

TOWARDS A NOVEL PREDICTIVE ANALYSIS
FRAMEWORK FOR NEW-GENERATION
CLINICAL DECISION SUPPORT SYSTEMS

THOMAS MAZZOCCO

Doctor of Philosophy

Institute of Computing Science and Mathematics

University of Stirling

October 2014

Thomas Mazzocco: *Towards a novel predictive analysis framework for new-generation clinical decision support systems*, Doctor of Philosophy, © October 2014

CONTENTS

Abstract	13
1 INTRODUCTION	18
1.1 Structure of the thesis	18
1.2 Motivation and aims	19
1.3 Original contributions	20
1.4 Publications	21
2 STATE OF THE ART	23
2.1 Clinical decision support systems	23
2.2 The role of knowledge	27
2.3 Features of an effective system	30
2.4 An overview of data mining	31
2.5 Logistic regression models	33
2.6 Pitfalls in published predictive models	34
2.7 Towards a general framework	38
3 FRAMEWORK DESCRIPTION	40
3.1 Dataset preparation	41
3.1.1 Variables pre-processing	41
3.1.2 Collinearity among independent variables	43
3.1.3 Variables selection	44

3.2	Model construction	46
3.2.1	Regression	47
3.3	Model validation	50
3.4	Model evaluation	52
3.5	Misclassification analysis	56
3.6	Implementation	58
4	CASE STUDY 1: A 28-DAY MORTALITY MODEL FOR ACUTE ALCOHOLIC HEPATITIS	59
4.1	Background	60
4.2	Aims	62
4.3	Dataset preparation	63
4.3.1	Study sample	63
4.3.2	Candidate variables	64
4.4	Model construction	65
4.4.1	Logistic regression	65
4.4.2	Variables selection and model validation	65
4.4.3	Model evaluation	66
4.5	Results	67
4.6	Misclassification analysis	70
4.7	Implementation	74
4.8	Discussion	78
4.9	Conclusion	80

5	CASE STUDY 2: A PREDICTIVE MODEL TO AID THE DIAGNOSIS OF DEMENTIA	82
5.1	Background	83
5.2	Aims	86
5.3	Dataset preparation	87
5.3.1	Study sample	87
5.3.2	Collected variables	87
5.4	Model construction	90
5.4.1	Logistic regression	90
5.4.2	Variables selection	91
5.4.3	Model validation	91
5.4.4	Model evaluation	92
5.5	Results	92
5.6	Misclassification analysis	101
5.7	Implementation	107
5.8	Discussion	111
5.9	Conclusion	113
6	CASE STUDY 3: A SIDE-EFFECTS MAPPING MODEL IN PATIENTS RECEIVING CHEMOTHERAPY	115
6.1	Background	116
6.2	Aims	119
6.3	Dataset preparation	120
6.4	Model construction	121

6.4.1	Pre-modelling	121
6.4.2	Regression	123
6.4.3	Model evaluation and validation	124
6.5	Results	127
6.5.1	Area under ROC curve	127
6.5.2	R^2 and p-values	128
6.5.3	Analysis	130
6.6	Implementation	133
6.7	Discussion	136
6.8	Conclusion	137
7	CONCLUSIONS AND FUTURE WORK	139
7.1	Conclusions	139
7.2	Future work	141
A	INFORMATION AND COMMUNICATION TECHNOLOGY USAGE IN PATIENTS	144
A.1	Communication technologies for healthcare	145
A.2	Methodology	147
A.3	Results from questionnaire	148
A.4	Analysis of results	150
A.5	Conclusion	155
B	MATHEMATICAL FORMULATION OF REGRESSIONS AND GRADIENT DESCENT	156
B.1	Linear regression	156

B.2	Logistic regression	157
B.3	Gradient descent	157
BIBLIOGRAPHY		159

LIST OF FIGURES

Figure 3.1	Examples of ROC curves	55
Figure 3.2	Schematic representation of the model building process	57
Figure 4.1	Comparison of accuracies with/without validation strategy	67
Figure 4.2	Distribution of predictions and misclassifications (function of SCORE/100)	72
Figure 4.3	Distribution of predictions and misclassifications (function of M)	72
Figure 4.4	Distribution of predictions and misclassifications for mDF	74
Figure 4.5	Input screen for ALD mortality predictor	76
Figure 4.6	Output example for ALD mortality predictor	77
Figure 5.1	TPR and TNR values for different thresholds	99
Figure 5.2	PPV and NPV values for different thresholds	100
Figure 5.3	Distribution predictions/misclassifications $\tau = 0.5$ (function of P(Y))	103
Figure 5.4	Distribution predictions/misclassifications $\tau = 0.5$ (function of SCORE)	103
Figure 5.5	Distribution predictions/misclassifications $\tau = 0.7$ (function of P(Y))	104

Figure 5.6	Distribution predictions/misclassifications $\tau = 0.7$ (function of SCORE)	104
Figure 5.7	Distribution predictions/misclassifications $\tau = 0.9$ (function of $P(Y)$)	105
Figure 5.8	Distribution predictions/misclassifications $\tau = 0.9$ (function of SCORE)	105
Figure 5.9	Input screen for dementia risk calculator	109
Figure 5.10	Output example for dementia risk calculator	110
Figure 6.1	Chosen functions	123
Figure 6.2	Comparison of performance for each symptom	127
Figure 6.3	Representation of probability given by the model (nausea for breast cancer)	132
Figure 6.4	Representation of probability given by the model (mucositis for lung cancer)	132
Figure 6.5	Input screen for chemotherapy side-effects model	134
Figure 6.6	Output example for chemotherapy side-effects model	135

LIST OF TABLES

Table 2.1	Studies which satisfy criteria for quality of logistic regression models	37
-----------	--	----

Table 4.1	Analysis of coefficients from logistic regression model	69
Table 4.2	Comparison of scoring systems in patients with AH, who died and survived	70
Table 4.3	Distribution of misclassifications (function of SCORE/100)	71
Table 4.4	Distribution of misclassifications (function of M)	71
Table 4.5	Distribution of misclassifications for mDF	74
Table 5.1	Results from logistic regression using expert driven vari- ables selection	95
Table 5.2	Results from logistic regression using statistics driven variables selection	97
Table 5.3	Summary of regressions with three subsets	98
Table 5.4	Distribution of misclassifications (function of $P(Y)$)	101
Table 5.5	Distribution of misclassifications (function of SCORE)	102
Table 6.1	Regression coefficients for breast cancer regression model	125
Table 6.2	Regression coefficients for colorectal cancer regression model	126
Table 6.3	Regression coefficients for lung cancer regression model	126
Table 6.4	AUC for current and previous models	128
Table 6.5	Comparison of R^2	129
Table 6.6	P-values for breast cancer regression model	129
Table 6.7	P-values for colorectal cancer regression model	130

Table 6.8 P-values for lung cancer regression model 131

ACRONYMS

AH Alcoholic hepatitis

ALD Alcoholic liver disease

AUC Area under curve

BBN Bayesian belief network

CDSS Clinical decision support system

CPS Child-Pugh Score

DSM-IV 4th diagnostic and statistical manual of mental disorders

FN False negative

FP False positive

HTML Hypertext markup language

ICT Information and communication technologies

ICU Intensive care unit

MDF Modified Maddrey discriminant function

MELD Model for end stage liver disease

MMSE Mini-mental state examination

NPV Negative predictive value

PHP PHP: hypertext preprocessor

PPV Positive predictive value

ROC Receiver operating characteristic

SAMP Stirling ALD mortality predictor

SVM Support vector machine

TN True negative

TNR True negative rate

TP True positive

TPR True positive rate

VIF Variance inflation factor

ABSTRACT

The idea of developing automated tools able to deal with the complexity of clinical information processing dates back to the late 60s: since then, there has been scope for improving medical care due to the rapid growth of medical knowledge, and the need to explore new ways of delivering this due to the shortage of physicians. Clinical decision support systems (CDSS) are able to aid in the acquisition of patient data and to suggest appropriate decisions on the basis of the data thus acquired. Many improvements are envisaged due to the adoption of such systems including: reduction of costs by faster diagnosis, reduction of unnecessary examinations, reduction of risk of adverse events and medication errors, increase in the available time for direct patient care, improved medications and examination prescriptions, improved patient satisfaction, and better compliance to gold-standard up-to-date clinical pathways and guidelines.

Logistic regression is a widely used algorithm which frequently appears in medical literature for building clinical decision support systems: however, published studies frequently have not followed commonly recommended procedures for using logistic regression and substantial shortcomings in the reporting of logistic regression results have been noted. Published literature has often accepted conclusions from studies which have not addressed the ap-

appropriateness and accuracy of the statistical analyses and other methodological issues, leading to design flaws in those models and to possible inconsistencies in the novel clinical knowledge based on such results.

The main objective of this interdisciplinary work is to design a sound framework for the development of clinical decision support systems. We propose a framework that supports the proper development of such systems, and in particular the underlying predictive models, identifying best practices for each stage of the model's development.

This framework is composed of a number of subsequent stages: 1) *dataset preparation* insures that appropriate variables are presented to the model in a consistent format, 2) the *model construction* stage builds the actual regression (or logistic regression) model determining its coefficients and selecting statistically significant variables; this phase is generally preceded by a pre-modelling stage during which model functional forms are hypothesized based on a priori knowledge 3) the further *model validation* stage investigates whether the model could suffer from overfitting, i.e., the model has a good accuracy on training data but significantly lower accuracy on unseen data, 4) the *evaluation* stage gives a measure of the predictive power of the model (making use of the ROC curve, which allows to evaluate the predictive power of the model without any assumptions on error costs, and possibly R^2 from regressions), 5) *misclassification analysis* could suggest useful insights into determining where the model could be unreliable, 6) *implementation* stage.

The proposed framework has been applied to three applications on different domains, with a view to improve previous research studies.

The first developed model [1] predicts mortality within 28 days of patients suffering from acute alcoholic hepatitis. The aim of this application is to build a new predictive model that can be used in clinical practice to identify patients at greatest risk of mortality in 28 days as they may benefit from aggressive intervention, and to monitor their progress while in hospital. A comparison generated by state of the art tools shows an improved predictive power, demonstrating how an appropriate variables inclusion may result in an overall better accuracy of the model, which increased by 25% following an appropriate variables selection process.

The second proposed predictive model [2] is designed to aid the diagnosis of dementia, as clinicians often experience difficulties in the diagnosis of dementia due to the intrinsic complexity of the process and lack of comprehensive diagnostic tools. The aim of this application is to improve on the performance of a recent application of Bayesian belief networks using an alternative approach based on logistic regression. The approach based on statistical variables selection outperformed the model which used variables selected by domain experts in previous studies. Obtained results outperform considered benchmarks by 15%.

The third built model [3] predicts the probability of experiencing a certain symptom among common side-effects in patients receiving chemotherapy. The newly developed model includes a pre-modelling stage (which was based

on previous research studies) and a subsequent regression. The computed accuracy of results (computed on a daily basis for each cycle of therapy) shows that the newly proposed approach has increased its predictive power by 19% when compared to the previously developed model: this has been obtained by an appropriate usage of available a priori knowledge to pre-model the functional forms.

As shown by the proposed applications, different aspects of CDSS development are subject to substantial improvements: the application of the proposed framework to different domains leads to more accurate models than the existing state-of-the-art proposals. The developed framework is capable of helping researchers to identify and overcome possible pitfalls in their ongoing research works, by providing them with best practices for each step of the development process.

An impact on the development of future clinical decision support systems is envisaged: the usage of an appropriate procedure in model development will produce more reliable and accurate systems, and will have a positive impact on the newly produced medical knowledge which may eventually be included in standard clinical practice.

Ut in omnibus glorificetur Deus

INTRODUCTION

This thesis is the result of a PhD research project funded by the Division of Computing Science and Mathematics (School of Natural Sciences) at the University of Stirling, carried out under the supervision of Professor Amir Hussain (principal supervisor) and Dr David Cairns (second supervisor).

1.1 STRUCTURE OF THE THESIS

The next chapter 2 gives a general introduction to clinical decision support systems: moving from definitions and motivations for the work, the role of knowledge in decision support and techniques used are presented. A brief analysis of key features and benefits of such systems follows. Finally, the chapter focuses on logistic regression models which have been developed in the recent past, identifying common pitfalls and suggesting a possible general framework for developing and evaluating such systems. Further literature review is presented throughout the thesis for each specific model which has been developed.

Following chapters are built around the author's published work. Chapter 3 introduces the proposed framework and the following chapters describe the development of three predictive models used to validate the proposed framework: specifically, chapter 4 presents a model developed to predict mortality within 28 days of patients suffering from acute alcoholic hepatitis [1], chapter 5 describes the development of a predictive model to aid the diagnosis of dementia [2], and chapter 6 presents the development of a model for predicting and monitoring the side-effects in patients receiving chemotherapy [3]. The development of such models also includes the implementation of web-based CDSSs built on developed models and different attempts to improve their performance.

Concluding remarks and ideas for future work are presented in chapter 7.

Appendix A presents an investigation on information and communication technology usage in patients [4] (the author's contribution is the statistical data elaboration and the subsequent interpretation of results) and appendix B describes the mathematical formulation of linear regression, logistic regression and gradient descent.

1.2 MOTIVATION AND AIMS

The increased demand for medical care (also due to the increase in average life expectancy) and the need to keep health costs under control, makes the exploration of new ways of delivering healthcare crucial.

Clinical decision support systems (CDSSs) are able to aid in the acquisition of patient data and to suggest appropriate decisions on the basis of the acquired data. Literature has identified many improvements which are envisaged from the adoption of such systems, including improved patient safety, better quality of care and enhanced efficiency in healthcare delivery.

However, published studies have frequently not followed commonly recommended procedures for model development and substantial shortcomings in reporting results have been ascertained.

The first objective of this interdisciplinary work is to design a sound framework for the development of clinical decision support systems. We propose a framework supporting the proper development of such systems, and in particular the underlying predictive models, identifying best practices for each stage of the model's development. Another main aim of this thesis is to build some applications based on the proposed framework, able to outperform state of the art models. As a peripheral objective, we also want to investigate possible barriers to the delivery of CDSS to patients using information and communication technologies (ICT).

1.3 ORIGINAL CONTRIBUTIONS

A schematic view of the original contributions of this thesis is hereby reported.

1. Chapter 3: a framework supporting the proper development of clinical decision support systems, and in particular the underlying predictive models, is proposed.
2. Chapter 4: a new mortality risk model to identify patients suffering from acute alcoholic hepatitis is developed, outperforming by 25% available state of the art tools. [1]
3. Chapter 5: a predictive model to aid the early diagnosis of dementia is developed, outperforming the considered benchmark by 15% and providing an analysis of the variables significance. [2]
4. Chapter 6: a side-effect model for cancer patients receiving chemotherapy is developed, showing the enhanced predictive power compared to the previous models. [3]

1.4 PUBLICATIONS

The following papers have resulted from the research presented in this thesis.

- T. Mazzocco and A. Hussain, "A side-effects mapping model in patients with lung, colorectal and breast cancer receiving chemotherapy," in *13th IEEE International Conference on e-Health Networking Applications and Services (Healthcom)*, pp. 34-39, IEEE, 2011.

- T. Mazzocco and A. Hussain, "Novel logistic regression models to aid the diagnosis of dementia," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3356-3361, 2012.
- T. Gandiya, A. Dua, G. King, T. Mazzocco, A. Hussain, and S. J. Leslie, "Self-reported 'communication technology' usage in patients attending a cardiology outpatient clinic in a remote regional hospital," *Telemedicine and e-Health*, vol. 18, no. 3, pp. 219-224, 2012.
- T. Mazzocco, A. Hussain, S. Hussain, and A. A. Shah, "A novel mortality model for acute alcoholic hepatitis including variables recorded after admission to hospital," *Computers in biology and medicine*, vol. 44, pp. 132-135, 2014.

STATE OF THE ART

This chapter introduces a general definition of a clinical decision support system and an analysis of the motivations for the present work. The role of existing medical knowledge in such systems and the way to extract new knowledge using data mining techniques are then presented. Finally a review of the features of effective clinical decision support systems and their envisaged benefits in delivered clinical practice concludes the chapter.

2.1 CLINICAL DECISION SUPPORT SYSTEMS

The idea of developing automated tools able to deal with the complexity of clinical information processing dates back to at least 1969: a work by Goertzel [5] highlighted the increasing need for better medical care due to the rapid growth of medical knowledge and the shortage of physicians. This work gave the definition of clinical decision support system (CDSS) as a "computer program to aid in the acquisition of patient data and to suggest appropriate decisions on the basis of the data thus acquired". So two main functions are identified for a CDSS:

- to help with data acquisition processes (clinical variables may include boolean variables, numeric values and free text descriptions);
- to assist health professionals with decision making tasks (e.g. diagnosis, suggested treatments, further examinations).

Almost 50 years later, the ideas behind clinical decision support systems are still the same: the increased demand for medical care, also due to the increase in average life expectancy, and the need to keep control of health costs make the adoption of new ways to deliver healthcare crucial.

Many improvements are envisaged from the use of such systems including: reduction of costs by faster diagnosis, reduction in unnecessary examinations, reduction of the risk of adverse events and medication errors, increase in available time for direct patient care, improved medication and examination ordering, improved patient satisfaction, and application of gold-standard up-to-date clinical pathways and guidelines. Potential benefits from CDSSs may then be summarized in three areas [6]:

- improved patient safety;
- improved quality of care;
- improved efficiency in health care delivery.

CDSSs have the potential to meet these needs, and the huge amount of clinical data which is collected and recorded nowadays seems to pave the way for a radical change in medical practice.

Nowadays a huge amount of information is collected in the healthcare environment, which could still be defined as 'information rich' yet 'knowledge poor'. This is due to a lack of effective tools able to analyze and discover underlying relationships and patterns in clinical data, while data mining and knowledge discovery have found many applications in other scientific domains [7].

Different studies have shown that, for different reasons, health care is suboptimal. From a very general point of view it has been noted [8] that while healthcare sectors require person-to-person interaction for treatment assessment, diagnosis, planning and decision making, the delivered decision support does not need such interaction. Information services in healthcare, and specifically decision support systems, should consider the need for a much greater degree of interpersonal communication compared to other sectors. Also, healthcare information services should be modelled on a national scale, in order to integrate data from different providers, while generally attempts are made to build solutions for smaller units (e.g. hospital, surgery, clinic).

Another study [9] has shown that while overall differences between primary care quality indicators in deprived and prosperous communities were small, both clinical and non-clinical indicators suggested that there is scope for focused interventions on healthcare delivery with a view to improving the quality of primary care in deprived areas.

Evidence [10] has shown that practitioners do not always adhere to recommended clinical practice. However, compliance improves following a

structured intervention, even if research studies do not indicate for how long the intervention affects practitioner compliance: appropriate decision support tools may help to increase practitioners' compliance to recommended practice.

In fact, to improve the chance that a decision support system has a positive influence, it has been found [11] that providing personalized feedback to practitioners was effective in order to improve adherence to recommendations. Also, requiring practitioners to acknowledge received reminders improved their adherence to recommendations [12].

Due to the discrepancy between clinical care actually being delivered and optimal patient care, alternative care models in traditional primary care are being actively explored. Published studies of clinical decision support systems (CDSSs) are increasing and their quality is also improving [13]: such systems can enhance clinical performance for drug dosing and prescribing, preventative care, diagnosis, disease management and other aspects of medical care.

Another research study [14] on the effects of computerized CDSSs on practitioner performance concluded that such systems may improve practitioner performance and identified few barriers to its implementation in clinical practice including: failure of physicians to use the system, poor integration into practitioner workflow, poor usability of the tools, non-acceptance of computer recommendations.

2.2 THE ROLE OF KNOWLEDGE

An important categorization among clinical decision support systems can be based on the "reasoning" engine of the system [15]: a *knowledge-based* CDSS is a CDSS with a knowledge base that consists of compiled information (often presented as a number of if-then rules) used to associate the input received from the user with the output provided to the user.

Examples of such CDSS can be found online based on the well-known Framingham study dating back to 1976 for general cardiovascular risk prediction [16]: this risk model used four boolean variables (sex, smoker, treatment for hypertension, diabetes) and three numerical variables (age, blood pressure, body mass index) to calculate a risk score of cardiovascular events.

Online tools collect necessary information from the user (and needed explanations) and are able to provide an immediate answer; this could be directed to healthcare professionals, who may consider the risk score as one of the preliminary indicators of patient's health condition, but also to the user of the system who may be then guided to appropriate actions to improve their health condition (e.g. modifying lifestyle) or redirected to general practitioners.

On the contrary, a *non-knowledge-based* CDSS do not rely on previous knowledge and uses a number of data points to provide predictions on new unseen cases. These rely on standard statistical and/or more sophisticated machine learning algorithms in order to make the system learn from available

clinical data, recognize patterns and improve predictions based on new annotated data.

For example, a recent PhD thesis [17] from Computer Science and Artificial Intelligence Laboratory (CSAIL) of Massachusetts Institute of Technology (MIT) developed some logistic regression models to detect hazardous episodes for patients in intensive care units. A number of real-time mortality models have been developed exploiting data coming from the Multi-parameter Intelligent Monitoring for Intensive Care (MIMIC) database, which was created to facilitate the development and evaluation of intensive care units (ICU) decision-support systems. Data have been pre-processed before developing predictive models which have been subsequently validated.

