area under the curve. However, this criteria does not perform a comparison between impurity of the parent node relatively to the weighted impurity of children nodes after splitting.

Several authors have performed comparative studies of the criteria described above (and presented on Table 6), using methods such as Permutation Statistics[(76)], mean posterior improvements and the use of the hyperbolic distribution measures[(77)], but most of the comparisons are based on empirical results, although some theoretical conclusions were obtained. Most of the authors state that the choice of the splitting criteria will not make much difference on tree performance. The criteria that would improve the tree´s performance dramatically would be a multivariate splitting criteria, where several attributes may participate in a single node split test. Most of the multivariate split criteria are based on linear combinations of the attributes, which can be performed by a linear discriminate analysis.

Table 6: Main earlier and actual tree growing algorithms.

Algorithm	Local Split	Dependant variable		Splitting criteria		
		quantitative	qualitative	Association	Purity	p-value
Belson	Binary		X	X		
AID	Binary	X		X		
MAID	Binary	X		X		
THAID	Binary		X	X	X	
CLS-1 to 9	n-ary		X	X		
ELISEE	Binary		X	X		
IDEA	n-ary	X	X	X		X
CHAID	n-ary	X	X	X		X
CART	Binary	X	X		X	
QUEST	Binary	X	X		X	

The tree will continue to grow until some of the following conditions is achieved:

- All instances in the training set belong to a single value of y
- The maximum tree depth is reached
- The number of cases in a terminal node is less than the minimum of cases for parent nodes
- If the node were split, the number of cases in one or more child nodes would be less than the minimum number of cases for child nodes
- The best splitting criteria is not greater than a certain threshold

However, using tightly stopping criteria tends to create small and under-fitted decision trees, while using loosely stopping criteria tends to generate larger trees that are over-fitted to the training set so, pruning methods originally suggested by Breiman[55] were developed for solving this problem. It has been suggested that a loosely stopping criteria should be used, letting the decision tree over-fit the training set, and then the over-fitted branches should be cut in order to create a smaller tree without he branches that are not contributing to the generalization accuracy. There are various techniques for pruning decision trees, most of them performing a top-bottom or bottom-up transversal to all nodes, where a node is pruned if this operation improves a certain criteria. Several studies aimed to compare the performance of different pruning methods, but results indicate that no pruning method tends to over-perform the others; in fact, while some methods (cost-complexity or reduced error pruning) tend to create smaller and less accurate trees (over-pruning), other methods (error-based, pessimist error and minimum error pruning) tend to be under-pruning.

PART III

MODEL DEVELOPMENT, APPLICATION AND ASSESSMENT

CHAPTER 5

MATERIAL AND METHODS

1. Collecting data – general procedures

Data were collected under the scope of the Diamarker project "Genetic susceptibility of multi-systemic complications of diabetes type 2 novel biomarkers for diagnosis and monitoring therapy", under the supervision of the principal investigator Miguel Castelo-Branco.

This project is a part of a bigger project DoIT – Development and Operationalization of Translational Research, promoted by Portugal Health Clusters and supported by QREN nº 13853, with a total number of 21 partners among companies, Research and Development institutions and hospitals.

This is an ongoing project which intends to recruit 300 type 2 diabetics and 300 controls in order to characterize phenotypes of diabetic retinopathy progression using multimodal imaging, and also other systemic complications with an emphasis on imaging of ocular, cardiac, brain and liver tissues, in correlation with clinical and biochemical signatures of diabetes type 2.

This study was designed as a pilot, observational and prospective study with one visit where controls and type 2 diabetes mellitus patients performed multimodal imaging examinations, namely, ophthalmological, psychophysical, heart, liver and cerebral imaging, after signing an informed consent and being previously evaluated for eligibility.

Inclusion criteria were as follows: men and women aged between 40 and 75 years, functionally independent, capable to provide written consent after proper education and discussion with the treating physician and/or the research physician, with type 2 diabetes for the diabetic group and without any type of diabetes for the control group. Exclusion criteria defined for the study were: history of neuropsychiatric, renal, heart, ocular or any other

severe non-age disease unrelated to diabetes, pregnancy or lactation.

At baseline visit, eligible patients were asked to participate in the study and signed the informed consent form. Then, subjects completed a questionnaire on lifestyle, cardiovascular risk factors and family history of diabetes, current medication, physical activity, dietary pattern and quality of life.

Afterwards, at the clinic, height, weight, waist and blood pressure were measured, and blood samples were collected for analysis of glucose markers, lipids, inflammatory and other biomarkers, and DNA analysis. An urine sample was also collected in the diabetic group.

Thus, the assessment schedule comprised the following procedures:

- Informed consent
- Patient identification, demographic details and medical history
- Inclusion/exclusion criteria
- Concomitant medications and non-drug therapies
- Blood collection
- Urine collection
- Vital signs (blood pressure)
- Ophthalmic examination including best corrected visual acuity and intraocular pressure measurements
- Visual psychophysical tests (Speed, Colour and Contrast Discrimination)
- Multimodal imaging
- Occurrence and details of adverse events
- Study discharge

Concerning multimodal imaging, cerebral, heart, liver and ophthalmological scanning was performed, according to the following procedures:

- Cerebral Imaging
    - Arterial Spin Labelling (ASL) and Blood oxygen level dependent (BOLD) contrast
    - Magnetic resonance spectroscopy
    - Fluid attenuated inversion recovery (FLAIR)
    - Magnetization-prepared radio-frequency pulses and rapid gradient echo (MR RAGE)
    - Time of flight magnetic resonance (TOF MR) angiography
- Heart Imaging
    - Ultrasound thickness of the intima-media complex in the carotid arteries

  - Triglyceride accumulation (TG) spectroscopy
  - Calcium score
  - True fast imaging with steady state precession (True-FISP)
- Liver Imaging
  - Gradient echo (GRE)
  - Intra-voxel Incoherent Motion Diffusion Weighted Imaging (IVIM DWI)
  - ME-GRE (Multiecho)
  - Triglyceride accumulation (TG) spectroscopy
- Ophthalmological Imaging
  - Colour Fundus Photography
  - Optical Coherence Tomography (OCT)

2. Selection of patients and data management

The selection of patients and controls was performed by the University Hospital of Coimbra (CHUC) and the Unit of Research and Development of the CHUC that synchronized this process with all the hospital centres involved and with the Institute for Biomedical Imaging and Life Sciences (IBILI) articulated with the Faculty of Medicine of the University of Coimbra (FMUC). There was no randomization since it is an observational study.

A database was developed in order to store all the data acquired, maintaining all the necessary confidentiality. Clinical data for all participants included in the study were storage in a SQL database with restrict access to the project investigators. Control of database was performed with am authentication login for users where the system would verify, using a login and password, the access credentials. The system administrator, under the supervision of the principal investigator, was responsible for the management of the database users, was authorized to insert new users, edit or remove actual users, and also to determine the access profile for users, so that investigators with edition profile were authorised to create, edit or eliminate data from the database, while visualization users could only use and perform queries about data stored, in an anonymous form, being denied the access to edition data pages as well as contact information, names and identification numbers.

3. Sample: train sample and test sample

The sample used for this study consisted on data available at the database in the 31th of December 2013 and was used as a training sample for the development of classification models. All the data inserted after that date formed a new sample, named test sample, on which developed classification models were tested.

The training sample considered 96 subjects, equally distributed by gender (55 males and 41 females), aged between 40 and 73 years, of which 49 (51.04%) were type 2 diabetics, diagnosed between one and 39 years before, and 47 (48.96%) were controls for this disease. In the diabetic group, 40 patients had ETDRS grading of diabetic retinopathy performed and registered in the database. Half of them (20 subjects) did not have diabetic retinopathy and the other half had non-proliferative diabetic retinopathy. None of the cases were diagnosed as having proliferative diabetic retinopathy.

The test sample was composed by all the subjects that entered onto the database after the first of January 2014, and 57 subjects were considered, 30 of them controls (52.63%) and 27 diabetics (47.37%), aged between 41 and 73 years old

4. Variables measured in the training sample and measurement instruments

For the training sample, data related to heart, liver and cerebral imaging, at the database, was still incomplete. We therefore decided to focus only blood sample measures, ophthalmic examination, visual psychophysical tests and eye imaging. Urine sample data was not used since they were only collected on type 2 diabetics.

Patients were tagged by an identification number, and demographic details such as age, gender, medical history (namely for hypertension in order to identify subjects with diagnosed hypertension), family history of diabetes and vital signs were collected.

Height (m) and weight (kg) were used in terms of body mass index (m/kg^2) and abdominal perimeter was discarded of the analysis since it was measured/recorded for all the diabetic subjects, but only in seven controls at the time of data extraction. For the same reason, pulse, systolic and diastolic blood pressure and bioimpedance were discarded from the analysis. The number of controls with registers for those variables was, respectively, fifteen, nineteen and one, at that time.

Blood and urine sample collection were for the analysis of metabolic and waste biomarkers

at the Coimbra Hospital and University Centre (CHUC), while genetic characterization is being performed at BIOCANT (Centre for Innovation in Biotechnology).

Blood tests performed for all subjects were:

- Blood glucose (mg/dL) and glycosylated haemoglobin. Glycosylated haemogloblin is reported in terms of the National Glycohemoglobin Standardization Programme (NGSP), expressed as percentage of HbA1C, and in terms of the International Federation of Clinical Chemistry Working Group (IFCC), expressed in mmol/mol. Although IFCC results are accuracy-based, and highly correlated with NGSP results ($NGSP = 0.018148 \times IFCC + 2.152$), the later ones can be directly related to clinical outcomes and diabetes care goals, so both are presented;
- Creatinine values (mg/dL) as a measure of renal function;
- ALT, AST, alkaline phosphatase and gamma GT as measures of liver function. Alanine transaminase (ALT) and Aspartate transaminase (AST) are enzymes that measure hepatic lesion since they appear augmented in blood when there is a lesion, although the second one in not liver specific, since it also appears in red cells, skeletal and cardiac muscles;
- Cholesterol (total, HDL and LDL), atherogenic index and triglycerides;
- Apolipoprotein A1, B100 and their relation (A1/B100), and Lipoprotein. Apolipoprotein is a protein that binds lipids and is associated to cholesterol. It is classified into two types: the apolipoprotein A1 is synthetized in the liver and in the small intestine and is a part of the HDL cholesterol and has the role of simplifying the transportation of the HDL cholesterol to the liver; the apolipoprotein B100 is synthetized in the liver and is a part of the LDL cholesterol that is responsible for joining it to cellular receptors and may lead to atherosclerosis if accumulated in the arteries. Though, the ratio between those two apolipoproteins (B100/A1) may reflect the risk of developing cardiovascular disease;
- Cell blood counts in cytometry: Leucocytes, Erythrocytes, Haemoglobin and Haematocrit, mean corpuscular volume (MCV), Mean corpuscular haemoglobin (MCH), Mean corpuscular haemoglobin concentration (MCHC), Erythrocyte variation coefficient (EVC), Platelet, Mean platelet volume (MPV), Plateleocrit and Platelet variation coefficient;
- Hormonology measured Thyroid stimulating hormone (TSH) and connecting peptide (C-peptide) since they may affect diabetes control.

Phenotyping of diabetic retinopathy includes ophthalmic characterization, psychophysical, and also optical coherence tomography and colour fundus photography.

The best corrected visual acuity (BCVA) was performed in both eyes, according to the Early Treatment Diabetic Retinopathy Study (ETDRS) protocol. The sequence of testing begun at 4-meters, first in the right and then in the left eye, continuing if the eye or eyes with the BCVA measured at 4-meters was worse than 20/100 Snellen equivalent (fewer than 20 letters read correctly on the 4 largest lines of the chart). In this case, the eye or eyes should be tested at 1 meter.

The Colour Fundus photographs were taken with a resolution of at least 768x576 pixels, after pupil dilatation in the study eye for each patient to evaluate diabetic retinopathy and perform the ETDRS classification. Two 45º angular field-of-view images were acquired: one covering the macular region, centred on the fovea (Field 2) and one centred on the optic disc (Field 1M), as presented on Figure 12. At this point, intraocular pressure was also measured.

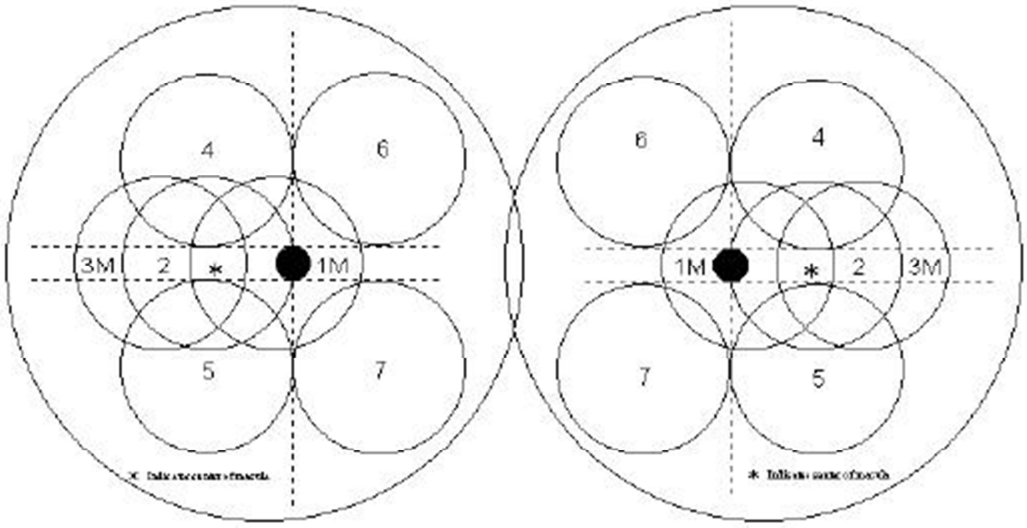

Figure 12 – Modified 7-standard Fields Colour Photographs. Figure obtained from the Study Protocol. Field 1M (Disc), Field 2 (macula), Field 3M (temporal to macula), Field 4 (Superior Temporal), Field 5 (Inferior Temporal), Field 6 (Superior Nasal), Field 7 (Inferior Nasal). Font: Diamarker Study Protocol.

Optical Coherence Tomography (OCT) was performed as a diagnostic imaging tool of the posterior segment eye structures, using low coherence interferometry to produce cross-sectional tomograms in those structures. An 840 nm light source emitted a probe beam of infrared light spitted between the eye and a reference mirror at known spatial locations, generating two beams that are reflected back to a photo detector. Those two beams are reflected back to a spectrometer, and thickness data of retina is obtained by measuring the time of flight delay of light back scattered from different layers in retina. These data are processed in an internal processor to produce enhanced images after adjusting for the movement of the eye and intraocular pressure variations and retinal

thickness is finally determined. An algorithm is used to determine the inner and outer retinal boundaries for each scan (several A-scans are performed, for each subject, along six B-scans in order to determine retinal thickness)

The Frequency Domain Spectralis OCT (Heidelber Engineering, Heidelber, Germany) was used for this procedure, with software 5.4.6, and both eyes were used, at maximum dilatation to help insure optimal quality scans. The Macular Thickness Map was acquired for a volume scan 20ºx20º (in 25 sections, 10 frames, HS 512 A-scans) for each eye, and also an Optic Disc RNFL (retinal nerve fiber layer) thickness map was obtained (100 frames, HS RNFL Single Exam Report with FoDi).

Optical Coherence Tomography is able to provide either qualitative information on morphology and reflectivity or quantitative information on thickness, mapping and volume, in real time, and is a non-invasive technique that has revolutionized the evaluation, treatment and prognosis of diabetic retinopathy.

Volume scan density, in micrometers (μm), was acquired for the central subfield, within 1mm of the centre of the macula. We also obtained the volume scan density for nasal, temporal, superior and inferior quadrants in the inner region (within 1 and 3 mm of the centre of the macula) and in the outer region, comprised between 3 and 6 mm of the centre of the macula, as shown in Figure 13.

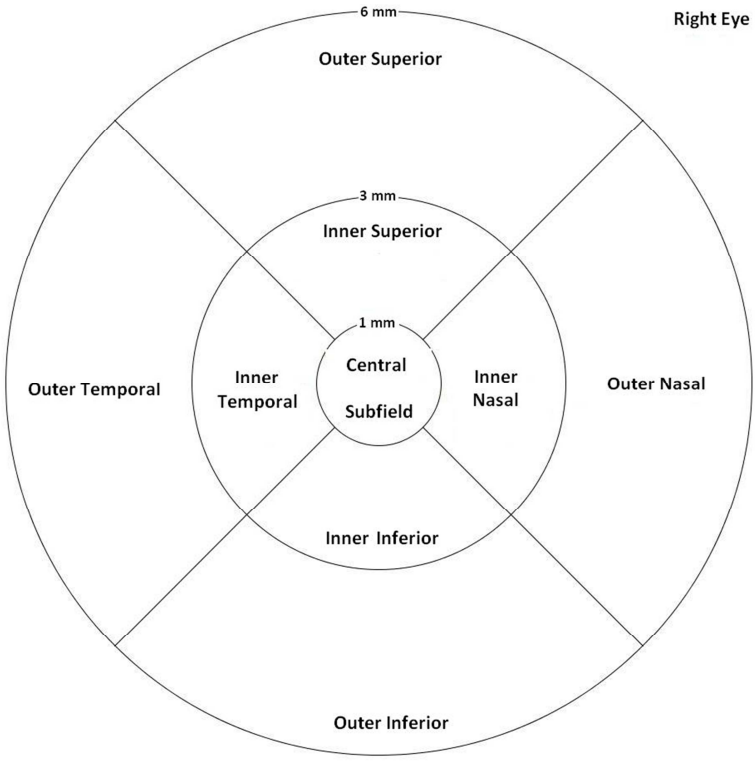

Figure 13 – Fields of volume scan density for Frequency Domain Spectralis OCT (Heidelber Engineering, Heidelber, Germany).

We were able to use, also, the retinal nerve fiber layers (RNFL) global thickness, on nasal and temporal quadrants, in micrometers (μm), and nasal superior and inferior, or temporal superior and inferior regions, obtained within π mm of the centre of the macula, as shown in the following Figure 14:

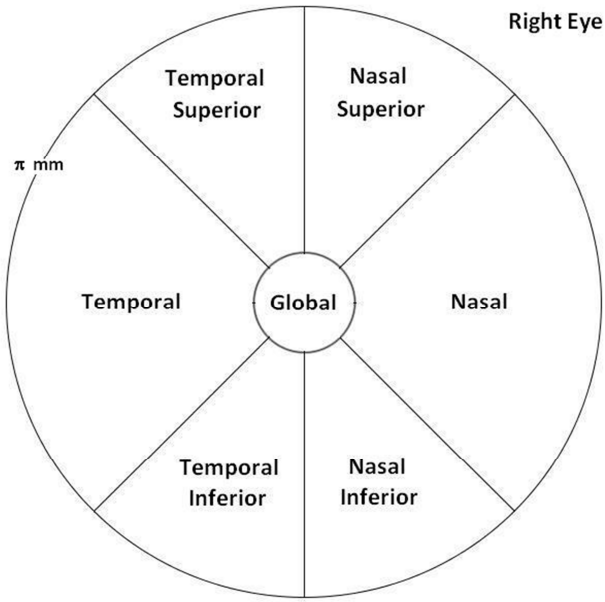

Figure 14 – RNFL quadrants for Frequency Domain Spectralis OCT (Heidelber Engineering, Heidelber, Germany).

Computerized psychophysical tests from the multifunctional module of the threshold of visual discrimination measure the ability of subjects to detect movement (Speed test), achromatic contrast (Achromatic test) and chromatic contrast (according to Protan, Deutan and Tritan axes corresponding to distinct cone populations). All the tests use lateral, randomly moving pairs of dots, one being a reference point within each meridian used. Peripheral presented stimuli are of short duration, between 400 and 900 milliseconds, and also of short dimension and reduced spatial amplitude (about 1 degree of the visual angle). Periphery distances to the fixation point are of 7.5 visual degrees if the selected meridian was the 0º or 90º, or of 10 or 15 visual degrees on meridians 45º and 135º, respectively. Central fixation was controlled by an eye-tracker device, and that information was used in real time to validate the trial. If there was no central fixation, the trial would be successively repeated until validation. Response to each trial was given after sound stimuli, which occurs at the end of vertical fixation. Properties as screen background point size and central cross remained constant for all the tests (speed, achromatic and chromatic), and the only property (dependent variable) that changed was the one being analysed at each case. These tests return a threshold that represents the minimal difference between the properties being analysed to the asymptotic value at chance level. The screen background is achromatic

(grey) and the luminance used had a sufficient magnitude to guarantee that the test occurred in conditions of photopic response, that is, 30 candelas/m^2.

In the speed test, both points correspond to stimuli of maximum achromatic luminance (white), and differ only on the movement speed (one has constant velocity while the other starts at maximum velocity and loses acceleration until it gets closer to the reference speed, until as the subject is able to discriminate the faster moving point.

Achromatic contrast discrimination or luminance test was performed defining stimuli as in the speed test, of short duration (400 to 900 milliseconds) and dimension, with short spatial amplitude (2º of visual angle). Speed of both points is equal and constant (5 visual degrees per second), and both are achromatic (grey), but they differ on the value of grey's luminance. One point has constant luminance (reference point), and the other starts clearly brighter, and successively loses luminance under a staircase procedure until it reaches the reference point luminance, until the subject can correctly identify the brighter point.

Chromatic contrast discrimination test was performed using the same peripheral stimuli as the speed discrimination test, with short duration (400 to 900 milliseconds) and dimension, with 2 degrees of visual angle amplitude of movement. The velocity of both points is equal and constant during the test, with the value of 5 visual degrees per second. The reference point consists in an achromatic constant stimulus (grey), while the test point has a very sharp colour at the beginning of the test, and successively begins to turn approximately with the same achromatic colour as the reference point, along the axes that isolate one type of cone (Protan, Deutan and Tritan, which respective deficiencies correspond to the patterns observed in Figure 15), until the subject correctly identifies the test point.

These tests were performed only in the dominant eye.

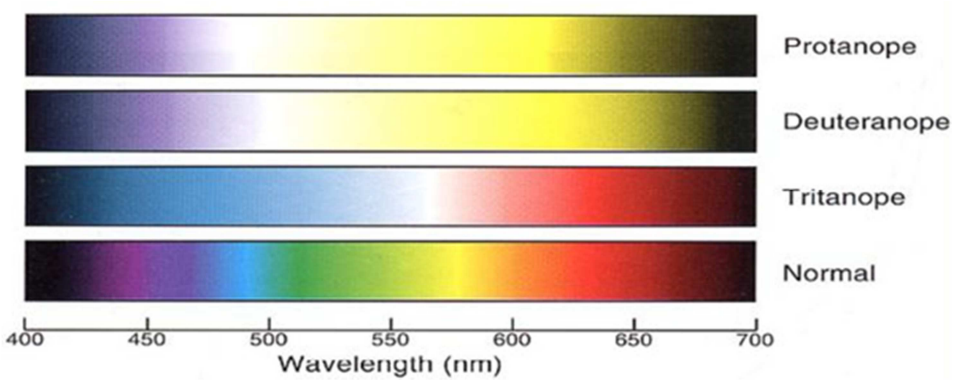

Figure 15 – Representation of normal and colour defects on chromatic vision.

5. Statistical methods

5.1 Handling correlated data from both eyes – measures and graphics of agreement

Naturally, as there was available data from optical coherence tomography (OCT) for both eyes, a critical decision should be made, along with the choice of unit of analysis.

Although it seems obvious that if we are looking for type 2 diabetes classifiers, the unit of analysis is the person and only one eye should be chosen, when we are trying to develop models for diabetic retinopathy classification, then some confusion may arise.

It seems that it is justifiable to waste one of the measurements since the unit of analysis is the person, not the eye so, another problem emerges: should we choose one eye, or the mean of both eyes? If we choose only one, which eye should we choose?

The use of the mean of both eyes can be tendentious, if outliers are present. In this particular case, the mean can be deviated from the expected mean thus it seems more appropriate to use one of the eyes, even though we lose some information.

Other criteria usually applied are the use of the best or of the worst eye. In this case, results may be biased, since we give an overestimation or an underestimation of the real values.

It seems more adequate to assign measures to an eye, such as the dominant eye, or even better, to randomly assign one eye to study. As psychophysical tests were performed only in the dominant eye, it seems acceptable to choose the dominant eye from OCT to perform statistical analysis. However, if data from both eyes are uncorrelated, and large differences occur between eyes, then both eyes information should be used. According to this, initially we analysed correlation between two eyes measurements by applying Spearman´s rank order correlation due to the lack of normality, and Wilcoxon Matched pairs test for comparison of both eyes.

However, and following Armstrong[32] recommendations, other measures of agreement should be used, such as the intra-class correlation coefficient or concordance correlation coefficient.

The **Intra-class Correlation Coefficient (ICC)** is recommended by Armstrong[32] and is used to assess the consistency, or conformity, of measurements made by multiple observers measuring the same quantity, thus, it is a measure of the reliability of measurements. Since each eye was measured by the same instrument, we have chosen to use the ICC as a measure of absolute agreement, where systematic differences are relevant, instead of using ICC as a measure of consistency (systematic differences between measures are irrelevant). The value of the population intra-class correlation coefficient is a measure of the

homogeneity of observations within the classes of a random factor relative to the variability of such observations between classes. It will have the value of zero when the estimated effect of the random factor is zero, and it will reach the unity when the estimated effect of error is zero, given that the total variation of the observations is greater than zero[79].

Intra-class correlation coefficient can be estimated through single or average measures. The single measure of ICC is an index for the reliability of the ratings for one, typical single measure or one ratter, which is uncommonly, used in clinical reliability studies; the average measures model is an index for the reliability of different measures or ratters averaged together. Therefore, this second estimative is more useful in the case that is being studied.

Three different models can be used to obtain that estimative: the two-way random model is used whenever we have random subjects and random items, measures or ratters (left and right eye measurements) while the two way mixed combines random subjects with fixed items, measures or ratters, as they are the only items, measures or ratters of interest; If we are interested in assessing each subject by a different set of randomly selected measures or ratters, then a one-way random model should be used. This model considers subject effects as random and its use is rare in clinical reliability studies. It seems that the two-way mixed model is the most appropriate for this study.

Reliability estimates produced under the fixed or random ICC models are numerically identical, but their interpretation is different. Results of an analysis under the mixed effects model cannot be generalized to other measures or ratters; since other measures concerns other eyes for each person, measured by OCT, which, in fact, do not exist.

The ICC is constructed to be applied to exchangeable measurements in which there is no meaningful way to order measures within a group as in this case; It does not matter which eye is measured first, the left or the right eye. Since ICC gives a composite of intra-observer and inter-observer variability, it can be difficult to interpret when observers are not exchangeable and alternative measures such as Cohen's Kappa statistic or Fleiss kappa or concordance correlation coefficient have been proposed as more suitable measures of agreement among non-exchangeable observers[80].

From those three measures, the only one that can be applied to quantitative measurements is the **Concordance correlation coefficient (CCC)**.

The concordance correlation coefficient ρ_c[81] evaluates the degree to which pairs of observations fall on the 45° line through the origin. It contains a measurement of precision ρ and accuracy C_b: $\rho_c = \rho C_b$, where ρ is a measure of precision since it is the Pearson

correlation coefficient, which measures how far each observation deviates from the best-fit line, and is a C_b is a measure of accuracy providing a bias correction factor that measures how far the best-fit line deviates from the 45° line through the origin.

The concordance correlation coefficient is nearly identical to some of the measures called intra-class correlations. Carol Nickerson[80] performed comparisons of the concordance correlation coefficient with an intra-class correlation on different data sets, and found only small differences between the two correlations in one case and on the third decimal.

In this particular case, we have used a pseudo-concordance correlation coefficient (pCCC) determined using ρ as the Spearman correlation coefficient. As Spearman rank-order correlation coefficient was used for analysis, instead of Pearson's Correlation Coefficient, due to the lack of normality in data distribution, a pseudo-Concordance correlation coefficient was determined replacing Pearson's by Spearman's correlation coefficient.

Hypotheses about the value of the population correlation coefficient ρ between a pair of variables can be tested using the Fisher transformation[82,83] applied to the sample correlation coefficient. This transformation is defined by

$$z = arctanh(r) = \frac{1}{2} ln\left(\frac{1+r}{1-r}\right)$$

And it is known that Z has normal distribution with mean $\frac{1}{2} ln\left(\frac{1+\rho}{1-\rho}\right)$ and standard error $\frac{1}{\sqrt{N-3}}$.

The procedure for computing a statistical test to compare two correlation coefficients can then be performed, after transforming each correlation coefficient into a Z score (Z_1 and Z_2) and testing the difference between Z scores using the combined standard error as $\sigma_{Z_1-Z_2} = \sqrt{\frac{1}{N_1-3} + \frac{1}{N_2-3}}$ where N_1 and N_2 are the number of pairs of scores used to determine Z_1 and Z_2, respectively. By doing this, it is easy to obtain a p-value for each comparison.

Using the same methodology applied to the intraclass correlation coefficient, to the concordance correlation coefficient and pseudo-correlation coefficient, we can compute a z statistic for the difference between each pair of measures, since their values are measured in the same scale as corelation coefficient. However, comparing each pair of the three coefficients, type I error increases so, althought conservative, a Bonferroni correction was applied to each comparison. Therefore, we preferred to compare these three coefficients using a mountain plot. The idea of comparing these coefficients is in order to show that

concordance between eyes exists, independently of the coefficient used.

A **Mountain plot**, or folded empirical cumulative distribution plot allows comparison of 2 or 3 measurements. It computes a percentile for each ranked difference between a new method and a reference method. To get a folded plot, all percentiles above 50 are converted to a new percentile determined as new_percentile= 100 – old_percentile, and these new percentiles are then plotted against the differences between the two methods[84]. In this graph, it is easy to find 95% of the data, even when data is not normally distributed, and different distributions can be easily compared.

Some graphical procedures can show this agreement between correlated data, such as eyes. Armstrong[32] proposed the **Bland and Altman Plot**, but **Youden Plot** analysis could also be performed.

Graphically, the Bland and Altman plot[44,85] is a statistical method to compare two measurements techniques. In this graphical method, the differences or, alternatively, the ratios between the two techniques are plotted against the averages of the two techniques.

Horizontal lines are drawn at the mean difference, and at the mean difference plus and minus 1.96 times the standard deviation (SD) of the differences. If the differences within mean ± 1.96 SD are not clinically important, the two methods may be used interchangeably.

The plot is useful to reveal a relationship between the differences and the sample averages, to look for any systematic biases and to identify possible outliers.

The Bland and Altman plot may be used to assess the repeatability of a technique by comparing repeated measurements using one single method on a series of subjects. In this case, the graph can also be used to check whether the variability or precision of a method is related to the size of the characteristic being measured. The original methodology of the Bland and Altman plot[44] uses the differences plotted against the mean difference between measures (left and right eyes, in this case) but they also proposed[85] two other methodologies, more useful when there is an increase in variability of the differences as the magnitude of the measurement increases: to plot differences as percentage of averages between measures or to plot ratios instead of differences. This last methodology will be used in this study, as there are no zero values.

The **Youden Plot** is a graphical method to analyse inter-laboratory data, where all laboratories have analysed 2 samples. The plot visualises within-laboratory variability, as

well as between-laboratory variability. In this case, left and right eyes can be considered as two different laboratories, where data are correlated.

For the original Youden plot[86], the two samples must be similar and reasonably close in magnitude of the evaluated property, since the axes in this plot are drawn on the same scale: one unit on the x-axis has the same length as one unit on the y-axis. This is a useful method to evaluate correlation, as well as differences between left and right eyes.

Each point in the plot corresponds to the results of one eye, and is defined by a first response variable on the horizontal axis and a second response variable 2 on the vertical axis.

A horizontal median line is drawn parallel to the x-axis so that there are as many points above the line as there are below it. The same is done for the y-axis. Note that outliers are not used in determining the position of the median lines. We can then define de Manhattan median as the intersection point of the two median lines.

A circle that should include 95% of the eyes is drawn, if individual systematic errors could be eliminated, and a 45º reference line is drawn through the Manhattan median.

Using this information, we can state that points that lie near the 45-degree reference line but far from the Manhattan median indicate large systematic error, and that points that lie far from the 45º line indicate large random error. Points outside the circle indicate a large total error.

It seems then that the Youden plot could be more likely a graphical interpretation of the concordance correlation coefficient.

Adapting what is necessary, we can use different laboratories as the eyes (right and left), and the two samples are the two types of subjects (patients and controls); we can visualize variability within subjects of each group and between groups. The measures used are right and left eyes.

To exemplify, we have randomly selected one patient and one measure from the OCT. Let A=(a_1;a_2) be the point that represents the values obtained respectively for the right (RE) and left (LE) eyes for the Nasal Superior region on the Retinal Nerve Fiber Layer measurements, for that given subject, and let the colour define if that subject is a control (blue) or a patient (red). This point A has coordinates (89; 113) and can be plotted on an orthonormal referential which as its origin (O) at the median values for the right and left eyes (the Manhathan's Median), for that variable, that is, O = (99; 111). The 45º straight is expressed generally by $r: y = mx + b, m \neq 0$, with $m = tg(45º) = 1$ and $b = Median_{LE} -$

$Median_{RE}$, hence the equation of the straight is $r: y = x + 12$ (Figure 16.A).

The Euclidian distance from the Manhathan's Median to A is the total error and is given by $\|\overrightarrow{OA}\| = \sqrt{(89-99)^2 + (113-111)^2} = 10{,}20$. As it is known, Total Error is the sum of random and systematic error, thus we need to determine the components of those errors (Figure 16.B).

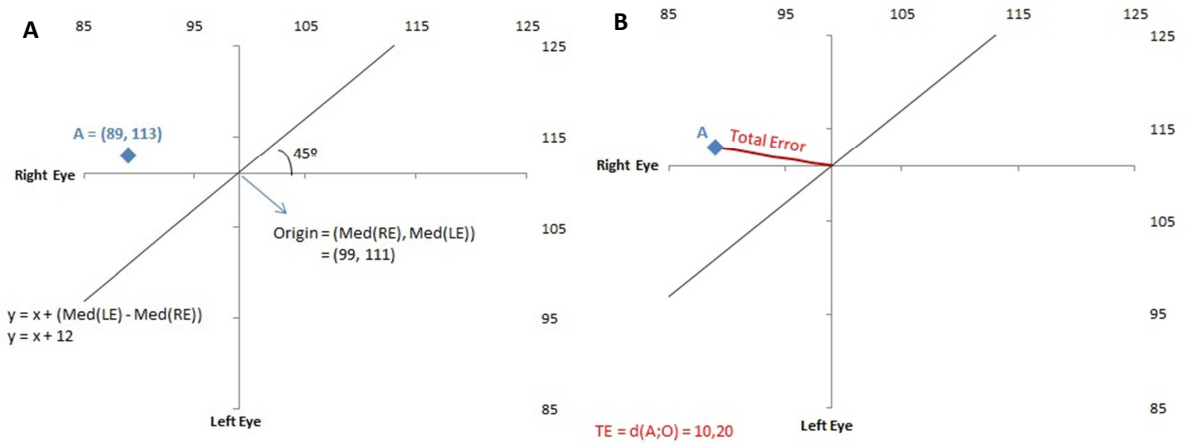

Figure 16 – Construction of a Youden plot for measurements performed in both eyes of the same subjects (A) and determination of the Total error of measurement between eyes (B).

The Random component of the error is given by the minimal distance from A to the 45º straight, let's say, the distance from *A* to *I*, with *I* being the interception point of the 45º straight $(y = x + 12)$ and a normal to this straight (let's say s) containing the point *A* $(s: y = -\frac{1}{m}x + b)$. This straight is given by $s: y = -x + (a_1 + a_2)$ and the Interception point *I* has general coordinates $I = \left(\frac{a_1 + a_2 + Median_{RE} - Median_{LE}}{2}; \frac{a_1 + a_2 - Median_{RE} + Median_{LE}}{2}\right)$.

For this example we have $s: y = -x + 202$ and $I = (95; 107)$ so the Euclidian distance from *I* to *A*, $\|\overrightarrow{IA}\| = \sqrt{(89-95)^2 + (113-107)^2} = 8{,}49$, represents the Random component of the error since it is the deviation from the point that is equally spaced from the Origin and so, it belongs to the 45º straight containing the Manhathan's Median (Figure 17.A). The distance from *I* to the origin (*O*) represents how far the point is from the median, if there was no random error, thus representing the systematic component of the error and is given by $\|\overrightarrow{OI}\| = \sqrt{(99-95)^2 + (113-111)^2} = 5{,}66$ (Figure 17.B).

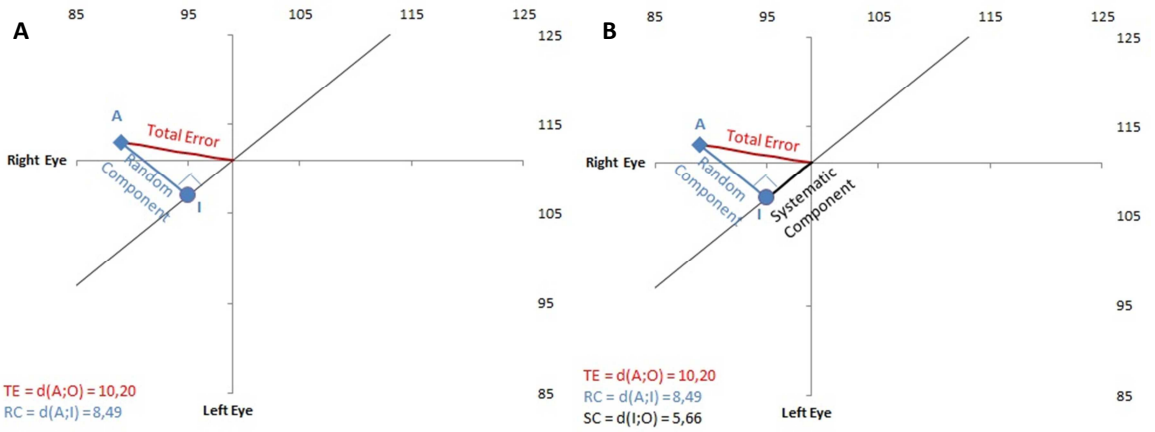

Figure 17 – Determination of the Random (A) and systematic (B) components of the error.

Since the Total Error is the sum of random and systematic errors, and as $\overrightarrow{OI} + \overrightarrow{IA} = \overrightarrow{OA}$, the projection of these components given by $\overrightarrow{OI}$ and $\overrightarrow{IA}$ onto the vector $\overrightarrow{OA}$ can be expressed as a percentage of the magnitude of the total error, as observed in Figure 18:

$$Random\ Error\ (RE) = \frac{\overrightarrow{IA}}{\overrightarrow{OI} + \overrightarrow{IA}} \times \left\|\overrightarrow{OA}\right\|$$

$$Systematic\ Error\ (SE) = \frac{\overrightarrow{OI}}{\overrightarrow{OI} + \overrightarrow{IA}} \times \left\|\overrightarrow{OA}\right\|$$

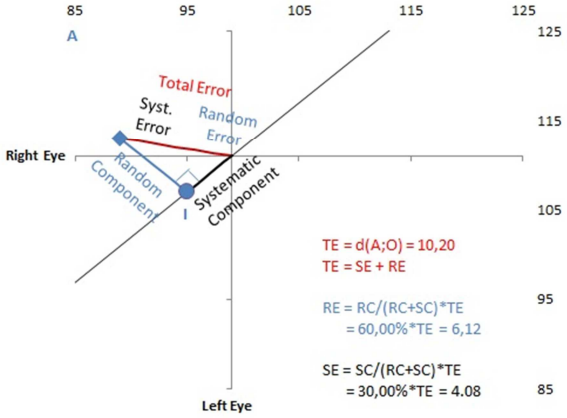

Figure 18 – Determination of the Random and systematic errors.

If the process is repeated for all the subjects in the sample, and if we sum the n squares of the random errors, we have a Total Variance in Random Error and the square root of the Variance, divided by n-1, represents the standard error of the mean random error, that is the standard error of the median values of right and left eyes for the Nasal Superior region measurements obtained by RNFL: $SEM_{RE} = \sqrt{\frac{\sum_{i=1}^{n} RE^2}{n-1}}$. It is then possible to determine a 95% confidence interval for the mean random error as the set of points that are at the same distance of the Manhattan Median, given by the circumference with centre in the Manhattan Median and radius $t_{0.975;n-1} \times SEM_{RE}$ (Figure 19) For this example, the radius of the circle should be 21.58 that is, the 95% confidence interval for the mean random error follows the condition $(x - 99)^2 + (y - 111)^2 = 21{,}58^2$, if all the systematic errors could

be eliminated.

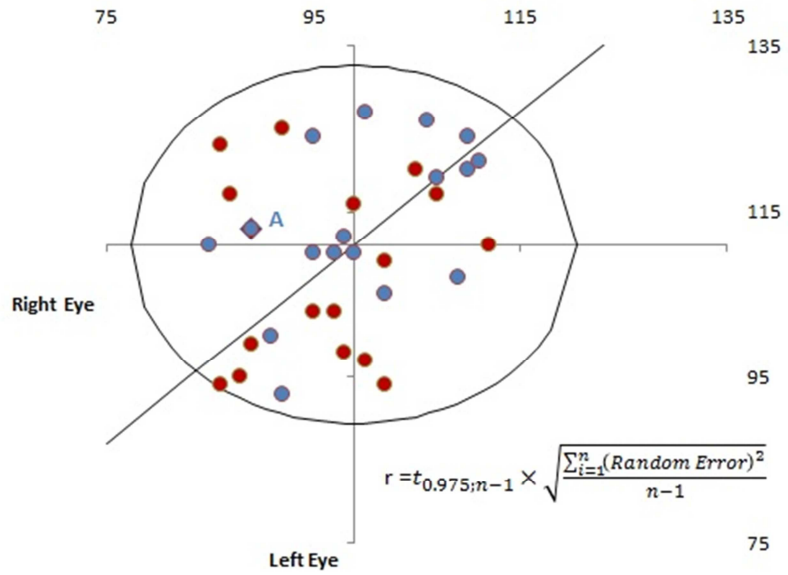

Figure 19 – Youden plot for measurements performed in both eyes of the same subjects.

As the Youden plot is based on medians of each eye and Bland and Altman plots depend on the mean differences between eyes $\pm$ 1.96 standard deviations, which suggests normal distributed data, we preferred the Youden plot method.

However, we used Youden plots on another setting. For each variable obtained with OCT, we obtained Youden plots for controls and type 2 diabetics in order to evaluate random error of each group, and compare them. Random errors, for each one of the variables studied, in each group, are Gaussian, thus the squared radius of circles in an Youden plot follow a Chi-square distribution and if we use the ratio of the squared radius of patients and controls, we obtain a F distribution, since other values of the circles radius are constant. The number of degrees of freedom depends on the number of pair of eyes in analysis. Then, the right-sided probability associated to that F distribution is the p-value for the comparison of random errors in measurement for the pairs of eyes between groups, and we may establish which group is more likely to present higher dispersion on results, and lower concordance between eyes, since systematic errors should be similar between groups.

This procedure was implemented just between controls and type 2 diabetics, and not between diabetic subjects with and without diabetic retinopathy, since it is expected that these two latest subgroups present less variability between them, as the systemic disease is present in both groups.

This procedure was developed with Microsoft Excel, and all the other methods referred were computed by MedCalc software (version 9.2.0.1, Frank Shoonjans, 1993-2006), and were evaluated at the significant level of 0.05.