At this point, it can be useful to note that non-knowledge-based CDSSs can also be exploited to generate new clinical knowledge: the analysis of the models obtained from clinical data suggests which factors may influence positively or negatively the provided output (e.g. a diagnosis, an expected side-effect, the necessity for further treatment).

It is then possible to identify a third role of a CDSS in the context of scientific research (along with tasks related to helping in data acquisition and decision making): in fact, to some extent, non-knowledge-based decision support systems have the capacity to learn from available data, leading to the identification of new relationships between provided input and desired output and, ultimately, facilitating the creation of new clinical knowledge [18].

Another recent example [19] of non-knowledge-based CDSS was developed by collecting data from an integrated sensor network which was installed in apartments for volunteer residents at a retirement community that allows residents to remain in their apartments even if their health deteriorates. The sensor networks supplement registered nurse care coordination by alerting them of changes in the normal sensor patterns. The analysis of data collected from such sensors (bed restlessness sensor, living room motion sensor, and bathroom visits for each resident) led to the discovery of anomalous behaviors which were strongly correlated to urinary tract infections. In this case, the sensor network was able to detect signs of illness earlier than traditional health care assessment. This example shows how heterogeneous and diverse information collected in the healthcare environment may lead to discovering hidden relationships with clinical conditions.

We can then design a CDSS which analyses a large amount of data, with a view to discovering hidden (and possibly complex) patterns. Also, we can use existing knowledge to design a CDSS, and then use a new set of observations to show how to refine gold standard practice or to identify possible conflicts with the existing knowledge base.

For the scope of this work, moving from available medical knowledge, we focus on building new CDSSs with a view to improving gold standard practice when possible, extracting also new information from available data.

2.3 FEATURES OF AN EFFECTIVE SYSTEM

Despite the huge number of developed CDSSs, only a few of them are then integrated into clinical workflows. A recent study [20] has estimated that such systems have improved clinical practice in about 68% of the considered cases. It also identified features of computer based decision support systems which significantly correlated with a positive impact on clinical practice:

- automatic provision of decision support as part of clinician workflow;
- provision of recommendations rather than just assessments;
- provision of decision support at the time and location of decision making.

Moreover, direct experimental evidence supported the importance of three further system characteristics:

- providing periodic performance feedback;
- sharing recommendations with patients;
- requesting documentation of reasons for not following recommendations.

Another recent study [21] confirmed that systems that required practitioners to provide reasons when overriding advice and, even more importantly, systems that provided advice concurrently to patients and practitioners were more likely to be effective in clinical practice.

Multiple studies have emphasized the importance of transparency in clinical decision support systems: it has been highlighted that such systems must

be implemented ensuring transparency so that the source and strength of evidence are fully disclosed to clinicians and other decisionmakers [22].

An effective clinical decision support system should be able to evaluate available patient data in real time, and to provide decision support (e.g. prompts, reminders, and suggestions for management, patient-specific recommendations) in a transparent way, so that suggestions can be easily linked to the source of evidence [23].

There is a strong requirement that the decision support in key areas of medication management and clinical care is provided in a way that allows healthcare professional to determine its credibility and validity by means of a transparent decision making process [24].

2.4 AN OVERVIEW OF DATA MINING

As previously discussed, the large growth of available medical databases has motivated the use of data with a view to discovering new medical knowledge from such databases. Data mining techniques could be employed to extract new clinical knowledge and then to provide better diagnostic capabilities and more effective patient care; potential uses of data mining techniques for medical diagnostics have been ascertained [25]. Knowledge management in the healthcare system is nowadays crucial in order to achieve high-quality cost-effective services and data mining techniques may help to exploit the full potential of collected data within an organization [26].

A group of heterogeneous algorithms are described as data mining techniques; they are used for different purposes but all with a view to discovering the hidden knowledge in the data. Generally, the algorithms try to fit a model which is the closest to the characteristics of the considered dataset, in order to build predictive or descriptive models, where predictive models are used to make predictions (e.g. to make a diagnosis for a specific disease) while descriptive models are used to identify patterns in data. For the scope of this thesis, we will focus on building predictive models.

The main data mining tasks were enumerated and described [27] as follow:

- *Classification* which classifies each data item into one of several predefined classes. It is possible to derive a set of classification rules from the classification model (based on the training dataset); this set of rules represents the knowledge extracted from the used dataset and can be used to classify unseen data items. This is probably the most important data mining technique, as medical diagnosis is an important application of classification.
- *Regression* is a method to map target data with a known type of function, with the aim of estimating an output value given unseen input values.
- *Time series analysis* concerns the study of an attribute value examined over a time period (generally at evenly spaced time intervals), with a view to predicting future values of the attribute.

- *Visualization techniques* are useful in order to discover hidden patterns in the considered dataset. Through the analysis of scatter diagrams (in a Cartesian plane) it is possible to identify interesting subsets of the initial dataset: other data mining techniques can then be applied on these subsets in order to discover further knowledge.
- *Association rules* is the discovery of rules which create associations among objects.
- *Clustering* is a set of techniques of multivariate data analysis aimed at the selecting and grouping of homogeneous elements in a dataset, based on measures of similarity between the elements within a multi-dimensional space. Common features of the objects in each cluster are then summarized to extract the class description used to classify unseen data points.

2.5 LOGISTIC REGRESSION MODELS

Logistic regression is a widely used algorithm which frequently appears in medical literature. It is a well established technique which is usable without any specific machine learning skill; also, many software tools are readily available to develop logistic regression models. It is a multivariable method which tries to establish a functional relationship between two or more predictor (independent) variables and one categorical outcome (dependent) variable in

a transparent way: which leads to an easy interpretation of developed models and may explain the broad diffusion of such technique. In logistic regression the probability of Y occurring is predicted given known values of X (vector containing predictors). The general form of the functional dependence given by the regression could be expressed as per formula 2.1:

$$P(Y) = (1 + e^{-(b_0 + b_1x_1 + \dots + b_nx_n)})^{-1} \quad (2.1)$$

where $P(Y)$ is the probability of Y occurring (or, in other words, of Y belonging to a certain class), x_n are predictor variables and b_n are coefficients to be determined by the logistic regression algorithm. The coefficients are estimated by fitting models, based on the available predictors to the observed data. The chosen model will be the one that, when values of the predictor variables are placed in it, results in values of Y being the closest to the observed values.

Further details about logistic regression models and techniques used to estimate coefficients will be provided in the following appendix B. Two applications presented in this thesis will use logistic regression which is one of the most used model for developing CDSSs.

2.6 PITFALLS IN PUBLISHED PREDICTIVE MODELS

A recent study [28] identified 10 criteria used to examine the quality of published logistic regression models; specifically:

- *selection of independent variables*: studies should explain how variables were selected for inclusion in the model. They may be chosen based on earlier research or in the presence of significant correlation with the dependant variable
- *coding of variables*: studies should provide a complete description of the coding scheme for independent variables (i.e. how they are recorded and coded). The coefficient for an independent variable and the subsequent interpretation strongly depend on such scheme which is crucial to the future usability of the model
- *sufficient events per independent variable*: the ratio between the number of events of the less frequent outcome and the number of model independent variables should be at least 10 to 1. If the number of events per independent variable is lower, the estimates of the regression coefficients may be unreliable and the related sample variance, as well as confidence intervals, may also be affected
- *collinearity*: studies should report results of undertaken tests for collinearity, i.e. should investigate whether two or more explanatory variables have an approximate linear relationship between them: such conditions may increase the variance of the coefficient estimates making them very sensitive to minor changes in the model
- *conformity with linear gradient for continuous variables*: models with continuous or ranked independent variables should assure conformity with

the linear gradient or check on the log-odds scale (this is not applicable to boolean predictor variables), i.e. any given change in a continuous predictor variable should have an effect on the log-odds of a positive outcome that is the same magnitude, regardless of the value of the predictor variable [29]

- *statistical significance*: statistical significance of each coefficient should be investigated and results reported
- *fitting procedure*: the procedure for entering variables (e.g. forward or backward selection) into the model should be explicitly stated
- *tests for interactions*: studies should include a discussion of reasons for including or not interaction terms, i.e. variables obtained as functions of two or more original predictors (e.g. x_1x_2)
- *validation*: models should present used validation procedures (e.g. cross validation, split-sample methods, bootstrapping) and discuss results: this step is crucial in order to correctly estimate the accuracy of a classifier
- *goodness-of-fit and discrimination measures*: goodness-of-fit measures and/or discrimination statistics (ROC curves) should be reported along with a description on how well the developed model matches the observed values

The study in [28] also provided an analysis which assessed a number of published papers on two medical journals: they concluded that published

<i>Criterion</i>	<i>% studies</i>
Selection of independent variables	81%
Coding of variables	10%
Sufficient events per independent variable	40%
Collinearity	17%
Conformity with linear gradient for continuous variables	19%
Statistical significance	100%
Fitting procedure	65%
Tests for interactions	39%
Validation	9%
Goodness-of-fit and discrimination measures	19%

Table 2.1: Studies which satisfy criteria for quality of logistic regression models

studies frequently did not follow commonly recommended procedures for using multivariable logistic regression. The results from this study is reported in the table 2.1, which shows the percentage of considered studies which satisfied the abovementioned criteria.

Another study [30] considered logistic regression models (which often find privileged positions in literature reviews) in a specific medical domain, i.e. genetic testing for cancer susceptibility. They found substantial shortcomings in the use and reporting of logistic regression results. Notable results were that no study reported any regression diagnostics or goodness-of-fit measures;

none of the studies reported validation analysis; most of the considered models had a ratio between events and independent variable near or below 10, suggesting that those models may be unreliable. Also, in the genetic testing for cancer susceptibility domain, published literature reviews have accepted conclusions from analyzed studies, which did not address the appropriateness and accuracy of the statistical analyses and other discussed methodological issues. This could lead to possible design flaws in these models and, while the conclusions of these studies may still be valid, once they are considered for inclusion into counseling guidelines and testing protocols, they may produce new clinical knowledge based on results that had not been rigorously tested following recommended statistical guidelines.

2.7 TOWARDS A GENERAL FRAMEWORK

The analysis presented in this chapter shows the potential of CDSSs and, at the same time, the criticalities found in the published literature highlighting the need for a general framework able to exploit such potentials avoiding common pitfalls. The idea developed within this thesis is then to design a flexible framework able to cope with different kinds of healthcare data - depending on application and performance requirements, required models can range from simple to complex - and to build effective models to be implemented as CDSSs.

In the following chapters of this thesis, a general framework will be developed: some applications will be proposed, developing predictive models and implementing at the same time (most of the) recommended practices previously discussed. The main objective is to design a general framework based on clinical and technological best practices, able to help researchers in this multidisciplinary domain develop effective clinical decision support systems.

FRAMEWORK DESCRIPTION

In the previous chapter, a number of pitfalls recurring in the published literature for developing CDSSs have been highlighted. There is then a need for a general framework able to exploit the potential of medical data avoiding common pitfalls which have been discussed in the previous section 2.6.

In this chapter, a framework for developing CDSSs, and in particular the underlying predictive models, is proposed: this will be able to deal with some of the most widely used kind of healthcare data and will help researchers to develop effective CDSSs.

Section 3.1 describes how to prepare the dataset which will be used for the model construction, which is detailed in section 3.2. A methodology to evaluate model performance is defined in section 3.3 and appropriate performance metrics for two-class classification problems are described in section 3.4.

Then, section 3.5 illustrates how to analyze model misclassifications, i.e. cases in which the classification proposed by the system differs from the real classification, and section 3.6 closes the chapter suggesting how to implement the model into an actual CDSS software application.

3.1 DATASET PREPARATION

As previously mentioned, a non-knowledge-based CDSS uses a known dataset to provide predictions on new unseen cases. The dataset is composed of a number of *data points*: each of them has a number of inputs expressed as *independent variables* (i.e. variables used to make predictions) and one output as the *dependent variable* (i.e. the associated outcome). This section examines how to pre-process the dataset in order to properly feed the learning algorithm used for model development.

3.1.1 *Variables pre-processing*

When developing a new model, a number of candidate variables are available: their collection is generally driven by previous research and/or by the expertise of domain specialists. Also, in the absence of such indications, the presence of a significant correlation with the dependent variable could be investigated by means of appropriate statistical tests.

Variables can be categorized into:

- *continuous variables*, which can assume any real value within given intervals;
- *categorical variables*, which can assume a finite number of values.

Categorical variables can be further classified into:

- *binary variables*, which can assume only two values;
- *nominal variables*, which can assume a finite number of values without an intrinsic order;
- *ordinal variables*, which can assume a finite number of values with an intrinsic order or ranking.

Continuous variables are normally used “as is” or after a normalization process, which consists in a transformation $\mathbb{R} \rightarrow [0, 1]$ or $\mathbb{R} \rightarrow [-1, 1]$ of the variable domain. Different methods are available for such transformations (e.g. min-max normalization, z-score normalization) and many artificial neural networks and classifiers based on distances require a normalization step in order to give consistent results [31]. While not necessary, this process is sometimes applied to regression and logistic regression as well (e.g. when the same dataset is also used with other techniques requiring normalization). In this case, it is crucial that the function used for normalization is explicitly reported when the model is described in order to consistently apply input normalization when the model is used.

Categorical variables cannot be used “as is”. They need to be coded into a series of $n - 1$ binary variables where n is the number of categories to be represented: it has to be noted that $n - 1$ binary variables are able to define exactly n categories, while using n binary variables would lead to a n -th variable which could be expressed as function of the other $n - 1$ ones causing problems to learning algorithms (i.e. making impossible the matrix inversion in the estimation algorithm), making the regression problem unsolvable [32]. This

coding is necessary to avoid the well known “dummy variable trap” which is special case of exact multicollinearity: if there is no omitted category, there is an exact linear relationship between the model constant and the dummy (binary) variables [33].

Generally speaking, model specifications should always explain how variables are collected (including units of measure), calculated and used in order to guarantee that the model will always be applied to datasets which are consistent with the one used for developing such models, i.e. variables are of the same kind and measured in the same unit.

3.1.2 *Collinearity among independent variables*

A preliminary test of collinearity should investigate whether two or more independent variables have a strong correlation: if there is perfect collinearity between independent variables it becomes impossible to obtain unique estimates of the model coefficients [34]. However, also high levels of collinearity present a problem for any regression analysis [35], increasing the probability that a good predictor (i.e. an independent variable which has good explanatory power) is considered not significant and then rejected by the model.

As already discussed, it is estimated that less than 20% of published literature on medical logistic regression models reported appropriate tests for detecting collinearity problems. The methodology proposed in this framework requires that an appropriate test is carried out to detect collinearity situations.

Various collinearity diagnostics are available: for example, the variance inflation factor (VIF) or the tolerance statistics (defined as $1/\text{VIF}$). VIF provides an estimate of how much the variance of an estimated coefficient is increased by the effect of collinearity [36]. Common criteria to determine if a collinearity problem is present are a tolerance value less than 0.1 [37] or, equivalently, a VIF value greater than 10 [38].

3.1.3 *Variables selection*

Stepwise regression is a procedure to select which variables should be included in a regression model. The idea is to build a model with a specified number of variables and then add (or remove) them one by one, according to a specified ranking, and then check whether the model significantly improved (or deteriorated) its performance. Such an iterative procedure terminates when adding (or removing) variables does not improve (or does deteriorate) the model accuracy.

Two approaches may be followed: backward stepwise elimination, which starts with all candidate variables and at each step one variable is removed evaluating if performance deteriorates, and forward stepwise selection, which starts with no candidate variables and at each step one variable is added evaluating if performance improves [39].

For the framework proposed in this thesis, the backward method is preferred because forward selection is more likely than backward elimination to exclude

independent variables involved in suppressor effects [34] (i.e. variables which increase the predictive power of other independent variables by their inclusion in a regression equation [40]). Forward selection may be used in an attempt to reduce computational time, if techniques involved in model construction are computationally heavy and/or the used dataset is big enough.

The implementation of variables selection strategies can lead to increased model performance, as will be shown in sections 4.5 and 5.5.

Stepwise regression suggests the best set of variables to fit points in a specified dataset according to a certain functional form. However, the number of variables which can be used in a model is limited by the size of the dataset used to estimate coefficients: introducing more variables will generally produce a better fit to the data but an excessive number of variables may overfit the dataset, leading the model to lose its generalisation power.

In order to create a model able to show similar accuracy on an unseen dataset, a dataset of adequate size, which could provide reasonably accurate estimates of the regression coefficients, is necessary.

If too many degrees of freedom are used, i.e. if too many coefficients are estimated with respect to the number of data points contained in the dataset, the resulting model will overfit the considered sample: it will include predictor variables and identify complex relations between input and output of the model that exist in the considered sample (leading to an overestimation of model accuracy) but not in the population, and this is detrimental to the real accuracy of the model. In the case of regression models, if the number of

the coefficients to be estimated is equal to the number of data points on the considered sample, the model may perfectly fit the sample data, even if all the predictors are totally unrelated to the response variable (i.e. they are noise) [41].

Summing up, although it is important that the model includes all relevant variables, it is also important that the model does not start with more variables than those that are justified for the given number of observations [29, 42, 43].

Given a specific dataset, for logistic regression it is recommended that the ratio between the minimum number of data points belonging to a class and the number of independent variables used for the model should be at least 10 [44]: if such criterion is not met, results should be taken with care and a larger dataset should be used to strengthen findings. This rule will help selecting the appropriate set of independent variables as in the first case study described in paragraph 4.4.2.

3.2 MODEL CONSTRUCTION

After dataset preparation, a number of data points are available for the model construction phase.

The vector of selected candidate independent variables is called X and B is a vector of coefficients. The aims of the step described in this section are 1) to find a function $\hat{y} = f(B, X)$ able to approximate the real relationship $y(X)$ between X and the dependent variable y and 2) to determine vector B .

The first element of the vector X is set to $x_0 = 1$, in order to include in the developed models a possible constant term while keeping a compact notation.

The form of function f is based on the available knowledge about the relationship between the model input and output. Also, continuous variables can be used stand-alone or as arguments in functions of one (e.g. $\sin x_1$) or more variables (e.g. $x_1 x_2$), again based on the available knowledge about the problem. An appropriate usage of a priori knowledge will result in more accurate models as highlighted in paragraph 6.4.1.

3.2.1 *Regression*

Depending on desired output y , in most cases, linear and logistic regression are able to provide models with a reasonable level of accuracy. These techniques will be used in the following chapter 4 and 5 (logistic regression), and 6 (linear regression).

In particular, if $y \in \mathbb{R}$, linear regression could be used. Otherwise, considering classification problems (i.e. $y \in \{0, 1\}$), logistic regression should be selected; in this case, the model output $\hat{y} = f(B, X)$ is the probability of an input data point belonging to a certain class. A threshold is generally applied to the probability calculated from the model in order to predict the class to which the data point is expected to belong. Besides being needed in the practical usage of the model, the threshold is also commonly used to quickly evaluate the accuracy of the model (i.e. once a threshold has been chosen, the

accuracy is easily evaluated as described in the following section 3.4). However, the predictive power should be evaluated regardless of an arbitrarily chosen threshold using receiver-operating characteristic (ROC) curves, as detailed hereinafter.

A linear regression model will assume the form:

$$f(\mathbf{B}, \mathbf{X}) = b_0 + b_1x_1 + b_2x_2 \cdots + b_nx_n = \mathbf{B}^T\mathbf{X} \quad (3.1)$$

A logistic regression model will instead assume the form:

$$f(\mathbf{B}, \mathbf{X}) = \frac{1}{1 + e^{-(b_0 + b_1x_1 + b_2x_2 \cdots + b_nx_n)}} = \frac{1}{1 + e^{-\mathbf{B}^T\mathbf{X}}} \quad (3.2)$$

After having defined the functional form of the model, an algorithm is necessary to determine the coefficients vector \mathbf{B} . Generally, this is estimated in the training phase of the model minimising an opportune cost function $J(\mathbf{B})$ defined for the problem.

It is worth noticing that the most commonly used cost functions tend to fit data points, without weighting differently false positives and false negatives which could have different costs. For the scope of this thesis, this is acceptable as an investigation about costs of false positives and false negatives was not available. However, it is recommended that cost functions are, whenever possible, modified in order to take care of such cost differences.

Appendix B provides further details of mathematical formulation of linear and logistic regression, including commonly used cost functions, as well as a description of a gradient descent algorithm which could generally be applied to minimise cost functions.

A statistical p-value is associated to any coefficient of linear and logistic regression. For each coefficient, the p-value tests the null hypothesis that the coefficient is equal to zero (i.e. it has no effect on the model).

Importantly, it is required that the p-value should be clearly reported for every coefficient. A low p-value (e.g. smaller than 0.05) indicates that the null hypothesis can be rejected (i.e. a predictor with a low p-value is likely to be a meaningful addition to the model because changes in the predictor's value are related to changes in the response variable), while a larger p-value implies that changes in the independent variable are not associated with changes in the output. Such p-values are used as ranking criterion in the stepwise regression procedure described in the previous paragraph 3.1.3.