5.2 Computing a global measure for data obtained from each meridian in psychophysical tests

Psychophysical tests (speed, achromatic and chromatic sensitivity measured in Protan, Tritam and Deutan axes) were measured in 4 different directions, or meridians, according to a given degree: at 0º, 45º, 90º and 135º. In order to have a global measure for each component (speed, achromatic and chromatic sensitivity), we plotted each one of the measures in a polar coordinate referential, obtaining four points, so that each point would have coordinates (ρ,θ) where ρ is the value obtained for the test at the meridian with θ degrees.

For instance, if a given subject has the values of 0.88, 0.83, 0.82 and 0.90 in the speed test for the meridians 0º, 45º, 90º and 135º, respectively, we can plot these values on a polar coordinate system, obtaining Figure 20:

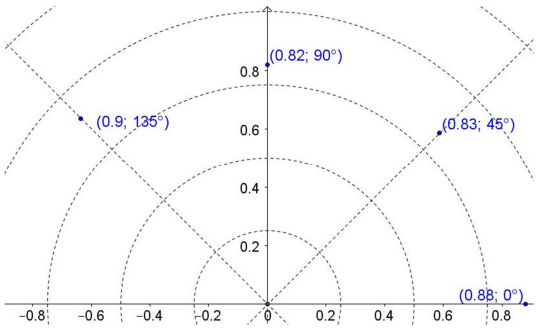

Figure 20 – values measured for the Speed test (º/s) for each one of the meridians, plotted in a polar referential.

If we join the point with line segments, passing at the origin of the referential, in order to obtain a polygon which area rises whenever a value in one of the meridian rises, we can trace a five-sided polygon like the one presented in the Figure 21:

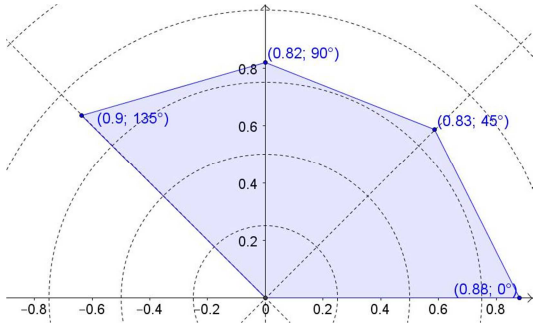

Figure 21 – 5-sided polygon obtained by joining the measure obtained for each meridian of the speed test (º/s), which represents the vertices, and the origin.

We may determine the area of those 5-sided polygons using the following theorem[87]:

Let P be a simple polygon with n positive oriented vertices $v_i, i = \overline{0, n-1}$ such as $P = (v_0, v_1, \dots, v_{n.1})$ since $v_n = v_0$. Let p be any point in the plan. Then, if $v_i = (x_i . y_i)$ for $i = \overline{0, n-1}$, in cartesian coordinates, then

$$A(P) = \frac{1}{2}\sum_{i=0}^{n-1}(x_i \times y_{i+1} - y_i \times x_{i+1})$$

Thus, after transforming values on polar coordinates to Cartesian coordinates, it is possible to have a global measure of each contrast sensitivity test.

Cartesian coordinates are determined using the classical formula:

$$\begin{cases} x = \rho sen(\Theta) \\ y = \rho cos(\Theta) \end{cases}$$

Where ρ is the radius corresponding to the value measured in each meridian, and Θ is the angle or the meridian for which the radius ρ was obtained.

Polygon figures were designed using GeoGebra – Dynamic Mathematics for Everyone, version 4.4, a free package developed by the International GeoGebra Institute in Austria.

5.3 Data reduction for classification

At the beginning, we had one hundred variables in analysis for diabetes classification, and 103 variables in analysis for diabetic retinopathy classification, as described in chapter 4. For the first goal, 96 cases were studied, and for the second aim we had 40 cases available to study on the training sample.

This was the first problem since multivariate data analysis requires more cases than variables. It is methodologically incorrect to study more variables at once than cases available.

Therefore, we started to reduce the number of variables in the analysis, by performing an univariate analysis for each variable in order to identify which variables could differentiate either diabetes presence or diabetic retinopathy presence. This was done applying an independent Student's t test or a Mann-Whitney test to each one of the variables, according to its distribution fit to a normal, considering type (control or type 2 diabetic) or considering ETDRS grading divided into two categories (DR absent or DR present) as the independent variable. The adjustment to normal distribution was performed by the Kolmogorov-Smirnov test with Lillefors correction whenever we had at least 25 cases in the group and by the

Shapiro-Wilk test otherwise. We decided to use two independent variables instead of one variable with three levels (control / diabetic without DR / diabetic with DR) since there were some variables measured only for the diabetic group, as duration of the disease, and ETDRS grading.

Those tests were performed through the Statistical Package for Social Sciences software (SPSS), version 20.0 (IBM Corporation, 1989-2011), and were analysed at a 5% significant level, although some graphics were obtained using STATISTICA, version 10, from the StatSoft Inc., 1984-2011, or using Microsoft Excel 2007.

After identifying which variables were significantly different between groups, we performed a Receiver Operating Characteristic Curve Analysis (ROC) in order to identify which variables could be used as binary univariate classifiers for the presence of diabetes and for the presence of diabetic retinopathy in the diabetic group. This was performed using SPSS, version 20.0 (IBM Corporation, 1989-2011) and using the MedCalc software (version 12.7.2.0, Frank Shoonjans, 1993-2013), particularly whenever was necessary to compare ROC curves for different variables, under no specific underlying distribution.

The **ROC curve** was firstly used in the signal detection theory, during the Second World War with the intention of analysing radar signals, especially after the attack on Pearl Harbour in 1941. Afterwards, ROC curves were applied in psychophysical to access human detection of weak signals and, nowadays, are extensively used in medicine, to evaluate diagnostic tests or in epidemiology and medical research simultaneously with evidence-based medicine. In radiology, it is a common method to evaluate and judge the accuracy of new radiology techniques.

It is a graphical procedure which plots the true positive rate, on the Y-axis of a Cartesian referential, to the false positive rate, on the X-axis of that referential, at different threshold settings, and may be used to illustrate the performance of a binary classifier. The maximum possible area obtained is 1, since both axes vary between 0 and 1, forming a square with an area of 1. If the area is 0.5, then the test has no discriminant power, since the true positive rate equals the false positive rate so, thus the performed test intends to ask the question: is the area under the ROC curve significantly different (higher) than 0.50?

By doing this, ROC analysis provides tools to select optimal models and discard the others, by using a 2x2 contingency table based on the number of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) cases, for each possible cut-off point, as follows (Table 7):

Table 7 - General 2x2 contingency table used for ROC analysis

		Test Result		Total
		Negative (T^-)	Positive (T^+)	
True condition (Gold Standard)	Disease Absent ($\bar{D}$)	TN	FP	$n_{\bar{D}}$
	Disease Present (D)	FN	TP	n_D
Total		n_{T^-}	n_{T^+}	n

Some efforts have been made to use ROC curves in problems with more than two groups, but they are still very complex, and do not apply to the goals of this study. However, for the three group problem, the intention is to create a volume function which may represent the accuracy of that variable.

We determined sensitivity (Sens), specificity (Spec) and the positive likelihood ratio (PLR) for variables that presented accuracy as univariate classifiers at the optimal cut-off, found by the determination of the maximum Youden Index (Y) which was calculated for different possible thresholds using the formula$Youden = Sensitivity + Specificity - 1$.

Sensitivity is the probability of getting a positive test result (T^+) in subjects with the disease (D), being computed as $Sensitivity = P(T^+|D) = \frac{TP}{n_D} = \frac{TP}{TP+FN}$, hence, it is related to the potential of a test to recognise subjects with the disease and discard with more certainty the presence of the condition since a test with high sensitivity is a test with few false negative results. Then, if a result is negative, it is most certainly a true negative case, and high sensitivity tests are usually applied to discard the presence of the conditions. However, if the test turns out to be positive, then there is a suspicious of the presence of the disease, and a more specific test must be performed. Therefore, tests with high sensitivity are used for screening of diseases.

Specificity is a measure that is complementary to sensitivity, since it is defined as the proportion of subjects with negative results in the test within controls, that is, $Specificity = P(T^-|\bar{D}) = \frac{TN}{n_{\bar{D}}} = \frac{TN}{TN+FP}$ and it is related to the test ability to exclude the condition of interest. However, we should note that a test with high specificity is a test with few false positive cases so, if a test returns a positive result, it is most certainly a true positive case and high specificity tests are usually applied to confirm the presence of the disease.

Neither sensitivity nor specificity is influenced by disease prevalence, so these parameters may be transposable for other populations. Unlike these measures, **predictive values** are dependant of the disease prevalence in the population and, therefore, predictive values

obtained in one study should not be used in other settings with different disease prevalence. The positive predictive value increases and negative predictive values decreases as the prevalence of the disease increases.

The positive predictive value (PPV) is defined as the probability of having the condition or disease in subjects which had a positive value in the test, whereas the negative predictive value (NPV) refers to the probability of being healthy or having the condition absent when the test result is negative. However, predictive values are derived from sensitivity and specificity using the Bays theorem. If we think of the prevalence, as the prior probability of a given subject having the condition, or disease, predictive values may be thought as posterior probabilities for that subject to have the disease, after knowing the test result. Generally, predictive values are defined as:

$$PPV = P(D|T^+) = \frac{TP}{n_{T+}} = \frac{TP}{TP+FP}$$

$$NPV = P(\overline{D}|T^-) = \frac{TN}{n_{T-}} = \frac{TN}{TN+FN}$$

If we want a measure of diagnostic accuracy not dependant from the prevalence of the disease, then it is useful to determine the **Likelihood Ratios**, since they reflect the link of the pre-test and post-test probability of a disease in a certain patient, as being defined as the ratio of the expected test result in subjects with a certain condition to subjects without the disease, quantifying how many times it is more likely that a test result is positive in subjects with the disease than in those without the condition. If both probabilities are equal, likelihood ratio is 1 and that test has no accuracy. We may then define Positive likelihood ratio (PLR) and Negative likelihood ratio (NLR) as:

$$PLR = \frac{P(T^+|D)}{P(T^+|\overline{D})} = \frac{Sensitivity}{1-Specificity}$$

$$NLR = \frac{P(T^-|D)}{P(T^-|\overline{D})} = \frac{1-Sensitivity}{Specificity}$$

Observing these formulas, the positive likelihood ratio (PLR) reflects how much more likely is that a test returns positive in patients with the disease than in patients without the disease and is, usually, higher than 1, being the best indicator for ruling-in diagnose. On the other hand, the negative likelihood ratio (NLR) represents the ratio of the probability that a negative test result would occur in subjects with the disease, to the probability that the same result would occur in a control subject, that is, how much less likely is that a test turns negative in a patient than in a subject without the disease and is usually less than 1.

5.4 Statistical classification

Statistical classification was performed using variables that presented statistical significant differences between groups, either concerning the absence or presence of diabetes, or concerning the absence or presence of diabetic retinopathy in the type 2 diabetic group.

For the first objective which intended to find a classification function or classification algorithm that could separate controls from type 2 diabetics, we used the variables collected on the:

- Subject
  - Age;
  - Body mass index;
  - Absence/Presence of diagnosed hypertension (blood pressure controlled by medication);
- Blood Samples
  - related to the liver and billiar ductus: ALT, alkaline phosphatase and Gamma GT;
  - related to lipids: cholesterol (total, HDL, LDL), atherogenic index, triglycerides and apolipoprotein A1;
  - cytometry parameters: leucocytes, haemoglobin, haematocrit and erythrocytes variation coefficient;
  - Hormonology: Peptide C;
- Eye
  - Best Corrected Visual Acuity;
  - Retinal Nerve Fiber Layer obtained with OCT: temporal quadrant;
  - Visual Psychophysical Tests
    - Speed test: all meridians and also the global area;
    - Achromatic contrast sensitivity: meridian 0º;
    - Chromatic contrast sensitivity
      - Protan: meridian 0º;
      - Deutan: meridian 0º and meridian 45º;
      - Tritan: all meridians (0º, 45º, 90º and 135º);

For the second objective which was to find a classifier for diabetic retinopathy, the variables used were:

- Subject
  - Duration of the disease;

- Blood Samples
  - cytometry parameters: erythrocytes, haemoglobin, and haematocrit;
- Eye
  - Volume Scan obtained by OCT: Inner Nasal region;
  - Visual Psychophysical Tests
    - Chromatic contrast sensitivity
      - Deutan: meridian 0º;
      - Tritan: meridian 0º, meridian 135º, and global area;

We performed classification using three different statistical methods, described later. Whatever the methodology used, for all classifiers obtained we determined group prediction based upon posterior probability for the presence of the condition (either presence of diabetes, in the first objective, or the presence of diabetic retinopathy, in the second purpose), using equal prior probabilities. Two of those methods were able to be used considering different prior probabilities. For instance, we could have used the prevalence of diabetes in the Portuguese population and the presence of diabetic retinopathy in the Diabetic Portuguese population as prior probabilities, but then we could not compare statistical classifiers since not all of them are able to consider different prior probabilities. On the other hand, by using equal prior probabilities, we are able to generalise statistical classifiers obtained for other populations or for changes on those values of prevalence.

The accuracy of classifiers obtained was evaluated comparing the area under the ROC curves drawn for them, using MedCalc software (version 12.7.2.0, Frank Shoonjans, 1993-2013) through the methodology of DeLong[88] for the calculation of the standard error of the area under the curve (AUC) and for the difference between AUC's and, consequently, the p-values obtained (which were considered to be statistical significant if lower than 0,05), and with the determination of binomial exact confidence intervals for the AUC. Sensitivity, specificity and positive likelihood ratios were determined for the cut-off value of each classifier. We are mostly interested in getting a classifier with maximum positive likelihood ratio, despite the negative likelihood ratio, so that it can be used for screening.

All the classifiers that presented a good performance on predictions were tested in a new sample, the test sample, described on chapter 3, and once more its performance was compared in this sample. This procedure could be done for type 2 diabetes classifiers, but not for the obtained diabetic retinopathy classifiers, since they could only be applied to 5

cases of the test sample, all without diabetic retinopathy.

For the best statistical classifier obtained, either for type 2 diabetes, or for diabetic retinopathy, positive and negative predictive values were determined according to the respective disease prevalence.

5.4.1 Development of the statistical classifiers

Group classification was performed under three different statistical methodologies: discriminant analysis, logistic regression and decision trees. Each one was applied to determine the posterior probability for the presence of type 2 diabetes, or for presence of diabetic retinopathy in the type 2 diabetes group, according to the aim that is being considered.

Discriminant analysis and Logistic regression methods were obtained using SPSS, version 20.0 (IBM Corporation, 1989-2011), using a forward stepwise procedure, hence that each variable entering in the model would reflect the variable with more classification accuracy, within the group of variables left to enter. SPSS uses a general forward stepwise method on discriminant analysis, based on the probability of the F test (the variable is included if the model improves with a p-value smaller than 0.05 and the variable is excluded if that probability is higher than 0.10). Regarding logistic regression the forward stepwise method may be based upon the same general procedure, or based on the Likelihood Ratio or on the Wall Statistic. The best of these models of logistic regression was chosen to continue in analysis.

Decision tree analysis was performed in STATISTICA (version 10, StatSoft Inc., 1984-2011) using the CART algorithm, the CHAID and Exhaustive CHAID algorithms and the QUEST algorithm. The best of models obtained was chosen to continue in analysis.

The prior advantage of logistic regression models and decision tree analysis to discriminant analysis is that there are no prior assumptions in terms of distribution of the sample, but discriminant analysis may turn more powerful if those assumptions are met.

5.4.1.1 Discriminant analysis

As the intended classifiers are binary, only one discriminant function was obtained so the model is the Fisher's Linear Discriminant Function, using the Wilk's lambda test for stepwise analysis, using the SPSS package.

The discriminant function (and standardized discriminant function) was obtained in order to write the model:

$$F(x_1, x_2, \dots x_n) = b_0 + \sum_{i=1}^{p} b_i \times x_i$$

$$\left(standardized\ function: F^*(x_1, x_2, \dots x_n) = \sum_{i=1}^{p} \beta_i \times x_i\right)$$

where p is the number of variables of the model and the B matrix (1xp) is estimated in order to maximize the variability of the scores of the discriminant function between groups and minimize it within the groups, that is, in order to maximize:

$$\lambda = \frac{SSF(D)}{SSE(D)}$$

The classification of new and old cases may be performed using the closeness to group centroid, which may be done by dividing the discriminant function in two mutually exclusive subspaces separated by the frontier line defined previously, in Equation 4, as

$$f = \frac{n_1 \bar{d}_1 + n_2 \bar{d}_2}{n_1 + n_2},$$

where $\bar{d}_1$ and $\bar{d}_2$ are the centroids for groups 1 and 2, respectively, and n_1 and n_2 the number of cases in each group.

The accuracy of classifications was tested on the training sample by ROC analysis, and posterior probability function was developed in order to classify any new subject.

Hence, for each one of the subjects in the training sample (x_i) without missing values on the variables identified to belong to the discriminant function, the value obtained in that function ($f_i = f(x_i)$) was normalized in order to determine its Mahalanobis distance to each group centroid $\left(\overline{d_j},\ j = 1,2\right)$, considering the variance of the discriminant function for each group ($s^2_{f_j}$), using the equation:

$$d_{ij} = d\left(f_i, \overline{d_j}\right) = \sqrt{\frac{\left(f_i - \overline{d_j}\right)^2}{s^2_{f_j}}}\ , j = 1, 2$$

For each subject, two distances were obtained. For each group (G_j), we know that the squared distances (D^2) follow a chi-square distribution with one degree of freedom ($D^2 \sim \chi^2_{(1)}$) so, we me may calculate the probability of obtaining that distance, given that the subject is classified as belonging to the group j, as:

$$P(D|G_j) = P(\chi^2_{(1)} > D_j^2)$$

Furthermore, by the Bayes theorem, we may obtain posterior probabilities for each group classification, using equal prior probabilities:

$$P(G_j|D) = \frac{0{,}5 \times P(D|G_j)}{\sum_{j=1}^{2} 0{,}5 \times P(D|G_j)}, j = 1{,}2$$

For classification purposes, the subject is classified as belonging to a group if that subject is on the left side or the right side of the frontier line, but we are also interested in knowing the posterior probability of belonging to the disease condition (which will be higher than 50% in the group where the subject is classified into). We applied a ROC analysis to evaluate if there was a better frontier line for classification, and corresponding posterior probability.
These probabilities either on the training sample or for new subjects may be automatically computed in a worksheet designed for the effect with Microsoft excel.

Assumptions of discriminant analysis were evaluated either for the provenience of the samples from a multivariate normal distribution, or for the homogeneity of covariance matrices between groups. The first assumption was evaluated applying the Kolmogorov-Smirnov test when the number of cases was at least 25, or the Shapiro-Wilk test when the number of cases was below 25, assuming that we had a multivariate normal distribution if all the variables is both group were normally distributed. If this assumption is violated, discriminant analysis may be performed, according to Sharma[51], since it does not affect type I error; it may affect type 2 error (and consequently, power) and rates of misclassification, in case of small samples.
Homogeneity of covariance matrices was evaluated through the Box's M test. The violation of this assumption may affect the type I error if groups do not have identical dimensions[51]. Two groups are said to have identical dimensions if the rate between the size of the biggest group and the size of the smallest group is less than 1.5, which is the case that is being studied, either for classification of type 2 diabetes, or classification of diabetic retinopathy.

5.4.1.2 Logistic regression analysis

The advantage of logistic regression to discriminant analysis is the lack of assumptions about normality and homogeneity of variance matrices. Even though, logistic regression performs better if independent variables are dicotomic, when compared to logistic regression models

that use quantitative independent variables. Therefore, variables identified as possible discriminators of the state being studied were dichotomized according to the cut-offs obtained by the application of ROC analysis, when data reduction was being performed.

We used a forward stepwise method based on the probability of F distribution to enter a variable (< 0.05) or to remove a variable (> 0.10), and based on the Likelihood Ratio and on the Wall Statistic. We evaluated the adjustment of the model to data with the Hosmer and Lemeshow test, and the quality of the regression by the Nagelkerke R^2. Is some overfitting was detected, either by the excellent values obtained in the two methods referred for regression evaluation, or because of the lack of significance on regression coefficients in a given step after they were significant at 5% level, the models were evaluated and some iterations were discarded, even if we had to lose some variables of interest.

For this analysis, the probability of a given subject to have the condition of interest (coded with the value 1), as:

$$P(Y=1|x)=\frac{e^{b_0+\sum_{i=1}^{k} b_i \times x_i}}{1+e^{b_0+\sum_{i=1}^{k} b_i \times x_i}} \left(or\ P(Y=1|x)=\frac{e^{b_0} \times \prod_{i=1}^{k} (e^{b_i})^{x_i}}{1+e^{b_0} \times \prod_{i=1}^{k} (e^{b_i})^{x_i}} \right)$$

And the subject is classified as having the condition of interest if that probability is higher than 0.50. We performed ROC analysis on these probabilities to detect if there was a better cut-off for that probability.

5.4.1.3 Decision Tree analysis

Decision trees, as logistic regression, are widely used for its lack of assumptions on data distribution. Nowadays, the most popular algorithms for decision trees are the CART[55], the CHAID[56] or Exhaustive CHAID algorithms, and the QUEST algorithm[57], thus those were the algorithms that we applied to grow the trees. When it was possible to define, we used equal prior probabilities so that algorithms could be compared, as well as with discriminant analysis results. On the other hand, decision tree may identify cut-offs for variables, in a multivariate context, which may correspond, or not, to those previously identified by ROC analysis.

We used the quantitative variables, as in the discriminant analysis procedure, so that we could identify cut-offs for all identified variables in the model. Posterior probabilities were calculated based on the ratio of cases for each group in nodes composition.

5.4.2 Testing developed statistical classifiers

Accuracy of developed models was compared using ROC analysis applied to the posterior probabilities. For group classifications, we used the Kappa coefficient of concordance to evaluate the percentage of correct decisions which is not due to the chance, and applied the McNemar test in order to evaluate if ratios of incorrect decisions were in the same proportion, that is, if classifiers had equal ratios of false positive and false negative decisions. For type 2 diabetes classifiers, it was possible to evaluate classifiers on a test sample (new cases) but for diabetic retinopathy classifiers, at the moment, it was not possible to evaluate them in new cases.

Predictive values for the best classifiers, adjusted for disease prevalence, were determined.

CHAPTER 6

RESULTS

SECTION A

CORRELATION BETWEEN EYES

1. Evaluation of recommendations found in the Literature

At this study, the outcome measures depend in general on individuals as units of measurement, and several parameters were collected, such as laboratory findings, liver, heart and brain images, visual psychophysical test measures for the dominant eye, and ophthalmological parameters obtained by OCT for both eyes.

As findings will report to individuals and not to each eye, it seems adequate to use only one eye, and the same collected on visual tests, the dominant eye. However, we must evaluate correlation and differences between eyes collected for OCT. since we may be wasting useful data.

2. Correlation among measurements

Figure 21 shows that correlation between the left and right eye are strong, but not too close to one, and that some statistical significant asymmetries[89] are detected on the RNFL at the Temporal quadrant, and at Nasal-Inferior and Nasal-Superior quadrants (Figure 22).

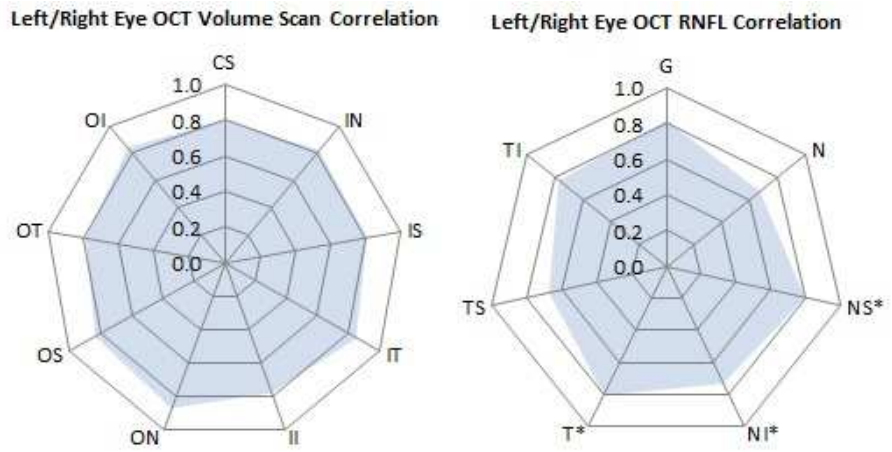

Figure 22 - Nonparametric Spearman Rank-Order Correlation Coefficient (*Statistical significant difference between eyes ($p < 0.05$) by Wilcoxon Matched-Pairs Test).

As correlation coefficients for measurements of OCT Volume Scan have values between 0.777 e 0.870 (respectively on the inner inferior and on the outer nasal quadrants) and for measurements of OCT RNFL are between 0.674 and 0.817 (respectively on the nasal quadrant and on the global measure), we can assume that there is a moderate to strong correlation between eyes, but not too close to one thus, according to Armstrong guidelines, measures of agreement between eyes should be used. On the other hand, no statistical significant differences between eyes were found in the Volume Scan, and few differences were found in RNFL (temporal, nasal-inferior and nasal-superior quadrants). In fact, estimates for the median difference between right and left eye, by the Hodges-Lehmann estimator, are +10.50 for the temporal quadrant, and -4.50 for the nasal-inferior quadrant and -3.00 for the nasal-superior quadrants.

3. Concordance among measurements

Armstrong guidelines suggest the use of measures of agreement such as the intra-class correlation coefficient (ICC) or the concordance correlation coefficient (CCC). As data do not follow a normal distribution, we used a pseudo concordance correlation coefficient (pCCC). These three measures of concordance are represented on Figure 23.

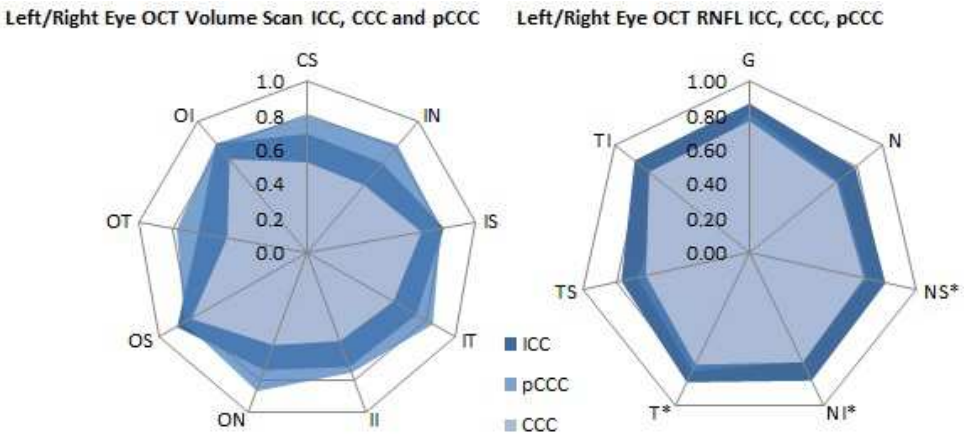

Figure 23 – Intra-class correlation coefficient, concordance correlation coefficient and pseudo-concordance correlation coefficient between left and right eyes on Volume Scan and RNFL. (*Statistical significant difference between eyes ($p < 0.05$) by Wilcoxon Matched-Pairs Test).

The following mountain plots represented on Figure 24, shows that usually the pCC is closer to the ICC than the CCC, separately for coefficients obtained for volume scan and for RNFL of OCT:

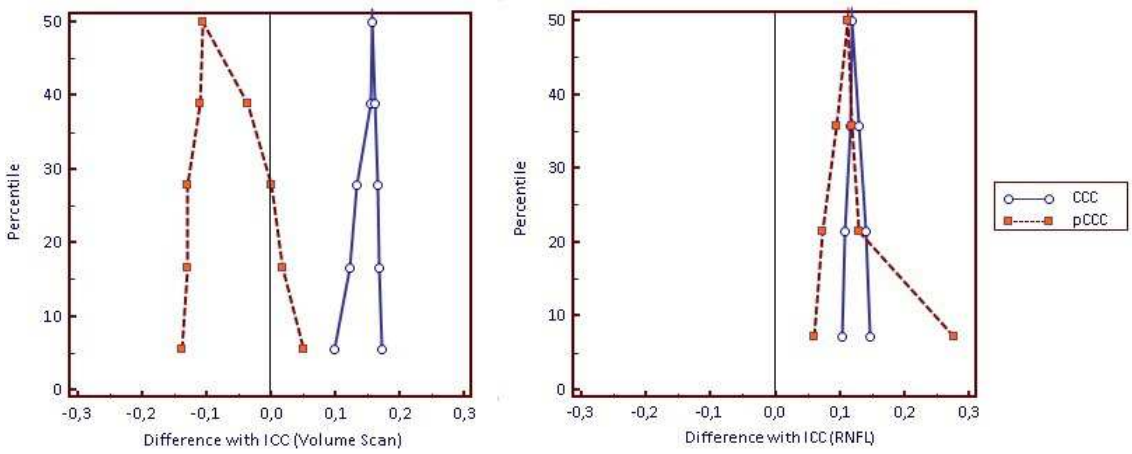

Figure 24 - Mountain plot for concordance correlation coefficient based on Pearson correlation coefficient (CCC) and on Spearman correlation coefficient (pCCC), compared to the Intra-class correlation coefficient separately for volume scan (A) and RNFL (B).

In fact, a mountain plot can show that almost all absolute differences between ICC and CCC, either on Volume Scan or RNFL lie between 0.1 and 0.2. In fact, only 11% of the absolute differences in volume scan are inferior to 0.1 and none in RNFL, while 43% of absolute differences between ICC and the pseudo-CCC in Volume Scan are inferior to 0.1 and 44% of the differences is RNFL are inferior to 0.1.

Comparing absolute differences between CCC and ICC (Δ_1) and CCC and pseudo-ICC (Δ_2), we found that ICC is closer to pseudo-CCC than to CCC in all the measures obtained for volume scan, and on 71.4% (5 out of the 7) regions of the RNFL analysed (table 8).

Table 8 - Comparison of measures of concordance.

						ICC vsersus CCC or Pseudo-CCC				CCC versus Pseudo-CCC			
			n	ICC	Measure	Δ_1	Z	p	Adj. p	Δ_2	Z	p	Adj. p
OCT – Volume Scan		Central Subfield CCC	98	0.690	0.524	0.166	1.834	0.067	0.200	-0.272	-3.486	< 0.001	0.001
		Central Subfield pCCC			0.796	**-0.106**	-1.652	0.099	0.296				
	Inner	Nasal CCC	98	0.679	0.512	0.167	1.804	0.071	0.213	-0.307	-4.055	< 0.001	0.000
		Nasal pCCC			0.819	**0.140**	-2.250	0.024	0.073				
		Superior CCC	98	0.809	0.677	0.132	2.072	0.038	0.115	-0.115	-1.745	0.081	0.243
		Superior pCCC			0.792	**0.017**	0.326	0.744	1.000				
		Temporal CCC	98	0.735	0.579	0.156	1.920	0.055	0.165	-0.267	-4.004	< 0.001	< 0.001
		Temporal pCCC			0.846	**0.111**	-2.084	0.037	0.111				
		Inferior CCC	98	0.715	0.554	0.161	1.883	0.060	0.179	-0.197	-2.420	0.016	0.047
		Inferior pCCC			0.751	**0.036**	-0.537	0.591	1.000				
	Outer	Nasal CCC	98	0.735	0.580	0.155	1.909	0.056	0.169	-0.286	-4.510	< 0.001	< 0.001
		Nasal pCCC			0.866	**0.131**	-2.601	0.009	0.028				
		Superior CCC	98	0.880	0.783	0.097	2.224	0.026	0.078	-0.047	-0.931	0.352	1.000
		Superior pCCC			0.830	**0.050**	1.293	0.196	0.588				
		Temporal CCC	98	0.647	0.476	0.171	1.739	0.082	0.246	-0.301	-3.584	< 0.001	0.001
		Temporal pCCC			0.777	**0.130**	-1.845	0.065	0.195				
		Inferior CCC	97	0.830	0.708	0.122	2.091	0.037	0.110	-0.122	-2.091	0.037	0.110
		Inferior pCCC			0.830	**0.000**	0.000	1.000	1.000				
OCT- RNLF		Global CCC	97	0.869	0.766	0.103	2.183	0.029	0.087	-0.044	-0.798	0.425	1.000
		Global pCCC			0.810	**0.059**	1.385	0.166	0.499				
	Nasal	Nasal CCC	97	0.790	0.651	0.139	2.018	0.044	0.131	-0.020	-0.244	0.808	1.000
		Nasal pCCC			0.671	**0.119**	1.775	0.076	0.228				
		Superior CCC	97	0.815	0.686	0.130	2.066	0.039	0.116	-0.017	-0.225	0.822	1.000
		Superior pCCC			0.703	**0.112**	1.841	0.066	0.197				
		Inferior CCC	96	0.835	0.717	**0.118**	2.066	0.039	0.116	0.011	0.152	0.879	1.000
		Inferior pCCC			0.706	0.129	2.218	0.027	0.080				
	Temporal	Temporal CCC	96	0.846	0.731	0.115	2.121	0.034	0.102	-0.043	-0.678	0.498	1.000
		Temporal pCCC			0.774	**0.072**	1.443	0.149	0.447				
	Temporal	Superior CCC	96	0.769	0.622	0.147	1.975	0.048	0.145	-0.052	-0.612	0.540	1.000
		Superior pCCC			0.674	**0.095**	1.363	0.173	0.519				
		Inferior CCC	96	0.250	0.141	**0.109**	0.774	0.439	1.000	-0.382	-2.990	0.003	0.008
		Inferior pCCC			0.523	0.273	-2.217	0.027	0.080				

As observed in Table 8, we can only find statistical significant difference between the concordance correlation coefficient based on Pearson's correlation coefficient and on Spearman's correlation coefficient, but not between each one of these and the intra-class correlation coefficient.

Whatever the concordance method used, we can assume that we can use the pseudo-concordance correlation coefficient and that both eyes measurements are concordant, besides being correlated.

4. Graphical evaluation of random errors between controls and type 2 diabetics as a measure of concordance and accuracy of data for analysis

Armstrong also suggests the use of Bland and Altman plots to evaluate concordance. However, data are not normally distributed and, consequently, the representation of mean differences and the interval of $\pm$ 1.96 standard deviations around the mean difference may not be the most appropriate. Therefore, we present Youden plots, which are centred on the median of each eye and compare the errors between measurements performed in both eyes.

Random errors between eyes are significantly higher in the diabetic group than in the control group for the inner-superior and inner-inferior subfields of the volume scan in OCT (Figure 25), but no other significant difference is found between groups in the random errors of measurements of both eyes, even on RNFL measures (Figure 26), meaning that the total error of measurement between eyes is similar between groups, with the exception of those regions on volume scan, since systematic error may be assumed to be constant.

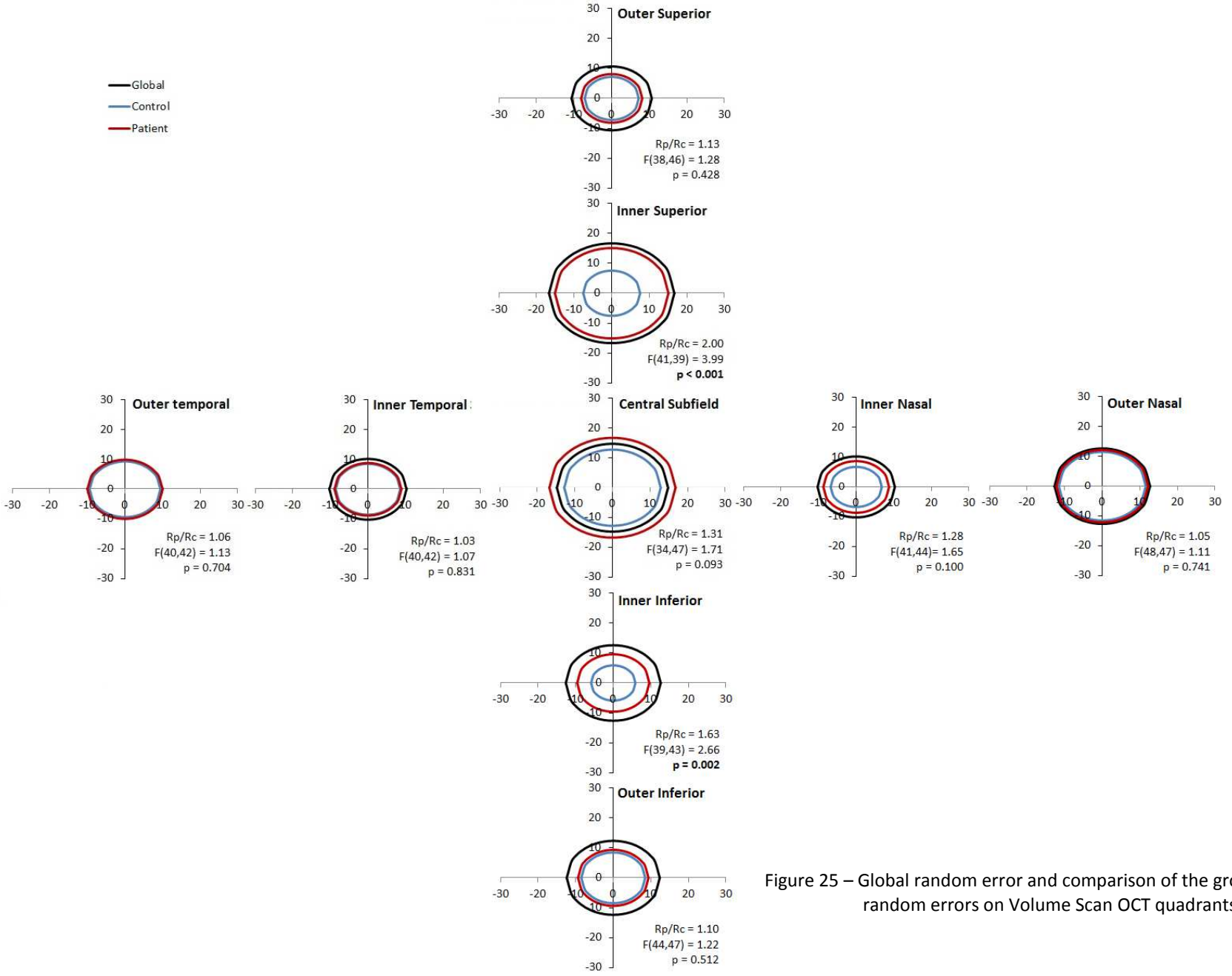

Figure 25 – Global random error and comparison of the group random errors on Volume Scan OCT quadrants.

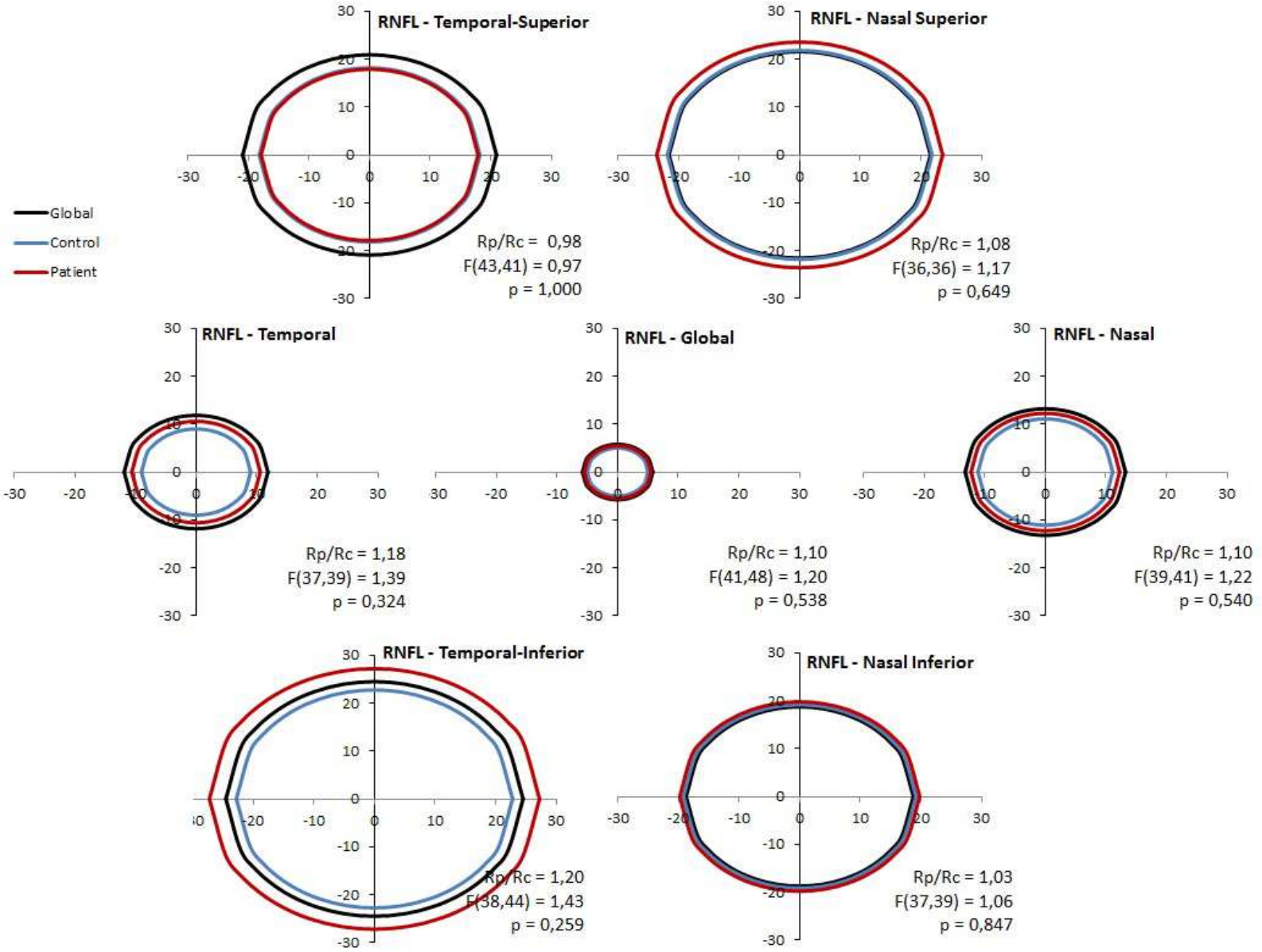

Figure 26 – Global random error and comparison of the group random errors on RNFL quadrants

SECTION B

STATISTICAL CLASSIFIERS FOR TYPE 2 DIABETES

1. Training sample description

For the training sample 96 subjects were studied, of which 49 (51.04%) were type 2 diabetics and 47 (48.96%) were controls for this disease (p = 0,919).

The dominant eye was chosen for analysis.

The best corrected visual acuity was measured in all the subjects and ranged between 0.20 and 1.30, with a mean of 0.92 ± 0.19, and at least 75% of the studied eyes had a minimum BCVA of 1.00.

Intraocular pressure was measured in 50 eyes and ranged between 8 and 23 mmHg, with a mean of 15.08 ± 3.56 mmHg, and 75% of the studied eyes had intraocular pressure below 18 mmHg.

Subjects were aged between 40 and 73 years at visit date, according to inclusion criteria, with a mean of 54.87 ± 9.34 years (Table 9).

Table 9 – Descriptive statistics on age and medical preliminary procedures measured in global sample.