Also, in a linear regression, the coefficient of determination (or R^2) is the proportion of variability in a data set that is accounted for by the statistical model and it varies between 0 and 1. So, the higher is the R^2 , the better is the goodness of fit. There is not a unique definition of R^2 coefficient for logistic regression models. It is then suggested that coefficients of determination are reported at least for linear regression models.

The general procedure described in this section could be generalized to some other techniques, taking care of adapting functional forms, cost functions and minimisation algorithms as appropriate. Adopted functional form, coefficients and possibly cost function (if not standard) should always be clearly reported on research studies detailing developed models.

3.3 MODEL VALIDATION

The aim of a classification task is to associate each element of a dataset to a class amongst a number of possible ones. Regression and logistic regression algorithms (as well as other supervised machine learning techniques) infer a model from labeled training data. The generated model is then evaluated on a separate testing set, which provides an estimate of the accuracy of the model.

A correct estimation of the accuracy of a classifier (in this context, also referred to as *model validation*) is crucial both to predict its future predictive power and to choose among a number of possible classifiers.

Despite the importance of correctly determining the accuracy of a model, as previously outlined in section 2.6, it is estimated that more than 90% of medical logistic regression models among published literature do not use an appropriate model validation strategy.

In order to correctly assess the accuracy of a classifier, an estimation method with low bias (i.e. the difference between real and estimated accuracy) and low variance (i.e. the variability of estimated accuracy when changing the used dataset) is desirable [45] and required by the framework proposed in this work. Three methods are recalled to estimate predictive power of a model: holdout, bootstrap and cross-validation methods.

The *holdout* method divides the available dataset into two partitions, the training and the testing set: the first one is used to build the model and the second one to evaluate its accuracy.

The *bootstrap* method [46] creates, given a dataset of size n , a number of testing sets by sampling n instances uniformly from data with replacements, i.e. some of the data points will appear in the bootstrap sample multiple times while other data points will not appear at all. Called p the probability of any given instance being chosen after n samples, b the number of bootstrap samples, α_i the accuracy computed on the i -th sample, and a the accuracy computed on the whole training set, the accuracy estimate can be expressed as $\frac{1}{b} \sum_i p \cdot \alpha_i + (1 - p) \cdot a$. The variance of the estimate is calculated as the variance of the accuracy estimates for the samples.

The *k-fold cross-validation* method divides the available dataset into k partitions of similar sizes. A model is built k times, using each time $k-1$ partitions as the training set and the remaining partition as the testing set. The estimation of the accuracy model is the average of the accuracies computed for the developed k models. If a dataset is composed of n data points, the n -fold cross-validation is also known as leave-one-out cross-validation. Cross-validation is defined as “stratified” if each partition reflects the distribution among classes of the original dataset.

A largescale experiment on real-world datasets [45] (following experimental results on artificial data and theoretical results in restricted settings) and a recent simulation study [47] compared the performance of these validation methods showing that:

- holdout will generally underestimate the accuracy of the model, hence inducing a large bias, because only a portion of data is available for the

learning process (since the accuracy of a model deriving from supervised learning algorithms increases when more data points are presented during training); on the other hand, using a test set with fewer data points will largely increase the variance of the accuracy estimation

- bootstrap has low variance, however, it may present an extremely large bias problem for both some large and small samples.
- cross-validation provides decreasing bias accuracy estimation while increasing k , reaching a virtually unbiased estimation with the leave-one-out cross-validation where almost all data points are used for training purposes; as k decreases, the accuracy variance increases due to the instability of the used training sets. When compared to regular cross-validation, stratification is generally a better scheme, both in terms of bias and variance

Based on such observations, for the framework hereby developed, stratified cross-validation is proposed for model selection and accuracy estimation due to its lower bias when compared to other proposed methods.

3.4 MODEL EVALUATION

CDSSs are often based on binary classification models, that is their aim is to classify each entry of a given dataset into two groups, *positives* and *negatives*, according to a classification rule. In the clinical domain, the choice of what to

define as positives and negatives is normally not arbitrary: a positive generally indicates an anomalous condition which needs to be addressed, such as the presence of a disease. As a consequence, the cost of missing a positive is generally higher than the cost of wrongly classifying a negative as a positive.

When the outcome of a model is the probability of an entry belonging to a group, such probabilities need to be converted into a boolean value using an appropriate threshold, in order to carry out the classification task. The performance of the developed CDSS models can then be evaluated; for such models, for each element of the dataset there are four possible outcomes with regard to the classification:

- true positive (TP): when a positive is correctly classified
- true negative (TN): when a negative is correctly classified
- false positive (FP), also known as type I error: when a negative is wrongly classified as positive
- false negative (FN), also known as type II error: when a positive is wrongly classified as negative

Different measurements of model performance can be subsequently defined. Given $P = TP + FN$ the number of real positives in the considered dataset, and $N = TN + FP$ the number of real negatives, it is possible to define [48]:

- sensitivity = true positive rate (TPR) = recall = TP/P
- specificity = true negative rate (TNR) = TN/N

- positive predictive value (PPV) = precision = $TP/(TP + FP)$
- negative predictive value (NPV) = $TN/(TN + FN)$
- accuracy = $(TP + TN)/(P + N)$

Such performance measurements are often used in literature. However, they do not properly and satisfactorily characterise the predictive power of a classifier. Indeed, false positive and false negative could have different kinds of implications in different domains, clear examples of which are shown in [49, 50, 51]. As a consequence, the decision threshold used to separate positive and negative outcomes is dependent on the assumed costs for type I and II errors, i.e. false positives and false negatives, and may vary depending on the acceptable trade-off between false positives and false negatives.

However, the predictive power of a model should be evaluated regardless of the chosen cut-off point. For this reason, receiver-operating characteristic (ROC) curves have been adopted for the framework proposed in this work as they provide an index of accuracy by determining the model's ability to discriminate between alternative states of health over the complete spectrum of operating conditions, i.e. different thresholds [52]. A ROC curve can also be thought as a plot of the probability of correctly classifying the positive cases against the rate of incorrectly classifying true negative ones: in other words, for each possible value of the decision threshold, a pair of true-positive and false-positive performance rates are represented on the ROC curve.

In figure 3.1, the diagonal line represents the ROC curve of a random classifier, the angular line shows the performance of an ideal classifier, and the

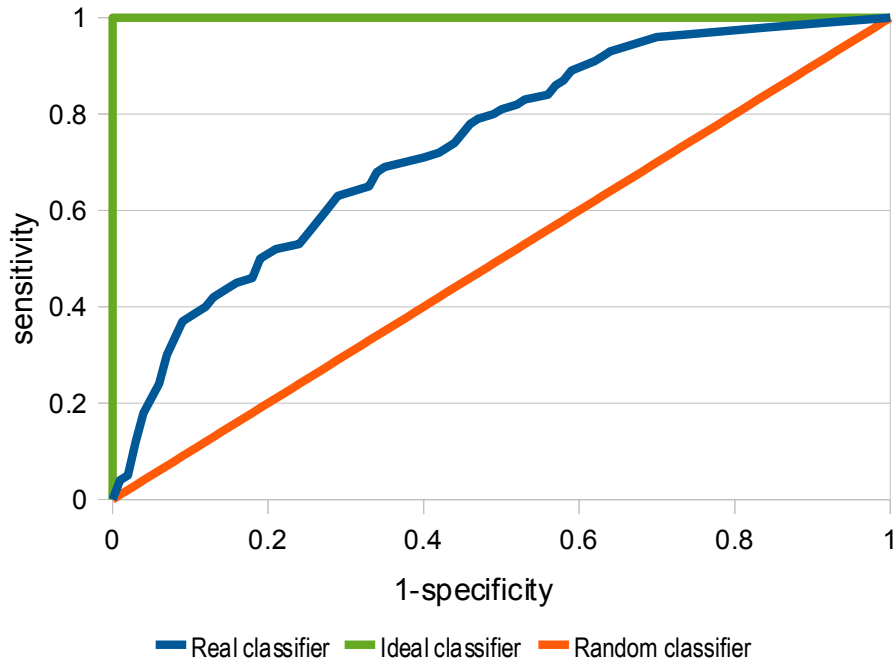


Figure 3.1: Examples of ROC curves

other line corresponds to an example of the ROC curve for an example model. The more the ROC curve tends to be near to the upper left corner, the better the performance is: the area under curve (AUC) is used as a performance metric and it usually varies from 0.5 for a random classifier to 1.0 for an ideal classifier (i.e. 100% of true positive and no false negative are detected).

As previously outlined [28], less than 20% among analyzed studies reported ROC curve analysis (i.e. used area under ROC curve as performance metrics).

Summing up, for classification problems, ROC analysis should always be reported while further performance metrics could be reported if comparison with previous work requires so.

3.5 MISCLASSIFICATION ANALYSIS

So far, some steps to assist the correct construction of a model have been identified: firstly, clinical information is used to identify relevant variables and hypothesize a model functional form; then the available dataset is used to select variables to be included in the model, to calculate coefficients for the model and to validate and evaluate the model itself.

Once that the model is developed, an output (e.g. a risk score, a classification, etc.) can be associated to any previously unseen data point. When using regression methods, this approach is completely transparent: indeed, it is quite easy to identify positive/negative correlations, and their magnitude, between each input of the model and the related output.

However, for classification problems, analyzing how misclassifications are distributed may suggest some insight into the predictive power of the model: it is sometimes possible to identify areas of input/output domains where the model could be unreliable. In this case, different strategies could be adopted: 1) a pre-clustering may be attempted in order to build different models for different areas of input domain; 2) where the model is not accurate enough, the transparency constraint could be relaxed in order to apply a “black-box” model, resulting in a model with improved performance while still reasonably transparent; 3) leaving the classification of such data points to a domain expert is also a possible choice, with the idea of using a decision support system only when it could guarantee reasonable accuracy.

The schematic representation of the proposed process for the model's development is reported in figure 3.2: available data and previous clinical knowledge are used for selecting variables and for building (regression) models; the dataset is then used to calculate the accuracy of the model and to identify areas of low accuracy and different strategies may be used to deal with such areas as previously described.

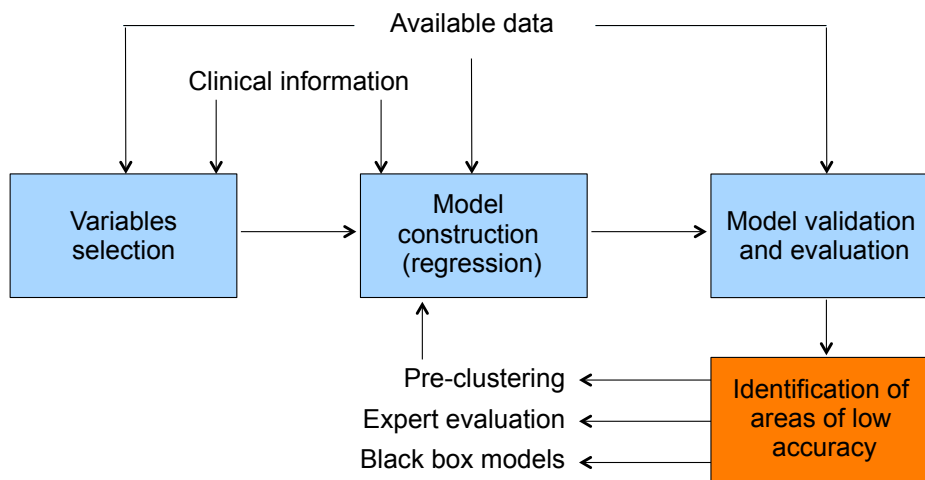


Figure 3.2: Schematic representation of the model building process

For the applications proposed in the following chapter 4 and 5, misclassification analysis is carried out. Because of the relatively small dataset sizes, areas with high misclassification rate will be highlighted, suggesting that such areas are dealt by domain experts.

The framework proposed in this thesis requires that an analysis of misclassification distribution is always carried out, at least when the output of the model is categorical: as will be shown in sections 4.6 and 5.6, such analysis could lead to increased accuracy (not providing predictions for a number of cases) selecting also cases (patients) where specialist prediction is needed.

3.6 IMPLEMENTATION

As previously highlighted, in order to have a positive impact on clinical practice, a CDSS should automatically provide decision support as part of a clinician's workflow at the time and location of decision making. This objective is reached implementing the model into a piece of software or a web application.

The implementation of the model could also be used to collect new data for further research work and/or to be used with an online training algorithm to improve performance of existing models.

The applications proposed in the next chapters have been developed as web-based prototype composed by PHP scripts, able to acquire data input from users and generate an HTML page containing the related output.

CASE STUDY 1: A 28-DAY MORTALITY MODEL FOR ACUTE ALCOHOLIC HEPATITIS

This chapter describes the first developed predictive model: its aim is to predict mortality within 28 days of patients suffering from acute alcoholic hepatitis. Severe forms of alcoholic hepatitis in patients with alcoholic liver disease are in fact associated with high mortality; it is therefore vital to identify those patients at greatest risk of mortality in 28 days as they may benefit from aggressive intervention. Applying the previously described framework, we propose a new predictive model that can be used in clinical practice to identify such patients and to monitor their progress while in hospital.

A cohort of 82 patients was selected and for each of them, a number of clinical findings and standard laboratory tests at the time of admission to hospital were recorded. Variables from currently used scoring systems are collected and, since studies have shown the usefulness of repeating scoring systems after one week of admission to predict outcome, some variables were collected up to 7 days after admission. It is expected that an appropriate variable selection approach could lead to a better accuracy of the developed model when compared with existing models, and yield new and potentially useful clinical insights.

The proposed logistic regression model selected four statistically significant predictors (namely, the level of creatinine on and after admission, the presence of encephalopathy and prothrombin time evaluated after admission). A comparison with the available mortality predictive scores showed an increase by 25% in predictive power, demonstrating increased accuracy in identifying these sick patients with alcoholic hepatitis in clinical practice.

4.1 BACKGROUND

Severe forms of alcoholic hepatitis (AH) in patients with alcoholic liver disease (ALD), characterized by jaundice, hepatocellular damage and fibrosis [53], are associated with high mortality especially in a younger population, but is also frequent in older people, who are more susceptible to the effects of excessive alcohol consumption [54]. It is estimated that severe alcoholic hepatitis has a death rate of up to 50% [55]: therefore, it is vital to be able to identify patients at greatest risk of mortality and in whom the therapeutic benefit/risk ratio is unfavorable as this group of patients may benefit from aggressive intervention.

Multiple prognostic factors have been studied over the last decade and a variety of scoring systems are used in clinical practice to assess the severity of acute AH. They include, Maddrey Discriminant Function (mDF) [56], Child-Pugh Score (CPS) [57, 58], Glasgow Alcoholic Hepatitis Score (GAHS) [59] and Model for End Stage Liver Disease (MELD) [60]. A recent study [61] has compared various prognostic scores used to evaluate the short-term mortality

in patients with acute-on-chronic liver failure. In these scoring systems a combination of various laboratory and clinical parameters are analyzed to determine the severity of AH in patients with ALD. The mDF was derived from observations of 55 patients: a stepwise discriminant analysis revealed a statistically significant association with death of prolongation of prothrombin time and height of serum bilirubin on admission to hospital. In the CPS, five variables (namely: bilirubin, serum albumin, prothrombin time, presence of encephalopathy and ascites) are used to predict the prognosis of chronic liver disease.

The GAHS was built using a stepwise logistic regression algorithm and was derived from few variables which were significant in respect of the patient's death risk: namely, age, serum bilirubin, and blood urea to predict 28-day mortality risk, and serum bilirubin, prothrombin time, and peripheral blood white blood cell count to predict 84-day mortality risk. Variables for the first model (28-day mortality) were observed on day 1, while for the second one (84-day mortality) they were measured on day 6-9.

In addition, various studies have suggested that a few individual parameters such as elevated serum bilirubin level, international normalized ratio of prothrombin time, white cell count and hepatic encephalopathy are associated with increased risk of mortality in patients with AH [62, 63].

Predictive factors of long-term survival in ALD have recently been studied [64]: only clinical factors were statistically significant prognostic factors while no histological features correlated with prognosis. Non-invasive methods are

also increasingly used to predict the prognosis of patients with chronic viral hepatitis, reducing the need for liver biopsy analysis and facilitating the better management of such patients [65].

A previous study [66] has shown that an increase of MELD score in the first week is an independent predictor of mortality. Consequently, the present study also considers the values of some selected variables evaluated days after admission to hospital, i.e. bilirubin, urea, creatinine and prothrombin time, were assessed also on following days up to day 7 after admission or to the time of death if it occurs within 7 days, and included in the pool of variables available for the model.

4.2 AIMS

The aims of this study are to propose a new enhanced predictive model which could increase the accuracy of the currently available tools, identify individual parameters that are associated with increased mortality and to assess if changes in the values of these variables after treatment were associated with bad prognosis in patients with AH. It is clinically crucial to identify patients with severe AH by using these individual parameters so that their treatment can be commenced and possibly adjusted early to reduce mortality. Moreover, the possible introduction of variables - re-evaluated days after admission to hospital - provides a way to constantly monitor the efficacy of administered treatments.

4.3 DATASET PREPARATION

4.3.1 *Study sample*

Patients with AH admitted to Crosshouse Hospital, Kilmarnock (UK) between April 2003 and July 2008 were retrospectively analyzed. The diagnosis of chronic ALD was made on the basis of a history of excessive intake of alcohol over a period of several years and by the exclusion of other causes of chronic liver disease either by blood tests or liver biopsy. In addition, the diagnosis of acute AH in these patients was made by the presence of one or more complications of chronic liver disease such as sepsis, upper gastro-intestinal bleeding, encephalopathy, hepato-renal syndrome and serum bilirubin level of more than 80 $\mu\text{moles/L}$ on admission.

Patients were excluded if they had evidence of co-existing viral hepatitis, auto-immune hepatitis, hepatocellular carcinoma or biliary obstruction. In addition, patients who died within 28 days of admission due to non-hepatic complications were also excluded from the study. The final dataset has 82 patients: 45 patients were still alive after 28 days of admission, while the remaining 37 patients succumbed to various complications of chronic liver disease. The study was approved by Clinical Effectiveness Department of Crosshouse Hospital, Kilmarnock (UK).

4.3.2 *Candidate variables*

For each patient, a number of clinical findings and standard laboratory tests at the time of admission were recorded, specifically:

- the age at acute diagnosis (expressed in years)
- aspartate aminotransferase (expressed in unit/L)
- alanine aminotransferase (expressed in unit/L)
- alkaline phosphate (expressed in unit/L)
- bilirubin and creatinine (expressed in $\mu\text{moles/L}$)
- albumin (expressed in g/L)
- prothrombin time (expressed in seconds)
- white cell count (expressed in billion/L)
- urea (expressed in millimoles/L)

A number of binary variables were also collected:

- gender of patient
- presence of sepsis
- presence of encephalopathy
- presence of upper gastrointestinal bleeding

In addition, serum bilirubin level, urea, creatinine and prothrombin time were recorded on day 7 of admission or at the time of death if a patient died within 7 days. Patients were graded according to mDF, CPS, GAHS at the time of admission.

Collinearity diagnostics have been run by means of the appropriate function of PASW Statistic (SPSS), returning tolerance and VIF values: criteria described in paragraph 3.1.2 are satisfied for all variables and then no collinearity issues are detected.

4.4 MODEL CONSTRUCTION

4.4.1 *Logistic regression*

For the purposes of this study, a logistic regression model was developed including variables selected with a backward selection algorithm among all the available ones (as detailed in the next paragraph 4.4.2); the performance of this model has been compared with the available mDF, CPS and GAHS models, which were evaluated on admission.

4.4.2 *Variables selection and model validation*

As discussed in paragraph 3.1.3, including extra variables in a predictive model should, in principle, increase the accuracy of the model itself (or, at

least, not decrease it). However, since we are building a predictive model based on a limited sample, by adding more variables, we face the risk of overfitting the dataset, posing some serious limitation to the scalability and the generalization ability of the model (i.e. performance will be sensibly lower using a bigger/different dataset). To overcome this problem, we have used a backward stepwise algorithm, as prescribed by the proposed framework.

We have then evaluated the accuracy (based on a threshold of 0.5) of logistic regression models using a 10-fold cross validation strategy, starting from a model including all considered variables and ending with a model with a single predictor. The number of variables to be considered for our final model corresponds to the case with the best accuracy. Also, should the best accuracy be reached in models with different number of variables, the one with the smallest number of variables will be considered, assuming that collecting less variables provides a more time and cost efficient approach.

4.4.3 *Model evaluation*

To quantify the performance of the proposed model compared to state-of-the-art models, such as mDF, CPS and GAHS, the Receiver Operating Characteristic (ROC) curves were used (evaluating the underlying area), which compare the specificity and sensitivity of a model without regard to the chosen threshold used to discriminate predicted outcomes [52].