	N	Min	Max	Mean	SD	P25	P50	P75
Age (visit)	96	40	73	54.88	9.35	47.00	54.00	62.00
Height (m)*	96	1.45	1.9	1.64	0.10	1.56	1.62	1.70
Weight (kg)*	96	45	115	74.65	14.23	63.15	73.70	84.78
BMI (kg/m^2)	96	18.5	43.7	27.86	4.89	24.40	26.95	31.05
AP (cm)*	54	71	140	100.28	13.44	91.75	100.00	109.25
Pulse (bpm)*	64	39	100	74.05	11.66	68.00	74.00	82.00
SBP (mmHg)	68	100	189	130.72	20.21	115.25	127.50	145.50
DBP (mmHg)*	68	46	100	75.37	10.66	69.25	76.50	82.00
Bioimpedance (%)*	40	12.3	61.9	35.32	10.44	27.30	34.10	41.23

* Normally distributed variables

BMI – Body mass index; AP – Abdominal perimeter; SBP – Systolic blood pressure; DBP – Diastolic blood pressure

The sample was homogeneous according to gender, eye dominance, previous family history of diabetes mellitus, and need for medication to control blood pressure, but, has expected, a predominance of the right hand for writing was observed, absence of previous gestational diabetes in women, and also a predominance of non-smokers, non-alcohol regular consumers or persons without regular exercise habits (Figure 27).

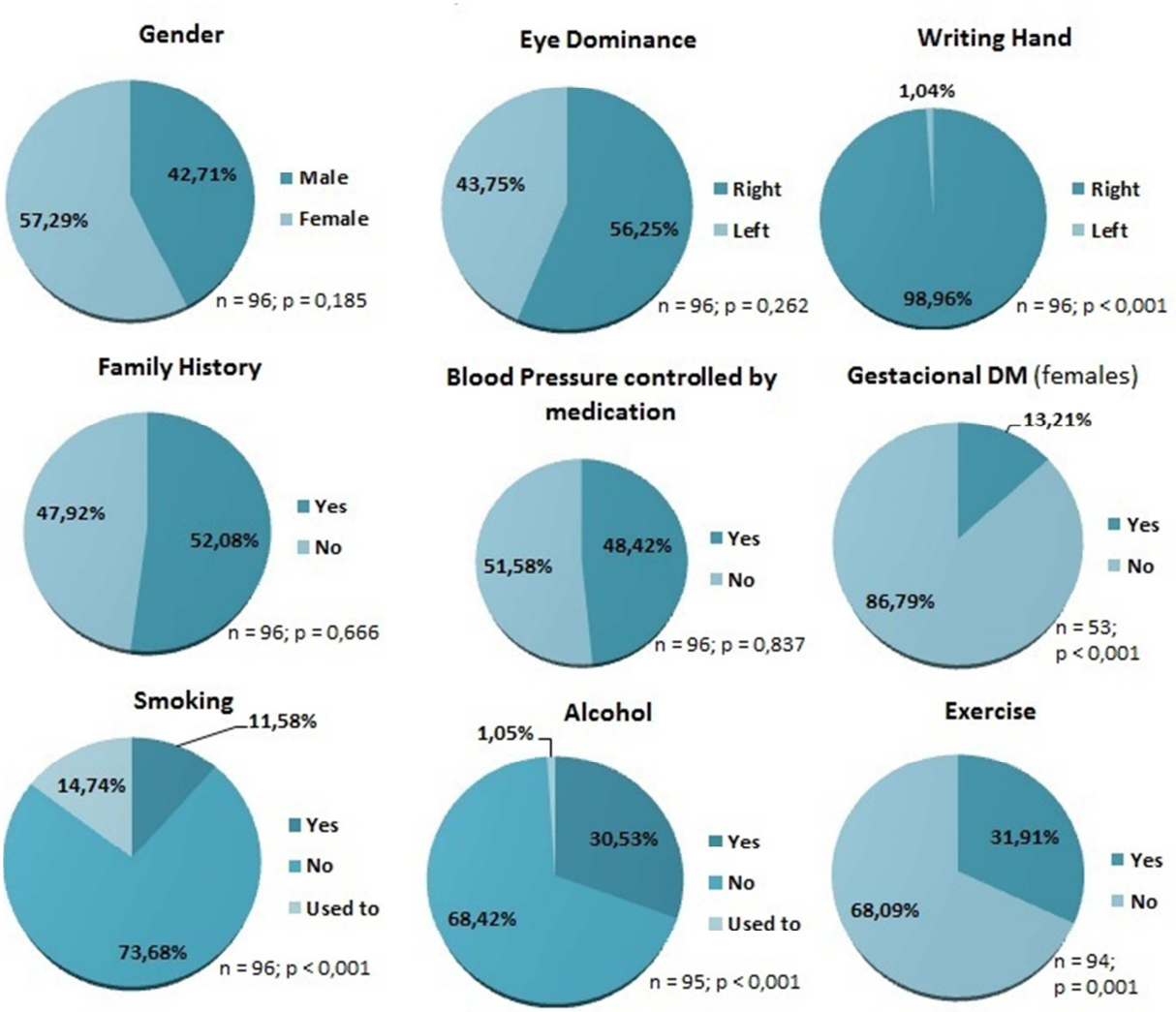

Fig. 27 – Distribution of sociodemographic charateristics (Binomial Test or Adjustement Chi-square test).

2. Variable reduction

2.1 Phase 1: Factors of differentiation in diabetes

2.1.1 Clinical and demographic assessment

There was no association between group type and characteristics such as gender, eye or hand dominance, and regular exercise practice but is more likely that patients with type 2 diabetes have previous history of diabetes in family and need medication to control blood pressure: it also seems that there exists a tendency for controls to be smokers and regular alcohol consumers than diabetics, although no statistical significant association is found at the 5% level. (Figure 28)

Figure 28 - Descriptive statistics and comparison of clinical and demographic measures assessed between controls and type 2 diabetics (Independence Chi-square; * Fisher exact test).

Groups were not matched for age (Figure 29), and statistical significant differences were found in height, weight, BMI and Systolic blood pressure, with controls being around 8 years younger and having 2 kg/m^2 less, in median, than patients (Table 10).

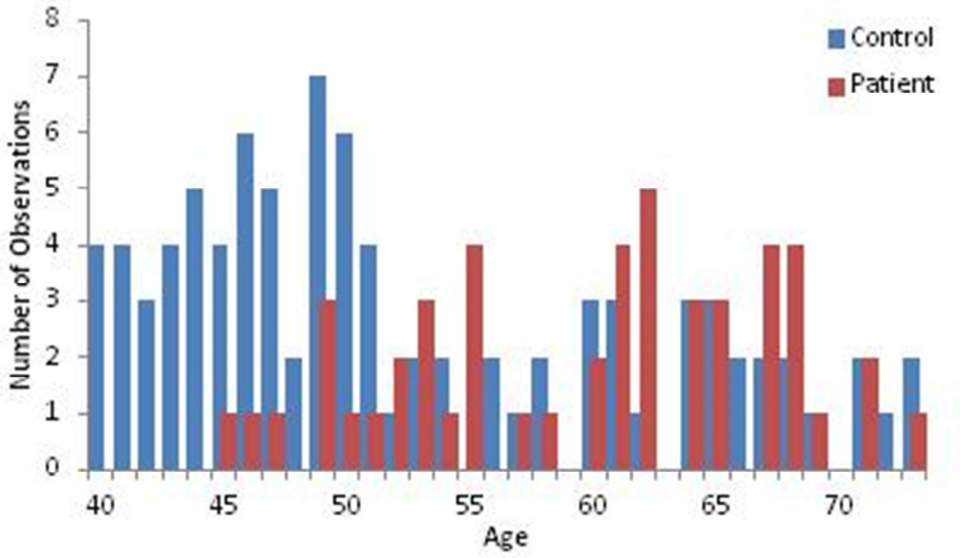

Figure 29 - age distribuion by group.

Table 10 - Descriptive statistics and group comparison between clinical and demographic variables measured between controls and type 2 diabetics.

	Type	N	Min	Max	Mean	SD	P25	P50	P75	p
Age (visit)	C	47	40.00	68.00	49.13	7.39	40.00	40.80	44.00	< 0.001**
	D	49	45.00	73.00	60.39	7.58	46.50	49.00	54.50	
Height (m)	C	47	1.50	1.90	1.66	0.10	1.53	1.55	1.58	0.006**
	D	49	1.45	1.87	1.62	0.10	1.45	1.50	1.53	
Weight (kg)	C	47	45.00	115.00	71.19	14.97	49.20	55.40	62.00	0.019*
	D	49	53.10	104.00	77.97	12.77	53.30	60.80	68.60	
BMI (kg/m^2)	C	47	18.50	33.40	25.63	3.32	20.22	21.06	23.70	< 0.001**
	D	49	21.00	43.70	30.00	5.21	22.40	23.30	26.25	
AP (cm)	C	7	74.00	106.00	91.14	11.61	74.00	74.00	80.00	0.053*
	D	47	71.00	140.00	101.64	13.26	76.40	85.00	94.00	
Pulse (bpm)	C	15	51.00	100.00	72.33	14.64	51.00	52.20	60.00	0.520*
	D	49	39.00	100.00	74.57	10.72	55.00	62.00	68.50	
SBP (mmHg)	C	19	100.00	146.00	118.58	10.75	100.00	104.00	110.00	< 0.001*
	D	49	101.00	189.00	135.43	21.11	105.00	110.00	116.00	
DBP (mmHg)	C	19	60.00	80.00	72.95	7.04	60.00	61.00	70.00	0.247*
	D	49	46.00	100.00	76.31	11.70	52.50	57.00	69.00	
Bioimpedance (%)	C	1	26.40	26.40	26.40	0.00	-	-	-	-
	D	39	12.30	61.90	35.55	10.47	21.20	23.90	28.80	

* Independent samples t-test; ** Mann-Whitney Test

BMI – Body mass index; AP – Abdominal perimeter; SBP – Systolic blood pressure; DBP – Diastolic blood pressure

Correlation analysis suggested similar profiles when analysing correlations between variables separately in control and patient groups.

In controls, statistical significant correlations were found, as expected, between height and weight ($r = 0.78$; $p < 0.001$) or DBP ($r = 0.50$; $p = 0.031$), between weight and BMI ($r = 0.89$; $p = 0.007$) or SBP ($r = 0.54$; $p = 0.017$) and between BMI and abdominal perimeter ($r = 0.92$; $p = 0.003$) or SBP ($r = 0.63$; $p = 0.004$).

On the diabetic group, the pattern was similar, weight was correlated with BMI ($r = 0.71$; $p <$

0.001) and abdominal perimeter (r = 0.72; p < 0.001), being these two also correlated (r = 0.69; p < 0.001). Moreover, in type 2 diabetics, systolic and diastolic blood pressure were also found to be correlated (r = 0.53; p < 0.001).

Age was not significantly correlated with any of these measures in both groups; in the patients group no correlation was found to be significant or above 0.40, in absolute value, as observed in Figure 30.

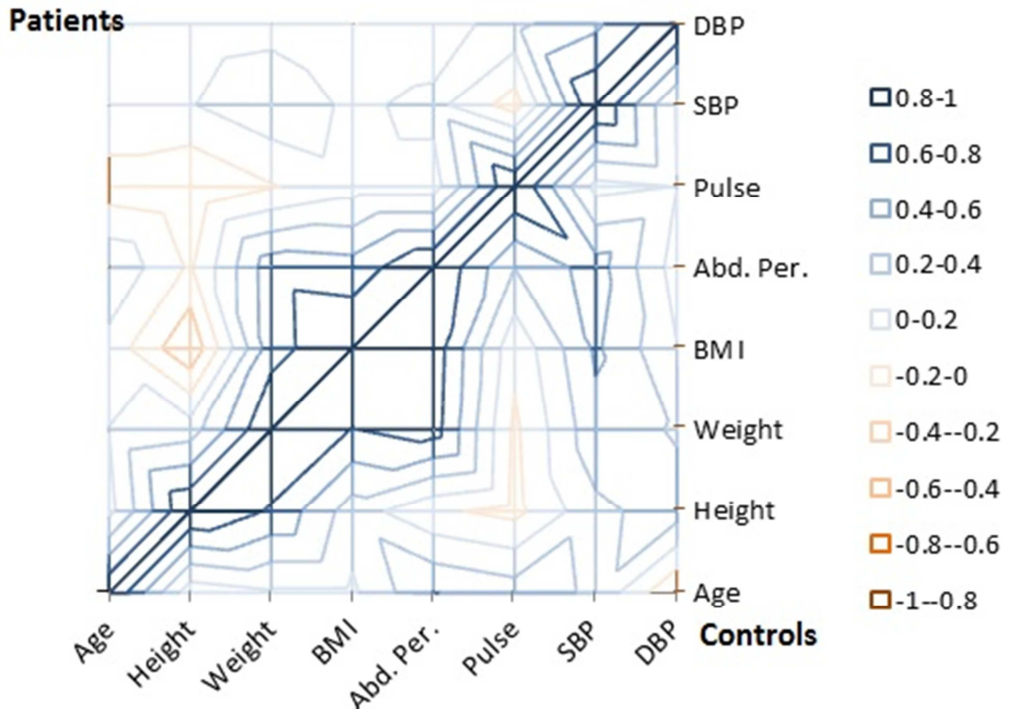

Figure 30 - Correlation between clinical and demographic variables measured in controls and in type 2 diabetics.

Aging has an important role in the natural decline of vision, and the study training sample was not homogenous between study groups according to age. In spite of this, we did not find any statistical significant correlation between age and blood tests values obtained in performed tests or between age and the evaluated ophthalmological procedures (OCT or visual psychophysical tests), either in the control group or the type 2 diabetic group.

Hence, with multivariate statistical procedures, we may be able to determine whether statistical significant differences between study groups found in age, with univariate analysis, are due to age or to the presence of type 2 diabetes. We should point out, once more, that univariate comparisons between groups were performed just as an exploratory method to conduct a variable reduction, in order to enable multivariate classification.

2.1.2 Blood Tests

2.1.2.1 Biochemistry

As expected, and being this one of the known parameters used for the diagnosis and monitoring of diabetes, in spite of he expected effects of therapeutic intervention, patients had significantly higher levels on glucose and glycosylated haemoglobin (Table 11). Glycosylated haemoglobin is presented in mmol/mol, according to the International Federation of Clinical Chemistry (IFCC) and in percentage (%), according to the National Glycohemoglobin Standardization Program. Measure of glycosylated hemoglobin expressed as a percentage have the advantage of being directly related to clinical outcomes, in spite of the agreement between the American Diabetes Association (ADA), the European Association for the Study of Diabetes (EASD) and the International Diabetes Federation (IDF) in reporting values of glycosylated hemoglobin in mmol/mol.

Table 11 - Descriptive statistics and group comparison of blood glucose between controls and type 2 diabetics.

	Type	N	Min	Max	Mean	SD	P25	P50	P75	P
Glucose	C	45	74.00	124.00	90.80	10.18	83.50	89.00	95.00	< 0.001**
	D	47	62.00	363.00	171.60	61.53	125.00	166.00	206.00	
HbA1C (NGSP)	C	45	4.80	7.20	5.49	0.43	5.30	5.40	5.70	< 0.001**
	D	48	5.20	17.30	9.46	2.38	7.70	9.30	10.88	
HbA1C (IFCC)	C	45	29.00	55.00	36.47	4.75	34.00	36.00	39.00	< 0.001**
	D	48	33.00	166.00	80.02	26.00	61.00	78.00	95.75	

* Independent samples t-test; ** Mann-Whitney U Test

Observing biochemistry values, it seems that renal function evaluated through creatinine levels is similar between groups (Table 12).

Table 12 – Descriptive statistics and comparison of creatinine values between controls and type 2 diabetics.

	Type	n	Min	Max	Mean	SD	P25	P50	P75	p
Creatinine	C	45	0.42	1.19	0.73	0.17	0.61	0.73	0.86	0.543**
	D	48	0.44	2.58	0.89	0.47	0.59	0.73	1.05	

* Independent samples t-test; ** Mann-Whitney U Test

Some parameters of liver function, as ALT and AST, were also evaluated and, although no statistical significant difference was found between patients and controls on AST, patients showed a significantly higher level on ALT. Note that ALT only exists in the liver and AST also exists on heart and muscles. Alkaline phosphatase and gamma GT, parameters related to the biliary ductus integrity, are significantly higher on type 2 diabetes mellitus (Table 13).

Table 13 - Descriptive statistics and group comparison of liver function parameters between controls and type 2 diabetics.

	Type	N	Min	Max	Mean	SD	P25	P50	P75	p
ALT	C	46	9.00	119.00	24.54	18.19	14.00	20.00	27.25	0.029**
	D	49	9.00	81.00	30.51	17.83	18.00	25.00	36.00	
AST	C	45	12.00	55.00	21.98	8.99	17.00	19.00	23.50	0.094**
	D	49	10.00	79.00	26.33	13.54	18.00	22.00	31.50	
Alkaline Phosphatase	C	46	29.00	134.00	64.07	20.14	51.00	60.00	73.25	0.004**
	D	49	37.00	164.00	79.16	28.13	59.50	75.00	91.50	
Gamma GT	C	46	9.00	83.00	28.85	18.95	15.75	22.50	34.50	0.020**
	D	49	8.00	223.00	44.94	46.13	21.00	30.00	48.00	

* Independent samples t-test; ** Mann-Whitney U Test

Concerning lipid parameters, we found that cholesterol levels, diabetics have better indicators than controls. The latters have significantly higher levels either of total cholesterol ($p = 0.001$), and of low density lipoproteins ($p = 0.003$). However, controls also showed higher levels on high density lipoproteins ($p = 0.003$) thus, probably, controls are at a higher risk but have nevertheless better levels for a putative indicator of protection concerning arteriosclerotic processes. The atherogenic index gives the coronary risk associated to problems with LDL cholesterol, and this is significantly higher in diabetics ($p = 0.028$). In fact, the atherogenic index represents the ratio between total and HDL cholesterol, so ideally, it should be below 5 units and, as we can observe in the following table, mean and median values in controls are respectively 5.5 and 3.35, and on diabetics this index has the values, respectively, of 9.30 and 4.00 (Table 14).

Triglycerides are essential in terms of energetic needs, but harmful if stored in high quantities since they are associated with atherosclerosis and cardiovascular disorders. Usually, high values of triglycerides are associated to high values of LDL cholesterol or low values of HDL cholesterol. In this study, patients have significantly higher levels of triglycerides, in spite of having lower values of LDL cholesterol and higher of HDL cholesterol.

Table 14 - Descriptive statistics and group comparison of lipid related parameters between controls and type 2 diabetics.

	Type	N	Min	Max	Mean	SEM	P25	P50	P75	P
Total Cholesterol	C	46	117.00	292.00	200.78	5.61	175.25	197.00	227.75	0.001**
	D	49	86.00	398.00	175.24	7.54	138.50	161.00	198.50	
Cholesterol HDL	C	46	28.00	87.00	57.30	2.16	45.75	54.50	68.25	< 0.001*
	D	49	14.00	65.00	42.18	1.63	34.00	41.00	50.00	
Atherogenic Index	C	46	2.10	5.50	3.68	0.14	2.90	3.35	4.53	0.028**
	D	49	2.30	9.30	4.38	0.22	3.20	4.00	5.15	
Cholesterol LDL	C	46	67.00	198.00	133.50	4.54	114.00	133.00	151.25	0.003**
	D	46	48.00	204.00	114.46	5.25	90.00	105.50	131.00	
Triglycerides	C	46	44.00	362.00	117.72	10.35	77.50	94.00	135.25	< 0.001**
	D	48	55.00	465.00	166.10	12.78	105.50	146.00	200.25	
Apolipoprotein A1	C	46	82.00	250.00	165.87	4.93	142.75	161.50	186.00	< 0.001*
	D	47	37.00	198.00	138.45	4.21	123.00	140.00	162.00	
Apolipoprotein B100	C	46	43.00	173.00	94.48	3.31	79.00	95.50	103.25	0.235**
	D	47	44.00	160.00	92.43	4.07	73.00	85.00	105.00	
B100/A1	C	46	0.25	1.01	0.59	0.03	0.48	0.54	0.72	0.073**
	D	47	0.30	2.74	0.72	0.06	0.51	0.58	0.84	
Lipoprotein	C	45	2.33	68.00	20.13	2.47	9.31	12.10	28.05	0.133**
	D	48	2.33	166.00	33.16	5.27	9.31	21.15	42.60	

* Independent samples t-test; ** Mann-Whitney U Test

Apolipoprotein are families of proteins that joins lipids and are associated to cholesterol, and are classified into two main types: the apolipoprotein A1 is synthetized in the liver and in the small intestine and is a part of the HDL cholesterol and its role to facilitate the transportation of the HDL cholesterol to the liver; apolipoprotein B100 is synthetized in the liver and is a part of the LDL cholesterol that is responsible for joining it to cellular receptors and may lead to atherosclerosis if accumulated in the arteries. Thereby, the ratio between those two apolipoproteins (B100/A1) may reflect the risk of developing cardiovascular disease. Apparently, there is no distinction between groups either on apolipoprotein B100 levels ($p = 0.235$), or in the ratio of apolipoproteins ($p = 0.073$), although is this last case there is a marginal tendency for diabetics to have higher values, which is related to the lack of difference in apolipoprotein B100 values and on the statistical difference on apolipoprotein A1 values ($p < 0.001$), which are lower in diabetics.

2.1.2.2 Cell Blood Count Cytometry

Circulating leucocytes are part of the immunological system acting and participating in the combat to eliminate microorganisms and chemical structures alien to the body partly through the generation of anti-bodies. We found that diabetic had significantly higher

number of leucocytes (p = 0.024), although no association was found between the type of diabetes and normal values of leucocytes (chi-square test using Monte-Carlo simulation: $p = 1.000 \in (1.000; 1.000)$) since only two controls and three diabetic had the leucocytes counts below normal and only three controls and four diabetic patients had leucocytes counts above the usually defined as normal cut-off values (Table 15).

Table 15 - Descriptive statistics and group comparison of leucocytes between controls and type 2 diabetics.

	Type	n	Min	Max	Mean	SEM	P25	P50	P75	P
Leucocytes	C	46	3.90	15.50	6.37	0.33	4.98	5.95	6.85	0.024**
	D	49	0.90	18.40	7.04	0.37	5.65	6.80	8.15	

* Independent samples t-test; ** Mann-Whitney U Test

Erythrocytes counts are similar in controls and diabetics (p = 0.078), as observed in table 16, although there is a tendency for controls to have slightly higher number, with mean and median values within normal values for erythrocytes count and patients with mean and median slightly below the lower limit for normal values (women: 4.5 x 10^6/mm^3; men: 5 x 10^6/mm^3). In fact, only one of the 46 controls (2.17%) presented an erythrocyte count below normal, being this number about eight times higher in diabetics (eight patients, 16.33%); on the other hand, the percentage of cases with erythrocytes counts larger than normal was 36.96% (17 cases) in controls and 16.33% in diabetics. However, age may be acting as a confounding variable, thus further considerations will be evaluated with multivariate analysis.

Table 16 - Descriptive statistics and group comparison of red cell counts between controls and type 2 diabetics.

	Type	n	Min	Max	Mean	SEM	P25	P50	P75	p
Erythrocytes	C	46	3.76	5.67	4.53	0.06	4.16	4.44	4.89	0.078*
	D	49	3.39	6.13	4.36	0.08	4.02	4.29	4.67	
Haemoglobin	C	46	11.10	16.90	14.10	0.18	13.18	14.30	14.93	0.004*
	D	49	10.30	15.70	13.29	0.20	12.25	13.10	14.65	
Haematocrit	C	46	32.80	49.30	41.48	0.54	38.38	41.60	44.03	0.006*
	D	49	30.70	48.00	39.18	0.60	35.75	38.50	43.30	
MCV	C	46	82.30	98.20	91.14	0.54	89.60	91.35	93.35	0.493**
	D	49	57.80	101.50	90.29	0.94	87.60	90.40	95.10	
MCH	C	46	26.80	33.70	31.16	0.20	30.40	31.25	32.10	0.242**
	D	49	18.70	35.00	30.65	0.34	29.50	30.90	32.00	
MCHC	C	46	31.90	35.40	34.02	0.11	33.68	34.05	34.53	0.546*
	D	49	32.40	35.60	33.92	0.12	33.35	34.00	34.50	
EVC	C	46	11.80	15.80	13.22	0.11	12.70	13.10	13.63	0.043**
	D	49	11.50	17.60	13.64	0.17	12.90	13.50	14.00	

* Independent samples t-test; ** Mann-Whitney U Test

MCV – Mean corpuscular volume; MCH – Mean corpuscular haemoglobin; MCHC – Mean corpuscular haemoglobin concentration; EVC – Erythrocytes variation coefficient

Nevertheless, 64.21% of all cases presented normal counts of erythrocytes, 60.87% in the control group and 54.10% in diabetic group, suggesting an association between the presence of type 2 diabetes and a lower number of erythrocytes.

In fact, the haemoglobin and the haematocrit, the ratio between the volume of all erythrocytes in a blood sample and the total volume of that blood sample are significantly lower in patients, when compared to controls ($p = 0.004$ and $p = 0.006$, respectively), as well as the erythrocytes variation coefficient (EVC) or the red cell distribution with, an index that measures variation in size, is significantly higher in patients ($p = 0.043$).

However, mean corpuscular volume (MCV), mean corpuscular haemoglobin (MCH), and mean corpuscular haemoglobin concentration (MCHC) are similar between groups.

Platelets were analysed regarding to their size and quantity. We found no statistical significant differences (Table 17) in platelets counts, mean platelet volume (MPV) and plateleocrit, or even on platelet variation coefficient (PVC).

Table 17 - Descriptive statistics and group comparison of platelet between controls and type 2 diabetics.

	Type	N	min	Max	mean	SEM	P25	P50	P75	P
Platelet	C	46	141.00	461.00	233.33	10.14	186.50	222.00	248.25	0.650**
	D	49	81.00	610.00	225.41	11.63	183.50	209.00	266.50	
MPV	C	46	7.20	11.30	9.19	0.16	8.40	9.20	9.80	0.068*
	D	49	7.10	13.20	9.64	0.19	8.90	9.70	10.40	
Plateleocrit	C	46	0.13	14.00	0.51	0.30	0.18	0.20	0.24	0.872**
	D	49	0.06	0.48	0.21	0.01	0.18	0.21	0.24	
PVC	C	46	16.00	18.00	16.35	0.08	16.00	16.00	17.00	0.285**
	D	48	16.00	18.00	16.50	0.09	16.00	16.00	17.00	

* Independent samples t-test; ** Mann-Whitney U Test

MPV – Mean platelet volume; PVC – Platelet variation coefficient

There was also no association between presence of diabetes and each one of the parameters referred categorised according to CHUC reference values, since almost all the values were classified within normal range of values.

2.1.2.3 Hormonology

Thyroid stimulating hormone (TSH) values were found to be identical between study groups ($p = 0.184$) but controls have significantly higher levels of the connecting peptide (C-peptide), as expected, since it serves as a linker between the A and B chains of insulin and facilitates assembly, folding and processing of insulin in the endoplasmic reticulum. It can be used as a marker of insulin secretion for the study of the pathophysiology of type 1 and type

2 diabetes (Table 18).

Table 18 - Descriptive statistics and group comparison of TSH (3rd generation) and Peptide C between controls and type 2 diabetics.

	Type	n	min	max	mean	SD	P25	P50	P75	p
TSH	C	44	0.01	5.70	1.80	0.17	1.10	1.45	2.10	0.184**
	D	48	0.10	5.10	2.06	0.17	1.13	1.95	2.48	
C-Peptide	C	42	0.90	4.00	2.02	0.10	1.58	1.90	2.50	0.002**
	D	47	0.10	4.80	1.56	0.18	0.70	1.30	2.10	

* Independent samples t-test; ** Mann-Whitney U Test

In fact, 89.36% of the cases had normal values of TSH (41 controls and 43 patients), which represents respectively 91.11% and 87.7% of controls and diabetics. In the control group, only two subjects had lower values of TSH and other two had higher values of TSH, and in the diabetic group two and four subjects had, respectively, lower and higher values of TSH thus, no association was fount between diabetes and thyroid dysfunction (chi-square test using Monte-Carlo simulation: $p = 0.877 \in (0.868; 0.885)$).

Concerning the C-peptide, it is more probable to find diabetics with lower than normal levels than controls ($\chi_1^2 = 9.42; p = 0.002$). In fact, 95.24% of controls have normal values of C-peptide and 70.21% of diabetics have normal values.

2.1.3 Ophthalmological tests

Intraocular pressure was identical between controls and type 2 diabetics (t_{48} = -0.878; p = 0.384) but patients presented significantly lower BCVA (p = 0.001), in spite of the median of 10/10 in both groups.

2.1.3.1 Optical Coherence Tomography

2.1.3.1.1 Volume Scan density

The retinal structures evaluated by volume scan density could not differentiate groups, since no statistical significant differences were found (Table 19). The central subfield, as well as all the quadrants in the inner and outer regions of the macula in diabetic patients presented similar results when compared with controls.

Table 19 - Descriptive statistics and group comparison of Volume Scan measured by OCT between controls and type 2 diabetics.

Volume Scan			N	Min	Max	Mean	SEM	P25	P50	P75	p
Central Subfield		C	46	242.00	318.00	280.91	2.68	265.00	283.00	293.00	0.459**
		D	49	174.00	416.00	289.41	5.98	268.00	284.00	302.00	
Inner	Nasal	C	46	313.00	388.00	346.24	2.44	336.50	347.00	358.00	0.803**
		D	49	242.00	415.00	344.71	4.03	335.00	345.00	356.00	
	Superior	C	46	250.00	391.00	341.98	3.12	332.75	345.50	353.25	0.809**
		D	49	303.00	433.00	346.92	3.73	332.50	345.00	355.00	
	Temporal	C	46	300.00	374.00	332.22	2.27	319.50	334.00	342.25	0.587**
		D	49	251.00	448.00	332.18	4.50	320.50	330.00	342.50	
	Inferior	C	46	313.00	385.00	341.26	2.38	329.75	341.00	353.00	0.379**
		D	49	275.00	424.00	338.04	4.14	325.00	339.00	348.00	
Outer	Nasal	C	46	127.00	346.00	309.39	4.74	302.25	313.00	323.75	0.994**
		D	49	262.00	416.00	315.00	3.59	302.00	311.00	323.50	
	Superior	C	46	269.00	324.00	298.39	1.93	291.75	298.50	305.25	0.687**
		D	49	256.00	358.00	301.27	2.84	289.00	298.00	311.00	
	Temporal	C	46	252.00	395.00	288.17	3.22	275.75	288.50	295.50	0.335**
		D	49	226.00	396.00	288.55	4.40	272.00	284.00	294.50	
	Inferior	C	46	252.00	320.00	286.54	2.49	274.75	286.50	298.00	0.331**
		D	49	237.00	394.00	284.39	3.91	268.50	284.00	292.50	

*Independent samples t-test; **Mann-Whitney U Test

2.1.3.1.2 Retinal Nerve Fiber Layer

However, it seems that subjects with type 2 diabetes have higher thickening of retinal nerve on the temporal field (p = 0.041), which is especially detected on the temporal inferior quadrant (p = 0.047) since no statistical significant differences were found on the temporal-superior quadrant (Table 20).

Table 20 - Descriptive statistics and group comparison of Retinal Nerve Fiber Layer measured with OCT between controls and type 2 diabetics.

RNFL			n	Min	Max	Mean	SEM	P25	P50	P75	P
Global		C	46	81.00	121.00	99.15	1.24	94.00	99.00	104.50	0.916**
		D	48	68.00	118.00	98.17	1.45	97.00	99.00	103.75	
Nasal		C	46	36.00	106.00	74.26	1.91	64.00	73.00	83.25	0.881*
		D	48	37.00	97.00	74.65	1.73	66.00	76.00	85.00	
Nasal	Superior	C	46	51.00	151.00	106.15	3.39	88.75	107.00	124.50	0.472*
		D	48	52.00	153.00	102.79	3.21	84.25	102.50	117.00	
	Inferior	C	46	87.00	175.00	125.96	3.30	112.00	125.00	140.00	0.757*
		D	48	91.00	162.00	124.67	2.57	111.25	121.50	140.00	
Temporal		C	46	46.00	106.00	68.78	1.64	61.00	67.50	76.00	0.041*
		D	48	32.00	160.00	75.17	2.59	67.00	75.00	82.75	
Temp.	Superior	C	46	80.00	192.00	140.59	3.32	127.75	142.50	156.00	0.546*
		D	48	51.00	187.00	137.65	3.54	124.00	135.50	153.50	
	Inferior	C	46	86.00	186.00	132.74	4.13	112.00	132.50	157.50	0.047*
		D	48	52.00	193.00	121.29	3.93	107.75	122.50	138.75	

*Independent samples t-test; **Mann-Whitney U Test

2.1.3.2 Psychophysical tests

Although aging has a role in the natural decline of vision, and sample of this study is not homogenous between study groups according to age, we found out that there are statistical significant differences between study groups on some of the areas. Later, it will be evaluated if differences are due to age or to the presence of type 2 diabetes.

Psychophysical visual tests were evaluated on Speed, achromatic and chromatic vision, and each one of these tests was performed on meridians 0º, 45º, 90º and 135º.

2.1.3.2.1 Speed

The speed test has showed statistical significant differences in all the four meridians evaluated, showing also that controls always have a better performance, as observed in Table 21.

Table 21 - Descriptive statistics and group comparison of Speed test measured in meridians 0º, 45º, 90º, 135º and global area generated by these meridians between controls and type 2 diabetics.

Speed			N	Min	Max	Mean	SEM	P25	P50	P75	p
Meridian	0º	C	44	0.15	2.90	1.03	0.10	0.56	0.88	1.47	0,001**
		D	45	0.16	9.69	1.97	0.25	0.75	1.62	2.56	
	45º	C	44	0.16	7.46	1.27	0.19	0.48	0.83	1.62	0,017**
		D	42	0.18	7.99	2.11	0.31	0.69	1.28	2.87	
	90º	C	43	0.16	4.48	1.09	0.13	0.47	0.82	1.38	0,002**
		D	46	0.15	7.88	1.99	0.25	0.85	1.41	2.90	
	135º	C	44	0.16	7.13	1.13	0.16	0.58	0.90	1.32	< 0,001**
		D	42	0.15	7.72	2.60	0.31	0.85	2.11	4.18	
Area		C	40	0.18	20.13	1.64	0.50	0.41	0.83	1.47	< 0.001**
		D	40	0.27	29.79	5.03	0.93	1.25	3.06	6.83	

* Independent samples t-test; ** Mann-Whitney U Test

The area of the polygons generated by the median points measured in each one of the meridians is significantly different between controls and patients ($p < 0.001$), as presented on Figure 31.

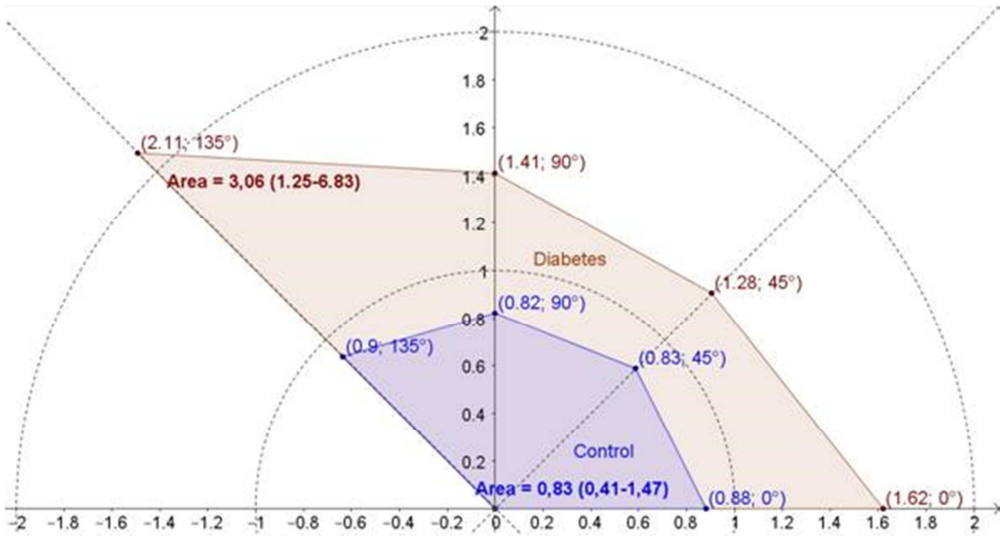

Figure 31 - Speed test on meridians 0º, 45º, 90º, 135º and global area generated by these meridians in controls and type 2 diabetics.

2.1.3.2.2 Achromatic contrast

Achromatic contrast is similar in both groups (Table 22), although type 2 diabetics tends to have less sensitivity to achromatic contrast along meridian 90º (p = 0,005).

Table 22 - Descriptive statistics and group comparison of the achromatic contrast test measured in meridians 0º, 45º, 90º, 135º and global area generated by these meridians between controls and type 2 diabetics.

			N	Min	Max	Mean	SEM	P25	P50	P75	P
Meridian	0º	C	46	1.00	5.44	2.59	0.13	1.93	2.58	3.10	0,415**
		D	45	1.00	10.47	2.74	0.27	1.55	2.31	3.30	
	45º	C	45	1.10	4.76	2.40	0.14	1.70	2.21	2.72	0,081**
		D	43	1.20	18.72	3.46	0.45	1.80	2.64	3.76	
	90º	C	46	1.00	4.85	2.34	0.16	1.48	2.17	3.14	0,005**
		D	46	1.00	8.52	3.19	0.23	2.21	2.84	3.91	
	135º	C	44	1.00	4.37	2.39	0.14	1.60	2.31	3.10	0,305**
		D	42	1.00	10.77	2.94	0.31	1.48	2.60	3.50	
Area		C	47	0.00	15.25	5.76	0.52	2.93	5.35	7.73	0.250**
		D	49	0.00	115.30	9.95	2.43	3.48	6.39	11.83	

*Independent samples t-test; **Mann-Whitney U Test

The areas of the polygons created by the medians of each meridian, representing total achromatic contrast sensitivity are similar in both groups (Figure 32).

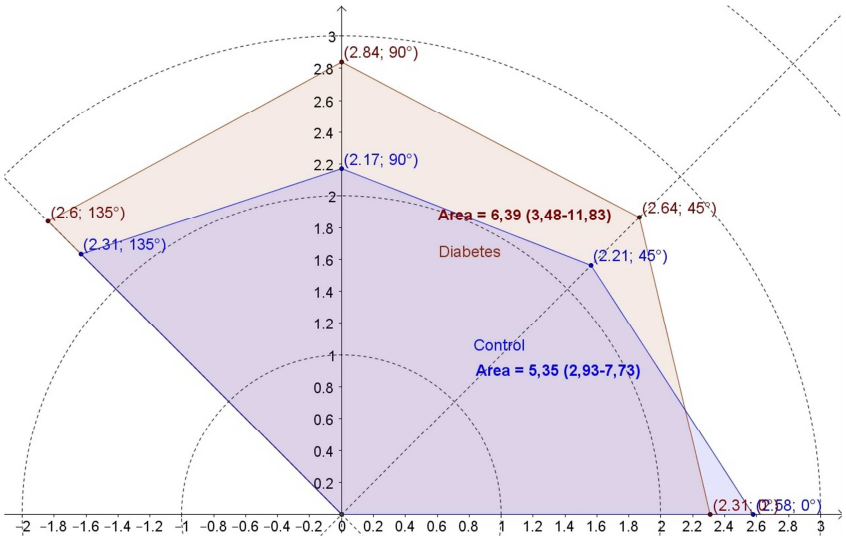

Figure 32 – Achromatic contrast test on meridians 0º, 45º, 90º, 135º and global area generated by these meridians in controls and type 2 diabetics.

2.1.3.2.3 Chromatic Contrast

Chromatic contrast is similar in controls and diabetics, regarding the measure obtained for the Protan axis, whatever the meridian evaluated, except for zero degrees, in which diabetics have a higher threshold for chromatic contrast sensitivity, in the Protan axis (Table 23, Figure 33).

Table 23 - Descriptive statistics and group comparison of Chromatic contrast test on the Protan axis, measured in meridians 0º, 45º, 90º, 135º and global area generated by these meridians between controls and type 2 diabetics.

Protan			n	Min	Max	Mean	SEM	P25	P50	P75	P
Meridian ($x10^{-3}$)	0º	C	46	1.23	4.95	2.40	0.17	1.24	2.46	3.08	0,043**
		D	45	1.23	9.49	3.09	0.26	1.55	3.07	4.31	
	45º	C	46	1.23	18.32	4.95	0.50	2.47	4.01	8.10	0,747**
		D	44	1.23	33.96	7.13	1.16	2.25	4.83	8.42	
	90º	C	46	1.23	10.76	3.83	0.29	2.32	3.71	4.94	0,555**
		D	46	1.23	27.34	4.61	0.77	1.85	3.48	5.10	
	135º	C	46	1.24	18.32	5.37	0.45	3.23	4.96	6.86	0,462**
		D	42	1.23	17.05	5.27	0.54	2.47	4.20	7.44	
Area ($x10^{-6}$)		C	46	2.71	84.80	18.95	2.12	9.76	15.20	23.35	0.980**
		D	42	2.42	453.00	31.25	10.76	7.87	15.90	27.75	

* Independent samples t-test; ** Mann-Whitney U Test

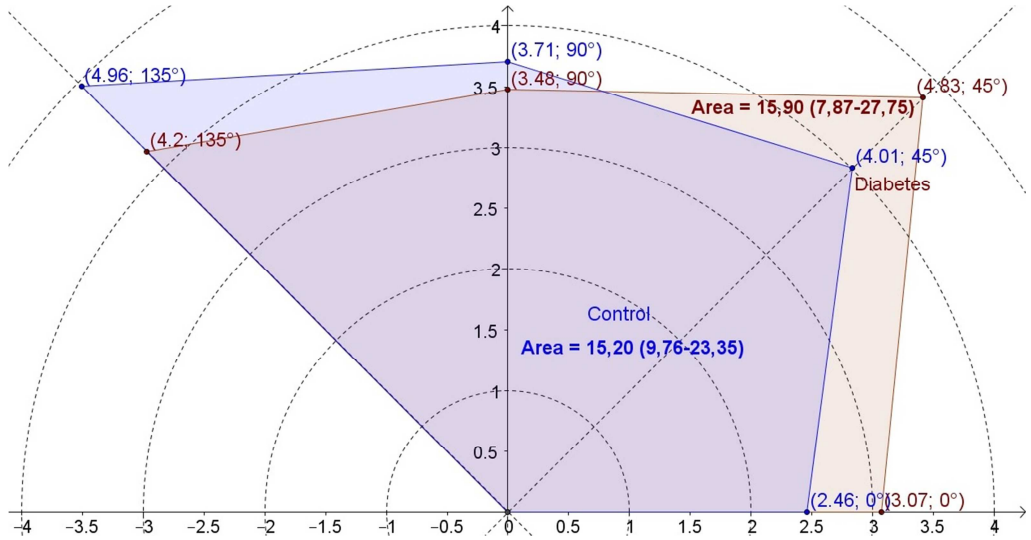

Figure 33 – Chromatic contrast test (Protan) on meridians 0º, 45º, 90º, 135º and global area generated by these meridians in controls and type 2 diabetics (meridian values should be read $x10^{-6}$; area values should be read $x10^{-6}$).