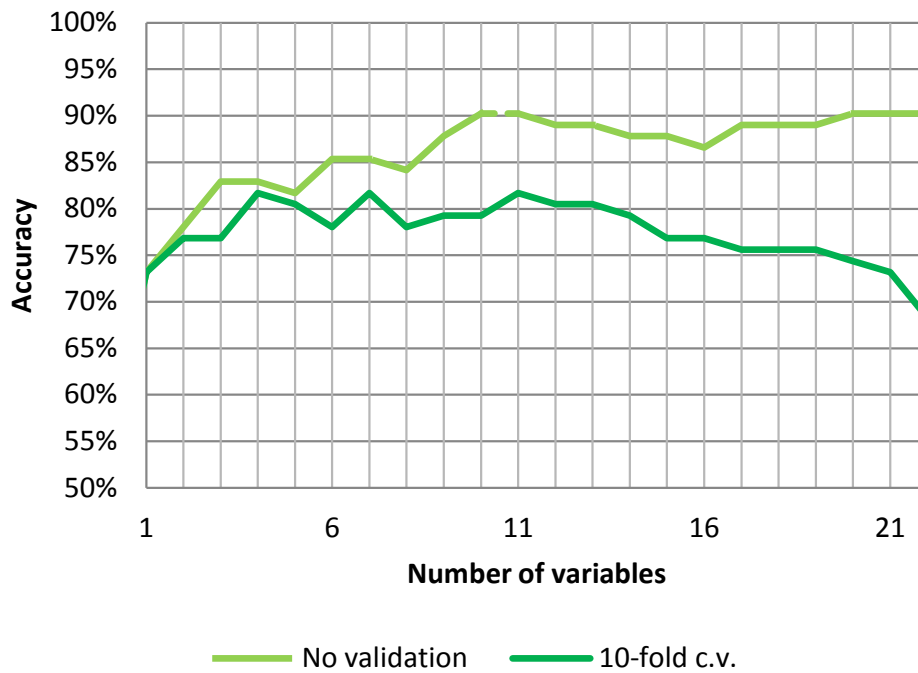


Figure 4.1: Comparison of accuracies with/without validation strategy

4.5 RESULTS

A logistic regression model was developed starting from all available variables, also including the ones recorded at day 7 (or at death if this occurs before day 7). We used a backward stepwise logistic regression algorithm to evaluate the optimal number of variables given the used dataset in order to avoid overfitting. The performance of each model, comprising a different number of variables, has been reported in the following figure 4.1, comparing the accuracy calculated using the whole dataset, and the accuracy implementing a 10-fold cross validation strategy.

From the analysis of the graph, while the best accuracy of the model with no validation strategy is reached when considering at least 10 variables, the best

accuracy of the model using a 10-fold cross validation approach is reached by including 4 variables in the model (the same accuracy is reached also including 7 and 11 variables, however it is believed that a model with fewer variables is cheaper to implement).

Then, our final model included four significant variables: namely, the level of creatinine on admission (CR) as well as after admission (CR7), the presence of encephalopathy (ENC, dummy variable) and prothrombin time evaluated after admission (PT7).

The resulting model is reported in the following formula 4.1 where the score, between 0 and 100, predicts the probability of mortality at day 28.

$$\text{SCORE} = 100 \cdot (1 + e^{-M})^{-1} \quad (4.1)$$

where

$$M = 0.046 \cdot \text{CR7} - 0.022 \cdot \text{CR} + 0.159 \cdot \text{PT7} + 1.390 \cdot \text{ENC} - 6.303 \quad (4.2)$$

A detailed analysis of the model coefficients is provided in table 4.1, where for each significant variable the coefficient, the standard error and the p-value deriving from logistic regression are tabulated.

A comparison of the different scoring systems based on variables collected on admission is tabulated on table 4.2: the performance of the three scoring systems as well as our proposed predictive model in these patients was individually analyzed using Student's t-test, to check if a significant difference exists between the average score for patients who died and patients who survived (low p-values confirm this hypothesis), and by ROC curve analysis

variable	coefficient	std. error	p value
CR	-0.022	0.010	0.033
CR7	0.046	0.013	<0.001
PT7	0.159	0.070	0.023
ENC	1.390	0.670	0.038
Constant	-6.303	1.630	<0.001

Table 4.1: Analysis of coefficients from logistic regression model

to evaluate the global predictive power (measured by area under ROC curve, indicated as AUC) of each model.

Scoring system	Alive after 28 days (n=45)	Died within 28 days (n=37)	p-value	AUC
mDF	37.2 ± 26.2	67.5 ± 56.9	<0.01	0.705
CPS	10.8 ± 1.6	11.8 ± 1.4	<0.01	0.681
GAHS	7.6 ± 1.6	8.7 ± 1.6	<0.01	0.687
Proposed model	24.5 ± 23.7	70.3 ± 30.7	<0.001	0.873

Table 4.2: Comparison of scoring systems in patients with AH, who died and survived

4.6 MISCLASSIFICATION ANALYSIS

The objective of this section is to determine whether the model could be unreliable under certain conditions. We carry out an a posteriori analysis of how misclassifications are distributed along a specified domain: in our case, we choose to study the distributions of predictions and misclassification as a function of the model output. Operatively, the entire range of outputs has been divided into 10 categories uniformly distributed between the minimum and maximum values of obtained output values: considering SCORE (as defined per formula 4.1) as model output, the interval to be divided into 10 categories is $[0, 1]$; considering M (as defined per formula 4.2) we focus on the interval identified by the extreme model output given using the considered dataset. For this analysis we have chosen a representative threshold of 0.5.

interval min	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
interval max	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
% misclassifications	3.3	16.7	0.0	0.0	50.0	66.7	0.0	14.3	0.0	5.0

Table 4.3: Distribution of misclassifications (function of SCORE/100)

interval min	-11.5	-9.0	-6.5	-3.9	-1.4	1.1	3.7	6.2	8.7	11.3
interval max	-9.0	-6.5	-3.9	-1.4	1.1	3.7	6.2	8.7	11.3	13.8
% misclassif.	4.5	0.0	14.3	50.0	9.1	0.0	50.0	0.0	0.0	0.0

Table 4.4: Distribution of misclassifications (function of M)

From the analysis of figure 4.2, it is clear that model predictions are concentrated in the first (close to 0) and last (close to 1) selected intervals; also, the proportion of misclassifications in such intervals are quite low as shown in table 4.3.

On the other hand, we note that there is a region (approximately between 0.4 and 0.6) where the ratio between misclassification and total predictions is greater than 50% suggesting that the developed model is not reliable at such an interval of output.

The same results are shown on figure 4.3 and the related table 4.4, where misclassifications are concentrated mainly at interval $[-3.92, -1.39]$. Excluding such interval (where 11% of predictions are concentrated), and assuming a threshold of 0.5, the accuracy increases from 88.7% to 93.7%.

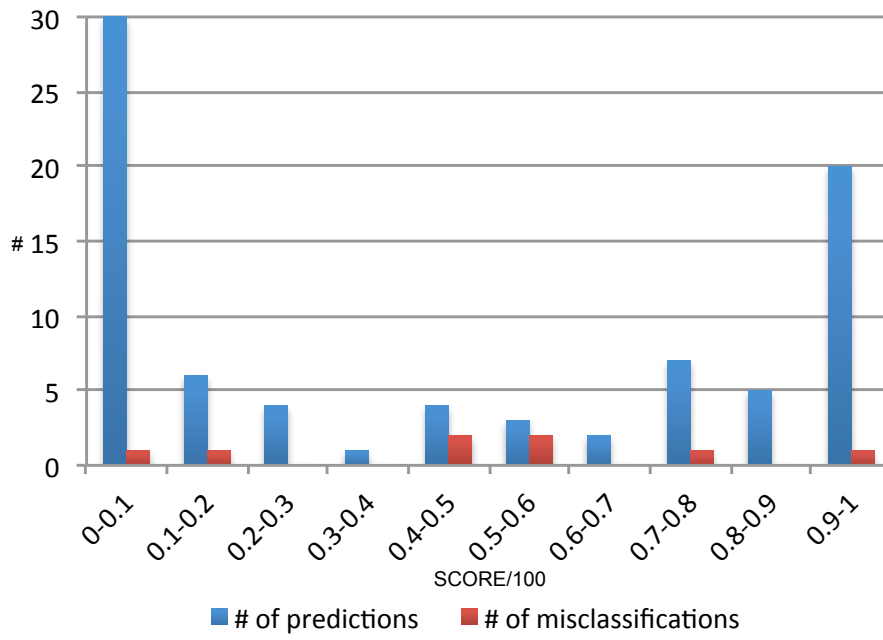


Figure 4.2: Distribution of predictions and misclassifications (function of SCORE/100)

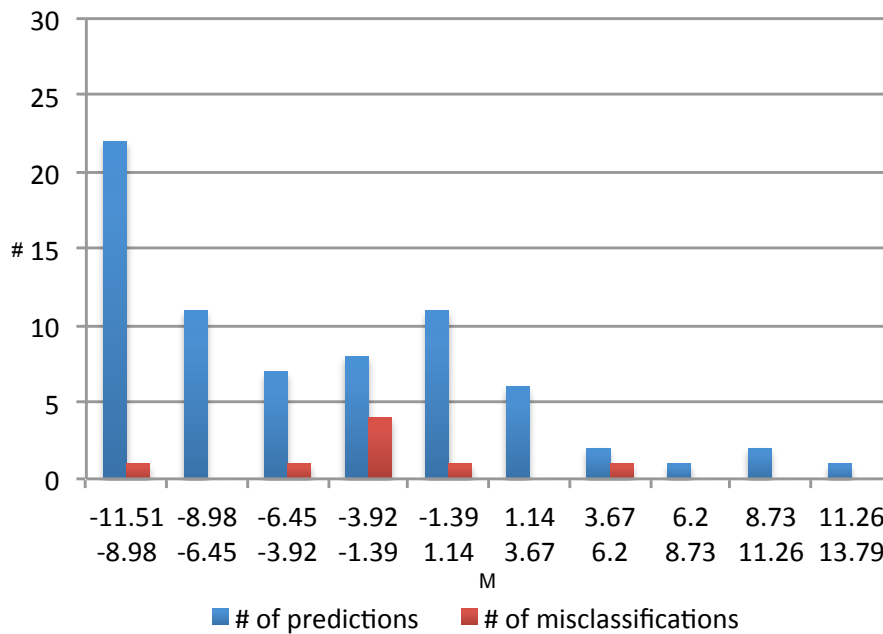


Figure 4.3: Distribution of predictions and misclassifications (function of M)

This information may be used to refine the model: for example, it is possible to design and consider the output of a “second step” model if the output of our proposed model is in the identified areas. Given the size of our dataset, we prefer at this stage to simply warn the user that the model could be unreliable when the output is in such an interval.

For comparison purposes, the analysis of misclassifications distribution has been carried out also for mDF model: the entire range of outputs has been divided into 10 categories distributed between the minimum and maximum values of obtained output values. Since only 3 data points had an output in the second half of the selected interval, they have been considered “outliers” and grouped in the tenth category. For this analysis we have chosen a threshold of 32 (giving an overall accuracy of 63.4%, close to the maximum attainable accuracy of 64.6% using a different threshold), as suggested by relevant literature [67]. Results are reported on figure 4.4 and on table 4.5.

Unlike the previous case, from the analysis of misclassifications it is not possible to isolate a single area in the output space where accuracy is definitely lower than the overall accuracy: a score between 32.8 and 74.5 (where 34% of predictions are concentrated) suggests that the model can be unreliable. Excluding such interval, the accuracy increases to 72.2%.

The comparison between mDF and the proposed model confirms that misclassification analysis has the potential to increase the performances of developed models. However, costs and benefits of strategies implemented to

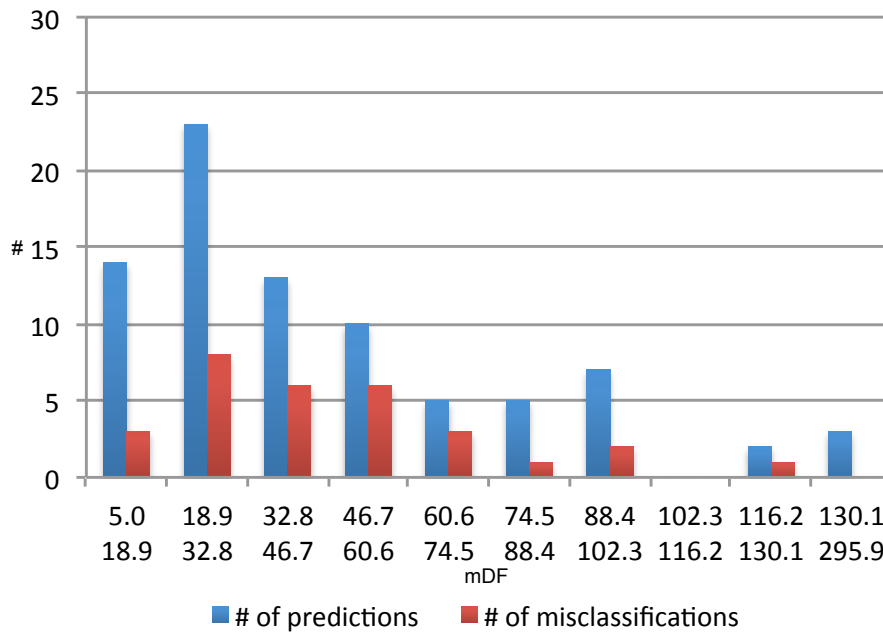


Figure 4.4: Distribution of predictions and misclassifications for mDF

interval min	5.0	18.9	32.8	46.7	60.6	74.5	88.4	102.3	116.2	130.1
interval max	18.9	32.8	46.7	60.6	74.5	88.4	102.3	116.2	130.1	295.9
% misclassif.	21.4	34.8	46.2	60.0	60.0	20.0	28.6	n/a	50.0	0.0

Table 4.5: Distribution of misclassifications for mDF

reduce misclassifications are dependent on the considered model and possibly on the used dataset.

4.7 IMPLEMENTATION

The developed model has been delivered as web-based prototype. The core of the developed prototype is a PHP script, which generates an HTML page


after data is collected from a HTML input form and then POSTed when user presses the "Calculate score" button.

This prototype, available at <http://bit.ly/ald-calc>, implements the mortality model for acute alcoholic hepatitis developed in the previous sections of this chapter, called *Stirling ALD mortality predictor (SAMP)*. It is intended to be used after 7 days from admission in hospital.

The user is asked to insert the level of creatinine at admission (CR) and on 7th day (CR7) in micromoles per litre, the prothrombin time (PT) in seconds and the presence of encephalopathy (ENC). The user interface is reported in figure 4.5.

The SAMP score is calculated using formula 4.2, and probability of mortality after 28 days from admission is computed by formula 4.1. An equivalence between score and probability is reported for specified significant values. An example of the output is reported in figure 4.6.

As discussed in the previous section 4.6, the user is warned about the unreliability of the model when the SAMP score is between -3.92 and -1.39.



**UNIVERSITY OF
STIRLING**

SCHOOL OF
NATURAL SCIENCES

Stirling ALD Mortality Predictor (SAMP)

Mortality risk model after 28 days from admission in hospital for patients suffering from alcoholic liver disease during acute hepatitis

Severe form of alcoholic hepatitis in patients with alcoholic liver disease is associated with high mortality; it is therefore vital to identify patients at greatest risk of mortality as they may benefit from aggressive intervention. This new predictive model, which uses four statistically significant predictors, could be used in clinical practice to identify such patients. The comparison with the available predictive scores showed an increase of 25% predictive power, demonstrating increased accuracy in identifying these sick patients with alcoholic hepatitis.

1. Level of creatinine at admission: micromoles per litre


2. Level of creatinine on 7th day: micromoles per litre

3. Prothrombin time: seconds

4. Is the patient suffering from encephalopathy?
 Yes No

Developed by [Thomas Mazzocco](#) and Amir Hussain, University of Stirling. All rights reserved.
Pilot prototype provided "as is" without any warranty.

Figure 4.5: Input screen for ALD mortality predictor



**UNIVERSITY OF
STIRLING**

SCHOOL OF
NATURAL SCIENCES

Stirling ALD Mortality Predictor (SAMP)

Mortality risk model after 28 days from admission in hospital for patients suffering from alcoholic liver disease during acute hepatitis

SAMP score: **-1.149**

The probability of death within 28 days from admission is about **24%**

A score of -4.6 or 4.6 corresponds respectively to a probability of death of about 1% or 99%.
A score of -2.9 or 2.9 corresponds respectively to a probability of death of about 5% or 95%.
A score of -2.2 or 2.2 corresponds respectively to a probability of death of about 10% or 90%.

[Back](#)

Developed by [Thomas Mazzocco](#) and Amir Hussain, University of Stirling. All rights reserved.
Pilot prototype provided "as is" without any warranty.

Figure 4.6: Output example for ALD mortality predictor

4.8 DISCUSSION

In the developed CDSS, we first determined the optimal number of variables, selecting four relevant variables: after applying a cross-validation strategy, as recommended in section 3.3, we had evidence that there is no benefit (in terms of model performance) in including more than 4 variables in the model, so we chose the smallest number of variables which can guarantee the highest predictive power. It has to be noted at this stage that, while the selected variables are definitely significant in predicting the outcome, nothing could be certainly stated about the excluded variables: the size of the considered dataset prevented us from including more variables in the model, as this could have led to an unreliable model. Further studies with datasets including more patients are now required to investigate if other variables may further improve the developed model as well as to validate the proposed model.

Previous studies [68, 66] have shown the usefulness of repeating scoring systems after one week of admission to predict outcome. Some variables recorded after admission have been included in the model: two variables (namely, creatinine level and prothrombin time) out of the four selected for the model were recorded after admission, confirming that a second evaluation of these parameters provided significant predictors for calculating 28 day mortality risk. Most of the patients with severe AH died within 7 days of admission (20 out of 37), as severe AH is associated with early mortality. The significance of re-evaluating these parameters is to continuously assess

effectiveness of treatment commenced on admission: the proposed score can be calculated up to 7 days after admission. If the score is improving then it is worth continuing treatment, otherwise depending on the general condition and other co-morbidities of the patients, a decision is taken either to withdraw the treatment or escalate to other major interventions such as liver transplantation.

In most studies (such as [56, 59]) 28-day survival is used to predict long term outcome in these patients: the comparison of the proposed score calculated for patient alive after 28 days and dead within 28 days highlighted a statistically significant relationship between the output of the proposed model and the actual mortality ($p < 0.001$).

While creatinine level and prothrombin time were already known to be predictors of 28 day mortality in patients with AH, and indeed they were also included in currently used predictive models, encephalopathy was not used as its assessment is very subjective especially at an earlier stage. However, since the presence of this condition has been selected as one of the four most significant predictors among all collected variables, it is believed that such variables should be included in the proposed model.

The comparison with the available predictive scores evaluated at admission showed an increase of 25% in predictive power (evaluated as suggested in section 3.4), suggesting that the inclusion in the model of variables evaluated after admission and the subsequent use of an appropriate variables selection procedure, as detailed in paragraph 3.1.3, may significantly improve the performance of such models.

Finally, according to the procedure described in section 3.5, an a posteriori analysis about misclassifications distribution was carried out: this identified a region in the output domain where the developed model is unreliable and its usage is discouraged.

4.9 CONCLUSION

Significant number of patients with AH succumb to early death because of hepatic inflammation and its complications, therefore early identification of these patients with the view of intensive intervention and management can lead to enhanced survival through the use of steroids or intensive treatment of complication of AH [69].

The proposed model combines for the first time variables collected at admission and up to 7 days after admission in order to build a more accurate prediction of mortality for patients during AH episodes. Also, the presence of encephalopathy has been included for the first time in the number of predictors. The final model, which includes four variables selected according to the best practices identified in paragraph 3.1.3, has shown a noticeable improvement in considered performance metrics outperforming the gold standard benchmark models.

In conclusion, the capability of employing logistic regression based mortality prediction model for patients with AH has been ascertained and a potential

impact is envisaged from this work both for clinical practice and further research.

A future large-scale study is required to clinically validate the results from this study and to assess the feasibility of deploying such a model in real clinical practice, additionally yielding new and potentially useful clinical insights (e.g. through the analysis of considered and excluded variables).

CASE STUDY 2: A PREDICTIVE MODEL TO AID THE DIAGNOSIS OF DEMENTIA

This chapter describes the second case study used to validate the proposed framework. The developed predictive model is designed to aid the diagnosis of early dementia, as clinicians often experience difficulties in such diagnosis due to the intrinsic complexity of the process and lack of comprehensive diagnostic tools.

Different models have been proposed to provide medical decision support in dementia diagnosis. The aim of this study is to improve on the performance of a recent application of Bayesian belief networks using an alternative approach based on logistic regression.

A pool of 14 variables has been evaluated in a sample of 164 patients suspected of dementia. First, a logistic regression model for dementia prediction is developed using all variables included in the previous model; then, a second model is built using a stepwise logistic regression starting with all collected variables and selecting the pool of the relevant ones. A range of performance metrics have been used to evaluate the developed models.

The new models have resulted in very good predictive power, demonstrating general performance improvement compared to a state-of-the-art prediction

model. Interestingly, the approach based on statistical variables selection outperformed the model which used variables selected by domain experts in the previous study.

5.1 BACKGROUND

Diagnoses of the common dementias of old age are operationally defined on the basis of different symptoms and neuropsychological profiles; in this process, clinicians use various sources of evidence in the reasoning process, which include evidence-based clinical guidelines, often supplemented by individual consultations [70].

General practitioners play a pivotal role in establishing diagnosis of dementia and in providing ongoing support and intervention. Nonetheless, substantial literature [71, 72] shows their difficulties in fulfilling this role, especially for early detection of dementia.

Since one of the reasons for this is the lack of readily available diagnostic instruments [73], efforts have been made to explore different screening measurements and methods.

Different models have been proposed to provide decision support for the diagnosis of dementia. An application of Support Vector Machines (SVMs) has been studied in a study [74] which compared a fully automated computer-based diagnosis system for dementia pathologies using neuroimages with the diagnostic classification made by six radiologists with different levels of

experience. Three different datasets were considered for comparative analysis between SVMs and radiologists; in the first two sets, the task was to detect sporadic Alzheimer's disease: SVMs correctly classified 95% of the cases versus the 65% to 95% of radiologists in the first set while in the second set SVMs showed an accuracy of 93% compared to the 80% to 90% scored by radiologists. Finally, the SVM was asked to separate patients with sporadic Alzheimer's disease from those with frontotemporal lobar degeneration: SVMs scored an accuracy of 89% versus the 63% to 83% of radiologists. These results are not directly comparable with the model object of the present study (whose aim is the diagnosis of generic dementia conditions), but they do demonstrate the potential of machine learning techniques applied to this medical domain (often outperforming specialists' predictions), showing also that no prior knowledge is required to be included in the model in order to construct a good prediction tool.

Another study [75] built a computer-based model which could provide a dementia diagnosis according to DSM-IV criteria formulated in 1994 by American Psychiatric Association and to the 10/66 dementia survey [76] in order to make a comparison between them. While the DSM-IV criteria are regarded as a gold-standard and used as a benchmark, the 10/66 dementia diagnostic algorithm takes into account a structured clinical mental state interview, a cognitive test battery, different informant interviews, a neurological assessment and a questionnaire to detect behavioural and psychological symptoms: the results from this screening tests are then used to make a prediction on dementia

diagnosis by means of a logistic regression equation previously developed in a pilot study. The tool has shown a sensitivity of 57.8% and a specificity of 98.3% using the implementation of the DSM-IV criteria; a sensitivity of 93.2% and a specificity of 96.8% was recorded using the 10/66 criteria. Although the very different number and type of collected variables prevent us from a comparison with the results of the present study (where a limited number of variables is collected), however, this application shows that a model based on logistic regression (and so without a prior built-in knowledge) could outperform the gold-standard guidelines.

A novel application (called DemNet) of Bayesian Belief Networks (BBN) to provide medical decision support for the diagnosis of dementia in primary care practice has been recently proposed [77]; this was built as a "hand-crafted" model, since it relies on domain experts' knowledge.

A BBN is a representation of complex domains that are characterised by uncertainty which enables inference of future uncertain events based on prior related known events. More formally a BBN includes a set of nodes (or vertices) that represent the domain variables, a set of directed edges which represent dependency relationships between the variables and a set of local probability tables, one table per variable, which quantitatively encodes the strength of each dependency relation [78]. Standard benchmarks have shown that good results are achieved in the diagnosis of dementia using BBN. The developed model was implemented in a software system designed to be used by clinical practice nurses involved in the primary level assessment of patients suspected

of having dementia; in an attempt to optimise user friendliness and utility in a busy primary care setting, the model sought to use numerically few parameters which are consistent with a reasonably high diagnostic accuracy [79].

The setting of this last application is quite agile (given the small amount of information required to make a prediction) when compared with the previously cited studies. However, they required both technical expertise and domain experts' knowledge for building the more sophisticated and complex Bayesian belief network model.

5.2 AIMS

The present study builds on the abovementioned previous research and its aim is to design, through an appropriate variables selection, a novel model that can outperform the considered benchmark hand-crafted model. Consequently, a performance improvement is expected using the same dataset and a common set of performance metrics.

5.3 DATASET PREPARATION

5.3.1 *Study sample*

The collection of data has been carried out as part of a previous research work [80]: 164 patient records from clinical practice were obtained from the Community Mental Health Team Elderly (CMHTE), Kildean Hospital, Stirling. A clinical protocol detailing the data requirements was developed and the necessary governance process was followed. Local Ethical Research Council approval was granted. Community Psychiatric Nurses (CPNs) from CMHTE agreed to collect the data, as it aligned with the diagnostic variables that they recorded during initial assessment of patients where dementia was suspected. Each completed record consisted of the CPNs initial assessment, as well as the actual diagnosis provided by a CMHTE diagnosing physician. It is worth noting that the data regarding 50 out of 164 patients were not fully provided: at least one value from collected variables was missing; when possible (i.e. there are no missing data for selected variables), records with missing data will be used in this study.

5.3.2 *Collected variables*

A set of 14 parameters has been evaluated within the considered sample of patients: specifically, 3 variables from standard tests (Mini-Mental State

Examination, Hachinski Ischemic Score and Clock Drawing Test), 8 qualitative variables (investigating the ability to carry out personal and domestic activities of daily living, the current and subtle functioning, the global severity and the possible presence of psychosis, memory impairment or tremors) and 3 variables about patients' clinical condition and history (age of each patient, duration of symptoms and whether they experienced a clear progression in symptoms) were collected. Collected variables from standard test are:

- cognitive impairment (CI), which represents the result of the Mini-Mental State Examination (MMSE), used for detection of dementia in individuals with suspected cognitive impairment [81], measured as the ratio between the score of the test and the maximum attainable score
- clock drawing test (CDT), which is used combined with the result of MMSE in screening for mild dementia [82]; the result of the test is mapped as a dummy variable which states if the patient could successfully complete the test or not
- Hachinski Ischemic Score (HI), which is generally used to discriminate Alzheimer's disease from vascular dementia [83]; the result of the test has been normalized between 0 and 1

Qualitative variables which have been collected are:

- domestic activities of daily living (DADL), which reflects the individual's ability to carry out activities such as shopping, housekeeping, finance management, food preparation and transportation

- personal activities of daily living (PADL), which reflects the individual's ability to carry out activities such as dressing, eating, ambulating and hygiene
- current functioning (CF), which reflects the individual's ability to function in daily life aggregating PADL and DADL
- subtle functioning (SF), which captures evidence of subtle changes in cognition such as progressive difficulty in balancing a cheque book
- global severity (GS), which represents the global severity of impairment; aggregates CF, CI and SF

These five categorical variables can represent three levels of impairment: severe, mild, none. Each of them has been mapped with two dummy variables, according to the procedure described in paragraph 3.1.1, the first one showing if at least a mild impairment was present (adding the suffix "m" to the variable name) and the second one stating if a severe impairment was present (adding the suffix "s").

The other recorded qualitative variables are:

- psychosis (PS), which captures non-cognitive symptoms associated with dementia such as impaired connection to reality, auditory or visual hallucinations and delusions; the variable is mapped with two dummy variables: the first one showing if symptoms are at least equivocal (PSe) and the second one stating if they are definitely present (PSp)

- memory impairment (MI), which records the possible presence of memory impairment (binary variable)
- tremors (TR), which records the possible presence of tremors (binary variable)

Finally, some data about history and clinical conditions of patient are recorded:

- age (AG), which records the age of patient expressed in years
- duration (DR), which records the duration of symptoms mapped with two dummy variables: the first showing at least a medium duration (DRm) and the second one a long duration (DRl)
- clear progression (CP), which records a clear progression in symptoms (binary variable)

5.4 MODEL CONSTRUCTION

5.4.1 *Logistic regression*

A logistic regression has been carried out to determine the vector of coefficients which in the considered dataset could associate each record (a patient) with the probability of experiencing dementia. For the purpose of this work, PASW Statistic (SPSS) has been used for the analyses. As required in paragraph

3.1.2, preliminary statistical tests were carried out in order to identify possible collinearity between variables.

5.4.2 *Variables selection*

Since the previous work - DemNet - initially made an expert-driven variables selection, logistic regression has been carried out in this study in two stages. In the first stage, only those variables involved in dementia prediction in the previous research were employed with the aim of investigating whether logistic regression could outperform the previous model. In the second stage, as recommended in paragraph 3.1.3, all available variables are used to carry out a backward stepwise logistic regression in order to select the most significant ones: the aim of this second step was to check whether the optimal subset was selected or if some of them need to be included or removed. This second model could possibly improve the global model performance and/or select a restricted pool of variables which could give comparable performance while requiring collection and elaboration of a smaller number of variables.

5.4.3 *Model validation*

As required by the proposed framework in section 3.3, the predictive power of the model has been assessed across different samples (using stratified cross-validation). The dataset has been split into three parts (evenly distributing

positive and negative outcomes) and two out of three thirds are cyclically selected to run the logistic regression and the remaining third is used to test the results: in this way, the whole dataset is used once to test the model.

5.4.4 *Model evaluation*

As prescribed by section 3.4, the area under ROC curve is used for this study as a performance metric. True positive/negative rates as well as positive/negative predicted values are also reported. Also, accuracy computed using a threshold of 0.5 is reported in order to allow comparison with previous studies.

5.5 RESULTS

In order to verify the presence of any candidate predictor which provides incomplete information with regard to the outcome, a cross-tabulation of all categorical variables has been built showing that CFm and GSm present only 0 and 1 cases respectively of patients affected by dementia who have these two variables equals to zero. For this reason, these two variables could make the model unreliable and this would be reflected in the abnormal standard deviation of the computed coefficients. These two variables have been dropped; however, since in the considered dataset they were combining effects from other variables, this removal is not expected to significantly decrease the model performance.

Collinearity diagnostics have been run by means of the appropriate function of PASW Statistic (SPSS), returning tolerance and VIF values. Since criteria mentioned in the previous paragraph 3.1.2 are satisfied for all variables, they appear not to be collinear.

The used dataset presents an uneven class distribution, since patients with dementia are 135 out of 164 and only the remaining 29 are not affected by this pathology. This means that the class distribution is around 1:5 and, in principle, this does not seem to be a problem for building a reliable classifier [84]. A check can be carried out in due course to verify that the dataset has a sufficient number of events per variable: as mentioned in paragraph 3.1.3, a rule of thumb suggests that the number of the less common of the two possible outcomes divided by the number of predictor variables should be at least 10; for a given set of data, introducing more variables will generally produce a better fit to the data but an excessive number of variables may overfit the dataset, leading the model to lose its generalisation power.

In the first stage, logistic regression has been carried out using only the variables employed in the previous model (DemNet) to investigate if this technique could provide an improvement.

The previous model used PADL, DADL, CF, CI, CDT, SF, GS, AG, DR, CP. Although variables like CI and AG are continuous and dividing them into subclasses could cause a consequent loss of information, in order to have the fairest possible comparison, for the purpose of this first logistic regression study, these variables have been divided as in the original DemNet model: CI

is thus mapped to two dummy variables, namely CI1 and CI2, denoting if the patient could score at least one or two thirds of the maximum attainable score of the MMSE test. AG is now mapped to 3 dummy variables, namely AG1, AG2 and AG3, representing if the patient is more than 64, 74 or 84 years old respectively. According to the new variable mapping scheme, the used variables are: PADLm, PADLs, DADLm, DADLs, CFs, CI1, CI2, CDT, SFm, SFs, GSs, AG1, AG2, AG3, DRm, DRL, CP.

Following cross-tabulation, it is seen that AG1 assumes the value 0 only three times within all records and never within patients without dementia; for this reason, this variable has been dropped. The result of the regression analysis is tabulated in the table 5.1.

The considered dataset has 164 observations, but only 29 of them are cases of patients without dementia; so the ratio between number of events and considered variables (16) for this particular dataset is $29/16 = 1.813$. Since this value is quite low the risk of overfitting data could be high: for this reason and in order to validate and test the new model, an appropriate validation strategy has been set up as previously discussed.

The accuracy (percentage of correct predicted outcomes, based on a threshold of 0.5) of the model computed on the unseen data is 86.0%, which outperforms the 75.0% accuracy score of the previous benchmark model. The AUC of the model is 0.783 again outperforming the previous model (0.764).

For completeness, it is also reported that the accuracy computed on the data used for computing the coefficients is 90.7% with an AUC of 0.907,

variable	coefficient	std. error	Wald	p value
DADLm	-.495	1.658	.089	.765
DADLs	-1.865	1.069	3.041	.081
PADLm	1.957	.857	5.220	.022
PADLs	-.482	1.269	.145	.704
CFs	2.523	1.201	4.410	.036
CI1	.092	1.378	.004	.947
CI2	-.838	.873	.922	.337
CDT	-1.702	.876	3.773	.052
SFm	-1.564	2.201	.505	.477
SFs	-1.221	1.188	1.055	.304
GSs	-1.015	1.130	.806	.369
AG2	1.598	1.052	2.307	.129
AG3	-1.886	.862	4.784	.029
DRm	3.129	1.687	3.439	.064
DRI	.231	.704	.107	.743
CP	3.376	.788	18.347	<.001
Constant	-1.954	2.453	.634	.426

Table 5.1: Results from logistic regression using expert driven variables selection

with evidence of a degree of overfitting. According to Cox and Snell [85] the averaged R^2 for the model computed on the whole dataset is 0.371, while using the R^2 measure proposed by Nagelkerke [86] the model scores 0.602.

In the second stage, stepwise logistic regression starting with all available variables has been carried out to perform a variables selection; the idea is to select only the variables which can improve significantly the model performance, leading to a more efficient and cost-effective model. Removal probability of the stepwise algorithm is fixed at 0.2 and variables AG and CI are included in the model as continuous variables to ensure no loss of information. The algorithm met the selection criteria after 12 iterations and selected 8 variables which are tabulated in the table 5.2 along with the coefficients and Wald statistics of the logistic regression executed only with those variables.

Comparing the pool of selected variables with the previously used one, we note that:

- CI, SF, AG, were included in the previous model but do not improve the logistic regression model's performance
- TR, not included in the previous model, significantly improved the model performance
- DADL, included in the previous model, is relevant only if reflects a severe impairment
- PADL, included in the previous model, is relevant only if reflects an impairment judged at least mild

variable	coefficient	std. error	Wald	p value
DADLs	-2.009	1.120	3.218	.073
PADLm	1.390	.745	3.477	.062
CFs	2.233	1.171	3.633	.057
GSs	-1.589	.941	2.851	.091
TR	-2.517	.942	7.141	.008
DRm	1.806	1.038	3.028	.082
CP	3.875	.800	23.474	<.001
CDT	-1.704	.833	4.182	.041
Constant	-2.348	1.113	4.449	.035

Table 5.2: Results from logistic regression using statistics driven variables selection

variable	coefficient		p value	
	average	std. error	average	std. error
DADLs	-2.315	1.724	.261	.273
PADLm	1.357	.312	.185	.175
CFs	2.494	2.362	.324	.426
GSs	-1.905	1.323	.278	.329
TR	-2.661	.846	.074	.098
DRm	2.529	1.902	.190	.047
CP	4.334	1.514	<.001	<.001
CDT	-2.091	.709	.089	.036
Constant	-3.009	1.873	.117	.014

Table 5.3: Summary of regressions with three subsets

- DR, included in the previous model, is relevant only if it reports at least a medium duration

The ratio between number of events and considered variables is higher than before ($29/8=3.625$) since less variables are involved, but it is still quite low. Again, to evaluate the model performance only with the selected variables, the above mentioned cross-validation approach has been implemented in order to avoid overfitting. Results from the different regression models using each of three datasets are averaged and summarized in the table 5.3.

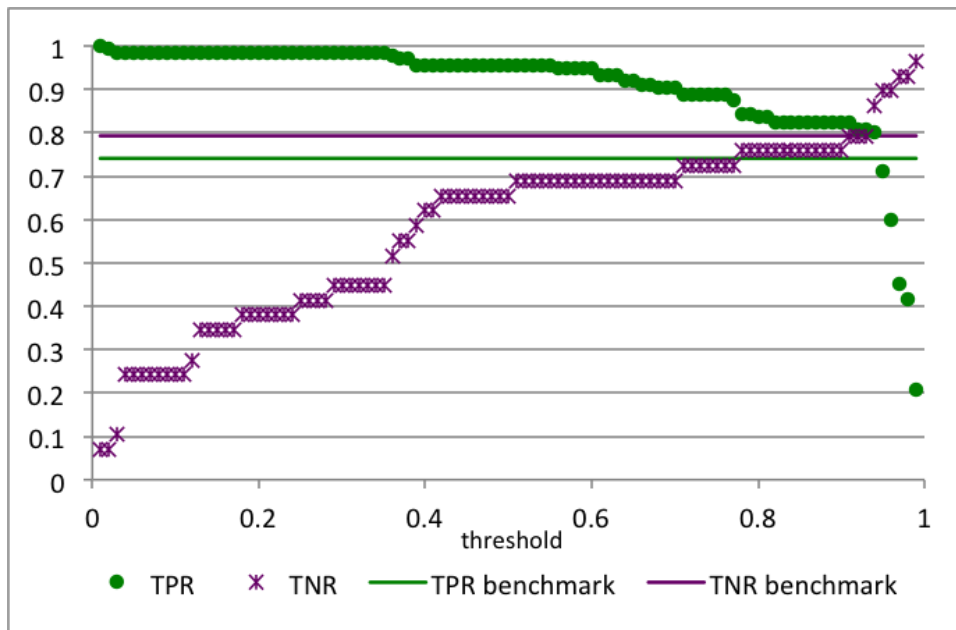


Figure 5.1: TPR and TNR values for different thresholds

The accuracy of the model computed on unseen data is 90.2% and the AUC of the model is 0.879, outperforming the logistic regression model with expert-driven variables selection. It is also found that the accuracy computed on the data used for computing the coefficients is 91.5% with an AUC of 0.905, outlining a possible certain degree of overfitting which has been avoided with cross-validation. The R^2 for the model is 0.365 according to Cox and Snell, while the model scored 0.601 according to Nagelkerke. Figures 5.1 and 5.2 report true positive/negative rates (TPR and TNR) and positive/negative predictive values (PPV and NPV) while the threshold is varied. The model is reported to outperform the benchmark model for threshold between 0.91 and 0.94.

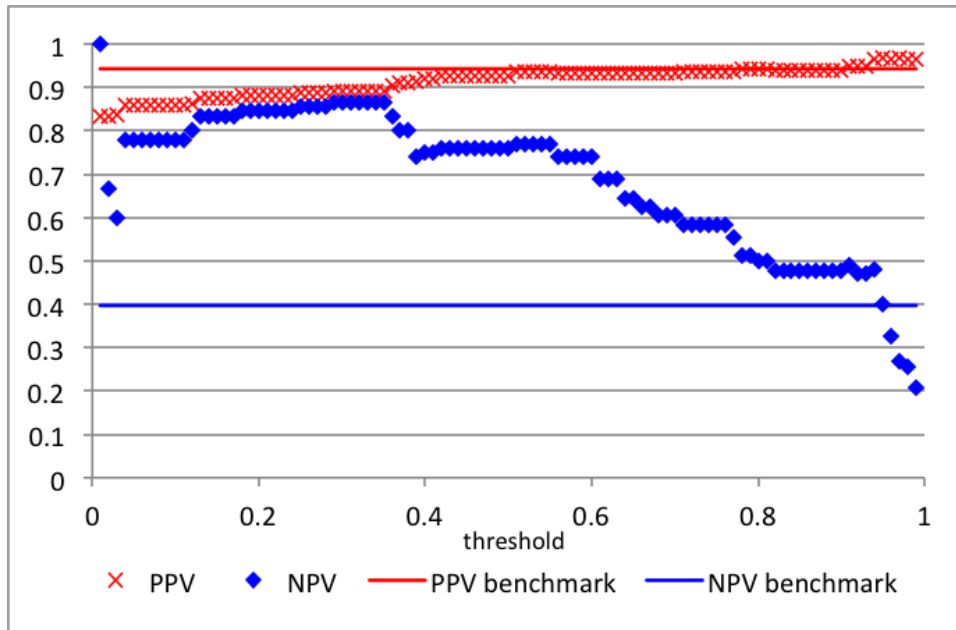


Figure 5.2: PPV and NPV values for different thresholds

Since in both models the selected variables are all categorical, the model to determine the probability of dementia conditions becomes

$$P(Y) = \frac{1}{1 + e^{-\text{SCORE}}} \quad (5.1)$$

where

$$\text{SCORE} = \text{constant} + \sum \text{coefficients of positive answers} \quad (5.2)$$

The final model consists then of a number (one per variable) of yes/no questions regarding the patient suspected of dementia. It is necessary to sum up all the coefficients associated with the questions where the answer is positive and add a constant in order to get the final score; this should be finally inserted in the formula 5.2 to get the probability of patient being affected with dementia.

5.6 MISCLASSIFICATION ANALYSIS

We investigate how predictions and misclassifications are distributed in order to determine whether the model could be unreliable under certain conditions. We choose to study such distributions as a function of the model output. Operatively, the entire range of outputs has been divided into 10 categories uniformly distributed between the minimum and maximum values of obtained output values: considering SCORE (as defined per formula 5.2) as model output, the interval to be divided into 10 categories is $[0, 1]$; considering $P(Y)$ (as defined per formula 5.1) we focus on the interval identified by the extreme model output given using the considered dataset. For this analysis we have chosen representative thresholds τ of 0.5, 0.7 and 0.9.