The Deutan axis can discriminate patients from controls, as observed in Table 24, since the threshold of contrast sensitivity is lower, therefore better, in controls, at least when measured across the 0º and the 45º meridians (respectively $p < 0.001$ and $p = 0.042$).

Table 24 - Descriptive statistics and group comparison of Chromatic contrast test on the Deutan axis, measured in meridians 0º, 45º, 90º, 135º and global area generated by these meridians between controls and type 2 diabetics.

Deutan		n	Min	Max	Mean	SD	P25	P50	P75	p
Meridian ($x10^{-3}$) 0º	C	46	1.23	9.15	2.83	0.24	1.25	2.48	3.09	< 0,001**
	D	43	1.24	55.95	6.59	1.41	2.48	4.33	7.08	
45º	C	46	1.24	83.48	14.94	2.91	2.93	6.63	20.05	0,042**
	D	41	1.24	364.53	29.21	8.98	4.99	12.15	31.43	
90º	C	46	1.23	28.93	6.10	0.83	2.80	4.03	8.44	0,120**
	D	45	1.23	86.43	14.74	3.08	2.79	4.95	15.26	
135º	C	46	1.23	72.73	14.22	2.29	4.89	9.57	17.52	0,379**
	D	40	1.23	69.83	21.24	3.39	4.38	10.40	35.71	
Area ($x10^{-6}$)	C	46	2.16	1450.00	128.78	41.18	11.83	32.15	83.85	0.013**
	D	39	4.38	8820.00	509.06	231.72	18.10	80.90	277.00	

* Independent samples t-test; ** Mann-Whitney U Test

Areas are also smaller in controls than in patients, with statistical significant differences ($p = 0.013$), thus contrast sensitivity on the Deutan axis is better in controls (Figure 34).

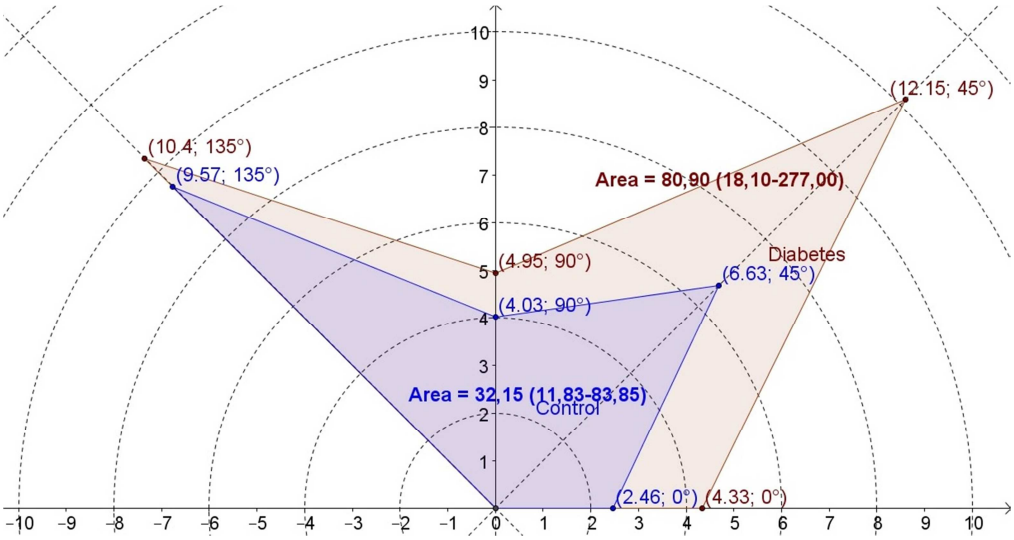

Figure 34 – Chromatic contrast test (Deutan) on meridians 0º, 45º, 90º, 135º and global area generated by these meridians in controls and type 2 diabetics (meridian values should be read $x10^{-6}$; area values should be read $x10^{-6}$).

All the meridians show higher threshold of contrast sensitivity on the Tritan axis, and the total contrast sensitivity on the Tritan axis, represented by the area of the polygon created by the medians measured in each one of the four meridians is also better in controls ($p < 0.001$), as related on Table 25 and Figure 35.

Table 25 - Descriptive statistics and group comparison of Chromatic contrast test on the Tritan axis, measured in meridians 0º, 45º, 90º, 135º and global area generated by these meridians between controls and type 2 diabetics.

	Tritan		N	Min	Max	Mean	SD	P25	P50	P75	p
Meridian ($x10^{-3}$)	0º	C	45	14.99	111.62	44.73	2.62	30.11	45.00	56.15	< 0,001**
		D	45	25.00	401.57	85.52	11.20	43.46	57.82	99.20	
	45º	C	45	25.00	305.13	101.65	10.37	48.88	81.11	139.89	0,005**
		D	42	25.00	411.62	154.10	15.38	67.32	134.06	207.98	
	90º	C	45	25.00	237.92	64.90	5.95	40.00	56.81	75.24	< 0,001**
		D	46	25.00	425.51	131.91	15.83	46.66	79.45	193.50	
	135º	C	45	14.99	184.35	75.16	6.30	45.76	66.96	99.35	0,001**
		D	41	35.23	568.07	160.58	19.29	54.48	129.54	238.77	
	Area ($x10^{-6}$)	C	45	1070	53200	6409.11	1209.81	2730	4360	6580	< 0.001**
		D	41	1080	140000	21004.15	4447.03	5755	9460	23350	

* Independent samples t-test; ** Mann-Whitney U Test

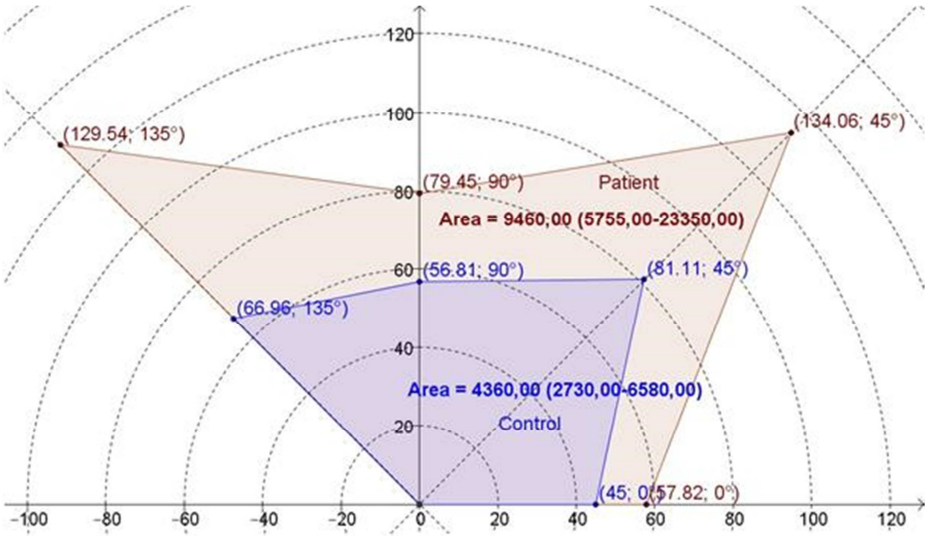

Figure 35 – Chromatic contrast test (Tritan) on meridians 0º, 45º, 90º, 135º and global area generated by these meridians in controls and type 2 diabetics (meridian values should be read $x10^{-6}$; area values should be read $x10^{-6}$).

2.2 Phase 2: Univariate classifiers of Diabetes

Receiver operating characteristic curves may be useful to detect which particular variables, one by one, may discriminate groups. In spite of being a procedure that does not enable the evaluation of interactions between variables in a set, it may be useful to identify clinically useful cut-offs isolated for each one of the continuous variables.

Using variables related with clinical and demographic (Table 10), and although a statistical significant difference between groups was found in height, groups cannot be discriminated due to that variable, since area under the Receiver-Operating Characteristic curve is 0.613, without reaching statistical significance at the 5% level (p = 0.057). In fact, if such variable was used to discriminate diabetes, the sensitivity value would be of 27.45%, which is not admissible for clinical discrimination. As the body mass index is capable of separating groups and is a measure that involves either weight or height, with acceptable values either for sensitivity and specificity, with an area under the ROC curve statistical significant, it at least is preferable to use this variable. It is important to point out that if variables are discriminatory, they might, in this context, not be specific (in the sense of the existence of other clinical entities and not in the sense of specificity as defined in ROC analysis). Note that abdominal perimeter may also discriminate groups, as systolic blood pressure (Table 26). However, these parameters were measured in few cases and, hence, they will not be used on multivariate classifiers. On the other hand, there is a large percentage of cases with blood

pressure controlled by medication, therefore variables related to blood pressure would insert a bias on the analysis, if used.

Table 26 - Accuracy of medical clinical outcome measures for univariate classification of type 2 diabetes.

Variable	AUC	SEM	P	LBCI	UBCI	Cut-off	Sensitivity	Specificity	+LR
Height	0.613	0.06	0.056	0.501	0.726	≤ 1.53	27.45	97.87	12.89
Weight	0.663	0.06	0.006	0.552	0.774	≥ 68.05	77.55	53.19	1.66
BMI	0.752	0.05	< 0.001	0.654	0.850	≥ 26.95	71.43	72.34	2.58
AP	0.733	0.09	0.049	0.549	0.916	≥ 93.50	78.72	71.43	2.76
Pulse	0.563	0.10	0.461	0.376	0.751	-	-	-	-
SBP	0.735	0.06	0.003	0.616	0.854	> 131.00	61.22	94.74	11.64
DBP	0.615	0.07	0.144	0.483	0.747	-	-	-	-

BMI – Body mass index; AP – Abdominal perimeter; SBP – Systolic blood pressure; DBP – Diastolic blood pressure.

Apparently, body mass index is the variable within this set that best discriminated diabetics, since it is the one with highest area under the ROC curve and smallest p-value, and it is the only one that presents acceptable values in all the four indexes presented (sensitivity, specificity, positive and negative predictive values). In terms of the area under the ROC curve (AUC), abdominal perimeter and systolic blood pressure present similar values, but in terms of indexes, weight is the only variable that has all four indexes above 50%. Comparing discriminative power between body mass index and each one of the referred variables, we can state that there is a significant difference between BMI and weight ($Z = 2.227$; $p = 0.026$), but not between BMI and abdominal perimeter ($Z=0.523$; $p = 0.600$) or BMI and systolic blood pressure ($Z = 0.579$; $p = 0.563$).

Concerning blood tests, as expected, either glucose or glycosylated haemoglobin have excellent discriminant power, with no statistical significant difference between them ($Z = 0.822$; $p = 0.411$), which means that any of them may be used to classify diabetes. Note that the positive likelihood ratio for glucose is higher than 49, which means that it is about 39 times more probable to have fasting blood glucose equal or higher than 114.50 in diabetics than is controls. The glycosylated haemoglobin, as a metabolic control of glucose parameter, appears abnormal much more frequently (21 times) in diabetics than in controls (Table 27).

Table 27 - Accuracy of blood glucose and glycosylated haemoglobin for univariate classification of type 2 diabetes.

Blood	AUC	SEM	p	LBCI	UBCI	Cut-off	Sens	Spec	+LR
Glucose	0.943	0.03	< 0.001	0.884	1.000	≥ 114.50	87.23	97.78	39.29
HbA1C (NGSP)	0.972	0.02	< 0.001	0.936	1.000	≥ 6.25	93.75	95.56	21.11
HbA1C (IFCC)						≥ 44.50			

Renal function evaluated through creatinine does not allow discrimination between groups (AUC = 0.537; p = 0.544) and liver function but liver function may separate diabetics from controls. In fact, either an ALT value not lower than 26.50, or and alkaline phosphatase value not lower than 64.50 or even a gamma GT value not inferior to 24.50 may classify a subject as having diabetes with a probability always higher than 62.50% (respectively, 68.57%, 66.67% and 62.75%), although less than 50% of diabetics have ALT values not inferior to 26.50. This fact is reflected on the positive likelihood ratio for these variables, which is around 2 for each one of them (Table 28).

Table 28 - Accuracy of liver function parameters for univariate classification of type 2 diabetes.

Variable	AUC	SEM	p	LBCI	UBCI	Cut-off	Sens	Spec	+LR
ALT	0.630	0.06	0.029	0.518	0.742	≥ 26.50	49.98	76.09	2.09
AST	0.600	0.06	0.095	0.486	0.714	-	-	-	-
Alkaline Phosphatase	0.670	0.06	0.004	0.561	0.778	≥ 64.50	65.31	65.22	1.88
Gamma GT	0.639	0.06	0.020	0.527	0.750	≥ 24.50	65.31	58.70	1.58

Within lipid related parameters, the ones that most separate groups are cholesterol HDL, apolipoprotein A1, triglycerides and total cholesterol, cholesterol LDL and atherogenic index, as presented in Table 29:

Table 29 - Accuracy of lipid related parameters for univariate classification of type 2 diabetes.

Variable	AUC	SEM	p	LBCI	UBCI	Cut-off	Sens	Spec	+LR
Total Cholesterol	0.702	0.06	0.001	0.595	0.809	≤ 182.50	69.39	69.57	2.28
Cholesterol HDL	0.789	0.05	< 0.001	0.699	0.879	≤ 45.50	69.39	76.09	2.90
Atherogenic Index	0.631	0.06	0.028	0.520	0.742	≥ 3.35	71.43	50.00	1.43
Cholesterol LDL	0.682	0.06	0.003	0.569	0.795	≤ 114.50	63.04	76.09	2.64
Triglycerides	0.704	0.05	0.001	0.598	0.811	≥ 119.00	68.75	69.57	2.26
Apolipoprotein A1	0.736	0.06	< 0.001	0.636	0.836	≤ 132.50	46.81	89.13	4.31
Apolipoprotein B100	0.571	0.06	0.235	0.452	0.690	-	-	-	-
B100/A1	0.608	0.06	0.073	0.493	0.723	-	-	-	-
Lipoprotein	0.590	0.06	0.136	0.473	0.707	-	-	-	-

Concerning cytometry and blood cell counts, diabetic patients tend to have higher values of

leucocytes and lower values of haemoglobin and haematocrit, but erythrocytes present higher variation coefficient in this group. Nevertheless, sensitivity of predictions based upon haemoglobin or haematocrit is inadequate, presenting more than 50% of false negative cases. The referred parameters have high specificity, which means that they may be used to classify presence but not absence of diabetes. Concerning the leucocyte and erythrocyte variation coefficient, the positive likelihood ratio is quite small compared to haematocrit and especially to the haemoglobin (Table 30).

Table 30 - Accuracy of Blood cell counts for univariate classification of type 2 diabetes.

Variable	AUC	SE	P	LBCI	UBCI	Cut-off	Sens	Spec	+LR
Leucocytes	0.635	0.06	0.024	0.521	0.748	$\geq$ 6.55	63.27	65.22	1.82
Erythrocytes	0.607	0.06	0.072	0.494	0.721	-	-	-	-
Haemoglobin	0.659	0.06	0.008	0.548	0.769	$\leq$ 12.65	38.78	93.48	5.95
Haematocrit	0.652	0.06	0.011	0.542	0.763	$\leq$ 38.25	48.98	80.43	2.50
MCV	0.46	0.06	0.439	0.427	0.665	-	-	-	-
MHC	0.570	0.06	0.242	0.454	0.686	-	-	-	-
MCHC	0.534	0.06	0.571	0.417	0.650	-	-	-	-
EVC	0.621	0.06	0.043	0.507	0.734	$\geq$ 13.25	63.27	63.04	1.71
Platelet	0.527	0.06	0.650	0.409	0.645	-	-	-	-
MPV	0.607	0.06	0.072	0.493	0.721	-	-	-	-
Plateleocrit	0.510	0.06	0.873	0.392	0.627	-	-	-	-
PVC	0.554	0.06	0.364	0.438	0.671	-	-	-	-

Hormonology parameters, such as thyroid stimulating hormone do not enable group discrimination (AUC = 0.590; p = 0.184) but peptide C levels can separate groups, mainly for confirming presence of type 2 diabetes rather than its absence, since the ratio of false negative cases is almost 50% (Table 31).

Table 31 - Accuracy of Hormonology for univariate classification of type 2 diabetes.

Variable	AUC	SE	p	LBCI	UBCI	Cut-off	Sens	Spec	+LR
TSH	0.590	0.06	0.184	0.462	0.699	-	-	-	-
C-Peptide	0.688	0.06	0.002	0.575	0.802	$\leq$ 1.35	51.06	88.10	4.29

Performing Receiver Operating Characteristic curve analysis on variables that presented significant differences, we find out that although there is a statistical significant difference on the retinal nerve fiber layer regarding the temporal-inferior quadrant, this measure cannot differentiate groups. Nevertheless, the thickening of the retinal nerve fiber layer on the temporal hemi field may separate diabetic patients from controls, mainly by excluding the presence of the disease, since the value obtained for specificity is unacceptable (Table 32).

Table 32 - Accuracy of OCT tests for univariate classification of type 2 diabetes.

	Variable			AUC	SE	p	LBCI	UBCI	Cut-off	Sens	Spec	+LR
	IOP			0.583	0.082	0.317	0.421	0.744	-	-	-	-
	BVCA			0.649	0.057	0.012	0.537	0.761	≤ 0.90	38.78	93.48	5.95
OCT	VS	CS		0.544	0.060	0.459	0.428	0.661	-	-	-	-
		Inner	N	0.485	0.060	0.803	0.368	0.602	-	-	-	-
			S	0.514	0.060	0.809	0.397	0.631	-	-	-	-
			T	0.468	0.060	0.587	0.350	0.585	-	-	-	-
			I	0.448	0.060	0.380	0.331	0.564	-	-	-	-
		Outer	N	0.500	0.060	0.994	0.383	0.618	-	-	-	-
			S	0.524	0.060	0.688	0.406	0.642	-	-	-	-
			T	0.443	0.060	0.335	0.326	0.559	-	-	-	-
			I	0.442	0.060	0.331	0.326	0.558	-	-	-	-
	RNFL	Global		0.506	0.060	0.916	0.387	0.626	-	-	-	-
		Nasal		0.524	0.060	0.683	0.407	0.642	-	-	-	-
		N	S	0.465	0.060	0.563	0.347	0.583	-	-	-	-
			I	0.483	0.060	0.782	0.365	0.602	-	-	-	-
		Temporal		0.642	0.060	0.018	0.530	0.754	≥ 67.50	72.92	50.00	1.46
		T	S	0.548	0.060	0.425	0.431	0.665	-	-	-	-
			I	0.599	0.060	0.100	0.484	0.713	-	-	-	-

Evaluating the area under the ROC curve for each one of these psychophysical parameters (Figure 36), we find out that diabetic patients can be discriminated by the speed test integrated on the psychophysical tests, according to all meridians, being the 135º meridian the one that presents higher accuracy in prediction (AUC = 0,731; p < 0,001), in spite of no statistical difference (p = 0,186) to meridian 45º, the measure with the worst accuracy. Note that all of these measures may separate groups, but each one of them only presents acceptable values either in sensitivity, or in specificity, but not in both. Thus, if used for detecting diabetes, they should be applied sequentially, that is, first we should look at meridians 90º and 135º. Afterwards, if they present abnormal values, equal or higher than cut-offs defined in Table 33, and if values are also higher in meridians 0º and 45º, probably we have a type 2 diabetic case. Thereby, speed area involving all this four meridians may be used to detect diabetes, with a positive likelihood ratio of 5.25.

Table 33 - Accuracy of Speed test for univariate classification of type 2 diabetes.

Speed		AUC	SE	p	LBCI	UBCI	Cut-off	Sens	Spec	+LR
Meridian	0º	0.707	0.06	0.001	0.599	0.814	≥ 1.59	55.56	81.82	3,06
	45º	0.649	0.06	0.018	0.533	0.765	≥ 2.17	35.71	88.64	3,14
	90º	0.687	0.06	0.002	0.577	0.797	≥ 0.86	76.09	53.49	1,64
	135º	0.731	0.0	< 0.0001	0.619	0.844	≥ 1.62	64.29	59.09	1,57
Area		0,728	0.06	< 0.0001	0.614	0.840	≥ 2.86	52.50	90.00	5.25

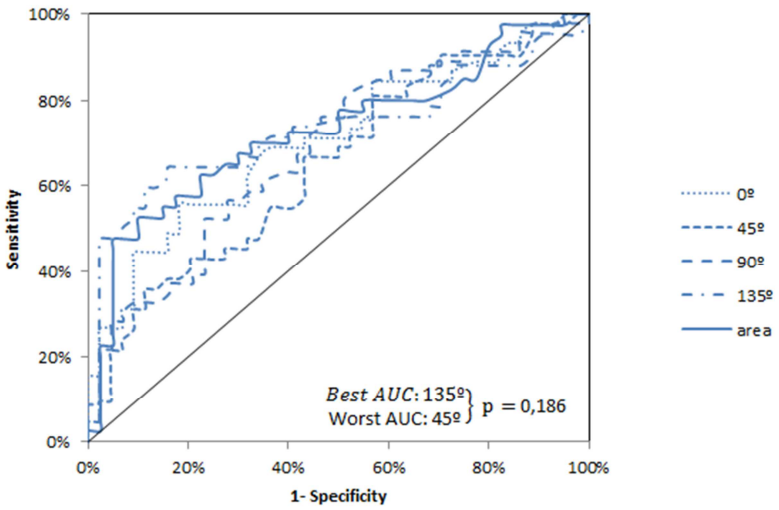

Figure 36 - ROC curve for Speed test.

As expected, since no statistical significant differences were found in achromatic vision in meridians 0º, 45º and 135º, none of the values measured in each one of these three meridians evaluated, or even in the area of the polygons generated by the median points of each meridian, are capable of separating controls from patients (Figure 37). However, for the 90º meridian, almost 70% of type 2 diabetics have values higher than 2.32, and 65% of controls present values below 2.32. In fact, it is two times more probable that a value equal or higher than 2.32 shows in a type 2 diabetic than in a control (Table 34).

Table 34 - Accuracy of Achromatic contrast sensitivity test for univariate classification of type 2 diabetes.

Achromatic		AUC	SE	p	LBCI	UBCI	Cut-off	Sens	Spec	+LR
Meridian	0º	0.450	0.062	0.416	0.329	0.572	-	-	-	-
	45º	0.608	0.061	0.082	0.488	0.727	-	-	-	-
	90º	0.671	0.056	0.005	0.560	0.782	≥ 2.32	69.57	65.22	2.00
	135º	0.564	0.063	0.306	0.440	0.688	-	-	-	-
Area		0.580	0.06	0.179	0.463	0.696	-	-	-	-

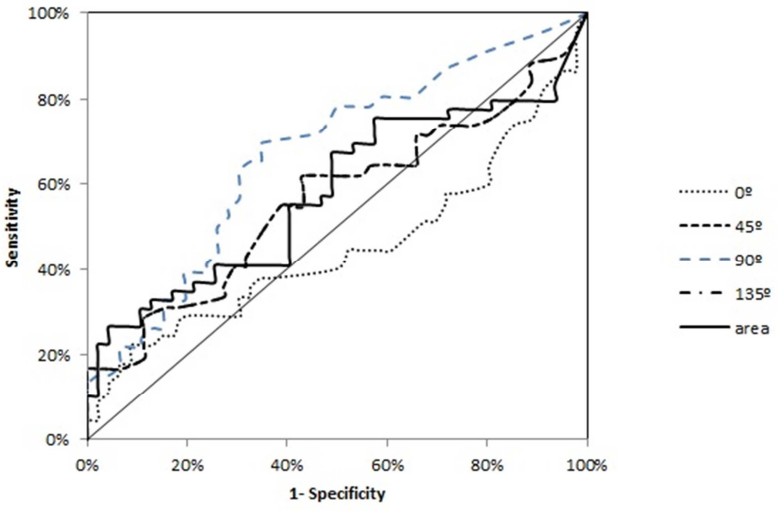

Figure 37 - ROC curve for Achromatic contrast sensitivity test.

Concerning chromatic contrast sensitivity, the Tritan axis is the one that most discriminates groups, since we can only find discrimination between patients and diabetics relatively to the Protan axis at the 0º meridian, and at the 0º and 45º meridians of the Deutan axis. In the Tritan axis, all the four meridians enable discrimination between groups, as well as the total area of the polygon generated by the medians of each one of the meridians (Table 35). Moreover, we should point out that positive likelihood ratio is 10 for the 0º meridian. ROC curves for this discriminating parameters are presented on Figures 38 (Protan), 39 (Deutan) and 40 (Tritan).

Table 35 - Accuracy of Chromatic contrast vision test for univariate classification of type 2 diabetes.

Axis/Merid.		AUC	SE	p	LBCI	UBCI	Cut-off	Sens	Spec	PPV
Chromatic Contrast	Protan 0º	0.623	0.06	0.043	0.508	0.738	$\geq 2.47 \times 10^{-3}$	69.57	65.22	2.00
	Protan 45º	0.520	0.06	0.747	0.397	0.643	-	-	-	
	Protan 90º	0.464	0.06	0.555	0.345	0.584	-	-	-	
	Protan 135º	0.454	0.06	0.462	0.330	0.579	-	-	-	
	Protan Area	0.499	0.06	0.987	0.377	0.621	-	-	-	
	Deutan 0º	0.725	0.05	< 0.001	0.619	0.832	$\geq 3.40 \times 10^{-3}$	65.12	78.26	3.00
	Deutan 45º	0.627	0.06	0.042	0.509	0.745	$\geq 8.34 \times 10^{-3}$	60.98	63.04	1.65
	Deutan 90º	0.595	0.06	0.120	0.477	0.713	-	-	-	
	Deutan 135º	0.555	0.07	0.380	0.428	0.682	-	-	-	
	Deutan Area	0.657	0.06	0.013	0.541	0.773	-	-	-	
	Tritan 0º	0.724	0.05	< 0.001	0.620	0.828	$\geq 67.20 \times 10^{-3}$	44.44	95.56	10.01
	Tritan 45º	0.674	0.06	0.005	0.561	0.787	$\geq 59.99 \times 10^{-3}$	90.48	40.00	1.51
	Tritan 90º	0.715	0.05	< 0.001	0.610	0.821	$\geq 77.13 \times 10^{-3}$	58.70	80.00	2.94
	Tritan 135º	0.716	0.06	0.001	0.605	0.828	$\geq 109.11 \times 10^{-3}$	60.98	82.22	3.43
	Tritan Area	0.731	0.06	< 0.001	0.624	0.829	$\geq 6.16 \times 10^{-3}$	75.61	71.11	2.62

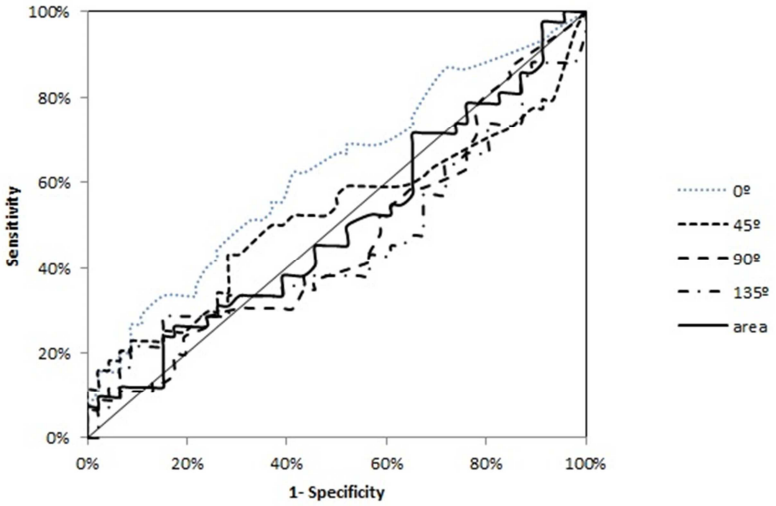

Figure 38 - ROC curve for chromatic contrast sensitivity test (Protan).

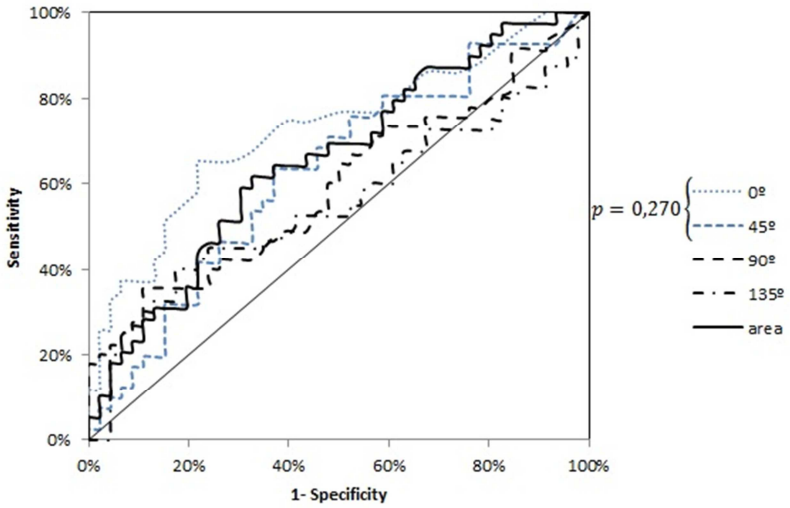

Figure 39 - ROC curve for chromatic contrast sensitivity test (Deutan).

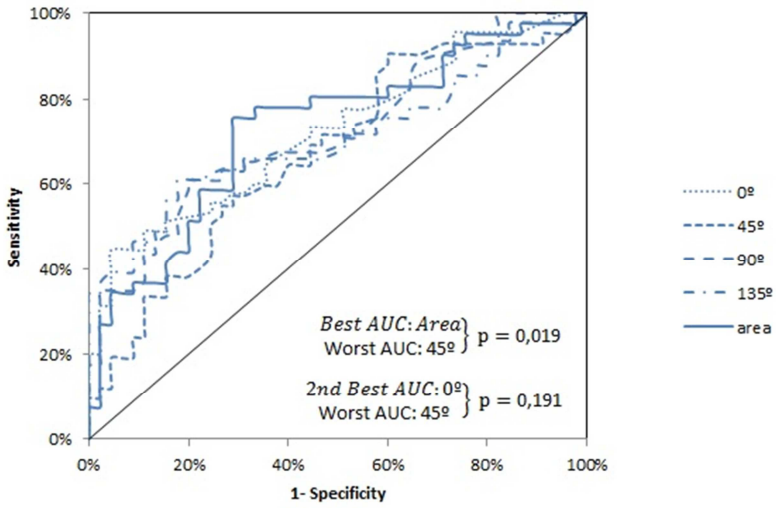

Figure 40 - ROC curve for chromatic contrast sensitivity test (Tritan).

3. Multivariate Models for Diabetes Classification

The following analyses were performed using as independent variables all the previous categorical variables that presented association with diabetes, such as blood pressure controlled by medication (diagnosed hypertension), considering age, body mass index and best corrected visual accuracy as covariates, and all numerical variables which presented statistical differences between groups and achieved statistical significance in area under the ROC curve, such as:

- parameters of the liver and biliar ductus: ALT, alkaline phosphatase and gamma GT;
- parameters associated to lipids: cholesterol (total, HDL and LDL), atherogenic index, triglycerides and apolipoprotein A1;
- cytometry parameters: leucocytes, haemoglobin, haematocrit and erythrocyte variation coefficient;
- retinal nerve fiber layer from OCT: temporal quadrant;
- visual psychophysical tests: speed (all meridians and global area), achromatic vision (meridian 0º), chromatic vision on Protan (meridian 0º), Deutan (meridians 0º and 45º) and Tritan (all meridians and global area) axes.

These parameters were dichotomized by determining the optimal cut-off to use on logistic regression, but discriminant and decision tree models used the quantitative variables, in order to evaluate and compare models and to reach the best one for diabetes classification.
Glucose levels and glycosylated haemoglobin were not considered since these were the ones used to diagnose diabetes.

3.1 Discriminant Function Analysis

One single discriminant function was obtained, using the Wilks' lambda method and a stepwise procedure based on the F probability (< 0.050 to enter; > 0.100 to remove), and classification was performed based on the minimization of the within groups covariance matrix.

Although numerical variables are not from a multivariate normal distribution (Figure 41) and covariate matrices are not homogeneous (Box's M $F(28, 13141) = 3.57$; $p < 0.001$), discriminant analysis may be performed, as explained in the methods section, with the possible consequence of increasing the number of cases classified as diabetic; however, since groups are distributed in identical proportions, that is unlikely to occur.

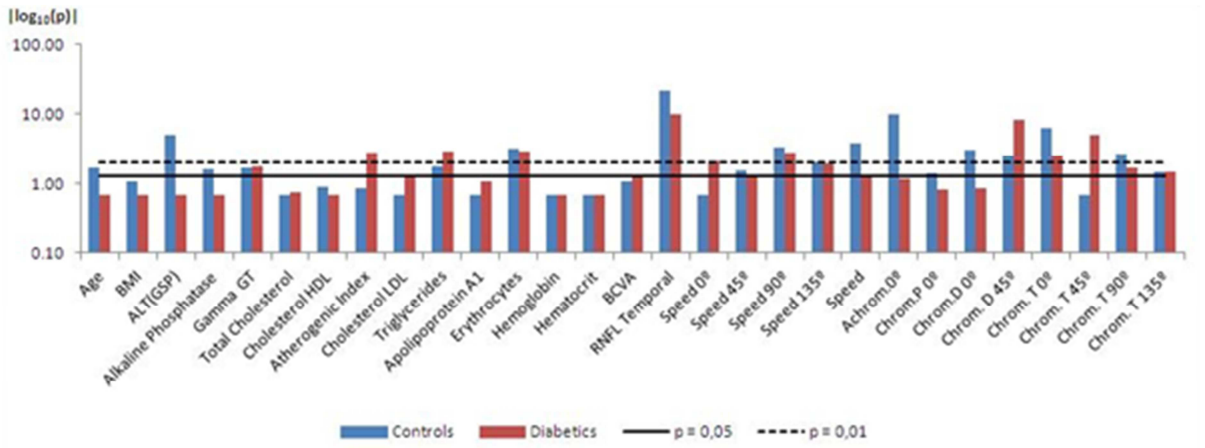

Figure 41 - p-values obtained from the Kolmogorov-Smirnov or the Shapiro-Wilk test to variables in analysis presented as $|\log_{10} p|$, in logarithmic scale. The horizontal lines reflect the values of 0.01 ($|\log_{10} 0.01| = 2.00$) and 0.05 ($|\log_{10} 0.05| = 1.30$) for type I errors. All bars below horizontal lines represent variables with normal distribution in the group.

Wilk's lambda identifies seven variables (Table 36) for group discrimination, and age was not identified as a separating variable, although it was initially considered as a potential discriminator, since groups were not matched for age. Thus, we may consider that coefficients of the identified variables are adjusted to age.

Table 36 - Variables included in the discriminant model (Wilks' Lambda method).

	Variables	Wilks' Lambda							
		Statistic	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
x_1	Hypertension	0.606	1	1	63	41.03	1	63	< 0.001
x_2	Chrom. Cont. Tritan (135º)	0.497	2	1	63	31.41	2	62	< 0.001
x_3	Cholesterol HDL	0.449	3	1	63	25.00	3	61	< 0.001
x_4	Peptide C	0.380	4	1	63	24.43	4	60	< 0.001
x_5	RNFL (Temporal)	0.341	5	1	63	22.77	5	59	< 0.001
x_6	Triglycerides	0.301	6	1	63	22.43	6	58	< 0.001
x_7	BMI	0.280	7	1	64	20.95	7	57	< 0.001

After seven steps, one discriminant function is obtained with an eigenvalue of 2.57, explaining 100% of the variance, and a canonical correlation of 0.849 (λ_{Wilks}=0.280: $\chi^2_{(7)}$=75.77; p < 0.001) between variables entering into the model and group classification.
The discriminant function can be written as:

$$F(x) = -3.233 + 1.548x_1 + 3.274x_2 - 0.039x_3 - 0.998x_4 + 0.034x_5 + 0.008x_6 + 0.090x_7$$

or, after standardizing coefficients,

$$F_S(x) = 0.606x_1 + 0.295x_2 - 0.490x_3 - 0.945x_4 + 0.548x_5 + 0.538x_6 + 0.345x_7$$

Function at group centroids assume the values of -1.462 for controls and +1.706 for diabetics, and pairwise group comparisons show significant differences between centroids ($F(7,57) = 20.95$; $p < 0.001$), meaning that the function can separate or discriminate groups and may be used for classification.

Classical classification may be performed using classification functions, obtained from the discriminant function, although we get no information about posterior probabilities. A case will be classified as diabetic as long as it has the value of the classification function for that group is higher than for the control group. The classification functions are given by:

$$\begin{cases} Class_C = -59.203 - 1.820x_1 + 5.168x_2 + 0.457x_3 - 1.761x_4 + 0.327x_5 + 0.051x_6 + 2.213x_7 \\ Class_D = -65.832 + 3.084x_1 + 15.539x_2 + 0.334x_3 - 4.992x_4 + 0.437x_5 + 0.075x_6 + 2.497x_7 \end{cases}$$

In order to obtain posterior probabilities, classification is performed so that a new or an old case is classified into the group which the centroid is closer and, in this specific case of two groups, it can be thought as dividing the discriminant space into two mutually exclusive regions, defining the frontier line by the weighted mean of the centroids, which leads to the value $f = +0.122$. The notion of closeness to centroid is performed by the determination of the Mahalanobis distance from the score obtained in the discriminant function to the centroid, and based upon this it is possible to improve the classification procedure since we become able to determine the probability of a given subject to be classified in a group, given the score obtained in the discrimination function. For points in the frontier line, this probability is 50%.

Using the discriminant function we can predict posterior probabilities for each subject and classify the subject, or new ones, according to the highest probability. For the determination of the posterior probabilities for a given subject, we need to obtain the squared Mahalanobis distances between the score obtained in the discriminant function for that subject, f(x), and each group centroid. These distances, d_C^2 and d_D^2, follow a chi-square distribution with one degree of freedom, and are given by:

$$\begin{cases} d_C^2 = \frac{f(x)+1.462}{0.892} \\ d_D^2 = \frac{f(x)-1.706}{1.099} \end{cases}$$

Posterior probabilities are given applying Bayes rule to the probability of obtaining that distance given that the subjects belong to a defined group and are defined as:

$$\begin{cases} P(G_1|d_C^2) = \frac{p_1 x P(\chi_1^2 > d_C^2)}{p_1 x P(\chi_1^2 > d_C^2) + p_2 x P(\chi_1^2 > d_D^2)} \\ P(G_2|d_D^2) = \frac{p_2 x P(\chi_1^2 > d_D^2)}{p_1 x P(\chi_1^2 > d_C^2) + p_2 x P(\chi_1^2 > d_D^2)} \end{cases}$$

Where p_1 and p_2 are the prior probabilities, which were assumed to be equal (0.50). The subject is classified as control or type 2 diabetic according to the highest posterior probability.

We can apply a ROC analysis either to the discriminant function, or to the posterior probabilities, and a cut-off of +0,264 for the frontier line is obtained as the optimal cut-off, corresponding to the posterior probability of 61.04% (Table 37). This means that we may improve the specificity of the classification, since the number of false positive cases decreases, without losing sensitivity; consequently, the positive likelihood ratio increases three times which is preferable. Note that concordance between models is excellent thus any of them may be used for classification.

Table 37 - Discriminant classifier accuracy using two different cut-offs for posterior probability: classical (50%) and obtained by ROC analysis (61,04%).

Model	AUC (p)	Cut-off Function	% Correct	k	p	between	McNemar (p)	Sens	Spec	+LR
D_F	0.985 (< 0.001)	0.122 (50.00%)	92.31%	0.846	< 0.001	0.949 (p < 0.001)	1.000	92.31%	92.31%	12.00
D_{ROC}		0.264 (61.04%)	94.87%	0.897	< 0.001		0.625	92.31%	97.44%	36.00

On the following scaterplott of the probability for group classification (Figure 42), we can observe the posterior probability of belonging to the control or diabetic group according to the value obtained in the discriminant function, and its distance to the correspondent centroid. Horizontal lines mark the cut-offs defined by discriminant analysis (50%) and ROC analysis (61.04%) for group classification while vertical lines mark the cut-offs defined by the definition frontier line (0.12) or obtained by ROC analysis (0.26), used for classification in the discriminant function.

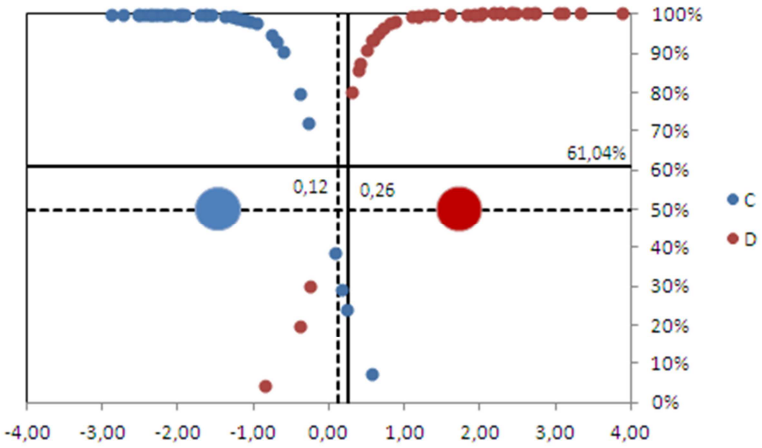

Figure 42 - Scaterplott of the probability for group (y) control (blue) or diabetic (red), based on the value of the discriminant funcion (d). Horizontal lines mark the cut-offs defined by discriminant analysis (50%) and ROC analysis (62,16%) for group classification while vertical lines mark the cut-offs defined by the definition frontier line (-0,12) or obtained by ROC analysis (-0,28), used for classification in the discriminant function; big circles mark the centroid for the discriminant function (at 50% probabilituy) for each group.

3.2 Regression procedures

When discriminant analysis fails the assumptions, it is usual to perform logistic regression. Logistic regression is a method designed to handle either numerical or categorical independent variables. Usually, logistic regression can not quantify differences in one unit of each independent variable, especially when there is a large dispersion, and it behaves better when categorical variables are used. Therefore, only categorised variables (according to cut-offs determined in ROC analysis) will be used, with exception to age at visit date, given the relevance of explaining away this variable. Initially, we intended to test both numeric and categorical variables in order to compare the performance of discriminant and regression procedures, but no model could be obtained when numerical variables were used. Three models are proposed in order to evaluate and compare classification power. The methodology was based on a forward stepwise procedure (with a probability of F to enter < 0.050 and probability of F to remove > 0.100), based either on the conditional statistic, or the Likelihood Ratio (LR) or the Wald statistic.

Basically, models obtained using the conditional statistic or the likelihood ratio lead to the same final results, after seven iterations, which are very similar to the results obtained when applying the Wald statistic and all models present statistical significant improvement in every step until the seventh step is reached. When variable age is included in the model, there is a significant improvement in models using conditional statistic or LR, but model which applies the Wald statistic looses significance (Table 38).

Table 38 - Significance of models and improvement, step by step, on forward stepwise logistic regression model (Conditional, Likelihood Ratio and Wald's methods).