The distribution of misclassifications is reported in the following tables 5.4 and 5.5.

interval min	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
interval max	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
% misclassif. $\tau = 0.5$	0.0	-	62.5	45.5	-	-	20.0	33.3	12.5	3.5
% misclassif. $\tau = 0.7$	0.0	-	62.5	45.5	-	-	80.0	33.3	12.5	3.5
% misclassif. $\tau = 0.9$	0.0	-	62.5	45.5	-	-	80.0	66.7	87.5	3.5

Table 5.4: Distribution of misclassifications (function of $P(Y)$)

interval min	-4.1	-3.0	-1.9	-0.8	-0.4	1.5	2.6	3.7	4.8	5.9
interval max	-3.0	-1.9	-0.8	-0.4	1.5	2.6	3.7	4.8	5.9	7.0
% miscl. $\tau = 0.5$	0.0	0.0	62.5	45.5	27.3	14.3	2.5	5.0	0.0	0.0
% miscl. $\tau = 0.7$	0.0	0.0	62.5	45.5	54.5	14.3	2.5	5.0	0.0	0.0
% miscl. $\tau = 0.9$	0.0	0.0	62.5	45.5	72.7	57.1	2.5	5.0	0.0	0.0

Table 5.5: Distribution of misclassifications (function of SCORE)

The distribution of misclassifications and predictions for the defined intervals are reported (respectively for t equal to 0.5, 0.7, 0.9) on figures 5.3, 5.5 and 5.7 for intervals of $P(Y)$, and on figures 5.4, 5.6 and 5.8 for intervals of SCORE.

From the analysis of the distributions function of $P(Y)$, it is clear that the first and last intervals has a lower proportion of misclassification than other intervals. More specifically, the interval of SCORE between -1.85 and 2.55 (or 1.45 depending on the chosen threshold) has a very high misclassification rate, suggesting that the developed model is not reliable at such an interval of output. Excluding the interval between -1.85 and 2.55 (where 27% of predictions are concentrated), and assuming a threshold of 0.5, the accuracy increases from 89.0% to 97.5%.

This information should be used to refine the model: in the implementation proposed in the next section 5.7, it is used to suggest the use of alternative tools to detect early dementia conditions. Further studies should collect more data, in order to reach a sufficient dataset size to build a possible second

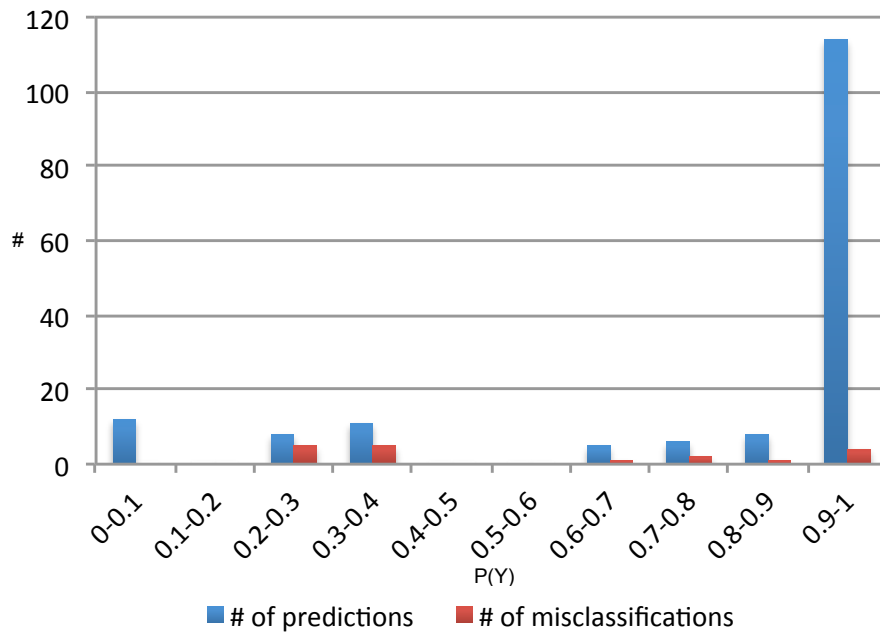


Figure 5.3: Distribution predictions/misclassifications $\tau = 0.5$ (function of $P(Y)$)

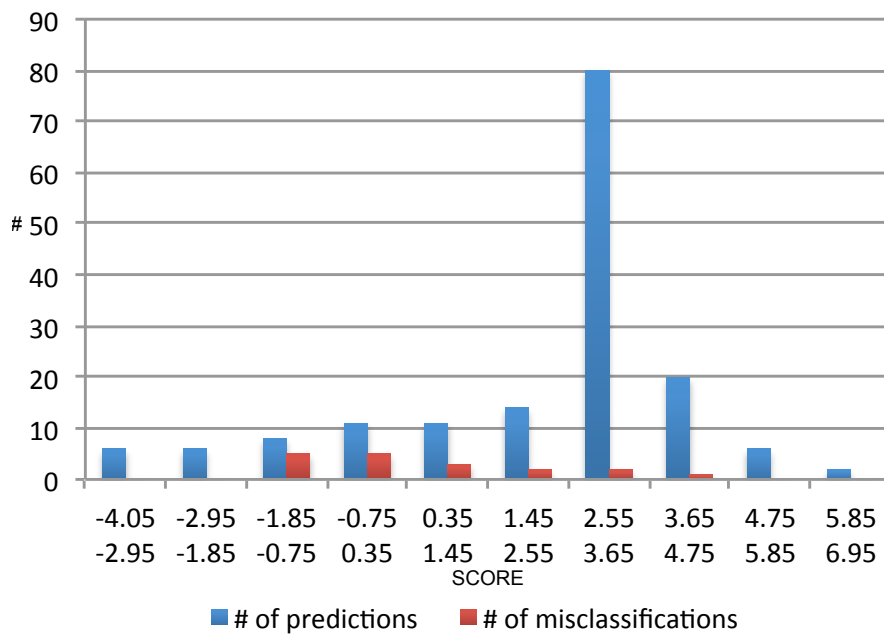


Figure 5.4: Distribution predictions/misclassifications $\tau = 0.5$ (function of SCORE)

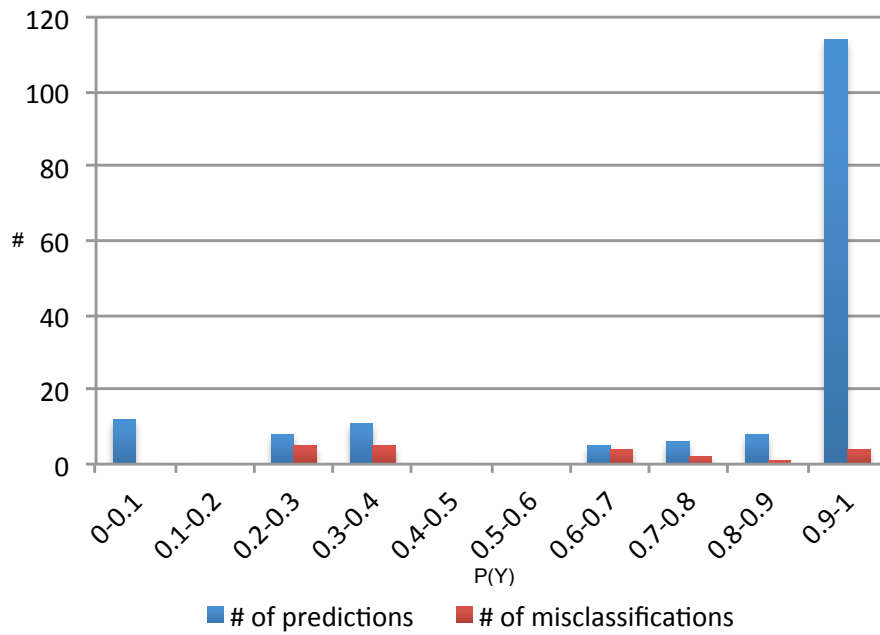


Figure 5.5: Distribution predictions/misclassifications $\tau = 0.7$ (function of $P(Y)$)

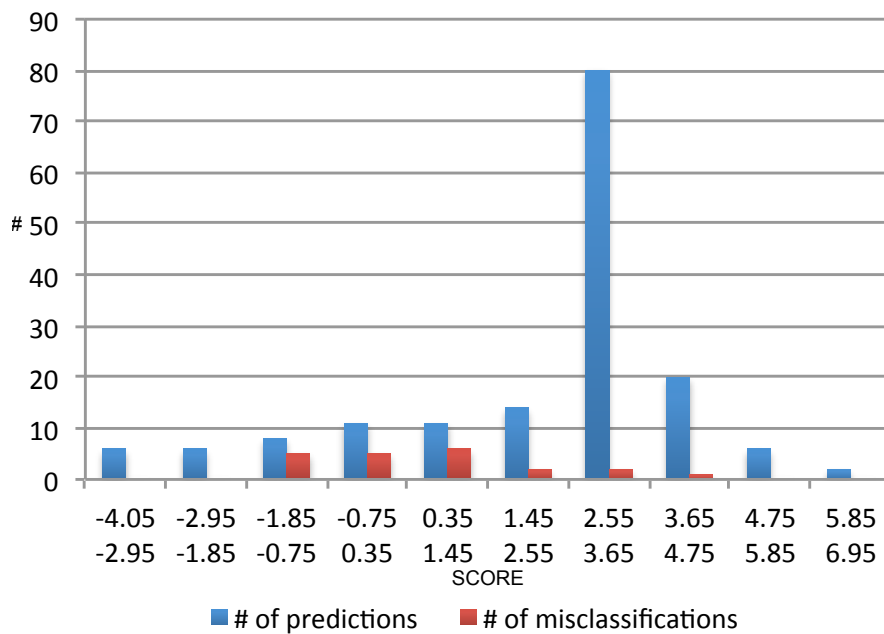


Figure 5.6: Distribution predictions/misclassifications $\tau = 0.7$ (function of SCORE)

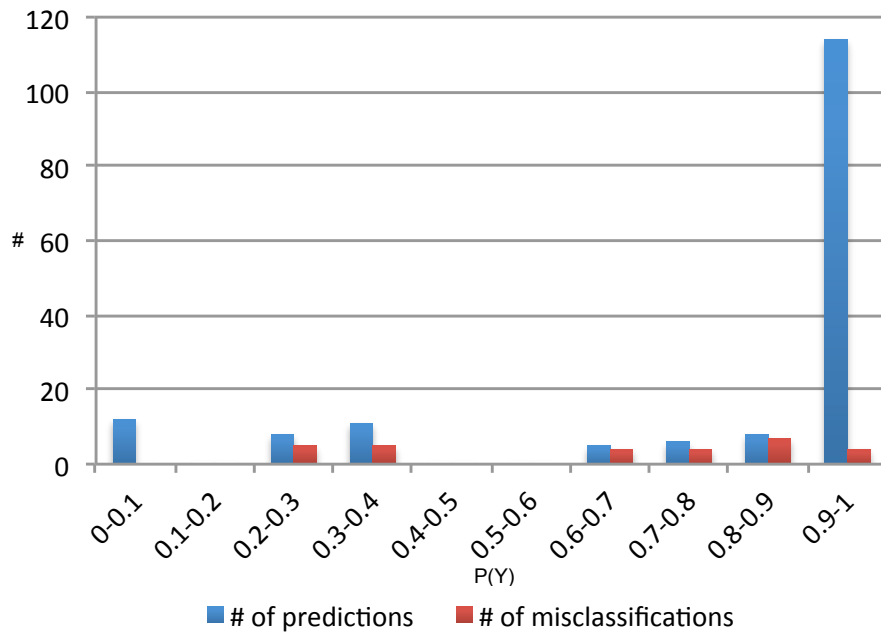


Figure 5.7: Distribution predictions/misclassifications $\tau = 0.9$ (function of $P(Y)$)

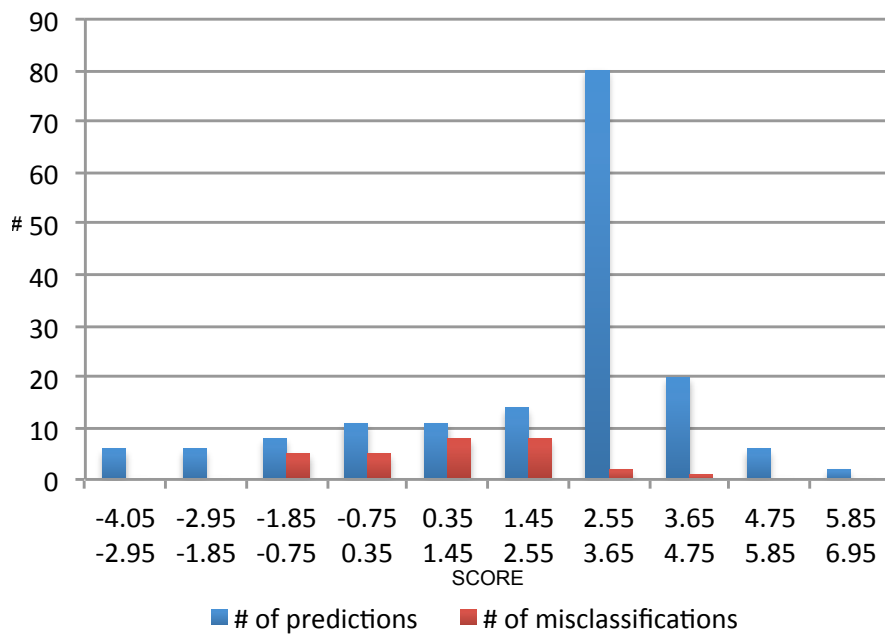


Figure 5.8: Distribution predictions/misclassifications $\tau = 0.9$ (function of SCORE)

model to be used downline of the proposed one, with a view to improve the accuracy of the model.

5.7 IMPLEMENTATION

The model developed within this chapter to aid the diagnosis of dementia has been implemented into a calculator named *Stirling dementia risk calculator*. It is a web-based prototype and its core is a PHP script, which generates an HTML page after data is collected from a HTML input form and then POSTed when user presses the "Calculate score" button.


The prototype, available at <http://bit.ly/dem-calc>, is intended to be used for patients suspected of having dementia. The user is asked to answer 8 categorical questions, specifically: level of impairment in domestic and personal activities of daily living, level of impairment in current functioning, level of global severity of impairment, the presence of tremors, the presence of a clear progression of symptoms, the ability of completing the clock drawing tests, the duration of symptoms (short, medium or long period). Levels to describe impairment are: severe, mild and none.

These variables are then mapped to identify the coefficients to be summed together as per formula 5.2. Specifically, with regard to figure 5.9 (which shows the user interface) and table 5.3 (which reports variables included in the model):

- DADLs coefficient is summed if answer to question 1 is "severe"
- PADM coefficient is summed if answer to question 2 is "severe" or "mild"

- CFs coefficient is summed if answer to question 3 is "severe"
- GSs coefficient is summed if answer to question 4 is "severe"
- TR coefficient is summed if answer to question 5 is "yes"
- DRm coefficient is summed if answer to question 6 is "medium" or "long"
- CP coefficient is summed if answer to question 7 is "yes"
- CDT coefficient is summed if answer to question 8 is "yes"

The tool calculate the probability of having dementia using formula 5.1, reporting it to user in a separate screen as in figure 5.10. As suggested from the analysis carried out in the previous section 5.6, the user is warned about model unreliability when the value of SCORE is between -1.85 and 2.55.



**UNIVERSITY OF
STIRLING**

SCHOOL OF
NATURAL SCIENCES

Stirling Dementia Risk Calculator

Risk model for patients suspected of having dementia

1. Is the patient showing impairment in domestic activities of daily living (ability to carry out activities such as shopping, housekeeping, finance management, food preparation and transportation)?
 Severely Mildly None

2. Is the patient showing impairment in personal activities of daily living (ability to carry out activities such as dressing, eating, ambulating and hygiene)?
 Severely Mildly None

3. Overall, is the patient showing impairment in current functioning, i.e. in general activities of daily living?
 Severely Mildly None

4. Overall, how would you rate the global severity of impairment?
 Severe Mild None

5. Is the patient experiencing tremors?
 Yes No


6. How long has the patient been showing symptoms for?
 Short period Medium period Long period

7. Did the patient show a clear progression in these symptoms?
 Yes No

8. Is the patient able to complete the clock drawing test?
 Yes No

Developed by [Thomas Mazzocco](#) and Amir Hussain, University of Stirling. All rights reserved.
Pilot prototype provided "as is" without any warranty.

Figure 5.9: Input screen for dementia risk calculator



**UNIVERSITY OF
STIRLING**

SCHOOL OF
NATURAL SCIENCES

Stirling Dementia Risk Calculator

Risk model for patients suspected of having dementia

The probability of suffering from dementia is **3%**

The most important factors with respect to the outcome are:

- the presence of tremors (negative correlation)
- the medium/long duration of symptoms (positive correlation)
- the impairment in personal activities of daily living (positive correlation)

[Back](#)

Developed by [Thomas Mazzocco](#) and Amir Hussain, University of Stirling. All rights reserved.
Pilot prototype provided "as is" without any warranty.

Figure 5.10: Output example for dementia risk calculator

5.8 DISCUSSION

The first fact which appears from the analysis of the results is that logistic regression clearly outperforms the performance of a Bayesian belief network, which was regarded as a state-of-the-art model. Unlike other artificial intelligent techniques commonly used for prediction known as “black-box” models (such as neural networks or support vector machines), logistic regression provides with a clear picture of which variable is significant (with the correspondent p-value) for the final predicted outcome, giving also a measure of the magnitude and the direction (reflected by the coefficient with its sign) of such impact.

The added value of the developed model which led to a performance increase is determined both by the selected technique and by the variables selection process, which followed all the indications suggested by the proposed framework in section 3.1.

The differences between variables selected by domain experts and by a statistical-driven approach have already been outlined. The most relevant difference is the inclusion of TR in the group of significant variables, with an inverse relationship to the probability of a positive diagnosis of dementia. The statistical-driven approach has shown the best performance in terms of predictive power and accuracy and a good R^2 coefficient. An approach with fewer variables (like the proposed statistical-driven one) could be cheaper to implement and employ: in this case, the workflow from the collection of data

to the provision of a predicted outcome will use fewer resources (in terms of time and money) than an approach which requires more data.

Comparing p-values provided on table 5.2 with the ones tabulated on table 5.3, it is clear how significance of each variable could be highly influenced by the size of the considered sample: considering the whole dataset, all the selected variables could be regarded as significant (when choosing a criterion of $p < 0.1$) while considering datasets composed by two thirds of the original dataset, only few variables could be said to be definitely significant. Again, looking at table 5.1, only a handful of variables could be regarded as definitely significant while nothing could be certainly stated about the significance of most coefficients; possibly, a bigger dataset could give further help in discrimination between significant and non significant variables. Also, since the criterion suggested in paragraph 3.1.3 about ratio between number of data points and number of independent variables is not met, a larger dataset should be used to strengthen findings.

The size of dataset is probably one of the main limitations of this work, which should then be regarded as a pilot study for further deeper researches. The final model (both with experts and with statistical driven variables selection) is relatively simple and could be easily implemented as a decision support system for carers as shown in the previous section 5.7. Probably the most important design decision to be determined will be the threshold used to transform the probability of a positive diagnosis in the actual predicted diagnosis: this choice

is highly dependent on the assumed costs for type I and II errors (i.e. false positives and false negatives).

5.9 CONCLUSION

The main aim of this study was to build an alternative model which could improve the results produced by a state-of-the-art Bayesian belief network model recently developed for dementia diagnosis. The proposed prediction models used a logistic regression algorithm to predict the diagnosis of dementia using variables selected either by domain experts or by a statistical driven procedure following the best practices identified in paragraph 3.1.3. Results have shown a noticeable improvement in considered performance metrics with the proposed models outperforming the benchmark hand-crafted model which required both technical expertise and domain experts' knowledge for building the more sophisticated and complex Bayesian belief network model. Interestingly, the approach based on statistical variables selection outperformed the model which used variables selected by domain experts in the previous study.

Some limitations should also be acknowledged: first, the considered sample of patients was quite small (164 patients) leading to a possible dependence on a specific dataset, with loss of generalization power; a second limitation relates to the fact that the available data were collected as part of a previous research, so the motivations behind collecting these specific variables and not others as well as the exact mapping procedure for qualitative variables are not

known in details. These two limitations are sufficient to prevent this model from being directly used in clinical practice; however, the new model offers fertile ground for further research and development.

In conclusion, whilst the preliminary results reported in this study should be taken with care, they do demonstrate the capability of employing relatively simple logistic regression based prediction models for dementia diagnosis and a range of contributions and potential impact is envisaged from this work both for clinical practice and further research.

A future large-scale investigation is required to determine the optimal approach for variables selection and to overcome the discussed existing limitations imposed by the size of the considered sample; this should also assess the feasibility of deploying such models in clinical practice.

CASE STUDY 3: A SIDE-EFFECTS MAPPING MODEL IN PATIENTS RECEIVING CHEMOTHERAPY

This chapter presents the third case study developed following the procedure described in the proposed framework: it is a predictive model used to predict the probability of experiencing a certain symptom among common side-effects in patients receiving chemotherapy.