Step	Variable	Improvement ($\chi^2_{(df)}$; p)			Model ($\chi^2_{(df)}$; p)		
		Cond.	LR	Wald	Cond.	LR	Wald
x_1	Hypertension	27.55(1); <0.001	27.55(1); <0.001	27.55(1); <0.001	27.55(1); <0.001	27.55(1); <0.001	27.55(1); <0.001
x_2	Achrom. 90º	13.68(1); <0.001	13.68(1); <0.001	13.68(1); <0.001	41.23(2); <0.001	41.23(2); <0.001	41.23(2); <0.001
x_3	BMI	9.24(1); 0.002	9.24(1); 0.002	9.24(1); 0.002	50.47(3); <0.001	50.47(3); <0.001	50.47(3); <0.001
x_4	Alipoprotein A1	8.2(1); 0.004	8.2(1); 0.004	8.2(1); 0.004	58.67(4); <0.001	58.67(4); <0.001	58.67(4); <0.001
x_5	RNFL (T)	9.83(1); 0.002	9.83(1); 0.002	9.83(1); 0.002	68.5(5); <0.001	68.5(5); <0.001	68.5(5); <0.001
x_6	C-Peptide	7.93(1); 0.005	7.93(1); 0.005	7.93(1); 0.005	76.43(6); <0.001	76.43(6); <0.001	76.43(6); <0.001
x_7	Age	13.29(1); <0.001	13.29(1); <0.001	-8.87(1)*; 0.003	89.72(7); <0.001	89.72(7); <0.001	67.57(5); <0.001

* a negative chi-square indicates that the chi-square value has decreased from previous step

Furthermore, there is an excellent adjustment to data in every step, given by the Hosmer and Lemeshow test which is always not significant, and the pseudo R-squared given by Nagelquerke R Square shows that there is a high degree of concordance and explained variability given by the selected variables. However, we should not consider steps six and seven, since there is a notorious over-fitting of the solution found with conditional and LR statistic, and a loss when using the Wald statistic (Table 39). This means that, despite of differences in age, between groups, there must be some correlation between age and values of C-peptide with the other variables previously included into the model, until step 5, which present higher risk for development of type 2 diabetes.

This means that despite of the differences in age, between groups, there must be some correlation between age and values of C-peptide with the other variables previously included into the model, which present higher risk for type 2 diabetes than age or C-peptide levels.

Table 39 - Adjustment of the model, step by step, to observed data, and overall correlation.

Step	Variable	Nagelkerque R square			Hosmer and lemeshow test ($\chi^2_{(df)}$; p)		
		Cond.	LR	Wald	Cond.	LR	Wald
x_1	Hypertension	0.462	0.462	0.462	0(0); -	0(0); -	0(0); -
x_2	Achrom. 90º	0.628	0.628	0.628	1.72(2); 0.423	1.72(2); 0.423	1.72(2); 0.423
x_3	BMI	0.721	0.721	0.721	2.37(6); 0.883	2.37(6); 0.883	2.37(6); 0.883
x_4	Alipoprotein A1	0.794	0.794	0.794	1.42(6); 0.964	1.42(6); 0.964	1.42(6); 0.964
x_5	RNFL (T)	0.870	0.870	0.870	0.51(7); 0.999	0.51(7); 0.999	0.51(7); 0.999
x_6	C-Peptide	0.924	0.924	0.924	1.94(6); 0.925	1.94(6); 0.925	1.94(6); 0.925
x_7	Age	1.000	1.000	0.864	0.00(5); 1.000	0.00(5); 1.000	1.89(6); 0.930

Note that the model with best adjustment to data is reached on step six, since there is a loss on data adjustment between step six and steps five and four. To join this, coefficients obtained show statistical significance until step five, inclusive, losing their statistical significance if iterations go beyond step five, thus the model that will be presented for classification is the model obtained after five iterations.

In fact, the performance of predictions is the presented in the nest table (Table 40), but it is clear that there is an over-fitting of the studied models. The model obtained at step five is more realistic, since it presents a good fitting and all coefficients are significant, therefore it is the one to be used.

Table 40- Evaluation of the accuracy of developed logistic regression models.

Step	Model	AUC (p)	p (AUC)	% Correct	k	p	McNemar (p)	Sens	Spec	+LR
Final	Cond./ LR	0.980 (< 0.001)	0.121	97.53%	0.951	< 0.001	1.000	97.56%	97.50%	39.02
	Wald	0.950 (< 0.001)		87.80%	0.755	< 0.001	0.344	92.86%	82.50%	5.31
5	All*	0.942 (< 0.001)	-	89.66%	0.793	< 0.001	0.508	83.87%	90.48%	8.81

* Cut-off for posterior probability obtained by ROC analysis presented the probability value equal or higher than 51.03%, resulting in the same sensitivity, specificity and positive likelihood ratios.

Thereby, to avoid over-fitting and keep statistical significance of the coefficients for classification, the model obtained after five steps for determination of the probability for type 2 diabetes is

$$P(D) = \frac{e^{-11.045+4.393x_1+3.959x_2+3.851x_3+4.002x_4+4.113x_5}}{1+e^{-11.045+4.393x_1+3.959x_2+3.851x_3+4.002x_4+4.113x_5}},\ x_i \in \{0,1\}, i = \overline{1,5}$$

Significance of model coefficients and confidence intervals for odds ratio for the model explicit on the previous equation are presented on the next table (Table 41):

Table 41 - Odds ratio and confidence intervals for variables identified on logistic regression model (step 5).

	B	SE	Wald	df	Sig.	Odds Ratio	95% CI for Odds Ratio	
							Lower	Upper
Hypertension	6.57	2.29	8.23	1.00	0.004	711.00	8.01	63144.17
Achrom. 90º	4.93	1.80	7.54	1.00	0.006	138.25	4.10	4664.25
BMI	3.85	1.59	5.89	1.00	0.015	47.06	2.10	1054.84
Alipoprotein A1	4.00	1.62	6.14	1.00	0.013	54.69	2.31	1296.07
RNFL (T)	4.11	1.75	5.52	1.00	0.019	61.14	1.98	1889.53

The traditional cut-off for this classification is the probability of 50%, thus, given a subject,

he will be classified as control or diabetic according to the following rule:

$$\begin{cases} Diabetic\ if\ P(D) \geq 0.5 \\ Control\ if\ P(D) < 0.5 \end{cases}$$

However, we may perform a ROC analysis on that probability, and the obtained optimal cut-off is 51.03%. By lowering the cut-off, there is no gain in sensitivity or in specificity, the models are exactly equal, presenting a perfect agreement with the one used with the cut-off of 0.50 (Cohen's kappa = 1.000; $p < 0.001$), hence the classical cut-off for classification will be used.

3.3 Decision trees

Decision trees are a multivariate process without assumptions on data distribution, except for multicolinearity, as logistic regression, but both handle quite well with this problem since decision trees are constructed based on a stepwise algorithm, as stepwise logistic regression.

Trees were constructed applying the CART, the CHAID or Exhaustive CHAID, and the QUEST algorithms, using continuous independent variables, exception made for diagnosed hypertension. Moreover, decision tree analysis may identify multivariate cut-off values for classification.

The obtained models were similar two by two, as observed on Figure 43:

- the CART and the QUEST algorithms use age as the first discriminating variable, thus predictions may be used separately according to the age group, and markers of type 2 diabetes are different according to age considering the cut-off of 51 years and 6 months.

  In fact, both algorithms leads to similar results in the older group (age $\geq$ 51.5), since the loss chromatic contrast sensitivity over the Tritan axis measured at 45º is chosen as a classifier for diabetes, with the same cut-off obtained by ROC analysis (59.99×10^{-3}), but while the QUEST algorithm stops splitting here, the CART algorithm splits the diabetic group into two more nodes, considering haemoglobin (whatever the gender) as a splitting variable.

  In the younger group splitting may be done using results on Speed test (CART) or body mass index (QUEST), and both identify diabetes with 80% probability.
- the CHAID and the Exhaustive CHAID algorithms choose the presence of hypertension as the first splitting variable, classifying a subject as type 2 diabetic with 83.61% probability

in the presence of hypertension. For normotensive subjects, both models are able to identify if value measures on the 135º meridian at the Tritan axis on chromatic vision exceed $134x10^{-3}$ and, therefore, the subject is classified as a type 2 diabetic with probability of 87.50%.

Models obtained using CART or QUEST algorithms perform quite well on the training sample, and with similar accuracy, evaluated by the area under the ROC curve, than models obtained using CHAID or Exhaustive CHAID algorithms. In fact, these two models are the ones that present higher sensitivity (Table 42), in spite of lower specificity which is concordant to statistical differences found with the McNemar test.

Table 42- Evaluation of the accuracy of developed decision tree models.

Model	AUC (p)	p	% Correct	k	p	McNemar (p)	Sens	Spec	+LR
T1	0.882 (< 0.001)	*	82.29%	0.629	< 0.001	0.332	82.86%	81.97%	4.59
T2	0.800 (0.048)		80.43%	0.615	< 0.001	< 0.001	94.29%	71.93%	3.36
T3	0.789 (< 0.001)		79.35%	0.591	< 0.001	0.004	91.43%	72.93%	3.26
T4	0.860 (< 0.001)		81.25%	0.619	< 0.001	0.031	88.57%	77.05%	3.86

* No statistical differences (De Long Test) were found between area under the ROC curve for posterior probabilities of all pairs of decision tree models; the minimum p-value obtained, unadjusted for multiple comparisons, was between T1 and T3 (p = 0.052).

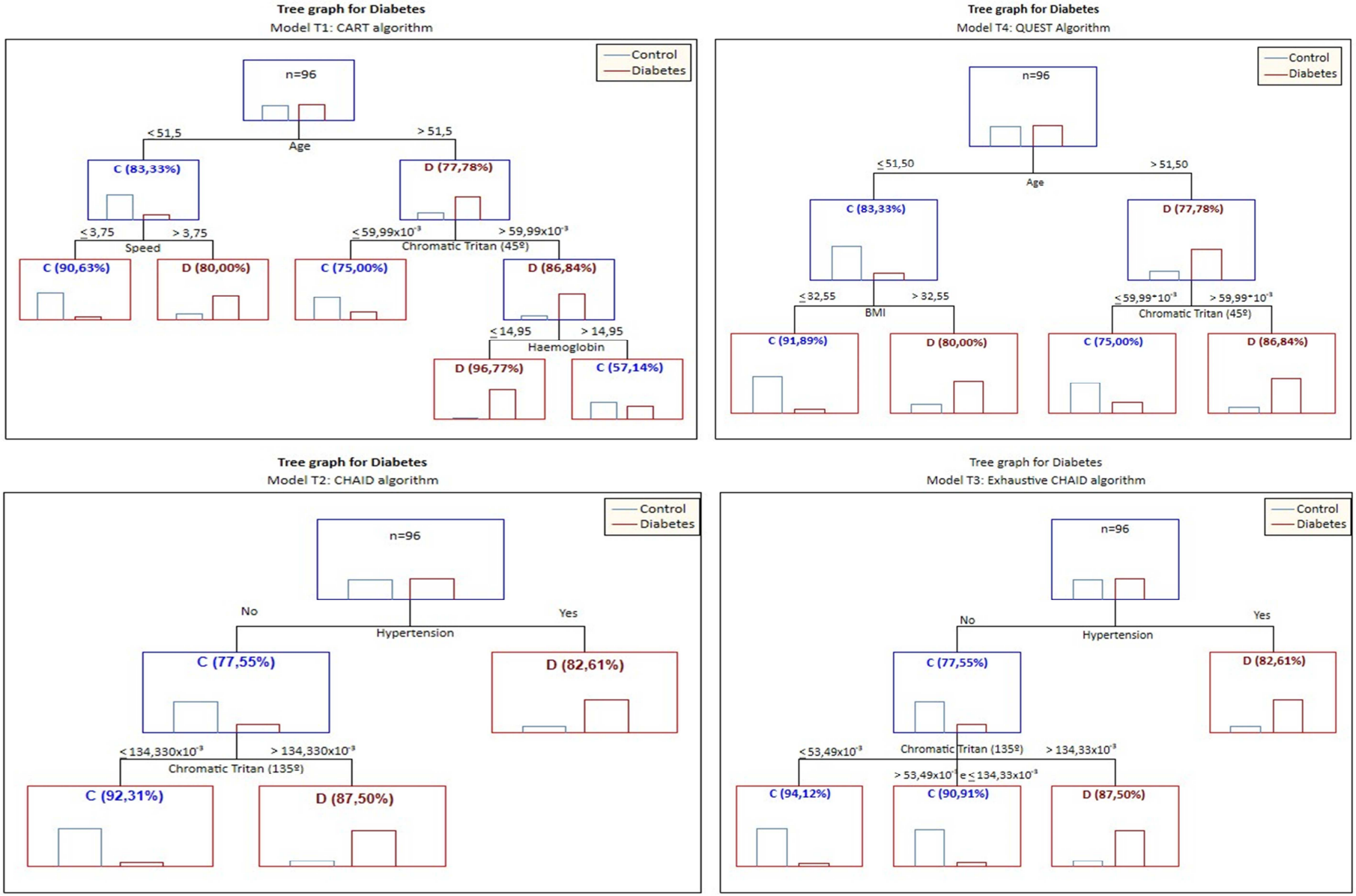

Figure 43 – Decision trees obtained with CART, CHAID / Exhaustive CHAID and QUEST algorithms.

1. Model Comparison – applying obtained models on a test sample

The presented models will be identified according to the used statistical methodology: D_f or D_{ROC} for models obtained by discriminant analysis; L_f or L_{ROC} for Logistic regression; T_1 and T_4 for Decision Tree analysis. Note that models D_F and D_{ROC}, or models L_F or L_{ROC} give the same formula for determining the posterior probabilities, but they differ due to the cut-off used for classification. In the training sample results were the same but they are not necessarily equal on the test sample.

The test sample considered 57 subjects, 30 of them controls (52.63%) and 27 diabetics (47.37%). Controls were aged between 41 and 72 years (53.53 $\pm$ 9.20), with median 50 years, and type 2 diabetics were aged between 47 and 73 years old (60.11 $\pm$ 7.90), with median 61 years. However, the application of the previously developed models was possible in fewer cases, since not all the data are available yet. Therefore, further confirmation must be obtained in the future.

Descriptive statistics obtained for the test sample are presented below (Table 43):

Table 43 - Descriptive statistics on the test sample.

		N	Min	Max	Mean	SD	P50	P25-P75	Low Risk	Hight Risk
Age	C	30	41.00	72.00	53.63	9.20	50.00	46.25-61.75	3 (13.64%)	19 (86.36%)
	D	27	47.00	73.00	60.11	7.90	61.00	53.00-65.00	16 (76.19%)	5 (23.81%)
BMI	C	23	19.90	33.30	24.55	3.39	24.10	21.70-27.30	17 (73.91%)	6 (26.09%)
	D	21	20.50	38.30	30.49	4.99	30.60	26.95-34.70	5 (23.81%)	16 (76.19%)
Cholesterol HDL	C	25	33.00	75.00	58.20	13.39	57.00	47.00-72.50	20 (80%)	5 (20%)
	D	26	23.00	56.00	36.04	9.14	35.00	28.00-40.25	4 (15.38%)	22 (84.62%)
Triglycerides	C	25	49.00	224.00	108.52	46.33	100.00	75.50-124.00	17 (68%)	8 (32%)
	D	26	75.00	318.00	138.96	52.77	122.00	103.25-158.00	10 (38.46%)	16 (61.54%)
Apolipoprotein A1	C	26	121.00	216.00	165.27	26.12	170.50	138.75-187.25	22 (84.62%)	4 (15.38%)
	D	27	88.00	170.00	120.63	20.23	116.00	108.00-132.00	6 (22.22%)	21 (77.78%)
Hemoglobin	C	24	7.90	16.70	14.13	1.74	14.60	13.30-15.00	22 (91.67%)	2 (8.33%)
	D	25	9.20	16.30	13.25	1.92	13.30	11.95-14.70	14 (56%)	11 (44%)
RNFL (Temporal)	C	13	57.00	124.00	77.62	19.17	79.00	60.00-86.50	5 (38.46%)	8 (61.54%)
	D	13	45.00	84.00	62.62	11.42	62.00	53.50-71.50	10 (76.92%)	3 (23.08%)
Achrom. Contrast (90º)	C	16	1.00	3.61	2.02	0.80	2.06	1.25-2.59	11 (68.75%)	5 (31.25%)
	D	20	1.51	5.92	3.18	1.09	3.10	2.26-3.93	5 (25%)	15 (75%)
Chrom. Cont. Tritan (45º)	C	17	0.03	0.22	0.09	0.06	0.08	0.04-0.12	7 (41.18%)	10 (58.82%)
	D	12	0.05	0.37	0.17	0.11	0.18	0.06-0.24	3 (25%)	9 (75%)
Chrom. Cont. Tritan (135º)	C	17	0.02	0.26	0.11	0.07	0.11	0.03-0.16	9 (52.94%)	8 (47.06%)
	D	11	0.05	0.24	0.15	0.07	0.15	0.08-0.21	4 (36.36%)	7 (63.64%)
Speed (Area)	C	14	0.12	29.42	2.95	7.63	0.85	0.61-1.43	13 (92.86%)	1 (7.14%)
	D	15	0.17	8.01	1.74	1.96	1.19	0.29-2.44	14 (93.33%)	1 (6.67%)

As observed in the previous table (Table 43), data were not available for all the subjects, considering all the variables, therefore, we tested discriminant analysis classifier in 23

subjects (7 controls and 16 diabetics), the logistic regression models in 33 subjects (10 controls and 23 diabetics) and decision tree models in 57 subjects (18 controls and 39 diabetics).

The percentage of agreement observed is good to very good. Agreement was not due to chance, being moderate when measured by Cohen's kappa when decision tree algorithms are applied, especially with the CART algorithm since the Cohen's kappa is 0.404. Discriminant analysis models are the ones with higher concordance, which can be stated as substantial to good (k = 0.620; p = 0.002), followed by models T3 and T4, with moderate to substantial concordance (Table 44). All the models present similar rates of false positive and false negative values, since McNemar test does not detect any significant differences.

Table 44 - Evaluation of developed models on the test sample – concordance and disagreement.

Model	% Correct	K	P	McNemar (p)
D_F	82.61%	0.620	0.002	0.625
L	81.82%	0.570	0.001	1.000
T1	71.93%	0.404	0.002	0.210
T2	86.67%	0.676	< 0.001	0.667
T3	86.67%	0.502	0.001	0.667
T4	80.70%	0.547	< 0.001	1.000

The model accuracy is presented by the area under the ROC curve determined for posterior probabilities and the 95% of expected values for sensitivity, specificity and positive likelihood ratio are determined (Table 45).

Table 45 - Evaluation of the accuracy of the developed models on the test sample.

Model	AUC (p)	Sensivity (95% CI)		Specificity (95% CI)		+LR (95% CI)	
D_F	0.937 (0.001)	81.25%	54.4%-96.0%	85.71%	42.1%-99.6%	5.69	0.90-35.40
L	0.935 (< 0.001)	86.96%	66.4%-97.2%	70.00%	34.8%-93.3%	2.90	1.10-7.60
T1	0.769 (0.001)	71.79%	55.1%-85.0%	72.22%	46.5%-90.3%	2.58	1.20-5.60
T2	0.894 (< 0.001)	87.88%	71.8%-96.6%	83.33%	51.6%-97.9%	5.27	1.50-18.80
T3	0.894 (< 0.001)	87.88%	71.8%-96.6%	83.33%	51.6%-97.9%	5.27	1.50-18.80
T4	0.822 (< 0.001)	87.18%	72.6%-95.7%	66.67%	41.0%-86.7%	2.62	1.30-5.10

Models developed by discriminant analysis or logistic regression procedures, based on different cut-offs for the probability of presence of type 2 diabetes, lead to the same solution when applied to this test sample. Decision tree algorithms lead to different

solutions and the model based upon the CART algorithm (T1) presents the lowest accuracy (AUC = 0.769), being also the one that shows lower positive likelihood ratio, although it is very similar to the tree generated by the QUEST algorithm. Note that all the models present reasonable to good values on observed specificity, but none of them shows statistical significance on this parameter, since confidence intervals include the 50% value, meaning that specificity may be below 50% on the population. On the other hand, models obtained by discriminant analysis methods do not reach significance on the positive likelihood ratio, since value one is included in 95% confidence intervals.

From this point of view, and as we intend to build the simplest classification model with the highest sensitivity for a screening purpose, as well as the highest positive likelihood ratio which is, in this case, related to the smallest false positive rate. Therefore, we may discard models obtained by discriminant analysis and decision trees based on the CART algorithm. On the other hand, the logistic regression model and the decision tree based on the QUEST algorithm (T4) present non-adequate 95% confidence intervals for specificity, which is an indicator that the false positive rate may be higher than 50%. These motives lead us to the choose the models obtained by decision tree analysis based on CHAID or Exhaustive CHAI algorithms, which have exactly the same solution.

In clinical practice, for screening purposes, we are interested in a model that optimizes positive predictive values rather than negative predictive values, that is, a model that give us a posterior probability higher than 50% when the disease is present. As this depends on prior probability, that is, on prevalence of the disease, then we will have the predictive values presented on Figure 43, according to group's prevalence for Portugal, published in 2013 by the National Observatory for Diabetes.

In fact, models with highest positive predictive value, whichever the prevalence, are the ones obtained by discriminant analysis (D), CHAID (T2) and Exhaustive CHAID (T3) algorithms, as observed in Figure 44. In fact, these classifiers may be used with higher accuracy in subjects on the age group of 60 to 75 years old, or with body mass index above 30 kg/m^2.

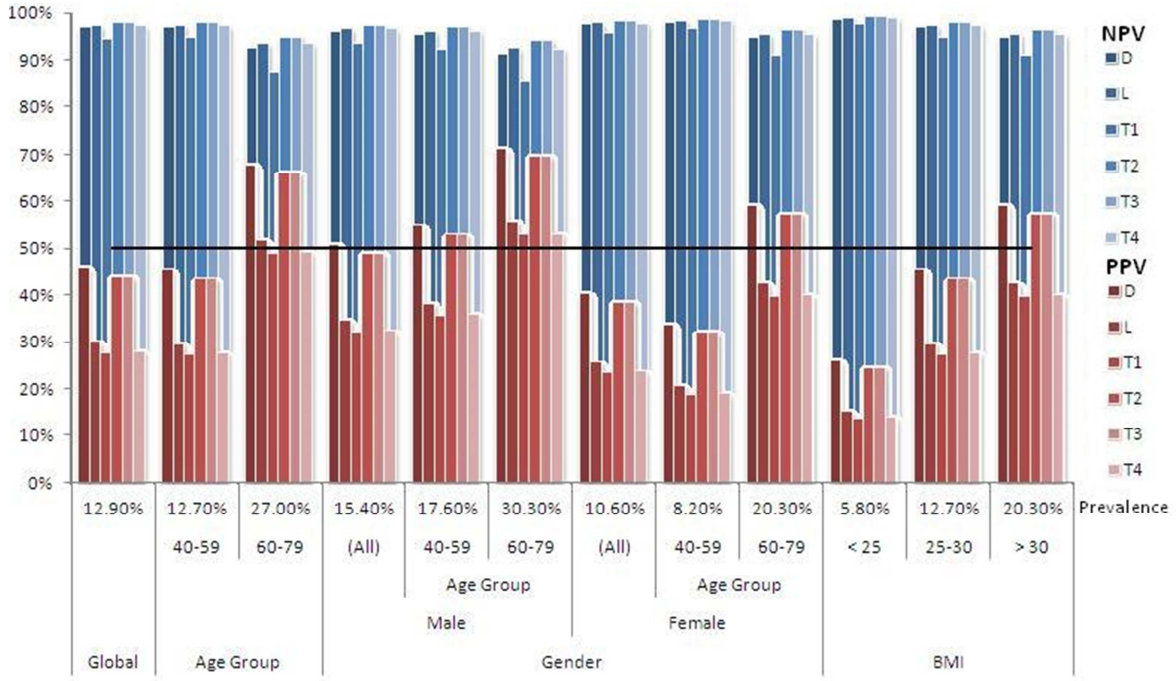

Figure 44 – Predictive values (positive – PPV and negative – NPV) according to disease prevalence (prevalence data published by the National Observatory for Diabetes, in 2013).

As explained before, discriminant analysis classifier will be dropped-down, and the proposed final model is the one obtained with CHAID or Exhaustive CHAID algorithms in decision tree analysis. We will use the CHAID model as it is simpler.

The model may be applied mainly in subjects with at least 60 years old or with at least 30 kg/m^2 of body mass index, as presented on Table 46. It classifies all subjects with diagnosed hypertension, under treatment for this condition, as type 2 diabetics with a posterior probability of 82.61%. For all the other subjects, the model may be written as follows:

$$Value\ measured\ on\ meridian\ 45º\ of\ the\ Chromatic\ Tritan\ axis$$

$$\begin{cases} \geq 134.33 \times 10^{-3} \Rightarrow P(D) = 87.50\% \\ < 134.33 \times 10^{-3} \Rightarrow P(D) = 7.69\% \\ Unknown \Rightarrow P(D) = 22.45\% \end{cases}$$

Table 46 - Expected predictive values for the final classifier of Diabetes (T2) and other indicators or accuracy,

			Diabetes Prevalence	Predictive Values				Accuracy and Concordance		Sensitivity and Specificity			Likelihood Ratios	
				Positive	(95% CI)	Negativ	(95% CI)							
Global			12.90%	43.84%	(37.27% - 50.42%)	97.89%	(95.99% - 99.79%)	AUC (p): 0.894 (< 0.001)	Cohen's kappa (p): 0.676 (p < 0.001)	Sens (95% CI): 87.88% (71.85% - 96.6%)	Spec (95% CI): 83.33% (51.60% - 97.90%)	McNemar p-value: 0.667	PLR (95% CI): 5.27 (1.50 - 18.80)	NLR (95% CI): 0.15 (0.06 - 0.40)
Age Group		40-59	12.70%	43.40%	(36.84% - 49.97%)	97.93%	(96.04% - 99.81%)							
		60-75	27.00%	66.10%	(59.83% - 72.37%)	94.90%	(91.98% - 97.81%)							
Gender	Male		15.40%	48.97%	(42.35% - 55.59%)	97.42%	(95.32% - 99.52%)							
	Age Group	40-59	17.60%	52.96%	(46.35% - 59.57%)	96.99%	(94.72% - 99.25%)							
		60-75	30.30%	69.62%	(63.53% - 75.71%)	94.05%	(90.92% - 97.19%)							
	Female		10.60%	38.46%	(32.02% - 44.91%)	98.30%	(96.59% - 100.01%)							
	Age Group	40-59	8.20%	32.01%	(25.83% - 38.19%)	98.72%	(97.23% - 100.21%)							
		60-75	20.30%	57.31%	(50.76% - 63.87%)	96.43%	(93.97% - 98.89%)							
BMI		< 25	5.80%	24.50%	(18.81% - 30.2%)	99.11%	(97.87% - 100.35%)							
		[25, 30[	12.70%	43.40%	(36.84% - 49.97%)	97.93%	(96.04% - 99.81%)							
		≥ 30	20.30%	57.31%	(50.76% - 63.87%)	96.43%	(93.97% - 98.89%)							

SECTION C

STATISTICAL CLASSIFIERS FOR DIABETIC RETINOPATHY IN TYPE 2 DIABETICS

1. Training sample description

From the training sample used for models obtained in Part II of the Results section, we performed a similar analysis for the diabetic group, in order to attempt to infer a model which enables classification of diabetic retinopathy. This was performed separately in order to use the duration of the disease, which is a known factor for retinopathy development, and that was only collected for these subgroups. Likewise, ETDRS grading was performed only for the diabetic subgroup, since the design of the study did not allow to attempt to identify a global diabetes and diabetic retinopathy classifier.

For the diabetic group, the duration of the disease follows a normal distribution, although it has a large dispersion: values range from 1 to 39 years with mean 14.13 $\pm$ 1.41 SEM.

Diabetic patients ranged age between 45 and 73 years old, with a mean of 49.98 $\pm$ 1.20 SEM and a median of 61 years old (inter-quartile range between 54.25 and 67 years old). ETDRS grading was performed in 40 of the 49 diabetic patients, and 20 of them had diabetic retinopathy. Considering the obtained sample, we may only attempt to identify predictors of the presence of diabetic retinopathy, but not of proliferative diabetic retinopathy since there were no cases with that condition; it was also not possible to attempt to discriminate subjects according to their ETDRS grading of non-proliferative diabetic retinopathy, since there were not enough data available in each group, as seen in Figure 45. All the subjects were right handed.

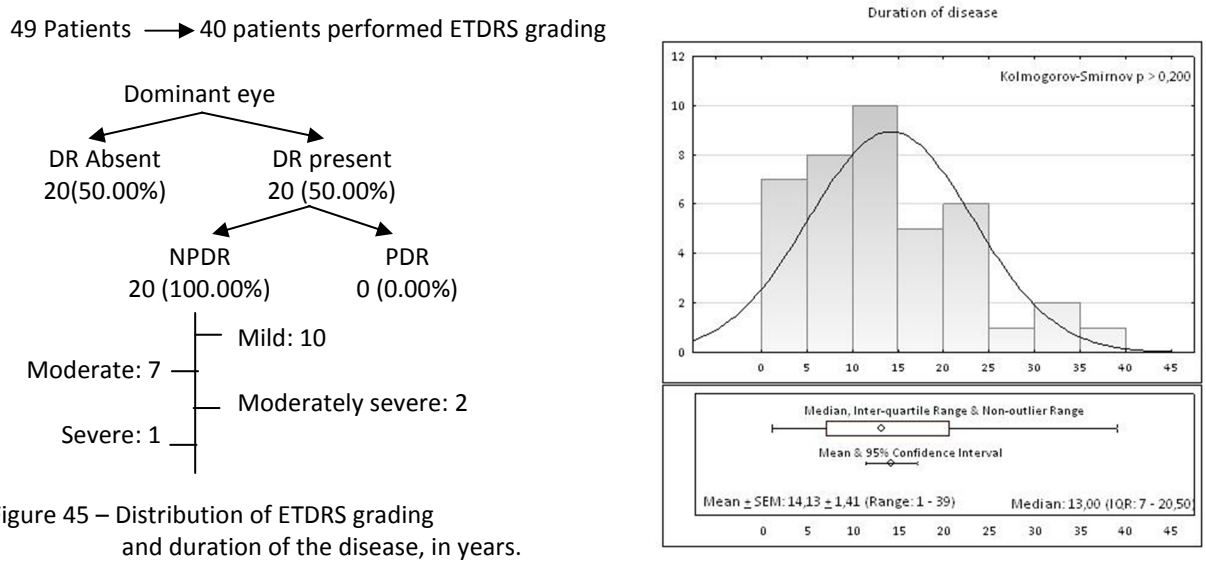

Figure 45 – Distribution of ETDRS grading and duration of the disease, in years.

According to gender, 45.00% of the cases in the sample (18) were male, and 55.00% of the cases (22) were female (Binomial test: p = 0.635). In this subgroup, only four cases (18.18%) have been previously diagnosed with gestational diabetes (Binomial test: p = 0.007), but there is no registry of the number of gestations for women. Hence this parameter will not be considered for classification models, which could be done for the female group.

The majority had right eye dominance (25 cases, 62.50%) and the other fifteen cases (37.50%) had left eye dominance (Binomial test: p = 0.155). Hereditary factor based upon family history of diabetes was present in 28 subjects, corresponding to 70.00% of the sample (Binomial test: p < 0.001).

Regarding daily habits, only three cases (7.50%) were regular smokers (Binomial test: p < 0.001), eight cases (20.00%) were regular drinkers (Binomial test: p = 0.003) and thirteen cases (32.50%) practiced regular exercise (Binomial test: p = 0.055). Three quarters of the sample (30 cases) had diagnosed hypertension, being currently medicated for that disease.

Concerning quantitative variables, all of them were normally distributed and descriptive statistics are presented in the following table (Table 47):

Table 47 – Descriptive statistics of clinical and demographic variables.

	N	Min	Max	Mean	SD	P25	P50	P75
Height (m)*	40	1.45	1.87	1.61	0.02	1.53	1.61	1.68
Weight (kg)*	40	53.10	104.00	78.11	2.03	68.38	80.60	86.00
BMI (kg/m^2)	40	22.10	43.70	30.24	0.85	26.20	29.95	33.50
AP (cm)*	38	71.00	140.00	101.34	2.28	94.00	100.00	110.00
Pulse (bpm)*	40	39.00	100.00	74.43	1.78	68.00	74.00	82.00
SBP (mmHg)	40	101.00	179.00	134.08	3.09	116.00	134.50	150.00
DBP (mmHg)*	40	46.00	100.00	76.73	1.91	69.25	78.00	85.00
Bioimpedance (%)*	31	21.20	61.90	34.89	1.82	26.80	32.80	40.40

* Normally distributed variables

BMI – Body mass index; AP – Abdominal perimeter; SBP – Systolic blood pressure; DBP – Diastolic blood pressure

2. Variable reduction

2.1 Phase 1: Factors differentiating Diabetic Retinopathy

2.1.1 Clinical and demographic assessment

Groups were homogeneous concerning gender and eye dominance. Note that groups were equally distributed regarding family history of diabetes (on both groups, the majority of the cases had previous family history of diabetes), diagnosed and treated hypertension and alcohol habits. Only a small part of the global sample corresponds to actual smokers, and subjects that practice regular exercise (Table 48).

Table 48 - Distribution of diabetic retinopathy for factor, and association with each factor (p-values for the independence Chi-square test).

		Diabetic Retinopathy		p
		No	Yes	
Gender	Male	7 (38.89%)	11 (61.11%)	0.204
	Female	13 (59.09%)	9 (40.91%)	
Eye dominance	Right	15 (60%)	10 (40%)	0.102
	Left	5 (33.33%)	10 (66.67%)	
Family history	No	6 (50.00%)	6 (50.00%)	1.000
	Yes	14 (50.00%)	14 (50.00%)	
HTA	No	5 (50.00%)	5 (50.00%)	1.000
	Yes	15 (50.00%)	15 (50.00%)	
Smoker	No	1 (33.33%)	2 (66.67%)	0.597
	Yes	19 (51.35%)	18 (48.65%)	
Alcohol	No	16 (50.00%)	16 (50.00%)	1.000
	Yes	4 (50.00%)	4 (50.00%)	
Exercise	No	13 (50.00%)	13 (50.00%)	0.821
	Yes	5 (50.00%)	5 (50.00%)	

For the female subgroup, despite the number of pregnancies, 2 cases (15.40%) from the subgroup of 13 females without diabetic retinopathy had gestational diabetes, and 2 cases within the other 8 (25.00%) with diabetic retinopathy had gestational diabetes.

Age distribution, height, weight, body mass index and bioimpedance were similar between type 2 diabetics without or with diabetic retinopathy. There was also no statistical significant difference between groups for pulse, systolic or diastolic blood pressure. Yet, and unsurprisingly, regarding the duration of the disease, the group with diabetic retinopathy had diabetes mellitus for a longer time (Table 49). In fact, the 95% confidence interval for the mean difference of duration of diabetes, between the group with diabetic retinopathy and the group without retinopathy, range between 2 and 12 years.

Table 49 - Descriptive statistics and group comparison of clinical and demographic variables measured between type 2 diabetics without and with diabetic retinopathy.

	DR	n	Min	Max	Mean	SEM	P25	P50	P75	P
Age (visit)	No	20	46.00	73.00	59.50	1.87	54.25	58.00	67.00	0.699*
	Yes	20	45.00	72.00	60.45	1.56	55.00	61.50	66.75	
Duration (Years)	No	20	1.00	31.00	10.50	14.50	4.75	9.50	14.50	0.008*
	Yes	20	3.00	39.00	17.75	30.60	19.50	21.75	30.60	
Height (m)	No	20	1.45	1.84	1.60	0.02	1.53	1.58	1.68	0.552*
	Yes	20	1.45	1.87	1.62	0.02	1.53	1.65	1.69	
Weight (kg)	No	20	53.10	93.70	74.25	3.04	61.60	78.15	84.98	0.108**
	Yes	20	63.60	104.00	81.98	2.47	71.40	82.50	88.78	
BMI (kg/m^2)	No	20	22.10	43.00	28.98	1.16	24.70	27.95	32.40	0.139*
	Yes	20	23.30	43.70	31.50	1.20	27.33	31.00	34.90	
AP (cm)	No	20	71.00	125.50	98.70	3.25	88.00	100.00	110.00	0.226*
	Yes	18	87.00	140.00	104.28	3.13	94.75	102.00	110.00	
Pulse (bpm)	No	20	39.00	100.00	73.70	3.00	67.25	75.50	81.50	0.689*
	Yes	20	59.00	90.00	75.15	1.99	68.50	73.50	82.00	
SBP (mmHg)	No	20	106.00	179.00	134.45	4.81	115.25	134.00	149.75	0.905*
	Yes	20	101.00	158.00	133.70	4.01	117.00	135.00	150.00	
DBP (mmHg)	No	20	46.00	100.00	78.40	2.62	70.50	79.00	85.50	0.388*
	Yes	20	51.00	94.00	75.05	2.81	65.25	76.50	85.00	
Bioimpedance (%)	No	15	21.20	61.90	35.75	3.12	25.70	31.00	47.00	0.662*
	Yes	16	24.30	57.80	34.09	2.07	29.20	32.95	37.23	

*Independent samples t-test; **Mann-Whitney U test

BMI – Body mass index; AP – Abdominal perimeter; SBP – Systolic blood pressure; DBP – Diastolic blood pressure.

2.1.2 Blood Tests

2.1.2.1 Biochemistry

Groups without and with diabetic retinopathy presented similar results regarding fasting blood glucose levels (p = 0.622), and glycosylated haemoglobin, with mean values of 9.41 and 9.44 on the NGSP scale (p = 0.967), respectively, and 79.40 and 79.84 on the IFCC scale (p = 0.959), respectively (Table 50).

Table 50 - Descriptive statistics and group comparison of Blood glucose and glycosylated haemoglobin values between type 2 diabetics without and with diabetic retinopathy.

	DR	N	Min	Max	Mean	SEM	P25	P50	P75	p
Glucose	No	19	96.00	291.00	166.11	11.44	125.00	164.00	210.00	0.622*
	Yes	19	62.00	363.00	176.21	16.82	125.00	169.00	225.00	
HbA1C (NGSP)	No	20	5.20	17.30	9.41	0.63	7.63	9.30	10.45	0.967*
	Yes	19	6.30	12.60	9.44	0.44	7.80	9.30	10.90	
HbA1C (IFCC)	No	20	33.00	166.00	79.40	6.87	60.25	78.00	90.50	0.959*
	Yes	19	45.00	114.00	79.84	4.78	62.00	78.00	96.00	

*Independent samples t-test; **Mann-Whitney U test

Creatinine values were similar between groups (p = 0.191), as well as the hepatic function and lipid related parameters, that showed no statistical differences between groups, as seen in the following tables (Table 51, 52 and 53).

Table 51 - Descriptive statistics and group comparison of creatinine values between type 2 diabetics without and with diabetic retinopathy.

	DR	n	Min	Max	Mean	SEM	P25	P50	P75	p
Creatinine	No	19	0.44	2.58	0.84	0.12	0.55	0.64	0.93	0.191**
	Yes	20	0.48	1.61	0.90	0.08	0.60	0.81	1.10	

*Independent samples t-test; **Mann-Whitney U test

Liver parameters evaluated do not enable group differentiation.

Table 52 - Descriptive statistics and group comparison of liver function parameters between type 2 diabetics without and with diabetic retinopathy.

	DR	N	Min	Max	Mean	SEM	P25	P50	P75	P
ALT	No	20	9.00	74.00	28.05	3.72	17.25	23.50	30.75	0.850**
	Yes	20	12.00	81.00	29.80	3.94	18.25	23.00	36.25	
AST	No	20	12.00	45.00	21.65	1.76	17.25	19.00	23.50	0.136**
	Yes	20	15.00	58.00	28.00	2.95	18.00	22.50	38.00	
Alkaline Phosphatase	No	20	44.00	137.00	76.65	5.97	57.25	71.00	87.00	0.440*
	Yes	20	40.00	164.00	83.85	7.03	60.25	79.00	103.00	
Gamma GT	No	20	8.00	190.00	36.75	8.76	17.75	25.00	40.75	0.064**
	Yes	20	12.00	223.00	50.10	10.25	28.00	37.50	61.75	

*Independent samples t-test; **Mann-Whitney U test

Parameters related to lipid metabolism are identical between groups.

Table 53 - Descriptive statistics and group comparison of lipid related parameters between type 2 diabetics without and with diabetic retinopathy.

	DR	N	Min	Max	Mean	SEM	P25	P50	P75	P
Total Cholesterol	No	20	86.00	398.00	190.10	15.38	150.50	177.50	236.50	0.126**
	Yes	20	116.00	229.00	161.65	7.91	131.75	154.00	187.00	
Cholesterol HDL	No	20	28.00	65.00	44.85	2.49	35.50	43.50	54.00	0.237*
	Yes	20	14.00	64.00	40.50	2.63	32.75	40.50	47.50	
Atherogenic Index	No	20	2.50	6.50	4.25	0.28	3.18	3.70	5.50	0.925**
	Yes	20	2.30	9.30	4.39	0.41	3.20	4.05	4.93	
Cholesterol LDL	No	19	48.00	195.00	119.89	8.96	94.00	111.00	159.00	0.223**
	Yes	19	71.00	175.00	105.42	6.77	79.00	100.00	123.00	
Triglycerides	No	20	63.00	465.00	168.55	21.13	100.00	143.50	200.25	0.527**
	Yes	19	55.00	386.00	160.05	22.07	103.00	123.00	170.00	
Apolipoprotein A1	No	19	112.00	186.00	145.63	4.91	129.00	147.00	164.00	0.166*
	Yes	19	37.00	198.00	132.00	8.30	108.00	134.00	159.00	
Apolipoprotein B100	No	19	44.00	155.00	94.74	7.57	70.00	86.00	125.00	0.479*
	Yes	19	50.00	142.00	88.16	5.22	74.00	85.00	97.00	
B100/A1	No	19	0.33	1.12	0.65	0.05	0.50	0.56	0.85	0.603**
	Yes	19	0.30	2.74	0.79	0.13	0.53	0.60	0.83	
Lipoprotein	No	20	2.33	154.00	34.49	8.67	9.31	21.35	51.40	0.430**
	Yes	19	2.40	117.00	24.59	6.12	9.31	15.20	35.90	

*Independent samples t-test; **Mann-Whitney U test

2.1.2.2 Cell Blood Count Cytometry

Leucocytes do not present statistical differences between groups (Table 54).

Table 54 - Descriptive statistics and group comparison of leucocytes between type 2 diabetics without and with diabetic retinopathy.

	DR	n	Min	Max	Mean	SEM	P25	P50	P75	P
Leucocytes	No	20	3.80	11.60	6.85	0.42	5.45	6.75	7.70	0.570**
	Yes	20	0.90	18.40	7.37	0.79	4.98	7.25	8.55	

*Independent samples t-test; **Mann-Whitney U test

However, the group with diabetic retinopathy presents significantly lower values for erythrocytes, haemoglobin and haematocrit than the group which do not have diabetic retinopathy (Table 55). No other measures related to red cell counts differentiate these groups.

Table 55 - Descriptive statistics and group comparison of red cell counts between type 2 diabetics without and with diabetic retinopathy.

	DR	N	Min	Max	Mean	SEM	P25	P50	P75	p
Erythrocytes	No	20	3.68	5.26	4.52	0.10	4.20	4.54	4.82	0.023**
	Yes	20	3.39	6.13	4.19	0.14	3.72	4.10	4.58	
Haemoglobin	No	20	11.20	15.70	13.75	0.28	12.73	13.55	14.90	0.010*
	Yes	20	11.10	15.20	12.70	0.27	11.85	12.35	13.33	
Haematocrit	No	20	34.50	45.90	40.25	0.78	37.63	39.65	43.55	0.014**
	Yes	20	32.90	45.20	37.41	0.84	34.58	36.10	39.48	
MCV	No	20	81.20	98.50	89.18	0.84	87.23	88.90	91.63	0.159**
	Yes	20	57.80	101.50	90.25	2.02	87.83	90.70	95.85	
MCH	No	20	27.60	32.80	30.45	0.33	29.33	31.00	31.28	0.304**
	Yes	20	18.70	35.00	30.68	0.74	29.75	30.70	32.88	
MCHC	No	20	32.50	35.50	34.14	0.19	33.55	34.20	34.83	0.501*
	Yes	20	32.40	35.60	33.97	0.17	33.63	34.05	34.40	
EVC	No	20	11.60	16.30	13.49	0.26	12.83	13.15	14.55	0.144**
	Yes	20	12.30	17.60	14.01	0.28	13.33	13.70	14.55	

*Independent samples t-test; **Mann-Whitney U test

MCV – Mean corpuscular volume; MCH – Mean corpuscular haemoglobin; MCHC – Mean corpuscular haemoglobin concentration; EVC – Erythrocytes variation coefficient

None of the measures related to platelet show statistical differences between groups, as observed in Table 56:

Table 56 - Descriptive statistics and group comparison of platelet between type 2 diabetics without and with diabetic retinopathy.

	DR	N	Min	Max	Mean	SEM	P25	P50	P75	p
Platelet	No	20	90.00	610.00	242.95	22.22	194.50	220.00	265.25	0.091**
	Yes	20	81.00	322.00	203.50	13.86	166.00	190.50	264.50	
MPV	No	20	7.90	11.20	9.44	0.18	9.00	9.30	9.88	0.583*
	Yes	20	7.10	13.10	9.66	0.35	8.38	9.60	10.63	
Plateleocrit	No	20	0.09	0.48	0.23	0.02	0.19	0.22	0.25	0.125**
	Yes	20	0.06	0.27	0.19	0.01	0.17	0.20	0.24	
PVC	No	20	16.00	18.00	16.35	0.13	16.00	16.00	17.00	0.066**
	Yes	20	16.00	18.00	16.70	0.15	16.00	17.00	17.00	

*Independent samples t-test; **Mann-Whitney U test
MPV – Mean platelet volume; PVC – Platelet variation coefficient

2.1.2.3 Hormonology

Likewise, hormonology parameters measured in these groups were similar (Table 57).

Table 57 - Descriptive statistics and group comparison of TSH and C-Peptide between type 2 diabetics without and with diabetic retinopathy.

	DR	N	Min	max	mean	SEM	P25	P50	P75	p
TSH	No	19	0.53	5.10	2.20	0.28	1.30	1.90	2.90	0.779**
	Yes	20	0.74	5.10	2.10	0.28	1.13	1.90	2.48	
C-Peptide	No	20	0.10	4.80	1.66	0.30	0.63	1.40	2.55	0.545**
	Yes	19	0.10	4.40	1.37	0.28	0.20	1.30	1.80	

*Independent samples t-test; **Mann-Whitney U test

2.1.3 Ophthalmological tests

Intraocular pressure was measured in 23 type 2 diabetics, 14 of which without diabetic retinopathy, and 9 with diabetic retinopathy. The first group ranged between 10 and 20 mmHg (mean 15.36 $\pm$ 0.75 mmHg), while the second one ranged between 8 and 23 mmHg (mean 15.78 $\pm$ 1.75 mmHg), and no statistical significant differences were found between groups (independent samples t-test p = 0.829).

Concerning the best corrected visual acuity, both groups presented a median value of 1.00, with and inter-quartile range from 0.80 to 1.00 so, groups were statistically identical (Mann-Whitney p = 0.925).

2.1.3.1 Optical Coherence Tomography

2.1.3.1.1 Volume Scan density

Groups presented similar values for volume scan, with exception of the Inner Nasal region, where the group with diabetic retinopathy is expected to have an increased density when compared to the group without diabetic retinopathy (Table 58). In fact, the expected median

increase lies between 1 and 28 units, with 95% confidence (Hodges-Lehmann confidence interval).

Table 58 - Descriptive statistics and group comparison of Volume Scan measured by OCT between type 2 diabetics without and with diabetic retinopathy.

Volume Scan		DR	N	Min	Max	Mean	SEM	P25	P50	P75	p
Central Subfield		No	20	227.00	329.00	276.90	5.41	262.50	278.00	293.75	0.090*
		Yes	20	174.00	416.00	300.15	12.07	272.00	291.00	319.25	
Inner	Nasal	No	20	242.00	384.00	336.30	6.17	330.50	339.00	351.75	0.026**
		Yes	20	289.00	415.00	353.85	6.74	338.25	354.00	369.00	
	Superior	No	20	316.00	378.00	342.35	3.88	326.50	342.50	351.00	0.157*
		Yes	20	303.00	433.00	354.30	7.31	334.25	349.00	373.25	
	Temporal	No	20	309.00	356.00	330.30	2.98	321.25	329.50	337.75	0.626**
		Yes	20	251.00	448.00	340.05	9.65	321.50	331.00	347.00	
	Inferior	No	20	275.00	422.00	336.50	6.49	325.00	335.50	346.50	0.267**
		Yes	20	282.00	424.00	344.50	6.87	336.00	341.50	355.50	
Outer	Nasal	No	20	290.00	383.00	314.35	4.55	299.00	311.00	323.75	0.850**
		Yes	20	270.00	416.00	317.85	7.02	302.25	309.00	332.75	
	Superior	No	20	269.00	335.00	299.05	3.34	292.25	298.50	307.75	0.336*
		Yes	20	256.00	358.00	305.45	5.64	289.00	300.50	323.75	
	Temporal	No	20	262.00	396.00	293.35	7.97	271.25	284.50	295.25	0.914**
		Yes	20	247.00	357.00	290.25	6.23	273.25	282.00	311.75	
	Inferior	No	20	262.00	394.00	289.05	6.25	273.00	286.00	296.50	0.579**
		Yes	20	239.00	369.00	284.50	6.48	264.25	281.00	296.50	

*Independent samples t-test; **Mann-Whitney U test

2.1.3.1.2 Retinal Nerve Fibre Layer

Retinal thickness presents similar values in both groups, as shown in the following table (Table 59).

Table 59 - Descriptive statistics and group comparison of Retinal Nerve Fibre Layer measured with OCT between type 2 diabetics without and with diabetic retinopathy.

RNFL		DR	n	Min	Max	Mean	SEM	P25	P50	P75	P
Global		No	20	77.00	113.00	98.45	2.02	97.25	99.50	103.00	0.833**
		Yes	19	68.00	118.00	97.16	2.85	93.00	101.00	104.00	
Nasal		No	20	57.00	88.00	73.10	2.49	61.25	74.00	84.75	0.536**
		Yes	19	37.00	93.00	74.63	3.08	66.00	78.00	86.00	
Nasal	Superior	No	20	61.00	127.00	102.20	3.96	89.75	108.00	116.75	0.849*
		Yes	19	52.00	142.00	103.47	5.38	82.00	111.00	118.00	
	Inferior	No	20	91.00	145.00	127.65	3.76	113.50	135.00	143.00	0.211**
		Yes	19	97.00	154.00	121.68	3.63	109.00	118.00	139.00	
Temporal		No	20	49.00	160.00	79.70	4.87	69.25	75.50	83.75	0.151**
		Yes	19	32.00	91.00	70.26	3.10	64.00	71.00	80.00	
Temp.	Superior	No	20	108.00	181.00	140.95	4.75	123.75	138.00	157.25	0.730*
		Yes	19	51.00	187.00	138.00	7.14	124.00	135.00	156.00	
	Inferior	No	20	65.00	170.00	115.70	5.53	101.00	115.50	132.00	0.475*
		Yes	19	52.00	193.00	122.11	7.00	110.00	124.00	140.00	

*Independent samples t-test; **Mann-Whitney U test

2.1.3.2 Psychophysical tests

Tests of contrast sensitivity produce identical results for the diabetic group, when comparing subjects with and without diabetic retinopathy. In fact, results are similar when evaluating speed and achromatic vision. Regarding chromatic contrast sensitivity, we could differentiate groups only at Deutan and Tritan axes of the chromatic contrast sensitivity function, as presented below.

2.1.3.2.1 Speed

Type 2 diabetics without and with diabetic retinopathy presented similar results for the speed test, whichever the meridian analysed. There was also no statistical significant difference between groups for the overall measure of speed discrimination (Table 60).

Table 60 - Descriptive statistics and group comparison for the speed test, measured in meridians 0º, 45º, 90º, 135º and global area generated by these meridians between type 2 diabetics without and with diabetic retinopathy.

Speed		DR	n	Min	Max	Mean	SEM	P25	P50	P75	p
Meridian	0º	No	19	0.50	5.64	2.03	0.30	0.85	1.77	2.73	0,753**
		Yes	18	0.16	9.69	2.24	0.54	0.52	2.08	2.72	
	45º	No	18	0.18	7.99	1.90	0.43	0.55	1.28	2.86	0,503**
		Yes	17	0.56	7.98	2.31	0.52	0.80	1.42	3.30	
	90º	No	19	0.32	3.12	1.53	0.20	0.99	1.30	2.25	0,620**
		Yes	18	0.15	7.88	2.33	0.53	0.84	1.42	3.67	
	135º	No	19	0.38	5.50	2.44	0.39	0.70	2.05	4.17	0,802**
		Yes	17	0.15	7.72	2.94	0.61	0.78	2.53	5.29	
	Area	No	18	0.55	18.36	4.40	1.04	1.32	3.61	5.63	0.597**
		Yes	16	0.27	29.79	6.44	1.94	1.42	2.74	10.34	

*Independent samples t-test; **Mann-Whitney U test

2.1.3.2.2 Achromatic contrast

There was no significant statistical difference between groups concerning achromatic discrimination test, meaning that this measure is not able to classify diabetic retinopathy (Table 61).

Table 61 - Descriptive statistics and group comparison for the achromatic vision test, measured in meridians 0º, 45º, 90º, 135º and global area generated by these meridians between type 2 diabetics without and with diabetic retinopathy.

Acrhom.		DR	N	Min	Max	Mean	SEM	P25	P50	P75	P
Meridian	0º	No	19	1.00	4.28	2.22	0.23	1.41	1.90	3.03	0,098**
		Yes	18	1.00	10.47	3.29	0.56	2.05	2.66	3.57	
	45º	No	18	1.20	8.37	3.17	0.48	1.75	2.39	3.79	0,938**
		Yes	18	1.20	7.03	2.97	0.37	2.05	2.62	3.26	
	90º	No	19	1.00	5.43	3.19	0.29	2.24	2.86	4.47	0,707**
		Yes	18	1.20	6.35	3.08	0.37	1.60	2.72	3.91	
	135º	No	19	1.00	10.77	3.32	0.51	1.41	2.90	4.61	0,217**
		Yes	16	1.00	4.47	2.45	0.24	1.63	2.50	3.11	
Area		No	18	3.82	39.36	10.79	2.25	4.48	6.51	14.85	0.597**
		Yes	16	3.14	18.02	8.54	1.18	3.76	7.99	12.13	

*Independent samples t-test; **Mann-Whitney U test

2.1.3.2.3 Chromatic Contrast

Chromatic contrasts discrimination may differentiate groups, but only in some measures. Measures obtained according to the Protan axis are similar between subjects without and with diabetic retinopathy (Table 62)-

Table 62 - Descriptive statistics and group comparison for the chromatic contrast test for the Protan axis, measured in meridians 0º, 45º, 90º, 135º and global area generated by these meridians between type 2 diabetics without and with diabetic retinopathy.

Protan		DR	n	Min	Max	Mean	SEM	P25	P50	P75	P
Meridian ($x10^{-3}$)	0º	No	19	1.23	4.52	2.51	0.24	1.25	2.47	3.09	0,258**
		Yes	18	1.23	9.49	3.38	0.50	1.25	3.07	4.78	
	45º	No	19	1.23	21.53	5.40	1.14	2.25	3.72	8.23	0,845**
		Yes	18	1.24	28.66	6.55	1.75	1.69	4.01	9.60	
	90º	No	19	1.24	18.32	4.39	0.88	1.85	3.49	5.54	0,988**
		Yes	18	1.23	27.34	5.67	1.69	1.85	3.59	5.10	
	135º	No	19	1.23	17.05	5.44	0.93	2.47	3.72	7.42	0,616**
		Yes	17	1.23	13.25	5.52	0.76	3.48	4.32	7.85	
Area ($x10^{-6}$)		No	19	2.00	105.00	22.63	5.45	6.00	19.00	29.00	0.731**
		Yes	17	4.00	453.00	40.00	25.88	9.00	12.00	19.00	

*Independent samples t-test; **Mann-Whitney U test

On the other hand, contrast sensitivity along meridian 0º of the Deutan axis may differentiate groups (p = 0.041). None of the other measures are possible classifiers of diabetic retinopathy (Table 63).

Table 63 - Descriptive statistics and group comparison of Chromatic contrast test for the Deutan axis, measured in meridians 0º, 45º, 90º, 135º and global area generated by these meridians between type 2 diabetics without and with diabetic retinopathy.

Deutan		DR	N	Min	Max	Mean	SEM	P25	P50	P75	P
Meridian ($x10^{-3}$)	0º	No	18	1.24	11.26	3.90	0.63	1.71	3.62	4.64	0,041**
		Yes	17	1.25	55.95	9.99	3.36	3.39	4.75	7.98	
	45º	No	18	1.24	74.08	15.17	4.28	2.48	7.71	25.30	0,096**
		Yes	17	1.24	76.32	24.10	5.17	7.99	19.71	37.00	
	90º	No	19	1.24	86.43	14.48	5.58	3.70	4.33	14.63	0,684**
		Yes	17	1.23	59.95	14.53	4.45	1.85	6.19	21.25	
	135º	No	18	1.23	69.83	20.87	5.06	5.20	14.05	27.91	0,878**
		Yes	16	1.85	57.12	20.72	5.26	4.32	9.78	46.85	
Area ($x10^{-6}$)		No	18	0.00	1616.00	204.78	93.66	14.75	80.00	153.50	0.506**
		Yes	16	6.00	2077.00	373.44	149.81	33.75	67.50	580.25	

*Independent samples t-test; **Mann-Whitney U test

Contrast sensitivity for the Tritan axis is the measure that possibly has more discriminative power (Table 64), since it presents statistical differences between groups for meridians 0º and 135º, and also for the global measure of the Tritan axis, given by the 5-sided polygon generated by the values measured for each one of the meridians and the (0,0) point.

Table 64 - Descriptive statistics and group comparison of chromatic contrast test for the Tritan axis, measured in meridians 0º, 45º, 90º, 135º and global area generated by these meridians between type 2 diabetics without and with diabetic retinopathy.

Tritan		DR	n	Min	Max	Mean	SEM	P25	P50	P75	p
Meridian ($x10^{-3}$)	0º	No	19	33.32	132.10	55.28	5.72	40.00	46.66	62.82	0.003**
		Yes	18	25.00	401.57	119.57	23.83	55.98	83.99	155.74	
	45º	No	18	25.00	310.36	132.75	20.16	66.24	127.01	192.81	0.339**
		Yes	18	53.48	411.62	172.52	26.12	86.90	134.06	271.83	
	90º	No	19	25.00	267.24	89.51	15.01	43.33	56.81	133.78	0.081**
		Yes	18	40.00	425.51	148.45	28.66	62.15	82.90	219.58	
	135º	No	18	38.48	391.88	107.50	21.54	45.01	63.35	148.88	0.002**
		Yes	17	35.23	568.07	229.62	32.46	135.69	198.36	298.41	
Area ($x10^{-3}$)		No	18	1.08	40.91	11.78	2.86	3.38	5.76	17.46	0.017**
		Yes	17	2.65	139.83	28.46	8.24	8.41	15.90	36.20	

*Independent samples t-test; **Mann-Whitney U test

2.2 Phase 2: Univariate classifiers of Diabetic Retinopathy

In this section, we will present the results of univariate classifiers for diabetic retinopathy. As we can observe in Tables 65 to 75, few variables from the previous identified variables as possible classifiers are identified.

From the sociodemographic parameters, duration of diabetes is the only one that may be identified as a possible discriminator (ROC curve presented in Figure 46), where a subject with less than 18.50 years of duration of the disease is classified has type 2 diabetic without

diabetic retinopathy with a probability of 90.00%. In fact, it is 5.50 times more likely that the duration of the disease is, at least, 18.50 years, in a subject with diabetic retinopathy than in a subject without diabetic retinopathy (Table 65) thus, longer duration of diabetes mellitus (type 2) represents an increased risk factor for the development of diabetic retinopathy.

Table 65 - Accuracy of medical preliminary procedures measured for univariate classification of diabetic retinopathy in type 2 diabetics.

Variable	AUC	SEM	p	LBCI	UBCI	Cut-off	Sensitivity	Specificity	+LR
Age	0.538	0.09	0.685	0.352	0.723	-	-	-	-
Height	0.564	0.09	0.490	0.380	0.747	-	-	-	-
Weight	0.649	0.09	0.108	0.478	0.820	-	-	-	-
BMI	0.646	0.09	0.114	0.473	0.819	-	-	-	-
AP	0.565	0.09	0.492	0.380	0.750	-	-	-	-
Pulse	0.525	0.09	0.787	0.343	0.707	-	-	-	-
SBP	0.510	0.09	0.914	0.327	0.693	-	-	-	-
DBP	0.576	0.09	0.409	0.394	0.758	-	-	-	-
Bioimpedance	0.502	0.11	0.984	0.283	0.721	-	-	-	-
Duration	0.748	0.08	0.007	0.593	0.902	$\geq$ 18.5	55.00%	90.00%	5.50

BMI – Body mass index; AP – Abdominal perimeter; SBP – Systolic blood pressure; DBP – Diastolic blood pressure.

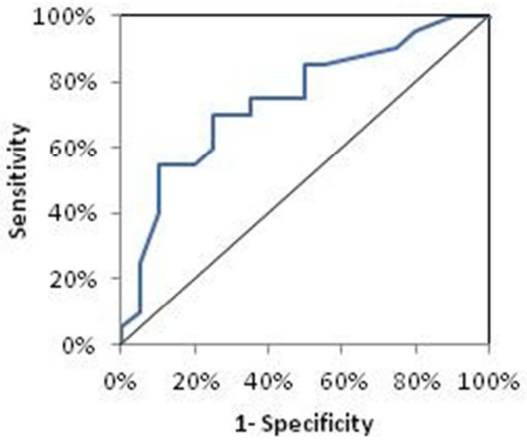

Figure 46 – ROC curve for duration of disease since diagnosis.

Parameters such as glycaemia and glycosylated haemoglobin, creatinine, or parameters related to the hepatic function or lipids are not capable of differentiating between subjects without or with diabetic retinopathy (Tables 66 to 69).

Table 66 - Accuracy of glycaemia and glycosylated haemoglobin for univariate classification of diabetic retinopathy in type 2 diabetics.

Blood	AUC	SEM	P	LBCI	UBCI	Cut-off	Sensitivity	Specificity	+LR
Glucose	0.536	0.10	0.704	0.348	0.724	-	-	-	-
HbA1C (NGSP) HbA1C (IFCC)	0.541	0.09	0.663	0.357	0.725	- -	-	-	-

Table 67 - Accuracy of creatinine values for univariate classification of diabetic retinopathy in type 2 diabetics.

Variable	AUC	SEM	p	LBCI	UBCI	Cut-off	Sensitivity	Specificity	+LR
Creatinin	0.622	0.09	0.191	0.443	0.801	-	-	-	-

Table 68 - Accuracy of hepatic function parameters for univariate classification of diabetic retinopathy in type 2 diabetics.

Variable	AUC	SEM	p	LBCI	UBCI	Cut-off	Sensitivity	Specificity	+LR
ALT	0.518	0.09	0.850	0.335	0.700	-	-	-	-
AST	0.638	0.09	0.137	0.463	0.812	-	-	-	-
Alkaline Phosphatase	0.570	0.09	0.449	0.390	0.750	-	-	-	-
Gamma GT	0.671	0.09	0.064	0.501	0.841	-	-	-	-

Table 69 - Accuracy of lipid related parameters for univariate classification of diabetic retinopathy in type 2 diabetics.

Variable	AUC	SEM	p	LBCI	UBCI	Cut-off	Sensitivity	Specificity	+LR
Total Cholesterol	0.641	0.09	0.126	0.467	0.816	-	-	-	-
Cholesterol HDL	0.593	0.09	0.317	0.415	0.770	-	-	-	-
Atherogenic Index	0.509	0.09	0.925	0.325	0.692	-	-	-	-
Cholesterol LDL	0.616	0.09	0.220	0.436	0.797	-	-	-	-
Triglycerides	0.559	0.09	0.527	0.375	0.743	-	-	-	-
Apolipoprotein A1	0.620	0.09	0.204	0.439	0.802	-	-	-	-
Apolipoprotein B100	0.537	0.10	0.693	0.349	0.726	-	-	-	-
B100/A1	0.551	0.10	0.589	0.364	0.739	-	-	-	-
Lipoprotein	0.574	0.09	0.431	0.392	0.755	-	-	-	-

However, subjects with diagnosed diabetic retinopathy are more likely to have lower values regarding erythrocytes, haemoglobin and, especially, haematocrit, as presented in Table 70. Parameters respecting hormonology do not differentiate groups (Table 71).

Table 70 - Accuracy of Blood cell counts for univariate classification of diabetic retinopathy in type 2 diabetics.

Variable	AUC	SE	p	LBCI	UBCI	Cut-off	Sensitivity	Specificity	+LR
Leucocytes	0.553	0.09	0.570	0.368	0.737	-	-	-	-
Erytrocytes	0.710	0.08	0.023	0.547	0.873	$\leq$ 4.23	65.00%	75.00%	2.60
Haemoglobin	0.746	0.08	0.008	0.591	0.902	$\leq$ 13.20	75.00%	70.00%	2.50
Haematocrit	0.728	0.08	0.014	0.569	0.886	$\leq$ 36.25	55.00%	90.00%	5.50
MCV	0.630	0.09	0.160	0.453	0.807	-	-	-	-
MHC	0.595	0.09	0.304	0.414	0.776	-	-	-	-
MCHC	0.583	0.09	0.372	0.399	0.766	-	-	-	-
EVC	0.635	0.09	0.144	0.457	0.813	-	-	-	-
Platelet	0.656	0.09	0.091	0.479	0.834	-	-	-	-
MPV	0.554	0.10	0.561	0.366	0.742	-	-	-	-
Plateleocrit	0.641	0.09	0.126	0.468	0.815	-	-	-	-
PVC	0.650	0.09	0.105	0.477	0.823	-	-	-	-

MCV – Mean corpuscular volume; MCH – Mean corpuscular haemoglobin; MCHC – Mean corpuscular haemoglobin concentration; EVC – Erythrocytes variation coefficient; MPV – Mean platelet volume; PVC – Platelet variation coefficient

Table 71 - Accuracy of Blood cell counts for univariate classification of diabetic retinopathy in type 2 diabetics.

Variable	AUC	SE	p	LBCI	UBCI	Cut-off	Sensitivity	Specificity	+LR
TSH	0.526	0.09	0.779	0.342	0.710	-	-	-	-
C-Peptide	0.557	0.09	0.546	0.373	0.740	-	-	-	-

In Figure 46, we may observe the ROC curves for each variable that have previously shown statistical differences between groups. From these, only the ones represented in blue are able to actually separate groups. Although haematocrit presents higher positive likelihood ratio, there is no statistical significant difference between haematocrit and the others regarding the classification accuracy (Figure 47).

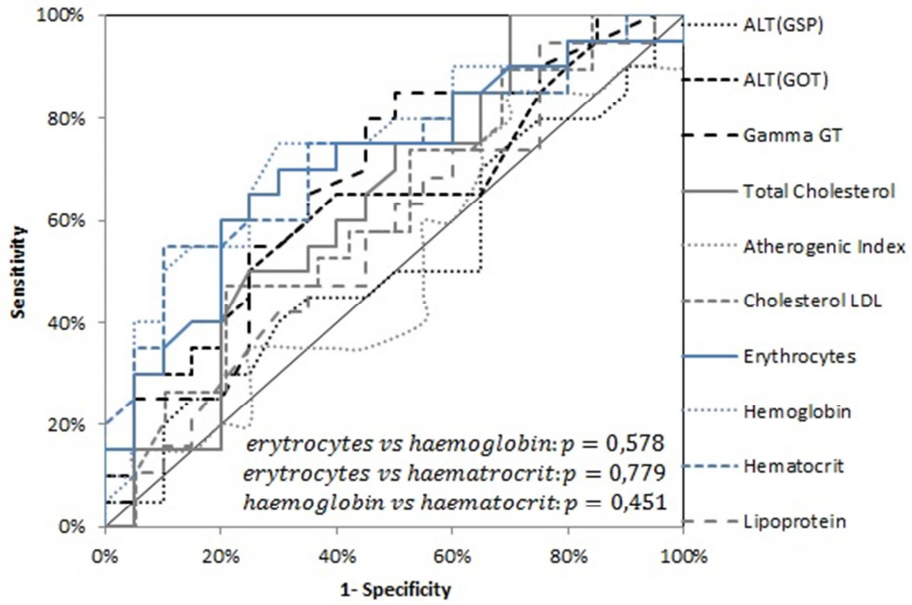

Figure 47 – ROC curves for biochemistry, cell blood counts cytometry and immunology parameters presenting statistical differences between groups. Curves plotted with blue present significant area under the ROC curve.

OCT values of volume scan density may separate groups for the measured values in the nasal quadrant, where it is about 5 times more likely that a subject with diabetic retinopathy presents values of, at least, 355.50 than subjects without diabetic retinopathy. This variable will be used for the development of a multivariate statistical classifier (Table 72).

However, the retinal nerve fibre thickness has no sufficient accuracy for classifying diabetic retinopathy.

Table 72 - Accuracy of OCT tests for univariate classification of diabetic retinopathy in type 2 diabetics.

Variable			AUC	SE	p	LBCI	UBCI	Cut-off	Sens	Spec	+LR
IOP			0.528	0.14	0.825	0.252	0.803	-	-	-	-
BVCA			0.508	0.09	0.935	0.326	0.689	-	-	-	-
OCT		CS	0.670	0.09	0.066	0.501	0.839	-	-	-	-
	VS	Inner N	0.705	0.08	0.027	0.539	0.871	≥ 355.50	50.00%	90.00%	5.00
		Inner S	0.610	0.09	0.234	0.432	0.788	-	-	-	-
		Inner T	0.545	0.09	0.626	0.362	0.728	-	-	-	-
		Inner I	0.603	0.09	0.267	0.424	0.781	-	-	-	-
		Outer N	0.518	0.09	0.850	0.334	0.701	-	-	-	-
		Outer S	0.549	0.10	0.598	0.362	0.735	-	-	-	-
		Outer T	0.490	0.09	0.914	0.306	0.674	-	-	-	-
		Outer I	0.551	0.09	0.579	0.368	0.734	-	-	-	-
	RNFL	Global	0.520	0.10	0.833	0.330	0.710	-	-	-	-
		Nasal	0.558	0.09	0.536	0.374	0.742	-	-	-	-
		N S	0.522	0.10	0.811	0.336	0.709	-	-	-	-
		N I	0.617	0.09	0.211	0.438	0.796	-	-	-	-
		Temporal	0.634	0.09	0.152	0.459	0.810	-	-	-	-
		T S	0.512	0.09	0.899	0.326	0.697	-	-	-	-
		T I	0.596	0.09	0.305	0.414	0.779	-	-	-	-

Results obtained for visual psychophysical tests, either on speed (Table 73), or in achromatic contrast (Table 74), are not able to separate between subjects without and with diabetic retinopathy.

Table 73 - Accuracy of the speed test for univariate classification of diabetic retinopathy in type 2 diabetics.

Speed		AUC	SE	P	LBCI	UBCI	Cut-off	Sens.	Spec.	+LR
Meridian	0º	0.532	0.10	0.738	0.336	0.728	-	-	-	-
	45º	0.569	0.10	0.488	0.372	0.765	-	-	-	-
	90º	0.550	0.10	0.605	0.357	0.742	-	-	-	-
	135º	0.526	0.10	0.788	0.328	0.725	-	-	-	-
Area		0.556	0.10	0.581	0.355	0.756	-	-	-	-

Table 74 - Accuracy of the achromatic test for univariate classification of diabetic retinopathy in type 2 diabetics.

Achromatic		AUC	SE	P	LBCI	UBCI	Cut-off	Sens.	Spec.	+LR
Meridian	0º	0.659	0.09	0.098	0.480	0.839	-	-	-	-
	45º	0.508	0.10	0.937	0.313	0.702	-	-	-	-
	90º	0.538	0.10	0.693	0.347	0.729	-	-	-	-
	135º	0.625	0.10	0.208	0.437	0.813	-	-	-	-
Area		0.556	0.10	0.581	0.356	0.755	-	-	-	-

From previously identified measures for chromatic contrast discrimination presenting statistical significant differences between groups, all of them showed sufficient accuracy for separating between subjects without and with diabetic retinopathy. The one with higher

positive likelihood ratio is contrast sensitivity along meridian 135º, on the Tritan axis (PLR = 6.35), as observed in Table 75, but the one with highest sensitivity is the global area generated by the 5-sides polygon formed by the values measured in each meridian and the origin (94.12%).

Table 75 - Accuracy of the chromatic vision test for univariate classification of diabetic retinopathy in type 2 diabetics.

Axis/Meridian		AUC	SE	P	LBCI	UBCI	Cut-off	Sens.	Spec.	PPV
Chromatic Contrast	Protan 0º	0.610	0.10	0.254	0.418	0.801	-	-	-	-
	Protan 45º	0.520	0.10	0.832	0.330	0.711	-	-	-	-
	Protan 90º	0.503	0.10	0.976	0.313	0.693	-	-	-	-
	Protan 135º	0.551	0.10	0.601	0.357	0.745	-	-	-	-
	Protan Area	0.534	0.10	0.727	0.336	0.732	-	-	-	-
	Deutan 0º	0.701	0.09	0.042	0.528	0.874	$\geq 4.46 \times 10^{-3}$	58.82%	77.78%	2.65
	Deutan 45º	0.665	0.09	0.096	0.483	0.847	-	-	-	-
	Deutan 90º	0.542	0.10	0.669	0.347	0.737	-	-	-	-
	Deutan 135º	0.517	0.10	0.863	0.318	0.717	-	-	-	-
	Deutan Area	0.568	0.10	0.501	0.371	0.764	-	-	-	-
	Tritan 0º	0.776	0.08	0.004	0.616	0.936	$\geq 63.25 \times 10^{-3}$	72.22%	78.95%	3.43
	Tritan *45º	0.594	0.10	0.335	0.405	0.783	-	-	-	-
	Tritan 90º	0.668	0.09	0.081	0.493	0.844	-	-	-	-
	Tritan 135º	0.792	0.08	0.003	0.633	0.952	$\geq 174.81 \times 10^{-3}$	70.59%	88.89%	6.35
	*Area	0.735	0.09	0.017	0.566	0.905	$\geq 5.94 \times 10^{-3}$	94.12%	55.56%	2.12

In Figure 48 we may observe the ROC curves for parameters which presented significant areas under the ROC curve.

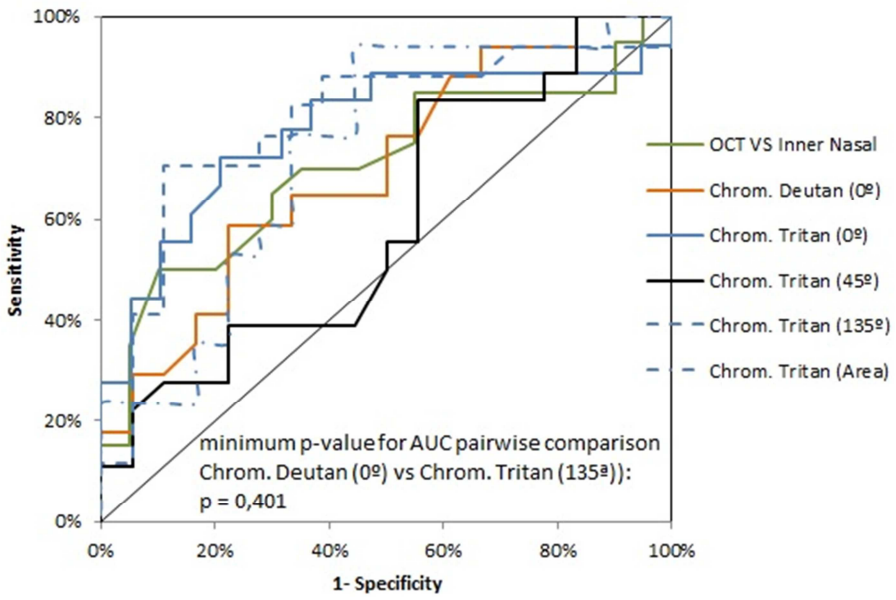

Figure 48 – ROC curves for OCT Volume Scan and Visual psychophysical tests that presented statistical differences between groups. Curves plotted in blue, green and orange present significant area under the ROC curve.

3. Multivariate models for Diabetic Retinopathy Classification

The following analyses were performed using as independent variables all the previous variables that presented association with the presence of diabetic retinopathy and achieved statistical significance in area under the ROC curve, such as:

- duration of the disease, in years;
- cytometry parameters: erythrocytes, haemoglobin and haematocrit;
- volume scan in the inner nasal quadrant;
- visual psychophysical tests: chromatic vision for The Deutan axis (meridian 0º) and for the Tritan axis (meridians 0º and 135º) and the area of the polygon generated by the four meridians and the origin for the Tritan axis.

Discriminant analysis and tree analysis used these quantitative variables, but on logistic regression they were dichotomized according to previous ROC results.

3.1 Discriminant Function Analysis

Some of the independent variables do not follow a normal distribution, as observed in Figure 49, but there is homogeneity on the covariate matrices (Box's M F(10, 4552) = 4.79; p = 0.942).

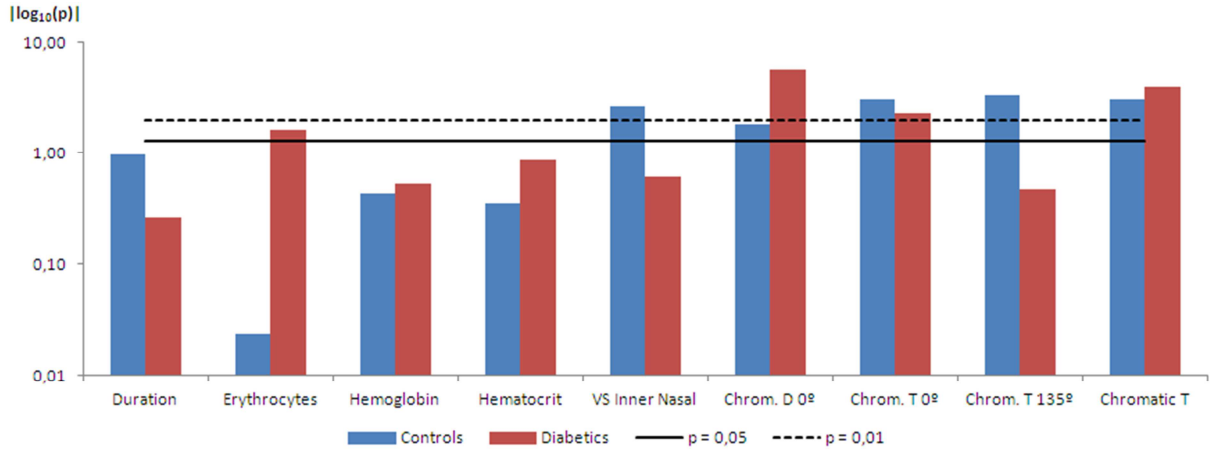

Figure 49 - p-values obtained from the Kolmogorov-Smirnov or the Shapiro-Wilk test to variables in analysis presented as $|\log_{10} p|$, in logarithmic scale. The horizontal lines reflect the values of 0.01 ($|\log_{10} 0.01| = 2.00$) and 0.05 ($|\log_{10} 0.05| = 1.30$) for type I errors. All bars below horizontal lines represent variables with normal distribution in the group.

Wilk's lambda identifies 4 variables (Table 76) as group discriminators:

Table 76 - Variables included in the discriminant model (Wilks' Lambda method).

Entered		Wilks' Lambda							
		Statistic	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
x_1	Chrom. Cont. Tritan (135º)	0.783	1	1	31	8.611	1	31	0.006
x_2	Duration	0.632	2	1	31	8.745	2	30	0.001
x_3	Haemoglobin	0.497	3	1	31	9.801	3	29	< 0.001
x_4	VS Inner Nasal	0.413	4	1	31	9.942	4	28	< 0.001

After four steps, one discriminant function is obtained with an eigenvalue of 1.42, explaining 100% of the variance, and a canonical correlation of 0.766 (λ_{Wilks}=0.280: $\chi^2_{(4)}$=25.63; p < 0.001) between variables considered for the model and group classification.

The discriminant function can be written as:

$$F(x) = -1.672 + 5.215x_1 + 0.109x_2 - 0.590x_3 + 0.021x_4$$

or, after standardizing coefficients,

$$F_S(x) = 0.613x_1 + 0.925x_2 - 0.678x_3 + 0.596x_4$$

Function centroids are significantly different (F(4,28) = 9.94; p < 0.001) and assume the values of -1.12 for subjects without diabetic retinopathy and +1.19 for those who have diabetic retinopathy. This means that the function can separate or discriminate groups and may be used for classification.

We obtained the classification functions, which may be used to classify, although that was not the methodology used for classification. Classification functions may be written as:

$$\begin{cases} Class_{DR\ absent} = -144.254 - 0.883x_1 + 0.409x_2 + 9.263x_3 + 0.457x_4 \\ Class_{R\ presentD} = -148.198 + 11.170x_1 + 0.660x_2 + 7.900x_3 + 0.506x_4 \end{cases}$$

The method for classification is based upon the closeness to the centroid, and distance from the frontier line, which has the value of$f = +0.002$. Based on this frontier, and on the Mahalanobis distance from the score in the discrimination function and the centroid, we may also obtain posterior probabilities and classify a given subject according to the highest probability.

As it was previously referred, Mahalanobis squared distances between the score in the discriminant function and group centroids follow a Chi-square distribution with one degree of freedom, which enables the determination of posterior probabilities. These distances,

$d^2_{DR\ absent}$ and $d^2_{DR\ present}$, in this particular case, are given by:

$$\begin{cases} d^2_{DR\ absent} = \frac{f(x)+1.121}{0.892} \\ d^2_{DR\ present} = \frac{f(X)-1.191}{1.099} \end{cases}$$

Posterior probabilities may be calculated applying Bayes' rule to the probability of obtaining each one of the distances, given that the subjects belongs to a defined group:

$$\begin{cases} P(DR_{Absent}|d^2_{DR\ absent}) = \frac{p_1 x P(\chi^2_1 > d^2_{DR\ absent})}{p_1 x P(\chi^2_1 > d^2_{DR\ absent}) + p_2 x P\left(\chi^2_1 > d^2_{DR\ present}\right)} \\ P\left(DR_{Present}|d^2_{DR\ present}\right) = \frac{p_2 x P(\chi^2_1 > d^2_{DR\ present})}{p_1 x P(\chi^2_1 > d^2_{DR\ absent}) + p_2 x P\left(\chi^2_1 > d^2_{DR\ pesent}\right)} \end{cases}$$

Where p_1 and p_2 are the prior probabilities, which were assumed to be equal (0.50). The subject is classified as DR absent or DR present according to the highest posterior probability.

We can apply a ROC analysis either to the discriminant function, or to the posterior probabilities, and a cut-off of +0.264 for the frontier line is obtained as the optimal cutoff, corresponding to the posterior probability of 58.46%. This means that we may improve specificity of classification, since the number of false positive cases decreases, without losing sensitivity; consequently, the positive likelihood ratio increases three times which is preferable. Note that concordance between models is excellent, thus any of them may be used for classification (Table 77).

Table 77 - Discriminant classifier accuracy using two different cut-offs for posterior probability: classical (50%) and obtained by ROC analysis (58.46%).

Model	AUC (p)	Cut-off Function (Prob)	% Correct	k	P	K between	McNemar (p)	Sens	Spec	+LR
D_F	0.931 (< 0.001)	0.002 (50.00%)	88.57%	0.770	< 0.001	1.000 (p < 0.001)	0.625	82.35%	94.44%	14.82
D_{ROC}		0.214 (58.46%)	88.57%	0.770	< 0.001		0.625	82.35%	94.44%	14.82

Similarly, Figure 50 represents the scatterplot of the probability for group classification, where blue points represent controls and red points represent type 2 diabtetics. There, we may observe the posterior probability of belonging to the control or diabetic group according to the value obtained in the discriminant function, and its distance to the correspondent centroid. Horizontal lines mark the cut-offs defined by discriminant analysis

(50%) and ROC analysis (58.46%) for group classification while vertical lines mark the cut-offs defined by the definition frontier line (0.002) or obtained by ROC analysis (0.214), used for classification with the discriminant function.

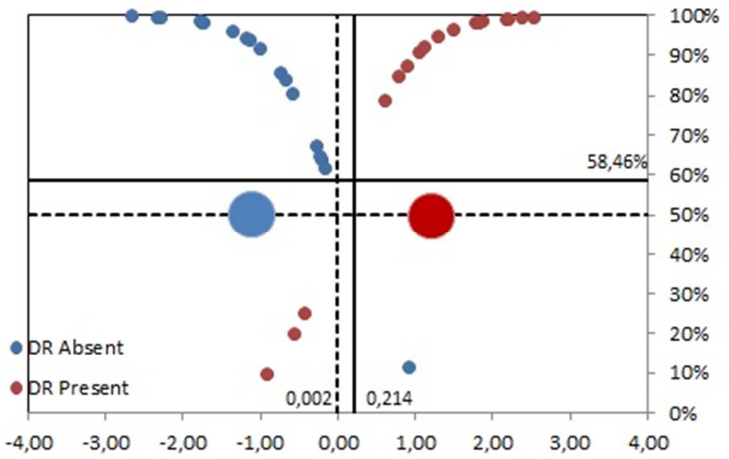

Figure 50 - Scaterplott of the probability for group (y) control (blue) or diabetic (red), based on the value of the discriminant funcion (d). Horizontal lines mark the cut-offs defined by discriminant analysis (50%) and ROC analysis (58.46%) for group classification while vertical lines mark the cut-offs defined by the definition frontier line (-0.12) or obtained by ROC analysis (-0.28), used for classification with the discriminant function; big circles mark the centroid for the discriminant function (at 50% probability) for each group.

Cut-offs defined lead exactely to the same solution, thus the 50% posterior probabily will be used. In the left half, cases are classified as type 2 diabetics without diabetic retinopathy, and in the upper-left quarter we have the true negative cases; on the other hand, true positive cases are represented in the right-upper quarter, and false positive cases in the right-bottom quarter of Figure 50.

3.2 Regression procedures

Logistic regression was performed using the same independent variables as the ones used with discriminant analysis, but they were previously dichotomised according to the cut-offs defined by ROC analysis. We attempted to build three models based on a forward stepwise procedure (with a probability of F to enter < 0.050 and a probability of F to remove > 0.100), using the conditional statistic, the Likelihood Ratio (LR) and the Wald statistic.

Curiously, the first two models identified the duration of the disease, haematocrit, and global chromatic contrast at the Tritan axis (area), while the last model identified only the duration of the disease as a discriminator variable. However, none of the models presented statistical significance regarding the coefficients of independent variables.