Cancer treatments are now more effective than ever and, as a consequence, cancer is becoming a chronic disease. Chemotherapy is a frequently used treatment in people with cancer and it can cause a number of side-effects which if not properly managed could have a negative impact on the patients' quality of life.

A sample of 56 patients receiving chemotherapy treatment for breast, colorectal and lung cancer is considered; each experienced side-effect is recorded during four consecutive treatment cycles (each lasting 14 days). Previous studies have used the same dataset to build side-effects predictive models for each symptom.

In this study, five of the most frequent side-effects (fatigue, nausea, mucositis, hand and foot sore, diarrhoea) are selected to build a comprehensive model which predicts the probability of experiencing a certain symptom on a specified

day of each cycle of therapy. An overall increase of predictive power is expected by using an appropriate pre-modelling strategy, i.e. by selecting a proper functional form of the model.

The computed accuracy of results shows that the newly proposed model has an enhanced predictive power compared to a state-of-the-art approach. The information gained from this study will help medical and nursing staff caring for such patients to more accurately predict the side-effects that patients will experience and therefore select appropriate help to minimise, whenever possible, the influence of those symptoms.

6.1 BACKGROUND

It is estimated that in 2007 almost 300,000 individuals in the United Kingdom were diagnosed with cancer and over the last 25 years, cancer incidents have considerably increased [87, 88, 89, 90].

Different treatments are available depending on the type and stage of cancer and the survival rates have been improving over the last 30 years; in particular, besides surgery an adjuvant chemotherapy treatment is often given: this helps to reduce the risk of cancer recurrence or death from microscopic spread of the cancer that is suspected (but cannot be detected) and also it may alleviate cancer related symptoms with a consequent improvement in patients' quality of life [91, 92].

However, it is to be noted that adjuvant chemotherapy exposes patients to risk of significant side-effects that could have a negative impact on patients' quality of life and daily living [93] and also on the maintenance of dose intensity treatment, which could influence the disease free and overall survival [94, 95].

A poor assessment and management of symptoms in patients with cancer have been ascertained [96]. It has also been observed that poorly informed patients are less likely to comply with treatment and are more likely to experience anxiety and hence a general reduction in their quality of life [97, 98]. As a consequence, an effective prediction of side-effects could help medical staff with better management of patients' needs, with special regard to discomfort minimization, unnecessary worry and anxiety reduction.

Different tools have been proposed as clinical decision support system in many clinical fields: with regard to cancer care, studies tend to focus on predictors of survival and life threatening toxicities [99, 100, 101].

In relation to the prediction of symptoms, only a few risk models have been presented [102]. The use of technology to communicate between healthcare professionals and patients may lead to improvements in quality of life and symptoms control, reductions in the rate of hospitalizations, emergency department visits and cost savings [103]. Patients also appear to have positive views of using this type of technology, reporting improvements in communication with healthcare providers [104].

The Advanced Symptom Management System (ASyMS©) has been developed and trialled as an example of the use of technology in cancer care [105, 106, 107, 108]. It has been built as a mobile telephone-based remote symptom monitoring system which can be used to register, monitor and predict the side-effects of chemotherapy while the patient is not with a healthcare professional [109].

First, patients using the system are asked to complete a symptom questionnaire on a mobile phone twice a day and sent this information directly to their hospital-based healthcare professional. Self-care advice is then given on the basis of the reported symptoms. Depending on their seriousness, an alert is generated to the healthcare professional via a 24 hour dedicated pager system. The healthcare professional is then informed of the symptoms that the patient has reported and may contact the patient if necessary. This system also allows nurses to monitor the symptoms remotely and facilitates the delivery of relevant and useful advice to the patient based on their current symptoms.

Next, the tool uses the patients' symptom history as well as a model developed based on a corpus of patients with similar medical conditions to predict the likely side effects a patient could expect over the course of treatment: patients are able to receive predictions concerning the possible symptoms they are going to experience throughout the course of treatment along with daily predictions that are updated as they enter data describing their own symptoms.

A diary is presented on patients' mobile phones where, for each day, a smiley, sad or neutral face is used to depict the overall side-effects situation predicted for that particular day: patients who wish to plan ahead can see at a glance which days they are more likely to feel well. Users may select any of the symptoms to see self-care advice on how to manage this symptom; they will also be able to see how many more days they are likely to experience each symptom.

6.2 AIMS

The aim of this study is to evaluate, improve and generalize the pilot model proposed by a previous study [110] using an enhanced (including different cancer conditions) dataset and according to a common set of performance metrics. In the previous study a number of different simple mathematical equations were developed to predict the probability of experiencing each symptom on a specified day of treatment, for patients with breast cancer. The idea was to build on the previously presented remote monitoring system for patients.

The objective of the present research is to generalise the previous model to build a more powerful and comprehensive side-effect risk model for patients with cancer undergoing adjuvant chemotherapy: this model is not limited to breast cancer but has been extended to cover colorectal and lung cancer conditions. A single mathematical model with an appropriate functional form

developed as recommended in section 3.2, which predicts on a day-by-day basis the symptoms that patients with cancer receiving chemotherapy are going to experience, is proposed.

The new model can be used as a tool to provide preparatory information to patients with cancer receiving chemotherapy and to their carers. An improvement in the patients' experience is expected by providing information on the side-effects that they are likely to experience on each day of treatment; furthermore, the provision of tailored information and possibly medications based on their individual needs can also be facilitated in the future by such a model.

6.3 DATASET PREPARATION

The collection of data has been carried out as part of previous research [111], over a 12-month period from June 2007 to May 2008: 56 patients' data from four clinical sites in Scotland were collected, although only 34 patients with breast cancer were considered within the cited study. Selected patients were diagnosed with breast, colorectal and lung cancer, starting a course of adjuvant chemotherapy, aged 18 years or over, able to read and write English and all deemed by members of the clinical team to be physically and psychologically fit to participate in the study. Ethical approval was gained from the study sites, and all patients provided written informed consent before their participation in the study. The observation of each patient involved treatment over four cycles,

each lasting 14 days, where treatment was administered at the beginning of each cycle.

For each patient the following independent variables are used for this study:

- number of the cycle (between 1 and 4), treated as ordinal variable
- number of the day within the cycle (1 to 14), treated as continuous variable

The dependent variable is the probability of experiencing symptoms among the five objects of our model. Patients are grouped as follows: patients with breast cancer (N=34), with colorectal cancer (N=9) and with lung cancer (N=13).

6.4 MODEL CONSTRUCTION

6.4.1 *Pre-modelling*

A previous study [110] has shown that the probability of experiencing a specific symptom is basically time dependent. In different ways, each of the five considered symptoms has two main tendencies over time that could be outlined: a 'peak effect', around the day in which the treatment is received by patients, and an 'inverted U-shape effect', rising from a low on the day after treatment to a peak around mid-cycle before falling again. In this study a more general model is proposed which combines these two effects. Moreover, since differences between cycles were outlined, a cycle-dependent coefficient

was added in order to capture those differences. Following this setting, a comprehensive model is proposed as per formula (6.1)

$$P(d) = a \cdot S(d) + b \cdot H(d) + \sum_{n=1}^4 c_n \cdot D_n \quad (6.1)$$

where:

- $P(d)$ is the probability of experiencing a specified symptom on day d ;
- D_n is a dummy variable which is set to 1 for the n -th cycle and 0 on other cycles;
- $S(d)$ is a function capturing the 'inverted U-shape effect';
- $H(d)$ is a function capturing the 'peak effect';
- a, b, c_n are coefficients determined for each symptom.

$S(d)$ is built so that $S(\text{first day}) = S(\text{last day}) = 0$ and $S(\text{middle day}) = 1$; in a similar way, $H(d)$ is built so that $H(\text{first day}) = 1$ and $H(\text{last day}) = 0$. Lots of functions could be adopted to be used as $S(d)$ and $H(d)$. For the purpose of this study a sinusoid function and a negative hyperbolic function have been chosen by trial and error.

It is worth noticing that c_n is treated as a categorical (ordinal) variable. As required by the procedure described in paragraph 3.1.1, it should have been split into 3 binary variables in order to avoid the "dummy variable trap". However, since the model does not include an additional constant and includes only one categorical (ordinal) variable, it is possible to use the equivalent codification with 4 binary variables without incurring in any numerical problem. This choice facilitates the analysis of the meaning of model coefficients.

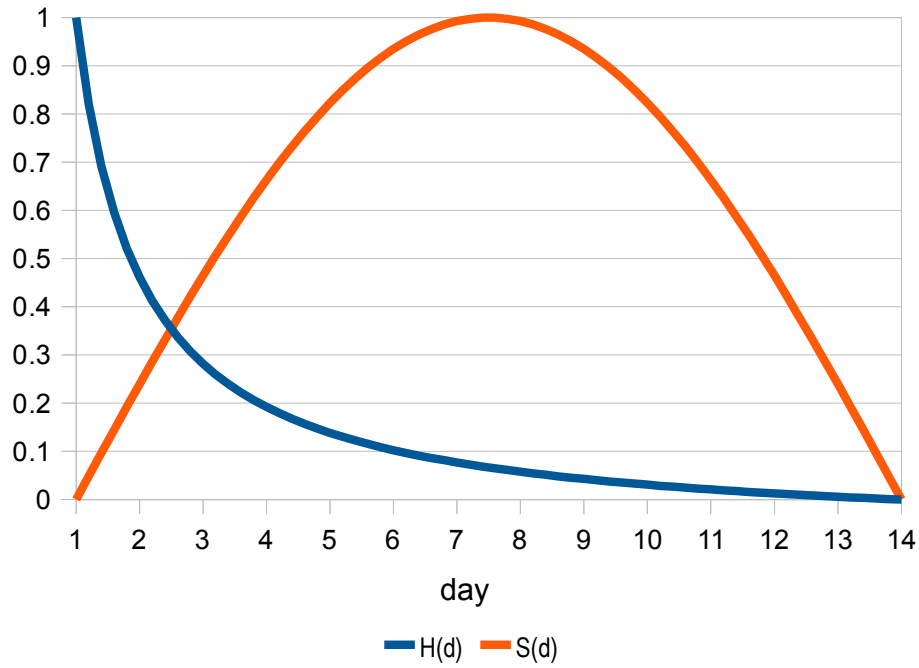


Figure 6.1: Chosen functions

Having 14 days within each cycle (and so $d_{\max} = 14$), the adopted formulas for these two terms are reported in equations (6.2) and (6.3).

$$S(d) = \sin\left(\frac{d-1}{d_{\max}-1}\pi\right) = \sin\left(\frac{d-1}{13}\pi\right) \quad (6.2)$$

$$H(d) = \frac{d_{\max}}{d_{\max}-1} \left(\frac{1}{d} - \frac{1}{d_{\max}}\right) = \frac{14}{13} \left(\frac{1}{d} - \frac{1}{14}\right) \quad (6.3)$$

The chosen functions are reported in figure 6.1.

6.4.2 Regression

Raw data collected from different patients about symptoms experienced in the same day of the same cycle have been grouped after dividing the dataset

into three subsets for breast, colorectal and lung cancer respectively. If they experienced a certain symptom on that day the considered output was 1, otherwise 0. After grouping the outputs, these were averaged for each group, giving a probability of experiencing that symptom at the corresponding time. It has to be noted that since the target output is a probability (i.e. a real number between 0 and 1), logistic regression cannot be used for tackling this problem.

Then, a linear regression (one each for breast, colorectal and lung cancer) based on functional form determined in the previous paragraph 6.4.1 was run in order to estimate the coefficients for the variables outlined above. As recommended in paragraph 3.2.1, p-values for each coefficient have been computed.

Coefficients for regression models are reported on tables 6.1, 6.2 and 6.3.

6.4.3 *Model evaluation and validation*

The outcome of the model is the probability of experiencing a specified symptom on a specified day and cycle.

For each symptom, the receiver operating characteristic (ROC) curve has been used to estimate the predictive power of the newly developed model, as prescribed by the proposed framework in section 3.4. Given the size of the dataset (especially for colorectal and lung cancer conditions), a leave-one-out cross-validation has been selected for model validation.

Table 6.1: Regression coefficients for breast cancer regression model

	Diarrhoea	H&F sore	Mucositis	Nausea	Fatigue
S	0.023	0.048	0.186	0.192	0.135
H	-0.025	-0.042	-0.263	0.474	0.085
D ₁	0.051	0.043	0.262	0.027	0.198
D ₂	0.065	0.071	0.124	0.051	0.287
D ₃	0.035	0.080	0.203	0.010	0.369
D ₄	0.073	0.070	0.311	0.075	0.286

The area under ROC curve has also been used to compare the performance of the developed models with the previously used ones. It has to be pointed out that since patients in the same group (i.e. during the same day of the same cycle) may have experienced different symptoms, the ideal AUC=1.0 cannot be reached.

Moreover, in order to measure the goodness of fit and to better explain the predictive power of the model, the standard coefficients of determination (R^2) are also computed for each model.

Table 6.2: Regression coefficients for colorectal cancer regression model

	Diarrhoea	H&F sore	Mucositis	Nausea	Fatigue
S	0.070	-0.072	0.010	-0.065	0.136
H	-0.098	0.022	0.025	-0.081	0.014
D ₁	0.117	0.083	0.031	0.089	0.102
D ₂	0.076	0.145	0.073	0.150	0.052
D ₃	0.122	0.215	0.116	0.206	0.040
D ₄	-0.004	0.081	-0.010	0.052	0.009

Table 6.3: Regression coefficients for lung cancer regression model

	Diarrhoea	H&F sore	Mucositis	Nausea	Fatigue
S	0.032	0.022	0.077	0.015	-0.006
H	0.014	0.031	-0.015	0.024	-0.155
D ₁	0.007	0.152	0.216	0.203	0.468
D ₂	0.056	0.139	0.316	0.193	0.588
D ₃	0.003	0.122	0.281	0.159	0.581
D ₄	0.027	0.029	0.100	0.193	0.708

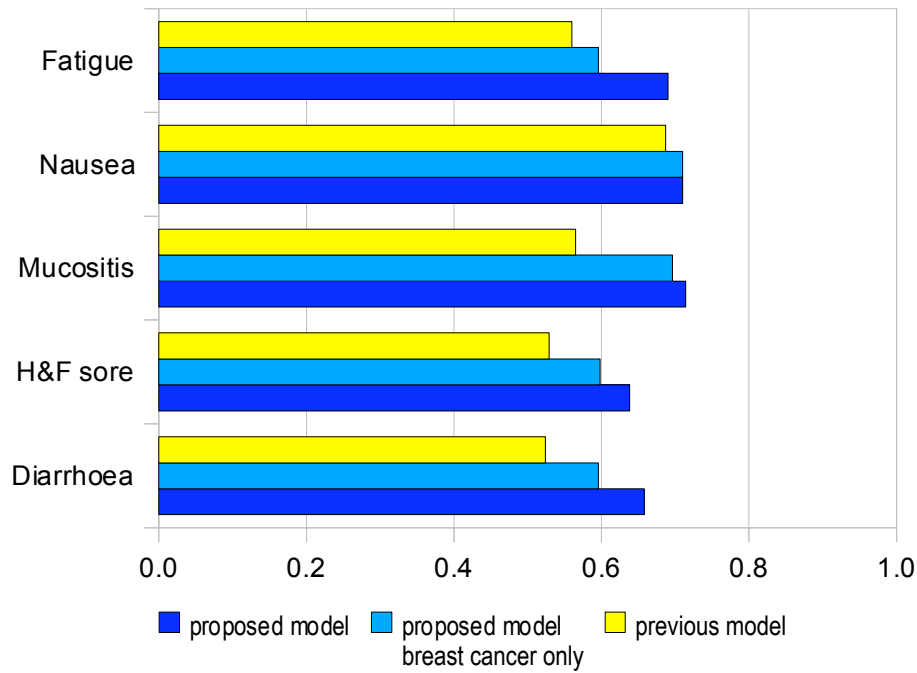


Figure 6.2: Comparison of performance for each symptom

6.5 RESULTS

6.5.1 *Area under ROC curve*

On table 6.4 the AUC for the proposed model and for the previously used model is reported for each symptom. In order to make a fair comparison with the available benchmark (which used only patients with breast cancer), the performance related to breast cancer are also evaluated separately. From the results tabulated on table 6.4 and also depicted in figure 6.2, it is clear that the proposed model significantly outperformed the previous model.

Table 6.4: AUC for current and previous models

#	Diarrhoea	H&F sore	Mucositis	Nausea	Fatigue
1	0.658	0.638	0.714	0.710	0.690
2	0.596	0.598	0.696	0.710	0.596
3	0.524	0.529	0.565	0.687	0.560

#1: proposed model - breast, colorectal and lung cancer

#2: proposed model - breast cancer only

#3: previous model - breast cancer only [110]

6.5.2 R^2 and p -values

Coefficients of determination for each symptom and type of cancer are given in table 6.5 (in this case, the R^2 coefficients measures the proportion of the variability in the dependent variable about the origin explained by regression - since the coefficients for identifying the cycle are treated as dummy variables - and so this coefficient cannot be compared to a similar one for models which include an intercept).

Each coefficient derived from the linear regression is associated with a p -value, showing the probability of observing the data if the associated coefficient

Table 6.5: Comparison of R^2

	Diarrhoea	H&F sore	Mucositis	Nausea	Fatigue
Breast	0.746	0.837	0.900	0.850	0.960
Colorectal	0.595	0.686	0.742	0.721	0.588
Lung	0.469	0.836	0.910	0.796	0.960

Table 6.6: P-values for breast cancer regression model

	Diarrhoea	H&F sore	Mucositis	Nausea	Fatigue
S	0.227	0.016	0.001	0.001	0.001
H	0.314	0.101	<0.001	<0.001	0.088
D ₁	0.006	0.022	<0.001	0.588	<0.001
D ₂	0.001	<0.001	0.012	0.305	<0.001
D ₃	0.057	<0.001	<0.001	0.836	<0.001
D ₄	<0.001	<0.001	<0.001	0.132	<0.001

was equal to zero. So, a high p-value indicates that the variable associated to the coefficient does not improve the global model.

P-values associated with each coefficient are tabulated on tables 6.6, 6.7 and 6.8.

Table 6.7: P-values for colorectal cancer regression model

	Diarrhoea	H&F sore	Mucositis	Nausea	Fatigue
S	0.149	0.051	0.655	0.024	0.938
H	0.125	0.639	0.392	0.035	0.671
D ₁	0.012	0.018	0.147	0.001	<0.001
D ₂	0.099	<0.001	0.001	<0.001	0.030
D ₃	0.009	<0.001	<0.001	<0.001	0.090
D ₄	0.934	0.020	0.626	0.054	0.702

6.5.3 Analysis

Data about cycle and day of treatment seem to have a good (and sometimes excellent, considering that the maximum attainable AUC is less than 1) predictive power for all the listed symptoms.

Also, the overall performance of the model seems to confirm that the effects observed for treatments of breast cancer (inverted U-shape and peak effects) may be re-usable for other kinds of cancer. However, as reflected by the lower R^2 (especially for colorectal cancer) these effects may be able to explain a smaller part of the variability of the proposed model.

Table 6.8: P-values for lung cancer regression model

	Diarrhoea	H&F sore	Mucositis	Nausea	Fatigue
S	0.209	0.454	0.075	0.760	0.915
H	0.672	0.423	0.795	0.704	0.038
D ₁	0.784	<0.001	<0.001	<0.001	<0.001
D ₂	0.024	<0.001	<0.001	<0.001	<0.001
D ₃	0.891	<0.001	<0.001	0.001	<0.001
D ₄	0.269	0.293	0.016	<0.001	<0.001

Moreover, p-values associated with coefficients show that the two main effects, while being statistically significant for many symptoms in patients with breast cancer, could not be generally seen as definitely relevant during the treatment of colorectal or lung cancer (assuming a significance level of 5%), except for both effects in determining nausea for patients with colorectal cancer and for inverted U-shape effect in prediction of fatigue for patients with lung cancer.

Two examples of the model's outcome for some symptoms are reported in figure 6.3 and figure 6.4.