The following tables present the adjustment and accuracy of the models (Tables 78 and 79). We may observe a statistical significant improvement with the inclusion of each variable in the model. We may also observe that the quality of the adjustment of the models obtained with conditional or likelihood ratio criteria for stepwise regression are quite good. However, Wald's criteria for stepwise regression do not allow the achievement of a good model (Table 78).

Table 78 - Significance of models and improvement, step by step, on forward stepwise logistic regression model (Conditional, Likelihood Ratio and Wald's methods).

Step/Variable		Improvement ($\chi^2_{(df)}$; p)			Model ($\chi^2_{(df)}$; p)		
		Cond.	LR	Wald	Cond.	LR	Wald
x_1	Chrom. T (area)	12.48(1); < 0.001	12.48(1); < 0.001	12.48(1); < 0.001	12.48(1); < 0.001	12.48(1); < 0.001	12.48(1); < 0.001
x_2	Duration	6.01(1); 0.014	6.01(1); 0.014	-	18.49(2); < 0.001	18.49(2); < 0.001	-
x_3	Haematocrit	4.75(1); 0.029	4.75(1); 0.029	-	23.24(3); < 0.001	23.24(3); < 0.001	-

Table 79 - Adjustement of the model, step by step, to observed data, and overall correlation.

Step/Variable		Nagelkerque R square			Hosmer and lemeshow test ($\chi^2_{(df)}$; p)		
		Cond.	LR	Wald	Cond.	LR	Wald
x_1	Chrom. T (area)	0.531	0.531	0.531	0(0); -	0(0); -	0(0); -
x_2	Duration	0.707	0.707	-	0(2); 1.000	0(2); 1.000	-
x_3	Haematocrit	0.818	0.818	-	0(3); 1.000	0(3); 1.000	-

However, as stated before, coefficients do not show statistical significance; in fact, they reveal there must be some redundancy between variables which may lead to some over fitting of the models, due to standard error obtained for the coefficients (Table 80).

Table 80 - Odds ratio and confidence intervals for variables identified on logistic regression model (models Conditional and Likelihood ratio).

	B	SE	Wald	df	Sig.	Odds Ratio	95% CI for Odds Ratio	
							Lower	Upper
Chrom. T (area)	21.621	1.57	3.663	1.00	0.056	20.00	0.93	429.90
Duration	2.996	14736.21	0.000	1.00	0.998	$1.17x10^{17}$	-	-
Haematocrit	39.299	14736.21	0.000	1.00	0.998	0.000	-	-

In fact, only the global area is nearly significant, and the following variables indicate redundancy. Hence, a new model was designed, using only the global area of the polygons generated by the origin and the four meridians measured at the Tritan axis, concerning chromatic contrast sensitivity. This variable explains 36.60% of the variance obtained for the probability of a type 2 diabetic person to develop diabetic retinopathy, since the Nagelkerke

r-squared is 0.366. The model may be written as

$$P(DR_{Present}) = \frac{e^{-2.303+2.996\times ChromT}}{1+e^{-2.303+2.996\times ChromT}} \text{ and } ChromT = \begin{cases} 0, if\ Area\ ChromT < 5.94\times 10^{-3} \\ 1, if Area\ ChromT \geq 5.94\times 10^{-3} \end{cases}$$

In fact, this variable presents a value of significance for the prediction of 0.008, and when the area of the polygon (Figure 51) generated by the values measured in the four meridians and the origin is, at least, $5.94x10^{-3}$, the mean risk of a type 2 diabetic to have diabetic retinopathy is 20 times higher, with a 95% confidence interval of 2.16 to 184.87.

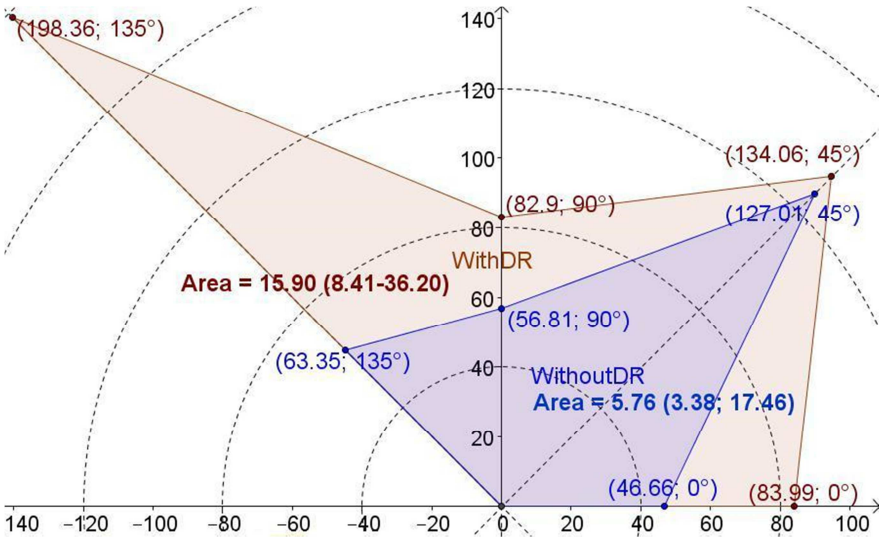

Figure 51 - Chromatic contrast test (Tritan) on meridians 0º, 45º, 90º, 135º and global area generated by these meridians in type 2 diabetes without and with diabetic retinopathy (meridian and area values should be read $x10^{-6}$; area values should be read $x10^{-6}$)

This model correctly classifies 73.61% at the training sample, although Kappa's Coefficient of concordance is only 0.456 ($p < 0.001$). The problem is that this model has a high false positive rate, statistically different from the false negative rate which is 5.88% (McNemar p-value < 0.001).

Using the cut-off of 37.88% for the predicted probability regarding diabetic retinopathy (identified has the optimal cut-off by ROC analysis on those probabilities), there is absolutely no gain, since classification is the same as the one obtained using the 50% value for cut-off, at the training sample (Table 81).

Table 81 - Logistic regression classifier accuracy using two different cut-offs for posterior probability: classical (50%) and obtained by ROC analysis (37.88%).

Model	AUC (p)	Cut-off (%)	% Correct	K	P	K between	McNemar (p)	Sens	Spec	+LR
L_F	0.748 (0.012)	50.00%	73.61%	0.456	< 0.001	1.000 (< 0.001)	< 0.001	94.12%	67.27%	2.88
L_{ROC}		37.88%	73.61%	0.456	< 0.001		< 0.001	94.12%	67.27%	2.88

3.3 Decision trees

As before, four algorithms were applied for growing decision trees: CART, CHAID, Exhaustive CHAID and QUEST. However, CHAID and exhaustive CHAID algorithms were unable to grow a tree, perhaps due to sample size, and CART and QUEST algorithms lead to the same solution, presented in Figure 52, identifying only chromatic contrast in the Tritan axis, at meridian 135º, as a classifier of diabetic retinopathy.

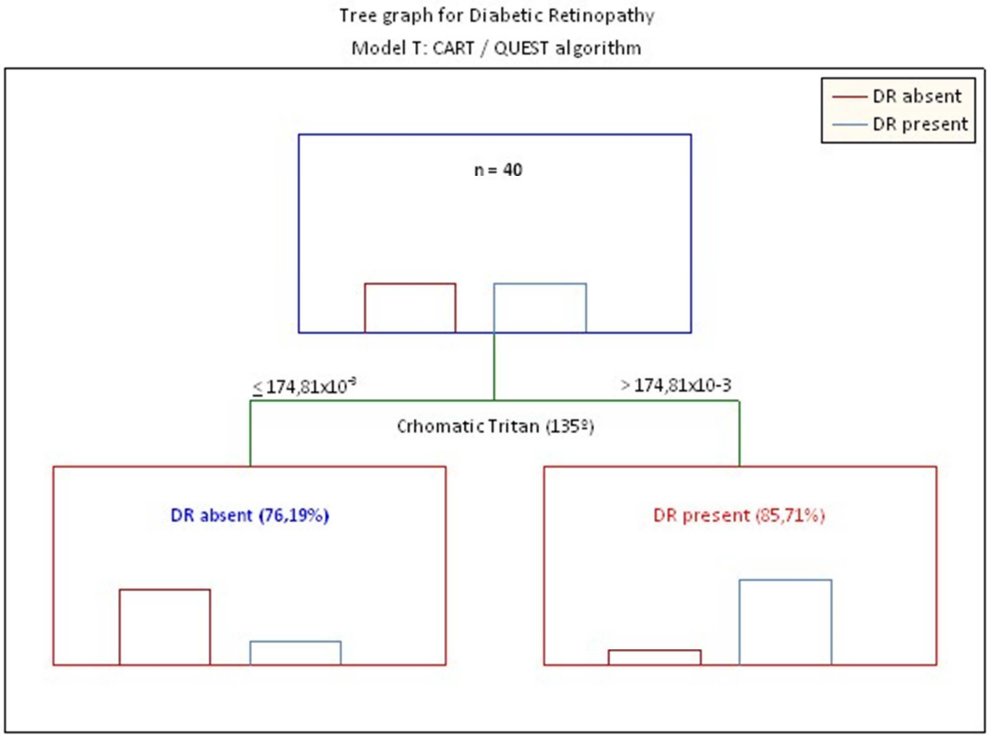

Figure 52 – Decision tree for classification of diabetic retinopathy.

The obtained model exhibited good accuracy, and an excellent positive likelihood ratio, as observed in Table 82.

Table 82 - Decision tree classifier accuracy.

Model	AUC (p)	% Correct	K	p	McNemar (p)	Sens.	Spec.	+LR
T	0.797 (0.003)	90.28%	0.713	< 0.001	0.453	70.59%	88.89%	18.41

4. Model Comparison

Despite the used methodology for evaluation of diabetes classification, where obtained classifiers were applied to a test sample, at this moment it still is not possible to use the same methodology for evaluation of the diabetic retinopathy classifiers that were

developed, since only five cases remain, all with diabetic retinopathy. Hence, further work must be done, and at the moment, classifiers performance will be evaluated only in the training sample.

All the models were applied to the same 35 cases, and the value of 50% probability of belonging to the group with diabetic retinopathy was used as cut-off. The model with highest sensitivity is the one obtained with logistic regression, which only uses chromatic contrast sensitivity on the Tritan axis for classification, and may be used as screening in a regular visit to the ophthalmologist.

On the other hand, the model obtained with discriminant analysis is the one with highest specificity, also using contrast sensitivity at the Tritan axis (meridian 135º), but also needs OCT acquisition and laboratory tests, namely values of visual scan in inner nasal region and haemoglobin, for classification. It also includes the duration of disease as a marker of diabetic retinopathy, which is an abstract parameter, difficult to measure with precision, since diabetes is a silent disease until diagnosed.

The model obtained with a decision tree algorithm is the one with worse sensitivity, and intermediate specificity, and is not much different from the logistic regression model, since it only uses chromatic sensitivity at the Tritan axis for classification. The decision tree algorithm uses the 135º meridian for classification, while the logistic model needs computation of the area of the polygon generated by the measures on the four meridians and origin, which is a simple process.

Models present moderate agreement, pair by pair, since Kappa coefficient between the discriminant model and the logistic model is 0.512 ($p = 0.001$); between the discriminant model and the decision tree model, the Kappa coefficient is 0.588 ($p < 0.001$) and between the logistic and decision tree models, kappa is 0.468 ($p = 0.001$).

Hence, the most complex model (the discriminant one) presents the higher accuracy, given by the area under the ROC curve for probabilities of presence of diabetic retinopathy, with statistical significant difference both to the logistic regression model ($p = 0,002$) and to the decision tree model ($p = 0.044$). These p-values were not corrected for multiple comparisons, thus by the Bonferroni rule, which is rather conservative, but may be applied here, the only difference that survives to multiple comparisons is the difference between the accuracy of the discriminant analysis model and the logistic regression model (Table 83).

Table 83 - Comparison of diabetic retinopathy classifiers on the training sample.

Model	AUC (p)	AUC between	% Correct	k	p	McNemar (p)	Sens	Spec	+LR
D	0.931 (< 0.001)	vs L: p = 0.002 vs T: p = 0.044	88.57%	0.770	< 0.001	0.625	82.35%	94.44%	14.82
L	0.748 (0.012)	vs T: 0.530	73.61%	0.456	< 0.001	0.039	94.12%	67.27%	2.88
T	0.797 (0.003)		90.28%	0.713	< 0.001	0.453	70.59%	88.89%	19.41

In fact, the expected values for population, with 95% confidence, and considering a prevalence of 34.6% on diabetic retinopathy among diabetics (not only type 2), getting the predictive values presented in Table 84.

Table 84 - Expected values of sensitivity, specificity, positive likelihood ratio and predictive values.

Model	Sensivity (95% CI)	Specificity (95% CI)	+LR (95% CI)	PPV (95% CI)	NPV (95% CI)
D	82.35% (56.6% - 96.2%)	94.44% (72.7% - 99.9%)	14.82 (2.2 - 100.9)	88,7% (56,0% - 99,5%)	91,0% (71,6% - 98,8%)
L	94.12% (71.3% - 99.9%)	67.27% (53.3% - 79.3%)	2.88 (1.9 - 4.3)	60,3% (43,2% - 75,8%)	95,6% (82,1% - 99,7%)
T	66.67% (41.0% - 86.7%)	96.61% (88.3% - 99.6%)	19.67 (4.8 - 79.8)	91,1% (69,0% - 99,2%)	86,1% (73,7% - 94,1%)

As the clinical importance of positive predictive value, in practice, is higher than sensitivity, and no statistical significant differences are found between the discriminant model and the decision tree model (adjusted p-value by the Bonferroni correction is 0.132), data presented on Table 84 and projected on Figure 53, suggest the use of the decision tree classifier, which is also a simpler method. However, models should be evaluated in a test sample before we decide for a definitive model.

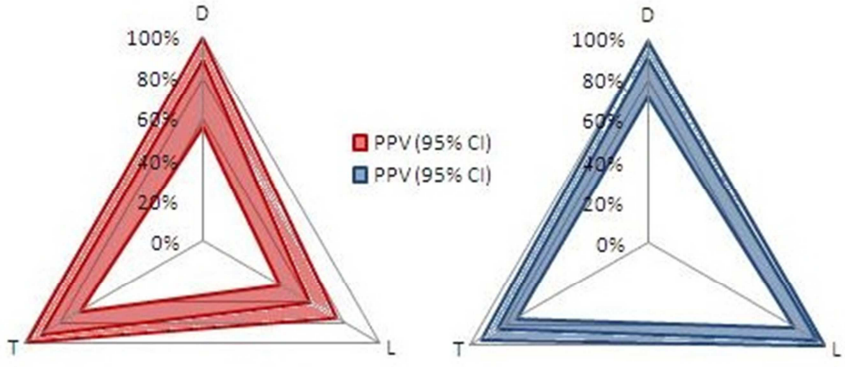

Figure 53 – Predictive values and 95% confidence interval relative to the three classifiers developed, assuming a prevalence of 34.6%..

CHAPTER 7

DISCUSSION

In order to identify possible type 2 diabetes markers that indicate the presence of diabetes, or the presence of non-proliferative diabetic retinopathy, we had several dilemmas to handle.

Perhaps the greatest dilemma was due to the fact of having age mismatch between original groups, when classifying type 2 diabetes, and the fact of having correlated data for the two eyes, whatever classification we were performing.

For age mismatch between controls and type 2 diabetics, the simplest solution was to enter age in the multivariate models in order to ascertain if that was a differentiation parameter. Naturally, vision is affected by the aging process, but we did not find any correlation between age and any of the clinical or demographic measures (Figure 30) or between age and blood test or eye related measures, whichever the group considered.

For duplicate data for the eyes, other considerations were made. In fact, OCT data were collected on both eyes, but visual psychophysical tests were performed only in the dominant eye. The simplest way to carry the analysis would be to discard collected data from the non-dominant eye, but then we could be wasting important information.
Armstrong[32] published, in 2013, guidelines about how to handle both eye data. The majority of studies that in literature that he revised considered, most of the times, only the right eye, or the dominant eye, in spite both eyes data were available. In cases where both eye information was collected and used, most of the studies in the three journals he revised (OVO, OVS, CBO) considered both eye but uncorrected for correlation between eyes.

As no association was found, by Armstrong, between the methods to select the eye for analysis and journals, it seems reasonably to say that there were no previous defined guidelines to handle this situation. Nevertheless, there seems to be present some heterogeneity between methodologies, within and between journals. A meta-analysis is a procedure that integrates quantitative findings from separate but similar studies and provides a numerical estimate of the overall effect of interest[90]. However, with only three journals revised it seems unreasonable to perform a meta-analysis[91,92]. Nevertheless, we may use a descriptive meta-analysis, that is, we may plot available data and observe variability without determining the usual Cochran's Q or I^2 statistics used to evaluate heterogeneity[93].

When only one eye data was collected, the proportion of expected dominant eyes used, under the random effects model, lies between 10.49% and 23.78%, with 95% confidence interval, being the right eye the one most used, as observed in Table 85 and Figure 54.

Table 85 - Data information from one eye – only one eye data collected.

Criteria (one eye data)		Sample size	Right Eye		Random Eye		Dominant Eye	
			Proportion (%)	95% CI	Proportion (%)	95% CI	Proportion (%)	95% CI
Journal	OVO	51	47.06	32.93 - 61.54	5.88	1.23 - 16.24	9.80	3,26 - 21,41
	OVS	62	30.65	19.56 - 43.65	17.74	9.20 - 29.53	19.36	10,42 - 31,37
	CBO	35	25.71	12.49 - 43.26	14.29	4.81 - 30.26	20.00	8,44 - 36,938
Total Effects (Random)		148	34.92	23.46 - 47.33	12.97	6.43 - 21.38	16.60	10.49 - 23.78

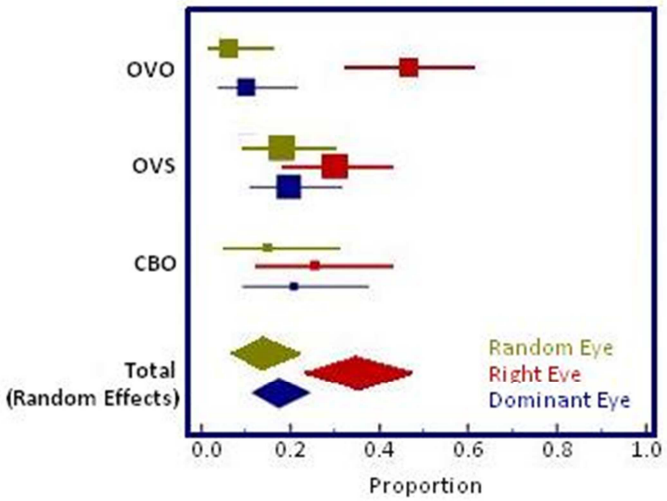

Figure 54 – Expected proportion of papers using right, random or dominant eye for analysis, in OVO, OVS, CBO and global measure of the total expected proportion.

When both eye data were collected, there was also much heterogeneity, as observed in Table 86 and Figure 55.

Table 86 - Data information from both eyes – two eye data collected.

Criteria (one eye data)		Sample size	One eye only		Both Corrected for Correlation		Both Uncorrected for Correlation	
			Proportion (%)	95% CI	Proportion (%)	95% CI	Proportion (%)	95% CI
Journal	OVO	19	15.79	3.38 - 39.58	10.53	1.30 - 33.14	31.58	12.57 - 56.55
	OVS	32	28.13	13.75 - 46.75	15.63	5.28 - 32.79	34.38	18.57 - 53.19
	CBO	31	19.36	7.45 - 37.47	9.68	2.04 - 25.75	38.71	21.85 - 57.81
Total Effects (Random)		82	22.74	14.51 - 32.20	13.37	7.02 - 21.38	35.85	26.05 - 46.29

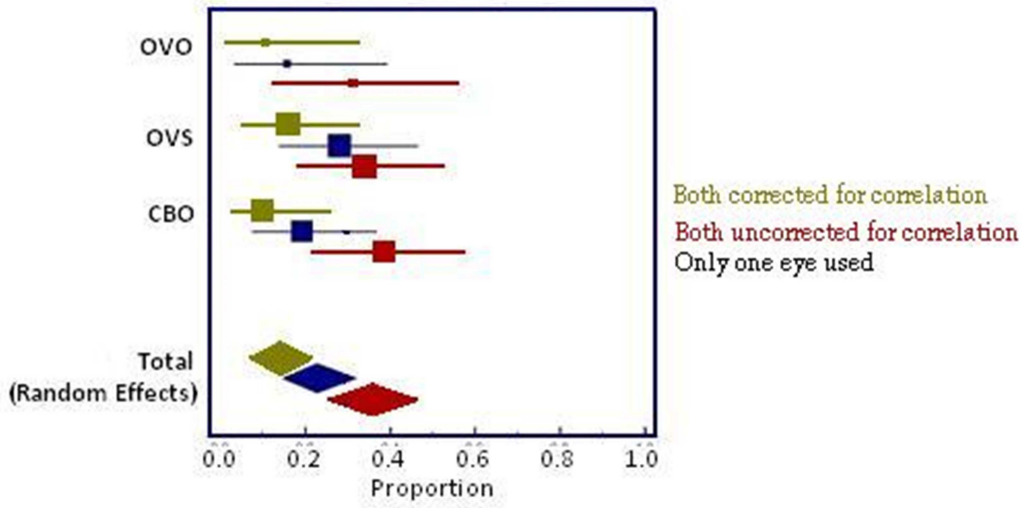

Figure 55 – Expected proportion of papers using Right, random or dominant eye for analysis, in OVO, OVS, CBO and global measure of the total expected proportion, when data from both eyes were available.

To join this, there were no criteria defined, in these journals, about when to use one or both eye information. This is why Armstrong's guidelines[32] are so important. Nonetheless, he does not refer guidelines for studies where we have data from one eye in some variables, and data from both eyes in other variables. Furthermore, no guidelines are defined about the use of both eyes information for classification problems, or for analysis of correlation when data are not Gaussian.

In fact, Armstrong[32] suggests the use of the intra-class correlation coefficient or the concordance correlation coefficient to evaluate correlation among measurements. However, in a certain way, both are based on the normal distribution assumptions, since the intra-class correlation coefficient is determined using mean or variance between measurements, and concordance correlation coefficient is based upon Pearson's correlation coefficient. Hence, we propose a pseudo-concordance correlation coefficient, based on Spearman's Rank order correlation, which have showed to be close to the other measures in the evaluation of correlation and concordance between eyes.

Armstrong[32] also suggests the use of Bland and Altman plots to measure agreement. Once again, this graphical procedure is based on the mean and standard deviation (SD), and its frontier lines are defined at the mean $\pm$ 1.96 SD of the differences, that is, by default, it assumes that data are normally distributed, which is often not the case. The Youden plot, which is centred on the median of each one of the measures, is an alternative procedure for graphical evaluation of correlation and concordance between measurements, without normality assumptions. However, as it also gives also a quantification of the random error between measurements, it may be used to access differences between groups and, by doing so we may access a global measure of precision between eye measurements in group observations, and a measure of precision of instrument measurement between eyes.

We decided to use only one eye in the study, as concordance between eyes was observed to be present, and there were no statistical significant differences between eyes in the majority of analyzed variables. In fact, differences were found only on the temporal quadrant of the RNFL ($p < 0.001$), and on the nasal-superior ($p < 0.001$) and nasal-inferior quadrants ($p = 0.002$), where no statistical differences in the random error between eyes were found. However, we found that the random error, which is not controlled as systematic error may be, presents statistical differences between controls and type 2 diabetics in the measurements made for the left and the right eyes, at the inner-superior and inner-inferior quadrants of volume scan density acquired by Spectralis OCT. At the end, there were no statistical significant differences between the eyes in these quadrants, thus random error was ignored. Likewise, as the methodology should be the same in all the study and as concordance was highly significant, being the object of study individuals and not eyes, we decided to use only one eye information.

As Armstrong[32] presented, the variability for the chosen eye for analysis is large. However, we had some directives to choose the eye, since data available for the visual psychophysical tests were for the dominant eye. Maybe in future studies the design should consider collecting data in visual psychophysical tests for both eyes in order to evaluate correlation, concordance and random error in visual psychophysical tests. Therefore, we used only the dominant eye data, even when we had both data available.

Another dilemma we had to handle was that we had as many variables as cases, but a data reduction based upon simple statistic methods was performed. These methods involved two group comparisons and posterior ROC analysis, which enabled us to discard useless variables, as they did not have any group separating property.

Hence, we identified a smallest subgroup of variables which allowed differentiation between controls and type 2 diabetics, as presented on chapter 1 of part II of the results, and a subgroup of variables which could separate subjects with diabetic retinopathy from others without diabetic retinopathy, within the type 2 diabetes group. Here, a question arises: why did we use two statistical classifiers and not just one which would discriminate between three groups, that is, between controls, type 2 diabetics without diabetic retinopathy and those who have diabetic retinopathy? The reduction of variables could be performed by similar methods, using univariate tests for independent samples that allow comparison between three groups, such as ANOVA or Kruskal-Wallis test as the first criteria for discarding some variables, and ROC analysis could be applied in order to discriminate between each pair of groups. Statistical classification could be performed using the same or similar methods since there are classification functions which permit discrimination between three groups. Discriminant function analysis maybe used for more than three groups, logistic regression should be replaced for ordinal regression, and decision tree algorithms can handle more than two groups. However, we could not use some variables that were measured only in the type 2 diabetic group, such as duration of the disease. This could be set to zero years, in controls, but we were biasing results. On the other hand, ETDRS grading for diabetic retinopathy was performed only in the diabetic group. Using binary variables for groups, we also have similar group dimension for the training sample (47 controls and 49 type 2 diabetics; 20 type 2 diabetics without and 20 with diabetic retinopathy), instead of having one variable with different distribution along groups. This fact is an asset, especially when discriminant analysis assumption fails, as the F distribution is very robust to the violation of multivariate normal distribution when groups have similar dimensions. However, it is not to reject the hypothesis of, when all data are available, developing a classification function that empowers classification into one of the three groups, simultaneously, or even into four groups if the sample will include cases with proliferative diabetic retinopathy, since we expect to have between 200 and 300 controls and a similar number of type 2 diabetics at the end of the inclusion process. At that time, data reduction will also be needed, since we may also use multimodal imaging results obtained for brain, heart and liver. Hence, this is an

ongoing process and these are the preliminary results in classification of type 2 diabetes or non-proliferative diabetic retinopathy using only blood sample and eye data.

Some analyses were performed in smaller subgroups, adding blood pressure plus systolic and diastolic volume obtained by heart imaging, which were found to be related with the presence of non-proliferative diabetic retinopathy. However, few cases had data inserted onto the database on these parameters and, therefore, multivariate tests could not include these variables since sample size would be inadequate for classification methods. Further work will be done when heart, brain and liver imaging data is processed and ready to be analyzed.

Considering classification results, they will be separated now into two different sections. Let us focus on type 2 diabetes classification.

According to sociodemographic parameters (Table 10 and Figure 26), we found that type 2 diabetics were significantly older ($p < 0.001$), presented significantly higher body mass index ($p < 0.001$), being heavier and smaller in height than controls, showed a tendency for differences for the abdominal perimeter, higher for this group ($p = 0.053$), as well as higher systolic blood pressure ($p < 0..001$) although no differences were found in diastolic blood pressure or pulse. Notice that the percentage of subjects medicated for blood pressure in this group was significantly higher than in the control group ($p < 0.001$), as well as the prevalence of family history of diabetes ($p < 0.001$).

Comparison between groups for blood glucose and glycosylated haemoglobin were obviously statistical significant (Table 11) and were not used as possible classifiers, since this were the parameters used as gold standard to confirm the presence of type 2 diabetics.

Considering all the other variables measured in blood samples, and presented on Tables 12 to 18, we were able to identify that type 2 diabetics had statistical significantly differences values for ALT ($p = 0.029$), alkaline phosphatase ($p = 0.004$), gamma GT ($p = 0.020$), total cholesterol ($p = 0.001$), cholesterol HDL ($p < 0.001$) and LDL ($p=0.003$), atherogenic index ($p = 0.028$), triglycerides ($p < 0.001$), apolipoprotein A1 ($p < 0.001$), leucocytes ($p = 0.024$), haemoglobin ($p = 0.004$), haematocrit ($p = 0.006$) and erythrocyte coefficient of variation ($p = 0.043$), C-peptide ($p = 0.002$), and the descriptive statistics allowed us to trace a preliminary profile of type 2 diabetics under treatment, since all of them had previously been diagnosed at least one year before: they showed some potentially liver damage, higher

risk of cardiovascular disease reflected by higher levels on triglycerides and atherogenic index, in spite of better control of cholesterol values (although they presented lower values either of LDL cholesterol or of HDL cholesterol), more prone to have lower values of haemoglobin, indicating lower oxygenation levels in this group, and lower values of haematocrit with higher values of erythrocyte coefficient of variation and, therefore, lower blood viscosity, perhaps due to the fact that the majority of type 2 diabetics are medicated for hypertension. As expected, type 2 diabetes subjects presented lower values of C-peptide, as it reflects the amount of insulin present in blood.

In fact, type 2 diabetics presented higher levels of ALT, an indicator of liver damage or injury, and higher levels of the enzymes alkaline phosphatase and gamma GT, related to all forms of liver disease. The maximum value for controls is within the normal range of alkaline phosphatase levels, but for type 2 diabetics it exceeded the normal maximum value (Table 13), and for those patients it may be an indicator of biliary obstruction. Gamma GT values in the blood are an indicator of the liver and biliary systems.

Parameters concerning cholesterol are significantly lower in the diabetic group, either the total cholesterol or the LDL and HDL cholesterol (Table 14). LDL cholesterol can build-up a lining over the walls of the arteries and increase the risk of heart disease, and should be below 129 mg/dL; HDL cholesterol protects against heart disease, by eliminating LDL cholesterol, and should be above 60 mg/dL. As apolipoprotein A1 is the principal protein component of the HDL cholesterol, it is also present in lower values on type 2 diabetics, although no statistical significant differences were found for apolipoprotein B100, present in LDL cholesterol. Total cholesterol is a measure of HDL, LDL and other lipid components. Type 2 diabetics show lower levels of mean and median cholesterol (respectively 175.24 and 161.00 mg/dL) than controls (200.78 and 197.00 mg/dL) but the range of values in type 2 diabetics (86.00 to 398.00 mg/dL) is much higher than in controls (117.00 to 292.00 mg/dL). Overall, desirable values should be under 200 mg/dL, but not too low, since cholesterol is necessary to build and maintain membranes, as it modulates membrane fluidity over the range of physiological temperatures. On the other hand, type 2 diabetics have higher risk of coronary or other cardiovascular disease, as they have higher values of triglycerides (mean 138.45; median 146.00) than controls (mean 117.72; median 94.00). Abnormal values (above 150 mg/dL), are present in both groups, as well for the atherogenic index, a parameter that reflects the ratio between triglycerides and HDL cholesterol as it is computed as $\log_{10}(triglycerides/HDL)$. In fact, type 2 diabetics present higher levels of triglycerides

and lower levels of HDL cholesterol, which is reflected on this ratio as that group has a higher risk of cardiovascular diseases.

Type 2 diabetics also presented higher values of leucocytes ($p = 0.024$) and lower values of haematocrit (the percentage of haemoglobin in total blood volume composed by red cells) and, consequently, lower values of haemoglobin which are related to a good oxidation as it is a transporter of oxygen in the organism.

As the extra-cellular and citoplasmatic life of insulin is very short, C-peptide is a marker for insulin values since it is connected to insulin forming pro-insulin and is released in blood in the same proportion as insulin. Moreover, C-peptide has a higher life-time than insulin thus, it was already expected that type 2 diabetics had lower values of C-peptide (Table 18), although this had to be confirmed.

Concerning ophthalmological tests, a type 2 diabetic is expected to present lower best corrected visual acuity ($p = 0.001$) than controls, similar volume scan density (Table 19), and differences on the temporal ($p = 0.041$) and temporal-inferior ($p = 0.047$) quadrants of the retinal nerve fibre layer thickness (Table 20), with higher thickness of the temporal quadrant but lower thickness at the inferior region of the temporal quadrant. This group of subjects has worst performance of the speed test (Table 21), whichever the meridian used, performed worse in the achromatic vision discrimination test (Table 22) along meridian 90º ($p = 0.005$), and also worst on chromatic contrast sensitivity test on the Protan axis, along the meridian 0º ($p = 0.043$), on the Deutan axis, either globally ($p = 0.013$) or along meridians 0º and 45º (respectively $p < 0.001$ and $p = 0.042$), as well as on the Tritan axis, whichever the meridian used or even the global measure, as observed in Tables 23 to 25.

When we performed a univariate classification on the above identified variables which had differences between the two subgroups, by a ROC analysis, we confirmed that, in fact, most of those parameters could be used as classifiers of type 2 diabetes since they presented good accuracy on prediction of type 2 diabetes, measured by the area under the ROC curve, as well as the variable positive predictive value (Tables 26 to 35). However, we did not posteriorly consider some of the variables (as weight and height, as they are included in the body mass index), and abdominal perimeter, pulse and systolic or diastolic blood pressure due to the lack of information in several cases.

Then, a more refined profile of a type 2 diabetic older than 40 may be traced as a subject whom will probably show one or more of the following conditions:

- Body Mass Index $\geq$ 26.95 kg/m^2;
- ALT $\geq$ 26.50 units per litre of serum;
- Alkaline Phosphatase $\geq$ 64.50 units per litre of serum;
- Gamma GT $\geq$ 24.50 units per litre of serum;
- Atherogenic Index $\geq$ 3.35;
- Total Cholesterol $\leq$ 182.50 mg/dL;
- Cholesterol LDL $\geq$ 114.50 mg/dL;
- Cholesterol HDL $\leq$ 45.50 mg/dL;
- Triglycerides $\geq$ 119.00 mg/dL;
- Apolipoprotein A1 $\leq$ 132.50 mg/dL;
- Leucocytes $\geq$ 6.55 ml/mm^3;
- Haemoglobin $\leq$ 12.65 g/100mL;
- Haematocrit $\leq$ 38.25%;
- C-Peptide $\leq$ 1.35 ng/mL;
- BCVA $\leq$ 0.90;
- RNFL on Temporal quadrant $\geq$ 67.50 μm;
- Speed test
  - Meridian 0º $\geq$ 1.59;
  - Meridian 45º $\geq$ 2.17;
  - Meridian 90º $\geq$ 0.86;
  - Meridian 135º $\geq$ 1.62;
  - Global area $\geq$ 2.86;
- Achromatic test
  - Meridian 90º $\geq$ 2.32;
- Chromatic test
  - Protan
    - Meridian 0º $\geq$ 2.47 x 10^{-3} (º/s);
  - Deutan
    - Meridian 0º $\geq$ 3.40 x 10^{-3} (candelas/m^2);
    - Meridian 0º $\geq$ 8.34 x 10^{-3} (candelas/m^2);
  - Tritan test
    - Meridian 0º $\geq$ 67.20 x 10^{-3} (ratio to maximum);
    - Meridian 45º $\geq$ 59.99 x 10^{-3} (ratio to maximum);

- Meridian 90º $\geq$ 77.13 x 10^{-3} (ratio to maximum);
- Meridian 135º $\geq$ 109.11 x 10^{-3} (ratio to maximum);
- Global area $\geq$ 6.16 x 10^{-3} (ratio to maximum).

These variables identified as univariate classifiers of diabetes were tested under multivariate techniques in order to evaluate their independent prediction of type 2 diabetes using discriminant analysis, logistic regression and decision trees algorithms.

Although classical assumptions of discriminant analysis were violated, the model could be developed and identified hypertension measured by blood pressure controlled by medication, body mass index, cholesterol HDL, triglycerides, C-peptide, retinal nerve fibre layer thickness in the temporal quadrant and chromatic contrast sensitivity for the Tritan axis along meridian 135º as classifiers of diabetes (Table 36), being the accuracy of predictions given by the area under the ROC curve for posterior probabilities of 0.985 ($p < 0.001$), at the training sample (Table 37). This model has high sensitivity (92.31%) and specificity (92.31%) if we consider the cut-off of 50% regarding posterior probability, and 97.44% if we consider the 61.04% cut-off, obtained by ROC analysis, for that probability), and its positive likelihood ratio is, respectively, 12.00 and 36.00. This model is very robust, but has few applications for screening of diabetes, since it needs parameters of blood tests, OCT and psychophysical tests.

A subject will be classified, according to this model, as a type 2 diabetic if $P(G_2|d_D^2) \geq 50\%$, given by

$$P(G_2|d_D^2) = \frac{0.5xP(\chi_1^2>d_D^2)}{0.5xP(\chi_1^2>d_C^2)+0.5xP(\chi_1^2>d_D^2)}, \text{ where} \begin{cases} d_C^2 = \frac{f(x)+1.462}{0.892} \\ d_D^2 = \frac{f(x)-1.706}{1.099} \end{cases} \text{ and}$$

$$F(x) = -3.233 + 1.548 \times HTA + 3.274 \times Tritan(135º) - 0.039 \times CholHDL -$$
$$-0.998 \times CPeptide + 0.034 \times RNFL(temp) + 0.008 \times Triglycerids +$$
$$+0.090 \times BMI$$

The variable HTA is set to 1 if the subject is medicated for hypertension, and 0 if not. All the other variables are numerical.

A simpler way to classify a subject as type 2 diabetic is if $F(x) > 0.122$, and determinate posterior probability later.

The model obtained by regression analysis (Tables 38, 39 and 41), using binary variables according to the cut-offs obtained by ROC analysis and previously presented, also identify hypertension, body mass index (≥ 26.95 kg/m^2) and thickness of the RNFL on the temporal quadrant (≥ 67.50 μm), but uses values of apolipoprotein A1 (≤ 132.50 mg/dL) and achromatic contrast sensitivity along the 90º meridian (≥ 2.32) as predictors of diabetes, with an accuracy of predictions based on the posterior probability of 0.952 ($p < 0.001$), which presents a sensibility of 86.05% and a specificity of 93.18%, with a positive likelihood ratio of 12.62. The gain in PLR, compared to the PLR obtained by discriminant analysis do not justifies the use of this model, as it loses sensitivity.
The logistic regression classifier is defined as $P(D) \geq 50\%$, where

$$P(D) = \frac{e^{-11.045+4.393\times HTA+3.959\times Achrom(90^{o})+3.851\times BMI+4.002\times AlipoA1+4.113\times RNFL(temp)}}{1+e^{-11.045+4.393\times HTA+3.959\times Achrom(90^{o})+3.851\times BMI+4.002\times AlipoA1+4.113\times RNFL(temp)}}$$

Each one of the variables assumes the value of 1 or 0 according to the dichotomization obtained by ROC analysis and presented before.

Applying decision tree algorithms, we obtained four models that do not differ in accuracy (Table 42; DeLong test minimum p-value, not adjusted for multiple comparisons, is 0.052). Models were evaluated by the application of ROC analysis to the probability of being diabetic. Although models obtained by the application of CHAID or Exhaustive CHAID algorithms are the ones with higher sensitivity (respectively 94.29% and 91.43%), the models obtained with CART and QUEST algorithms are the ones with higher positive likelihood ratio (respectively 4.59 and 3.86).
Decision tree models are simpler to apply in a routine ophthalmological exam, since they only use sociodemographic variables such as age (CART algorithm) or age and BMI (QUEST algorithm) or blood pressure controlled by medication (CHAID and Exhaustive CHAID algorithms), and values obtained in the speed test and Chromatic contrast over the Tritan axis or in meridian 45º (CART and QUEST algorithms) or in meridian 135º (CHAID and Exhaustive CHAID algorithms). CART algorithm considers haemoglobin but in the last node, so tree may be pruned in order to consider only non-invasive exams.

Comparing the performance of these models on the training sample, the one with highest positive likelihood ratio, as well as area under the ROC curve, is the one obtained with discriminant analysis, and the one with highest sensitivity is the one obtained with the CHAID algorithm on decision tree analysis, as summarized below (Table 87):

Table 87 - Accuracy of developed models measured in the training sample.

Model	AUC (p)	Cut-off Function (Prob)	% Correct	k	p	McNemar (p)	Sens.	Spec.	+LR
D_F	0.985 (< 0.001)	0.122 (50.00%)	92.31%	0.846	< 0.001	1.000	92.31%	92.31%	12.00
D_{ROC}		0.264 (61.04%)	94.87%	0.897	< 0.001	0.625	92.31%	97.44%	36.00
L	0.942 (< 0.001)	50.00%	89.66%	0.793	< 0.001	0.508	83.87%	90.48%	8.81
T1	0.882 (< 0.001)	50.00%	82.29%	0.629	< 0.001	0.332	82.86%	81.97%	4.59
T2	0.800 (0.048)	50.00%	80.43%	0.615	< 0.001	< 0.001	94.29%	71.93%	3.36
T3	0.789 (< 0.001)	50.00%	79.35%	0.591	< 0.001	0.004	91.43%	72.93%	3.26
T4	0.860 (< 0.001)	50.00%	81.25%	0.619	< 0.001	0.031	88.57%	77.05%	3.86

In fact, we find statistical significant difference between the area under the ROC curve obtained by discriminant model and decision trees based on CHAID, Exhaustive CHAID and QUEST algorithms (respectively p = 0.002, p = 0.001 and p = 0.012 by the DeLong test). The logistic regression model presents also better accuracy than trees obtained with CHAID and Exhaustive CHAID algorithms (respectively p = 0.027, p = 0.018). These p-values were not adjusted for multiple comparisons, thus, a simple but rather conservative procedure would be considering the Bonferroni correction, that is, those p-values should be multiplied by 15 (C_2^6) and the resulting p-values adjusted ($\hat{p}$) would be given by $\hat{p} = min\{1; p - value\}$. In this case, we can assume that the discriminant model accuracy is higher and statistical significant different from decision tree models obtained by CHAID and Exhaustive CHAID algorithms (respectively p = 0.030 and p = 0.015), and no other statistical significant differences are found between the other models.

Notice that all models, with the exception of T2, T3 and T4 present similar false positive and false negative rates, according to the McNemar test.

When we apply these models to new cases, that is, to the test sample, results are surprising since the worse models in the training sample became the best models on the test sample.

Note that (Table 87) the discriminant and logistic classifiers are the ones with highest area under the ROC curve (respectively 0.937 and 0.935, both with p values under 0.001), but both present lower limit of the 95% confidence interval for specificity below 50%, and the first one has also a lower limit on the 95% confidence interval for positive likelihood ratio below 1, meaning that in the extreme ranges it may be as probable to have a true positive classification as a false positive one. Moreover, decision tree T1 and T4 classifiers also present lower limits on 95% confidence interval above 50%, and are the ones with lower accuracy. Then, the chosen classifier for diabetes is the one obtained with the CHAID or with the exhaustive CHAID decision tree algorithms, which performance is equal when applied on the test sample, since they present good accuracy (AUC = 0.894, p < 0.001), both with expected sensitivity and specificity on population above 50% (respectively $87.88\% \in (71.80\%; 96.60\%)$ and $83.33\% \in (51.60\%; 97.90\%)$ with 95% confidence), and an expected positive likelihood ratio between 1.50 and 18.80, which means that is 1.5 times to about 19 times more probable to have a true positive classification than a false positive classification.

In clinical practice, and for screening purposes, we are interested in evaluating positive (PPV) and negative predictive values (NPV) in order to determine the posterior probability for a given subject to be a type 2 diabetic, when it is classified as that (PPV) or to be healthy in respect to type 2 diabetes when is classified as normal (NPV). As this values depend on disease prevalence, and diabetes prevalence is becoming higher every year, due to several factors such as the aging of populations, pair wise with the reduction of physical activity and the increase of obesity, among others, the values presented in Figure 55 are based on the estimates of diabetes prevalence from the year of 2012, and predictive values need to be updated as new data on prevalence becomes available.

The model presented may be written as the following, classifying a subject as diabetic whenever$P(D) \geq 50\%$, for any subject without diagnosed hypertension, since all subjects undergoing treatment for hypertension have higher risk of also have type 2 diabetes, with a probability of 82.61%:

$$Value\ measured\ on\ meridian\ 135º\ of\ the\ Chromatic\ Tritan\ axis$$

$$\Rightarrow \begin{cases} \geq 134.33 \times 10^{-3} \Rightarrow P(D) = 87.50\% \\ < 134.33 \times 10^{-3} \Rightarrow P(D) = 7.69\% \\ Unknown \Rightarrow P(D) = 22.45\% \end{cases}$$

This is a simple and non-invasive method to detect type 2 diabetes, that may be used in ophthalmological visits. If there is a suspicion of the presence of type 2 diabetes, then it should be confirmed with standard diagnostic tests, which are the gold standard.

However, if prevalence continues to grow, in Portugal, at the same rate as it grew between the years of 2009 and 2012 (1.2%), in a few years (Figure 56) the positive predictive value of this classifier will be above 50%, with similar negative predictive values. If, as expected, type 2 diabetes prevalence grows faster than 1.2% every three years, then rapidly we may get higher positive predictive values for this classifier, without losing its negative predictive value.

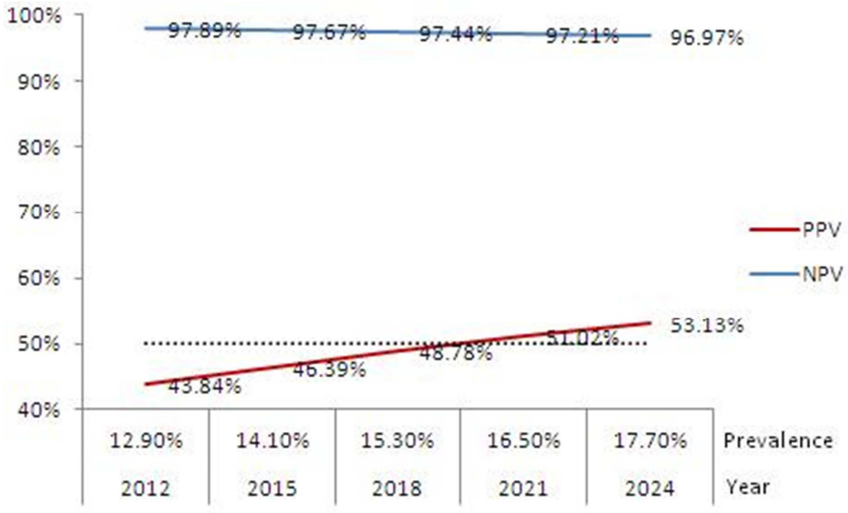

Figure 56 – 12 years prevision for predictive values of T2 classifier.

Concerning diabetic retinopathy, we were not able to apply exactly the same methodology used for the development of type 2 diabetes classifier, since we were not able to have, at this moment, a test sample with a reasonable number of cases to evaluate developed models based on the training sample. This will be done afterwards, and probably we will have enough data to test developed models within a few months.

The development of diabetic retinopathy classifiers using the training sample was obtained using the same methodology as before. Hence, we begun to compare the same variables as before, as well as family history of diabetes, duration of the disease, and compared groups according to ETDRS grading which was performed on 40 type 2 diabetics (20 with and 20 without diabetic retinopathy), and dichotomized as DR present or absent. Groups were matched for age and gender, as well as daily habits (tobacco and alcohol use, and regular exercise practice) and family history of diabetes. Likewise, no statistical significant differences were found in the percentage of cases with need for control of blood pressure with medication, height, weight, body mass index and bio impedance, pulse, systolic or diastolic blood pressure. However, subjects with diabetic retinopathy had type 2 diabetes for a longer duration than subjects without diabetic retinopathy (from 2 to 12 years longer), as presented in Tables 48 and 49.

Concerning blood tests (Tables 50 to 57), we did not find statistical significant differences for glucose levels ($p = 0.622$) or HbA1C levels ($p = 0.967$), but we were able to differentiate groups based upon creatinine ($p < 0.001$), erythrocytes ($p = 0.023$), haemoglobin and haematocrit (respectively 0.010 and 0.014). We may say that a type 2 diabetic with non-proliferative diabetic retinopathy has usually previous history of diabetes on family, higher levels of creatinine and less oxidation as he present lower values of haemoglobin and haematocrit.

Measures obtained for the dominant eye (Tables 58 to 64) are statistically different for the inner-nasal quadrant of volume scan density from OCT ($p = 0.026$), where subjects with diabetic retinopathy present higher volume scan density, without statistical significant differences regarding the retinal nerve fibre layer thickness. Speed, achromatic vision and chromatic vision for the Protan axis were also similar between groups. Yet, for the Deutan axis, over the 0º meridian ($p = 0.041$), and for the Tritan axis, over meridians 0º ($P = 0.003$) and 135º ($p = 0.002$), as well on the global measure for Tritan ($p = 0.017$) which was obtained by the 5-sided polygon generated by the median values of meridians and the origin, where type 2 diabetics without diabetic retinopathy showed to have changed chromatic discrimination.

Using Receiver Operating Characteristic curve analysis (Tables 65 to 75), we were able to reduce the set of variables, since we found that the only variables with significant area under

the ROC curve were the duration of the disease (AUC = 0.748; p = 0.007), erythrocytes (AUC = 0.710; p = 0.023), haemoglobin (AUC = 0.746; p = 0.008) and haematocrit (AUC = 0.728; p = 0.014), volume scan density on inner-nasal quadrant (AUC = 0.705; p = 0.027), discrimination for the Deutan axis over the 0º meridian (AUC = 0.701; p = 0.042), and for the Tritan axis over the 0º meridian (AUC = 0.776; p = 0.004), and the 135º meridian (AUC = 0.792; p = 0.003), as well as for the global area (AUC = 0.735; p = 0.017), which allowed us to identify the following cut-offs:

- duration of the disease $\geq$ 18.50 years;
- erythrocytes $\leq$ 4.23x10^6/mm^3;
- haemoglobin $\leq$ 13.20 g/100mL;
- haematocrit $\leq$ 36.25%;
- volume scan (OCT) on Inner-Nasal quadrant $\geq$ 355.50 μm;
- chromatic test
  - Deutan
    - meridian 0º $\geq$ 4.46 x 10^{-3} (candelas/m^2);
  - Tritan test
    - meridian 0º $\geq$ 63.25 x 10^{-3} (ratio to maximum);
    - meridian 135º $\geq$ 174.81 x 10^{-3} (ratio to maximum);
    - global area $\geq$ 5.94 x 10^{-3} (ratio to maximum);

As on the development of type 2 diabetes classification models, the variables identified as univariate discriminators of the presence of diabetic retinopathy were tested using discriminant analysis, logistic regression and decision trees algorithms.

Normal distribution for variables on each group failed mainly on eye measurements, and especially in the group without diabetic retinopathy but, as in the case of the development of a diabetes classifier, we proceeded with the application of discriminant analysis since groups had exactly the same dimension and the F statistic is very robust to normality deviations in these situations. On the other hand, we confirmed the homogeneity of covariance matrices since we obtained a p-value of 0.942 at the Box M test.

We obtained a classifier for the presence of diabetic retinopathy defined as:

$P\left(DR_{Present}\middle|d^2_{DR\ present}\right) \geq 50\%$, where

$$P(DR_{Present}|d^2_{DR\ present}) = \frac{0.5xP(\chi^2_1>d^2_{DR\ present})}{0.5xP(\chi^2_1>d^2_{DR\ absent})+0.5xP(\chi^2_1>d^2_{DR\ pesent})}$$

Where $\begin{cases} d^2_C = \frac{f(x)+1.462}{0.892} \\ d^2_D = \frac{f(x)-1.706}{1.099} \end{cases}$ and

$$F(x) = -1.672 + 5.215 \times Tritan_{M135º} + 0.109 \times Duration -$$

$$-0.590 \times Haemoglobin + 0.021 \times VS_{InnerNasal}$$

Or, simply ascertain whether $F(x) > 0.002$ or not.

The model obtained by regression analysis, using binary variables according to the cut-offs obtained by ROC analysis and presented above and in Tables 78 to 80, identifies only the global area of chromatic discrimination on the Tritan axis as separating subjects with diabetic retinopathy present from those where it is absent, and the classifier is defined as $P(DR_{Present}) \geq 50\%$, where

$$P(DR_{Present}) = \frac{e^{-2.303+2.996\times ChromT}}{1+e^{-2.303+2.996\times ChromT}}$$

$(ChromT = 1\ if ChromT \geq 5.94 \times 10^{-3}; otherwise, ChromT = 0\ )$

Considering decision tree algorithms, CHAID and Exhaustive CHAID were not able to grow the tree, and CART and QUEST algorithms lead to the same solution, which is

$Value\ measured\ on\ meridian\ 135º\ of\ the\ Chromatic\ Tritan\ axis$

$$\begin{cases} \geq 174.81 \times 10^{-3} \Rightarrow P(D) = 85.71\% \\ < 174.81 \times 10^{-3} \Rightarrow P(D) = 23.81\% \end{cases}$$

Comparing the obtained models at the training sample, we may observe that the all the three classifiers (discriminant, logistic and decision tree classifiers) are centered on chromatic discrimination over the Tritan axis, either on the global area or along meridian 135º. However, the discriminant model is more complex and also identifies volume scan

measured on the inner nasal quadrant, haemoglobin and duration of the disease as markers of diabetic retinopathy.

In fact, as observed on Table 83, the discriminant model has significantly higher accuracy measured by the area under the ROC curve determined for posterior probabilities than the logistic model (non-adjusted DeLong test p = 0.002) and also than the decision tree model (non-adjusted DeLong test p = 0.044). However, we should consider that difference only exists between the logistic regression classifier and the discriminant classifier due to multiple comparisons performed. Moreover, positive likelihood ratios of the discriminant model (14.82) and of the decision tree model (19.41) are much higher than the one obtained for logistic regression model (2.88), in spite of this model is the one with higher sensitivity ($87.88\% \in (71.80\%; 96.60\%)$ and $83.33\% \in (51.60\%; 97.90\%)$ with 95% confidence.

All these three models should be evaluated on a set of new cases, and will be, in a near future, but at the moment we can only compare their performance in the training sample.

With this data, and considering a prevalence of 34.6% for diabetic retinopathy within diabetics, the decision tree and discriminant models present good positive predictive values, as observed on Table 84 and Figure 49, with a lower bound on the 95% confidence interval above 50%, which does not happen with the logistic regression model. Note that the disease prevalence used was determined among all diabetics, and not only for type 2 diabetics. However, the estimates refer that diabetic retinopathy will affect 50% of diabetics and, if that really happens on type 2 diabetics, then the decision tree model is expected to have a positive predictive value of 95.16%, which means that in every 100 subjects with values for chromatic vision over the Tritan axis along meridian 135º of 0.17481 or higher, 95 will be, in fact, type 2 diabetics. With the actual values of prevalence, this is real for 91 in every 100 subjects.

The logistic model as high negative predictive value $95.6\% \in (82.1\%; 99.70\%)$ with 95% confidence, thus, as it is a simple model based also on chromatic discrimination over the Tritan axis, it may be applied to all cases that turned out to be classified as negative on the decision tree model, improving classification.

This may be a simple and non-invasive test to perform and that enables the standardization of different criteria for diabetic retinopathy classification. However, as it was previously referred, this model must be evaluated first on a test sample of adequate dimension.

CHAPTER 8

CONCLUSIONS

Correlation between eye measurements obtained by Optical Coherence Tomography is moderate to strong. In fact, concerning the volume scan density, the minimal correlation coefficient found was of 0.777 for the Inner Inferior quadrant, and concerning the retinal nerve fibre layer thickness, the minimal correlation coefficient obtained was of 0.674 in the nasal quadrant.

Whenever we had data not adjusted to normal distribution, we may use a pseudo-concordance correlation coefficient to evaluate concordance between eyes, as it uses non-parametric assumptions, and it is a closer measure to the intra-class correlation coefficient than the classical concordance correlation coefficient which is computed using Pearson's correlation though, based on parametric assumptions.

Besides being correlated, the two eyes also show great concordance. Hence, only one eye is sufficient for analysis.

In spite of hypertension is evaluated through the register of medication taken for blood pressure control, we found that type 2 diabetics have higher values of systolic blood pressure than controls, but no difference was found in the diastolic blood pressure.

Subjects with type 2 diabetics have higher risk concerning liver and biliary system damage, evaluated by levels of ALT, alkaline phosphatase and gamma GT in the blood, higher risk of

cardiovascular disease reflected by higher levels for triglycerides and atherogenic index. Furthermore, they have less oxygenation due to lower levels of haemoglobin, haematocrit and higher coefficient of variation on erythrocytes.
Oddly, type 2 diabetics present lower levels of total cholesterol, cholesterol LDL and apolipoprotein A1, but also lower levels of cholesterol HDL. However, when we analyse these variables in a multivariate context, interacting with other parameters, they became risk factors for type 2 diabetes.

Concerning vision, type 2 diabetics are expected to have a lower best corrected visual acuity, higher thickness of the retinal nerve fibre layer in the temporal quadrant, less perception of motion, less perception of colour (achromatic vision) along meridian 90º, and monochromatic vision along meridian 0º, with higher probability of having damages in all the photo pigment cones (Protan, Deutan and Tritan). Concerning the 135º meridian, vision is compromised since the vision is mostly dichromatic due to Tritan axis, and for the 45º meridian vision is usually atypical due to Deutan and Tritan axes.

The development of a simple global measure for speed, achromatic and chromatic vision tests, dependant of the four measures obtained in each one of the meridians, for each test, revealed to be useful, allowing discrimination of type 2 diabetes with the speed test where type 2 diabetics shown to be slower at movement detection, and on the Tritan axis of the chromatic test, where type 2 diabetics presented more difficulties. The algorithm for computing the area of the 5-side polygon is simple, and may be easily implemented.

Although discriminant analysis assumptions were violated, the model for type 2 diabetes classification is very robust either in the training sample or the test sample.
It identifies Hypertension, Body Mass Index, Cholesterol HDL, Triglycerides, C-Peptide, thickness in the temporal quadrant of the retinal nerve fibre layer of the dominant eye and chromatic contrast sensitivity at meridian 135º as markers of type 2 diabetes, which allows to define a profile for this subjects as individuals who need medication for controlling blood pressure, higher Body Mass Index, lower values of Cholesterol HDL and higher values of Triglycerides (indicating lower metabolic control on lipids and thus higher risk of

cardiovascular diseases), lower values on C-peptide (indicating lower values of insulin present on blood and of insulin production by the pancreas), and with a higher thickness of the retinal nerve fibre layer which may induce a lack of perception of light and thus may be related with the tritanope defect on meridian 135º.

The three logistic regression models lead to the same solution, whichever the method used. It identifies three of the previous variables as type 2 diabetic markers: Hypertension, Body Mass Index and thickness of the retinal nerve fibre layer of the dominant eye, in the temporal quadrant. Cholesterol HDL and Triglycerides, identified with the discriminant analysis model, were replaced by Apolipoprotein A1 (which has a specific role on the lipid metabolism) at the logistic regression model, having the role of increasing the risk of the presence of type 2 diabetes. Tritanope defect is replaced by the total loss on chromatic vision along meridian 90º Nevertheless, the profile of a type II diabetic given by this model is similar to the previous one, where diagnosed hypertension under medication and body mass index are considered as risk factors for type 2 diabetes.

Decision tree models are similar two by two; CART and QUEST algorithms base their primary decision on age, perhaps due to the fact of patients in the sample are older. On the other hand, models based on the CHAID algorithm have as first decision criteria the fact that subjects have their blood pressure controlled by medication, as discriminant and logistic regression models. All the models present good accuracy on the training sample, and share common criteria for splitting nodes based upon chromatic vision at Tritan axis.

The CART algorithm identifies the global measure for speed discrimination, chromatic vision on meridian 45º and values of haemoglobin as markers of type 2 diabetes. The profile for these subjects will be defined as a subject that is younger than 51.5 years and has a global value on speed of, at least, 3.75, or for older subjects with at least 51.5 years, presenting minimum values of 59.99×10^{-3} for the 45º meridian of the Tritan axis on chromatic vision. The accuracy of the classification for older subjects may be improved whenever subjects present haemoglobin values below 14.95 g/100mL, whatever the gender.
The QUEST algorithm also bases its classification of older subjects, with at least 51.5 years, on values measured for the 45º meridian of the Tritan axis, with the same cut-off, which was the one identified with univariate ROC analysis, but for younger subjects bases its

classification is based up on the Body Mass Index. Considering that a subject with less than 51.5 years and a Body Mass Index of, at least, 32.55 kg/m^2, there is 80% probability of being a type 2 diabetic.

For CHAID and Exhaustive CHAID algorithms, chromatic sensitivity over the meridian 135º of the Tritan axis is the criteria to classify subjects without diagnosed hypertension, when its value is, at least 134.33×10^{-3}, defining the probability for the presence of type 2 diabetes as 87.50%. If the subject has diagnosed hypertension and is being treated for this medical condition, then the probability that he or she also is, as well, type 2 diabetic, is 82.61%.

In spite of the violation of the assumptions for discriminant analysis, the models behaves quite well when applied to the test sample, showing higher concordance with the true result than the logistic regression model, which does not have so many requirements, but uses less information since it is based on dichotomized variables and not on quantitative ones. Notwithstanding, the decision tree model, namely the CHAID algorithm, has the better performance when applied to the test sample.

The posterior probability for the presence of type 2 diabetes on non-hypertensive subjects, or undiagnosed hypertensive subjects is 22.45%, similar to the type 2 diabetes prevalence for the Portuguese population for the age group over 60 years old.

The loss of the chromatic vision for the Tritan axis is a crucial marker for type 2 diabetes.

The classifier for type 2 diabetes based upon decision tree algorithm has a high positive predictive value, adjusted for actual prevalence values of this disease for the Portuguese population, especially concerning the age group from 60 to 75 years of age, particularly for males but also for females, or for obese subjects, with at least 30 kg/m^2 of the Body Mass Index.

The duration of diabetes is a known factor contributing for the development of diabetic retinopathy that was identified, once more, as a marker for the progression of this disease and, though, visual impairment.

Values of erythrocytes, haemoglobin and haematocrit may also be considered as possible markers of the presence of diabetic retinopathy.
In what concerns the eye, we found that volume scan density at the inner-nasal quadrant and chromatic vision on the Deutan (meridian 0º) and Tritan (meridian 0º, meridian 135º and global value of Tritan) axes may, negatively, discriminate the presence of diabetic retinopathy.

The model obtained with discriminant analysis, although failing the multivariate normality assumption, gives an accurate profile for patients with diabetic retinopathy based on higher duration of the disease, lower values on haemoglobin, and worse values of volume scan density at the inner-nasal quadrant and Tritanope presence at least over the meridian 135º.

The model obtained with logistic regression analysis only considers the overall measure for the Tritan axis as a classifier for the presence of diabetic retinopathy, although the adequate accuracy demonstrated.

Considering the decision tree model based on CHAID or Exhaustive CHAID algorithms, we are able to classify the presence of diabetic retinopathy based upon values of chromatic colour discrimination over the Tritan axis, on meridian 0º.

The model presenting higher positive predictive value, at least when applied to the training sample, was developed with decision tree algorithms.

The loss of the chromatic vision for the Tritan axis is a crucial marker for non-proliferative diabetic retinopathy in type 2 diabetes.

For the developed models, the ones that always performed better were the models based upon decision tree algorithms, without assumptions on data distribution. The logistic regression models do not have, also, distribution assumptions, but were the ones with worse performance, although its performance is adequate. The use of binary variables brings loss of information with impact on the accuracy of these models. Discriminant analysis models

are robust to the violation of assumptions on data distribution, and return classifiers with good accuracy on previsions, but also more complex than decision tree models.

We may present a classifier for type 2 diabetes subjects already tested on new cases, and based upon measures obtained for the dominant eye of subjects aged between 40 and 75 years old (to use in subjects not undergoing hypertension therapy):

$$Value\ measured\ on\ meridian\ 45º\ of\ the\ Chromatic\ Tritan\ axis$$

$$\begin{cases} \geq 134.33 \times 10^{-3} \Rightarrow P(D) = 87.50\% \\ < 134.33 \times 10^{-3} \Rightarrow P(D) = 7.69\% \\ \quad Unknown \Rightarrow P(D) = 22.45\% \end{cases}$$

The eye plays, though, an important role in the diagnostic of type 2 diabetes, giving important clues for diagnosing this systemic disease on subjects older than 40 years old.

A classifier for non-proliferative diabetic retinopathy which needs to be evaluated on new set of cases, also based on measures obtained for the dominant eye of subjects aged between 40 and 75 years old:

$$Value\ measured\ on\ meridian\ 135º\ of\ the\ Chromatic\ Tritan\ axis$$

$$\begin{cases} \geq 174.81 \times 10^{-3} \Rightarrow P(D) = 85.71\% \\ < 174.81 \times 10^{-3} \Rightarrow P(D) = 23.81\% \end{cases}$$

The Tritan axis is the most important marker identified. It enables classification of type 2 diabetes (meridian 45º) in subjects not undergoing treatment for hypertension, and also classification of non-proliferative diabetic retinopathy (meridian 135º) in type 2 diabetics. The identified marker is specific for subjects aged between 40 and 75 years, without neuropsychiatric, renal, heart, ocular or any other severe disease unrelated to the aging process.

A final model for screening each of the referred conditions may be proposed (Figure 57), in spite of the need for future confirmation (on an independent test sample) concerning the diabetic retinopathy classifier. Hence, Tritanope vision represents an augmented risk for both clinical classification frameworks.

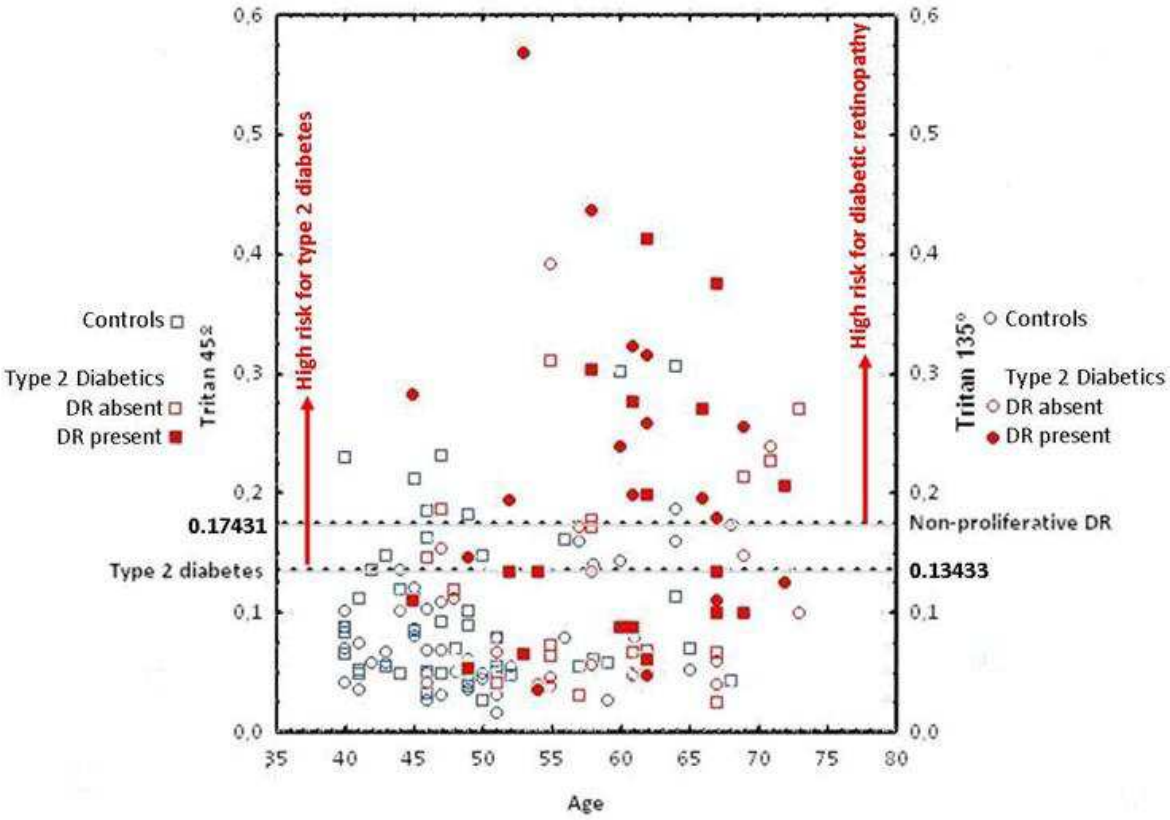

Figure 57 – Final classification model for type 2 diabetes and non-proliferative diabetic retinopathy in subjects aged between 40 and 75 years old

CHAPTER 9

FINAL CONSIDERATIONS

1. Study Limitations

Although they have already been mentioned in the discussion chapter, we leave here a summary of some limitations found during the analysis, as we intend to pursue the study and they will be the object of further work.

At the moment, not all of the cases have been included in the study, and type 2 diabetics and control groups were not matched for age. However, age was considered in all the multiple variable analyses for classification, thus so models were adjusted for age differences.

Furthermore, there were still few data available for analysis regarding multimodal imaging related to heart, liver and brain, hence these variables were not considered in analysis, as well as medical procedures such as blood pressure or abdominal perimeter measurements, which may be related to type II diabetes.

We should be able to evaluate the impact of gestational diabetes on type II diabetic females, which prevalence is known to be increasing from 3.4% in 2005 to 4.8% in 2012, representing an enhanced risk of the developing type 2 diabetes in the future years. However, we were not able to develop a classifier for women using this variable, since parity was not evaluated, nor the number of pregnancies and number of pregnancies with gestational diabetes.

Random error between left and right eye was significantly different for the inner superior and inner inferior quadrants. We decided to use only one eye in the study, since

concordance between eyes was high, in spite of some differences found in the Temporal, Nasal-Inferior and Nasal-Superior quadrants of the RNFL, as suggested by Armstrong[32]. We hope that, until the end of the study inclusion process, the random error found between groups in measurements performed on volume scan density in both eyes becomes identical, since the only way to improve random error is to increase the sample size.

When we presented classifiers for non-proliferative diabetic retinopathy, we were not able to evaluate the developed models into an independent test sample, since few data are still available. We consider these classifiers as preliminary classifiers, which will need to be tested in a sample of new cases. On the other hand, there were no cases in the sample with proliferative diabetic retinopathy thus we were not able to study this condition.

Overall, there are still many parameters with missing data in the database, not because they have not been measured, but because we are still waiting to be recorded in the database. This fact reduced the sample size of the data for training and for test samples in about 20% and we have collected only about less than 50% of the final sample size.

2. Further work

At the end of the study, we will have available 400 to 600 subjects, half on each group (controls and type 2 diabetics).

We intend to study with more detail the correlation between eyes, and especially the random error of measurements. With the increase of the sample size, it is expected that the errors in measurements between eyes may decrease due to the reduction of the random error, since systematic error should be controlled. However, it is of great interest, especially for methodology, that we compare random errors in measurements between groups, since it can be used as an assessment of the precision of the measurements. The random error comparison since it may be, by itself, a discrimination parameter between groups.

With the complete sample, we will be able to have all the data from all the tests performed and cross-correlation between organ dependent measures may be assessed. It is our

intention to use half of the final large sample as a training sample, and that developed models may be evaluated on new subjects. By then, we will be able to study the impact of diabetes not only on diabetic retinopathy, but also in liver and heart injury, as well as in brain.

Concerning statistical methods, all the classifiers presented may be used then, with few adaptations, for three group classification, that is, for discrimination between controls, type 2 diabetics without diabetic retinopathy and type 2 diabetics with diabetic retinopathy patients, or for discrimination between grading levels of diabetic retinopathy. However, ROC analysis is still used for discrimination between groups. Some work is already undergoing in order to obtain these curves for three group discrimination, using volume formulas instead or areas under the curve. In fact, their construction is somewhat similar to a three group discrimination function, where one function discriminates between one group and the other two, and the second function discriminates between the last two groups. This will always be an univariate procedure and it is being developed, but the intention is to integrate it with the multi-ROC procedures that are beginning to appear. In fact, multi-ROC procedures are in a development phase, although some have been recently published, using integration of linear combinations in a reduced space of the area under the ROC curve. It would be very interesting to develop ROC functions for three group discrimination, with cut-off definition for each group, and to integrate this in a multi-ROC procedure.

On the other hand, not related to statistical methods but with the clinical practice, if we become able to discriminate between this diabetic retinopathy grading, then we may propose a method for its classification based upon more objective measurements in order to standardize diabetic retinopathy grading and easily obtain an objective quantification of that grading.

REFERENCES

1 World Health Organization. The world health report 2003: shaping the future. Geneva: WHO, 2003.

2 Resnikoff S., Pascolini D., Mariottia S.P., Pokharela G.P.. Global magnitude of visual impairment caused by uncorrected refractive errors in 2004. Bulletin of the World Health Organization, 86 (1), January 2008.

3 World Health Organization. Prevention of blindness and deafness. Global initiative for the elimination of avoidable blindness. Geneva: WHO, 2000.

4 United Nations, Population Division. World population prospects — the 2002 revision. New York (NY), United Nations, 2003.

5 Murray C.J.L., Lopez A.D., editors. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020. Cambridge, MA: Harvard School of Public Health on behalf of the World Health Organization and the World Bank, (Global Burden of Disease and Injury Series, Vol. 1), 1996.

6 Murray C.J.L., Lopez A.D., Mathers C.D., Stein C.. The Global Burden of Disease 2000 Project: aims, methods and data sources. Geneva: World Health Organization, Global Programme on Evidence for Health Policy Discussion paper N. 36, 2001.

7 World Health Organization. Global data on Visual impairment 2010. Geneva: WHO, 2012.

8 Thylefors B., Negrel A.-D., Pararajasegaram R., Dadzie K.Y.. Global Data on Blindness. WHO Bulletin OMS. Vol 73, 1995.

9 Resnikoff S., Pascolini D., Etya'ale D., Kocur I., Pararajasegaram R., Pokharel G.P., Mariotti S.P.. Global data on visual impairment in the year 2002. Bulletin of the World Health Organization, 82 (11), November 2004.

10 Mayeaux E.J. Jr. Nail disorders. Prim Care;27: 333-51 2000.

11 Daniel C.R. 3rd, Sams W.M. Jr., Scher R.K.. Nails in systemic disease. Dermatol Clin, 3:465-83, 1985.

12 World Health Organization. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus. Geneva, 1999.

13 Mathers C.D., Loncar D.. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med*, 3(11):e442, 2006.

14 World Health Organization. Global health risks. Mortality and burden of disease attributable to selected major risks. Geneva, 2009.

15 Gardete-Correia L., Boavida J.M., Raposo J.F., Mesquita A.C., Fona C., Carvalho R., Massano-cardoso S.. First diabetes prevalence study in Portugal: PREVDIAB study. Diabetes Med. Aug; 27(8): 879-81, 2010.

16 Klein R. et al *in* The Wisconsin epidemiologic study of diabetic retinopathy. II. Prevalence and risk of diabetic retinopathy when age of diagnosis is less than 30 years. Archives of Ophthalmology, 102:520-526, 1984.

17 Figueira J., Nascimento J., Henriques J., Gonçalves L., Rosa P., Silva R., Henriques J.. RETINOPATIA DIABÉTICA Guidelines. Grupo Português de Retina-Vítreo, Grupo de Estudos em Retina, Sociedade Portuguesa de Oftalmologia, 2009.

18 World Health Organization. Prevention of Blindness from Diabetes Mellitus: report of a WHO consultation in Geneva, Switzerland, 9-11 November 2005.

19 DRS Study Group. Photocoagulation treatment of proliferative diabetic retinopathy. Clinical application of Diabetic Retinopathy Study (DRS) findings. DRS report number 8, Ophthalmology, 88:583-600, 1981.

20 Vine A.K.. The efficacy of additional argon laser photocoagulation for persistent, severe proliferative diabetic retinopathy. Ophthalmology, 932: 1532-1537, 1985.

21 ETDRS Study Research Group. Photocoagulation for diabetic macular edema. ETDRS report number 1. Archives of Ophthalmology, 103:1796-1806, 1985.

22 ETDRS Study Research Group. Early photocoagulation for diabetic retinopathy. ETDRS report number 9. Ophthalmology, 98 (Suppl 5):766-785, 1991.

23 Diabetic Retinopathy Vitrectomy Study. Early vitrectomy for severe vitrous hemorrage in diabetic retinopathy. Two-years results of a randomized trial. DRVS report 2. Archives of Opthalmology, 103: 1644-1652, 1985.

24 Diabetic Retinopathy Vitrectomy Study. Two-year course of visual acuity in severe proliferative diabetic retinopathy with conventional management. Diabetic Retinopathy Vitrectomy Study (DRVS) report no. 1. Ophthalmology, 92: 492-502, 1985.

25 Writing Team for the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group. The Journal of American Association, 287: 2563-2569, 2002.

26 Stratton I.M. et al. UKPDS 50: risk factors for incidence and progression of retinopathy in type II diabetes over 6 years from diagnosis. Diabetologia, 44: 156-163, 2001.

27 Mathews D.R. et al. Risk of progression of retinopathy and vision loss related to tight blood pressure control in type 2 diabetes mellitus. UKPDS 60. Archives of Opthalmology, 122: 1631-1640, 2004.

28 Royal College of Ophthalmologists. Guidelines for the Management of Diabetic Retinopathy. London, 1977.

29 Royal College of Ophthalmologists. Diabetic Retinopathy Guidelines. London, December 2012.

30 Panozzo G., Gusson E., Parolini B., Mercanti A.. Role of OCT in the diagnosis and follow up of diabetic macular edema. Semin.Ophthalmol, 18:74-81, 2003.

31 Murdoch I.. People and Eyes: Statistics in Ophthalmology. Community Eye Health, Vol 11, N 27: 43, 1998.

32 Armstrong R.A.. Statistical guidelines for the analysis of data obtained from one or both eyes. Ophthalmic and Physiological Optics, 33, 7–14, 2013.

33 Karakosta A., Vassilaki M., Plainis S., Elfaal N.H., Tsilimbaris M., Moschandreas J.. Choice of analytic approaches for eye specific outcomes: one eye or two. Am J Ophthalmol, 153: 571–579, 2012.

34 Glynn R.J., Rosner B.. Regression methods when the eye is the unit of analysis. Ophthalmic Epidemiol, 19: 159–165, 2012.

35 Rosner B.. Statistical methods in ophthalmology: an adjustment for the intraclass correlation between eyes. Biometrics, 38: 105–114, 1982.

36 Rosner B., Glynn R.J., Lee M.L.. Incorporation of clustering effects for the Wilcoxon rank sum test: a large-sample approach. Biometrics, 59: 1089–1098, 2003.

37 Armstrong R.A., Eperjesi F., Gilmartin B.. The application of analysis of variance (ANOVA) to different experimental designs in optometry. Ophthalmic Physiol Opt, 22: 1–9, 2002.

38 Rosner B., Glynn R.J., Lee M.L.. The Wilcoxon signed rank test for paired comparisons of clustered data. Biometrics, 62: 185–192, 2006.

39 Rosner B., Glynn R.J., Lee M.L.. A non-parametric test of observational non-normally distributed ophthalmic data with eye-specific exposures and outcomes. Ophthalmic Epidemiol, 14: 243–250, 2007.

40 Fleiss J., Levin B., Paik M.C.. Statistical Methods for Rates and Proportions, 3rd ed. Wiley and Sons: New York, pp 440 –461, 2003.

41 Bland J.M., Altman D.G.. Measurement error and correlation coefficients. BMJ, 313: 41–42, 1996.

42 Glynn R.J., Rosner B.. Accounting for the correlation between fellow eyes in regression analysis. Arch Ophthalmol, 110: 381–387, 1992.

43 Glynn R.J., Rosner B.. Comparison of alternative regression models for paired binary data. Stat Med, 13: 1023–1036, 1994.

44 Bland J.M., Altman D.G.. Statistical method for assessing agreement between two methods of clinical measurement. The Lancet, i: 307-310, 1986.

45 McAlinden C., Khadka J., Pseudovs K.. Statistical methods for conducting agreement (comparison of clinical tests) and precision (repeatability or reproducibility) studies in optometry and ophthalmology. Ophthalmic Physiol Opt, 31: 330–338, 2011.

46 Armstrong R.A., Davies L., Dunne M.C.M., Gilmartin B.. Statistical guidelines for clinical studies of human vision. Ophthalmic Physiol Opt, 31: 123–126, 2011.

47 Alpaydin E.. Introduction to Machine Learning, 2nd Edition, The MIT press, Cambridge, Massachusetts, London, England, 2010.

48 Fisher R.A.. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7: 179-188, 1936.

49 Fisher R.A.. The statistical utilization of multiple mesurements. Annals of Eugenics, 8: 376-386, 1938.

50 Maroco, J. Análise Estatística com utilização do SPSS, 2ª Ed. Edições Sílabo, Chapter 12, 2003.

51 Sharma, S.C.. Applied multivariate techniques. John Wiley & Sons, Inc., New York Chichester Brisbane Toronto Singapore, 1996.

52 Mathews D.E., Farewell VT. Using and understanding Medical Statistics, 3rd Ed. Karger, Chapter 11, 1996.

53 Greene W.H.. Econometric Analysis, fifth Edition, Prenctice Hall, 720-723, 1993.

54 Anderson T.W.. An introduction to Multivariate Statistical Analysis. Wiley, 1958.

55 Breiman L., Friedman J.H., Olshen R. A., Stone C.J.. *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1984.

56 Kass G.V.. An Exploratory Technique for Investigating Large Quantities of Categorical Data. Applied Statistics, Vol 29 No. 2: 119-127, 1980.

57 Loh W.Y., Shih X.. Split selection methods for classification trees. Statistica Sinica, 7: 815-840, 1997.

58 Belson W.A.. Matching and prediction on the principle of biological classification. Applied Statistics, 8(2): 65-75, 1959.

59 Morgan J.N. and Sonquist J.A.. Problems in the analysis of survey data, and a proposal. Journal of American Statistical Association, 58: 415-434, 1963.

60 Cellard J.C., Labbé B., Savitsky G.. Le programme ELISEE, presentation et application. Metra, 3(6): 511-519, 1967.

61 Sonquist J.A., Baker E.l., Morgan J.N.. Seraching for structure (Alias-AID-III). Survey Research Center, Institute for Social research, University of Michigan, Ann Arbor. 1971.

62 Bouroch, J.-M., Tenenhauss M.. Quelques méthods de segmentation. Revue française d'informatique et de recherce óperationelle, 4(2): 29-42, 1970.

63 Bouroche J.-M., Tenenhauss M.. Some segmentation methods. Metra, 7: 407-418, 1972.

64 Messenger R., Mandel L.. A modal serch technique for predictive nominal scale multivariate analysis. Journal of the American Statistical Association, 67(340): 768-772, 1972.

65 Morgan J.N., Messenger R.C.. THAID, a sequential analysis program for analysis of nominal scale dependent variables. Survey Research Center, Institute for Social research, University of Michigan, Ann Arbor, 1973.

66 Gillo M.W.. MAID, a Honeywell 600 program for automatized survey analysis. Behaviorial Science, 17(2): 251-252, 1972.

67 Gillo M.W., Shelly M.W.. Predictive modeling of multivariable and multivariate data. Journal of American Statistical Association, 69(347): 646-653, 1974.

68 Hunt E.B., Marin J., Stone P.J.. Experiments in induction. New York and London: Academic Press, 1966.

69 Press L.I., Rogers M.S., Shure G.H.. An interactive technique for the anaalysis of multivariate data. Behavioral Sciences, 14(5): 364-370, 1969.

70 Quinlan J.R.. Induction of decision trees. Machine Learning 1, 81-106, 1986.

71 Quinlan J.R.. C4.5: Programmes for Machine Learning. Morgan Kaufmann, Los Altos, 1993.

72 Rokach L., Maimon O.. Decicion Trees. *In* Data Mining and knowledge Discovery Handbook, 2nd Ed. Maimon O., Rokach L., Chpater 9, 2010.

73 Quinlan J.R.. Simplifying decision trees. International Journal of Man-Machine Studies, 27: 221-234, 1987.

74 Attneave F.. Applications of Information Theory to Psychology. Holt, Rinehart and Winston, 1959.

75 Dietterich T.G., Kearns M., Mansour Y.. Applying the weak learning framework to understand and improve C4.5. Proceedings of the Thirteenth International Conference on Machine Learning. San Francisco: Morgan Kaufmann, 96-104, 1996.

76 Kearns M., Mansour Y.. On the boosting ability of top-down decision tree learning algorithms. Journal of Computer Sciences, 58(1): 109-128, 1999.

77 Li X., Dubes R.C., Tree classifier design with a Permutation statistic. Pattern Recognition 19:229-235, 1986.

78 Taylor P.C., Silverman B.W.. Block diagrams and splitting criteria for classification trees.
In: Statistics & Computing, Vol. 4, p. 147 - 161, 1993.

79 Hays W.L.. Statistics. Holt, Rinehart, and Winston, 1988.

80 Nickerson C.A.E.. A Note on 'A Concordance Correlation Coefficient to Evaluate Reproducibility'. *Biometrics* (International Biometric Society), 53 (4): 1503–1507, 1997.

81 Lin L.I.-K. A concordance correlation coefficient to evaluate reproducibility. Biometrics, 45: 255-268, 1989.

82 Fisher R.A.. Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika* (Biometrika Trust) 10 (4): 507–521, 1915.

83 Fisher R.A.. On the `probable error' of a coefficient of correlation deduced from a small sample. *Metron* 1: 3–32, 1921.

84 Krouwer J.S., Monti K.L.. A simple, graphical method to evaluate laboratory assays. Eur J Clin Chem Clin Biochem, 33:525-527, 1995.

85 Bland J.M., Altman D.G.. Measuring agreement in method comparison studies. Statistical Methods in Medical Research, 8: 135-160, 1999.

86 Youden W.J.. Graphical diagnosis of interlaboratory test results. Industrial Quality Control, 15, 24-28, 1959.

87 Beyer W.H.. CRC Standard Mathematical Tables, 28th ed. Boca Raton, Fl, CRC Press. Pp 123-124, 1987.

88 DeLong E.R., DeLong D.M., Clarke-Pearson D.L.. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics, 44: 837-845, 1988.

89 Mateus C., Lemos R., Silva M.F., Reis A., Fonseca P., Oliveiros B., Castelo-Branco M.. Aging of Low and High Level Vision: From Chromatic and Achromatic Contrast Sensitivity to Local and 3D Object Motion Perception. PLOS One, Volume 8, Issue 1, 2013.

90 Petrie A., Bulman J.S., Osborn J.F.. Further statistics in dentistry. Part 8: systematic reviews and meta-analyses. British Dental Journal, 194:73-78, 2003.

91 DerSimonian R., Laird N.. Meta-analysis in clinical trials. Controlled Clinical Trials, 7: 177-188, 1986.

92 Borenstein M., Hedges LV, Higgins J.P.T., Rothstein H.R.. Introduction to meta-analysis. Chichester, UK, Wiley, 2009.

93 Higgins J.P., Thompson S.G., Deeks J.J., Altman D.G.. Measuring inconsistency in meta-analyses. BMJ, 327:557-560, 2003.