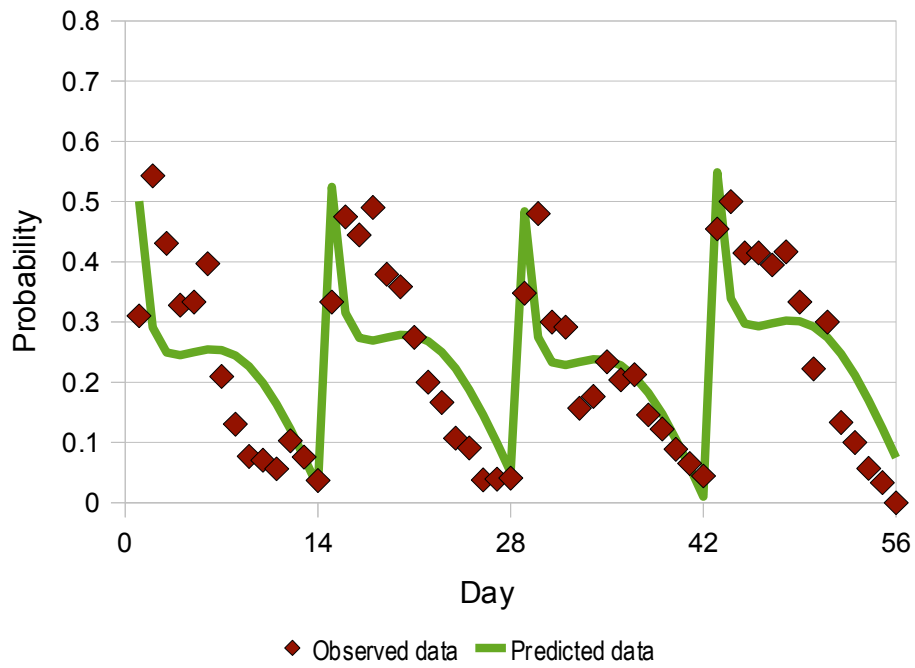


Figure 6.3: Representation of probability given by the model (nausea for breast cancer)

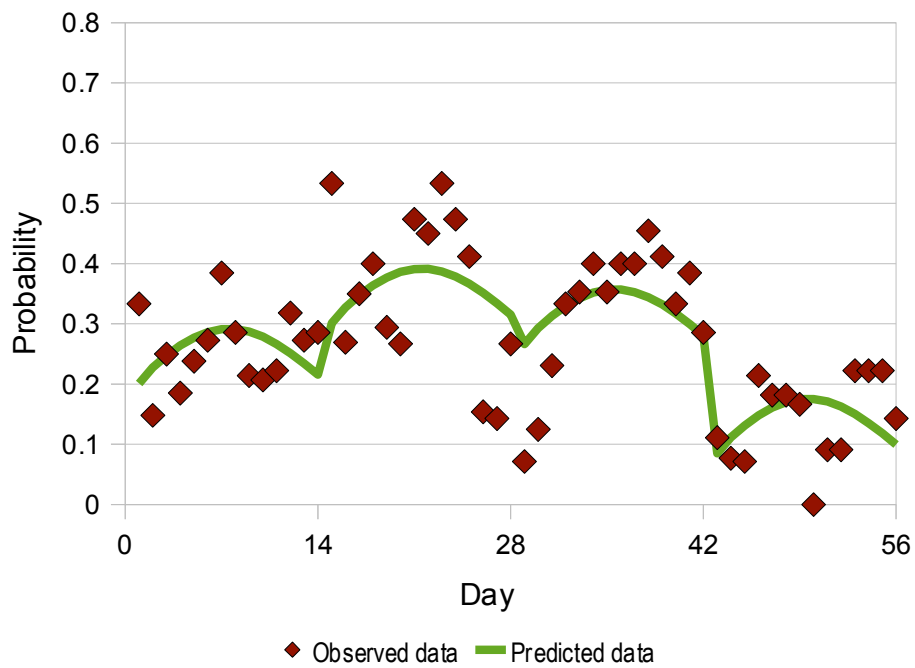


Figure 6.4: Representation of probability given by the model (mucositis for lung cancer)


6.6 IMPLEMENTATION

The improved model developed in this chapter has been turned into a web-based prototype composed by a PHP script, which generates an HTML page after data is collected from HTML input form and then POSTed when user presses the "Calculate score" button.

This prototype, available at <http://bit.ly/chemo-calc>, implements the side-effects mapping model in patients receiving chemotherapy discussed in the previous sections and it is intended to be used within the 4 cycles of chemotherapy for patients suffering from breast, colorectal and lung cancer.

The user is asked to specify the pathology, the cycle of treatment and the day from the beginning of cycle. The user interface is shown in figure 6.5.

Probabilities are then calculated using formula 6.1 for the specified day d (shown as *today*), and also for day $d + 1$ (shown as *tomorrow*) and $d + 2$ (shown as *the day after tomorrow*). Expected side-effects are shown if probability is greater than 20% and are classified as "very likely" if greater than 40% or "likely" otherwise. An example of the output is reported in figure 6.6.



UNIVERSITY OF STIRLING

SCHOOL OF NATURAL SCIENCES

Stirling Chemotherapy Side-Effects Model

Side-effects mapping model in patients with lung, colorectal and breast cancer receiving chemotherapy

Chemotherapy is a frequently used treatment in people with cancer and it can cause a number of side-effects which if not properly managed could have a negative impact on the patients' quality of life. Predicting side-effects will help medical and nursing staff caring for such patients and selecting appropriate help to minimise, whenever possible, the influence of those symptoms.

1. Select pathology:
 Breast cancer
 Colorectal cancer
 Lung cancer


2. Cycle of chemotherapy:
 First
 Second
 Third
 Fourth

3. Day:

Calculate

Developed by [Thomas Mazzocco](#) and Amir Hussain, University of Stirling. All rights reserved.
Pilot prototype provided "as is" without any warranty.

Figure 6.5: Input screen for chemotherapy side-effects model



**UNIVERSITY OF
STIRLING**

SCHOOL OF
NATURAL SCIENCES

Stirling Chemotherapy Side-Effects Model

Side-effects mapping model in patients with lung, colorectal and breast cancer receiving chemotherapy

Based on the information provided, expected side effects are reported below.

Today	<i>very likely:</i>	<ul style="list-style-type: none"> • nausea (50%)
	<i>likely:</i>	<ul style="list-style-type: none"> • fatigue (28%)
Tomorrow	<i>likely:</i>	<ul style="list-style-type: none"> • nausea (29%) • fatigue (26%)
The day after tomorrow	<i>likely:</i>	<ul style="list-style-type: none"> • fatigue (28%) • mucositis (27%) • nausea (24%)

Very likely side effects have a probability of occurring greater than 40%.
Likely side effects have a probability of occurring greater than 20%.
Side effects with a lower probability of occurring are not reported.

[Back](#)

Developed by [Thomas Mazzocco](#) and Amir Hussain, University of Stirling. All rights reserved.
Pilot prototype provided "as is" without any warranty.

Figure 6.6: Output example for chemotherapy side-effects model

6.7 DISCUSSION

The main aim of this case study was to improve results produced by previous state-of-the-art models for prediction of side-effect symptoms experienced by patients with breast cancer receiving adjuvant chemotherapy, extending the model also for colorectal and lung cancers.

The model combines, for the first time, the two empirically determined effects: specifically the peak effect and the inverted U-shape effect, for which two different functions were chosen by trial and error.

From the analysis of chosen performance metrics, the model seems to have generally reached the envisaged performance providing an average increase of 19% in comparison with the previous model: this demonstrates the potential of these kinds of models in the management of chemotherapy related toxicities within clinical practice. Some limitations should also be noted: first, the considered sample of patients was not so large and was not equally distributed between breast, colorectal and lung cancer; a second limitation relates to the fact that the available data covered just four cycles of chemotherapy, while most adjuvant breast cancer chemotherapy regimens consist of six to eight cycles; finally, data about different administered treatments are not available at present.

The proposed framework in section 3.5 recommended to carry out a misclassification analysis only for classification problems, so such analysis has not been performed for this case study.

A range of contributions and potential impact is envisaged from this work both for clinical practice and further research. On clinical practice, patients could know in advance which symptom they should expect and when, and health professionals could take appropriate action wherever possible in order to avoid or, at least, minimize expected discomforts.

From the point of view of future research in this interdisciplinary area, a comprehensive model has been proposed for time series symptom analysis: which is expandable with different non-linear basis functions and the applied framework has also shown how to select possible variables which require to be considered or excluded to improve the model giving, as a by-product, some new insights on symptoms' pattern within each cycle, between different cycles and between different treatments. This added-value aspect can also be further researched and new insights correlated with clinical findings.

6.8 CONCLUSION

This work successfully built on a previous state-of-the-art tool for side-effect modelling on patients with cancer undergoing adjuvant chemotherapy treatment, following recommendations given in section 3.2. The proposed model has been both generally improved and may be reusable in different contexts.

Whilst the encouraging results reported in this small-scale study should be taken with care, they do illustrate the potential of this kind of a time series modelling approach.

For future work, further large-scale investigations using larger datasets can be carried out and different modelling techniques can be applied (using different types of basis functions) with a view to providing a reliable model that could be potentially deployed in clinical practice.

CONCLUSIONS AND FUTURE WORK

7.1 CONCLUSIONS

This thesis presented a sound framework for the development of CDSSs, and in particular of the underlying predictive models, identifying best practices for each stage of the model's development.

As outlined throughout the thesis, clinical decision support systems have a huge potential for leading to better care and appropriate implementations of such systems will promote a better use of medical knowledge. Three applications of the proposed framework have been developed.

Every step of the framework addressed specific problems emerging from relevant literature. Whenever possible, adopted solutions have been validated within the proposed case studies.

The first developed model predicted mortality within 28 days of patients suffering from acute alcoholic hepatitis. A comparison generated by state of the art tools shows an improved predictive power, demonstrating how the inclusion of suitable variables leads to an overall improved accuracy of the model. In this case, adopting an appropriate selection of variables, as

suggested by the developed framework, increased the predictive power of the model by 25% when compared to the widely used mDF, CPS and GAHS scores.

The second developed predictive model was designed to aid the early diagnosis of dementia, improving on the performance of a recent application of Bayesian belief networks [80] by means of a novel approach based on logistic regression. Our model, which adopts a statistical selection of variables according to the proposed general framework, outperformed the model which used variables selected by domain experts in previous studies. The obtained results outperform considered benchmarks by 15%.

The third built model predicted the probability of experiencing a certain symptom among common side-effects in patients receiving chemotherapy. The newly developed model included a pre-modelling stage, which was based on previous research studies [110], and a subsequent regression. The accuracy of the predictive results (computed on a daily basis for each cycle of therapy) shows that the newly proposed approach has enhanced its predictive power by 19% when compared to the previously developed models. Such improvements have been obtained by an appropriate usage of available a priori knowledge to pre-model the functional forms, as suggested by the proposed framework.

The results reported in the three case studies above demonstrate the utility of employing the proposed framework to address different kinds of problems with the aim of improving existing models and developing new predictive models. Such improvements are clearly envisaged to have an impact both on

clinical practice and further research. In particular, in terms of clinical impact, the proposed applications are expected to 1) help identify patients suffering from acute episodes of AH who may benefit from aggressive intervention to improve their survival rate, 2) aid the early diagnosis of dementia conditions, and 3) deliver better symptom management for patients suffering from breast, colorectal and lung cancer. Also, these three case studies may also be extended clinically, to validate the inclusion/exclusion of specific variables and to improve the predictive power in identified areas with a high presence of misclassifications.

7.2 FUTURE WORK

In terms of future work, large-scale studies are required to clinically validate the results from the presented case studies and to assess the feasibility of deploying such models in real clinical practice.

Also, the adoption of different cost functions may enhance regression based systems: as discussed, type I and type II errors can have different costs. With reference to the proposed case studies, it is reasonable to hypothesize that the cost of not identifying a patient (suffering from AH) who may benefit from aggressive intervention can be different from the cost of administering an aggressive treatment to a patient who could be treated with standard medications: this difference could for example be estimated by the mortality rate among the two groups of patients receiving the wrong treatments. Also,

the cost of further investigating possible early dementia conditions could generally differ from the cost of treating the same conditions at advanced stages following an incorrect early diagnosis: this difference could for example be estimated by the monetary costs of different medical examinations and treatments, as well as by other indicators measuring the patient's expected quality of life.

Further investigations may be able to determine such cost differences and implement an appropriate cost function in order to minimize overall "costs" of system misclassifications: resulting models will take into account the costs of false positives and false negatives. It is worth noticing that for a logistic regression model, the modification of the cost function could be quite straightforward (e.g. by multiplying each of the two addends of the summation reported in formula B.4 for the actual misclassification cost).

Also, different strategies for dealing with misclassifications should be investigated: the limited size of the considered datasets did not allow for the design of ensemble classifiers system, i.e. systems with two or more classification steps. As shown, misclassifications may not be uniformly distributed and, when this occurs, it is possible to isolate areas where the developed model is not accurate: in such areas, the system does not provide the user with any recommendation. This strategy, while increasing the overall predictive power of the model by reducing the overall number of misclassifications, also reduces the number of cases in which the model itself could be used.

It may be possible to build more complex models (or based on different variables) in the identified areas with a view to identifying different relationships between input and output. This could have a locally better predictive power when compared to the original model: further investigations are required to study whether this strategy could lead to more powerful systems able to cope also with areas where regressions fail to give accurate results.

Finally, the work presented in this thesis focused on regression models. Other machine learning techniques were not covered in this work, as they have not been used in the applications developed within this thesis essentially due to limitations imposed by the size of available datasets. Also, some of these techniques (such as neural networks or support vector machines) may not be able to clearly justify the relationship between the provided input and the model output. Some other techniques (e.g. decision trees) may be able to overcome this limitation, however they require a level of technical expertise which may prevent clinical researchers from using them. Future research studies should investigate how to extend and adapt the proposed framework in order to be used with other machine learning techniques.

INFORMATION AND COMMUNICATION TECHNOLOGY USAGE IN PATIENTS

Chapter 2 explored few aspects of the development process of models behind CDSSs: specifically, the chapter presented an investigation of the role of knowledge in such systems and how new medical knowledge could be extracted from data mining process. Key features for the successful integration of CDSSs in the actual clinical workflows were then presented. Finally, the focus was on how such models have been developed in the recent past, specifically using logistic regression, identifying common pitfalls in published works and suggesting a possible general framework for developing and evaluating such systems.

This appendix investigates the possibility of deployment of CDSSs to patients in order to improve healthcare, for example leading to a better symptom management, improving patients' quality of life, reducing number of visits and hospitalization, saving money and time resources.

Specifically, this analysis presents the results of a study which assessed the perceived usage of, and attitudes toward, communication technologies (i.e. mobile phone and texting, e-mail, and the internet) in patients attending a cardiology clinic with a view to understanding whether CDSSs may be

successfully deployed to patients and then guiding future health service redesign.

This study was performed in a remote regional hospital serving both urban and rural populations. A questionnaire was completed by a sample of 221 patients attending a general cardiology clinic. The questions asked about patients' access to and use of technology at home. Data collected also included age, gender, travel time to the clinic, mode of travel, and whether the respondent was accompanied to the clinic. Appropriate statistical tests were used with significance taken at the 0.05 level.

As probably expected, it has been ascertained that age was the strongest predictor of use of communication technologies, with younger patients more likely to use e-mail, web, mobile phone, and texting. However, frequency of use of e-mail was not related to age and an encouraging 99% of patients used at least one communication technology.

A.1 COMMUNICATION TECHNOLOGIES FOR HEALTHCARE

Ensuring equal access to good quality healthcare should be a core aim of each national health system. Effective communication between patients and healthcare staff is key to the delivery of this aim. Traditionally, communication occurs by face-to-face contact or written correspondence. However, there has been increasing interest in the use of simple digital communication technologies (including mobile phone and texting, e-mail, and the internet).

These technologies have the potential to be more rapid, responsive, and cost-effective in improving the quality of healthcare by enhancing the level of communication across organizational boundaries of healthcare provision, such as the primary-secondary care interface. Communication problems across this interface have long been identified as a source of frustration to patients and clinicians alike [112]. Furthermore, studies have shown that using such technologies can produce care clinically similar to that from face-to-face consultations with health professionals, improve patients' access to care, and reduce hospital and travel costs [113]. Overcoming geographical barriers to access is one of the principal challenges for providing health services to remote and rural areas [114].

However, despite the promise of the "digital age", its full potential in terms of healthcare and improved access to healthcare is not currently being realized. This is likely because of a variety of reasons, including technology limitations, although lack (or perceived lack) of technology skills and confidence in both patients and healthcare providers remains a major barrier to more widespread use [115, 116].

There has been a considerable increase in the use of communication technologies within the general public, and while it might be assumed that general skill levels are high, there are few published data on patients' knowledge and skills in this area. Specific training may overcome skills deficiency, but it remains likely that the successful widespread implementation of new technologies

into healthcare services will depend in a large part on the existing abilities of patients and healthcare providers.

In this appendix, we aim to assess the perceived usage of, and attitudes toward, communication technologies (mobile phone and texting, e-mail, and the internet) in patients attending a cardiology clinic with a view to guiding future health service redesign.

A.2 METHODOLOGY

Raigmore Hospital is situated in the north of Scotland (Highland Region) and serves a population of over 300,000 dispersed over a large geographical area (10,085 square miles). Approximately 70% of patients live within 1-hour travel time from the major hospital with good roads and public transport, but much of the rest of the population have considerable geographical hurdles to attending the clinic, including transport by plane and ferry from island locations. There are no peripheral specialist cardiology clinics in this area, and therefore all patients travel to Raigmore Hospital for specialist cardiology review.

A self-completion questionnaire-based survey was conducted on a convenience sample of patients attending a general cardiology clinic between February and May 2009. All patients attending the clinic were included, and questionnaires were distributed by the clerical staff at the clinic. These were

adult patients (>16 years old) with a range of cardiologic conditions. This included "new" and "return" patients.

In the absence of a validated instrument a questionnaire was developed in several iterative stages including a pilot study. The results from this pilot study allowed further questionnaire redesign until the final version was agreed upon. The questions asked about patients' access to and use of technology at home. Most questions required a yes/no response or used a 4-point rating scale. Respondent data included age, gender, travel time to the clinic, mode of travel, and whether the respondent was accompanied to the clinic.

Data were transposed from self-completed paper questionnaires into a spreadsheet. The p-values using an appropriate statistical test were used to assess the influence of age (unpaired t-test), gender (chi-squared test), and distance from the hospital (unpaired t-test) on the use of the internet, mobile phone, text messages, and e-mail. For testing frequencies of usage for each technology analysis of variance was used. Significance was taken at the 0.05 level.

A.3 RESULTS FROM QUESTIONNAIRE

In total, 221 patient responses were studied: 124 of them (57%) were male. The average age was 62.1 ± 14.1 years (range: 16-89 years). The self-reported places of residence were as follows: countryside, 40 (18%); city, 77 (35%); town, 36 (16%); and village, 68 (31%). The median travel time to hospital was 30 min

(range: 0-1 h). The majority (172, i.e. 78%) of patients used private transport to attend hospital, 23 (10%) used public transport, and 7 (3%) used hospital transport or ambulance.

The majority of patients had a CD or DVD player at home: 166 (75%) and 163 (74%), respectively. Many patients (136, i.e. 62%) had a home computer, and 111 (50%) reported having broadband access. With regard to patients' overall use of the four means of communication technologies, 14.0% never use two out of the four, 14.7% never use three out of the four, and only 0.7% never use any.

Gender and distance from the main urban center were not related to use of the Internet or mobile phone usage (or mobile texting function) (all $p > 0.05$). However, age was closely correlated with usage of communication technologies. On average, participants who used the communication technology were younger than those who did not for the Internet (56 ± 14 versus 68 ± 10 ; $p < 0.001$), mobile phone (59 ± 14 versus 73 ± 12 years; $p < 0.001$), mobile texting function (54 ± 13 versus 68 ± 10 ; $p < 0.001$), and e-mail (56 ± 14 versus 69 ± 11 years; $p < 0.001$). Furthermore, age was found to be a key factor in determining the frequency of use of mobile phones ($p < 0.001$) and texting ($p < 0.001$) but not the frequency of using the internet ($p = 0.43$) and e-mail ($p = 0.76$).

Many respondents (80, i.e. 36%) reported a wish to contact a doctor or nurse between clinic appointments by e-mail, 93 (42%) did not wish to do this, and the remainder did not respond to this question. Patients who had a computer

at home were much more likely to wish to use e-mail between appointments compared with those patients without a computer at home (76/136, i.e. 56% versus 3/78, i.e. 4%; $p < 0.001$). Once again, age was a key factor in determining whether patients desired to use e-mail to contact a doctor or nurse ($p < 0.001$).

A.4 ANALYSIS OF RESULTS

Increased use of communication technologies has the potential to improve patient care [117]. Implementation of these can be difficult, and patient skills and confidence with them will vary. Within the variables investigated in the present study, age was the strongest predictor of use of communication technologies (mobile phone and texting, e-mail, and Internet). Age was also a predictor of frequency of use of mobile phone and texting but not for e-mail and Internet. Patients who lived in more remote areas were no more likely to use communication technologies than those who lived in more urban areas. Successful and widespread implementation of communication technologies in healthcare is likely to depend on part on the availability within the general public. In our cohort only 62% of patients had a home computer, and only 50% had broadband connectivity, and this is likely to limit the immediate ability to develop web or e-mail based healthcare solutions to the majority of our population. However, the proportion of younger patients with computers is much higher, and as this cohort of patient ages there will be an increase in the proportion of patients with access to computers and mobile phones and the

skills to use them. The potential benefits for the delivery of healthcare services by greater use of technology are considerable. Traditional face-to-face medical outpatient clinics are a common way of assessment and monitoring patients and have been in place for many years, but many patients are seen in the clinic with little or no positive outcome in terms of treatment decisions, with some patients being seen as a matter of routine [118]. Alternative methods for reviewing patients may include telephone consultations, telemedicine, e-mail, or other web-based communication. These are of particular interest in remote areas where long distances can exist between patients and their healthcare providers.

Access to healthcare professionals is more difficult in remote areas for a variety of reasons [119]. The most recent Scottish Household Survey highlights that less than half of people in remote rural areas find access to hospital outpatient departments "very or fairly convenient"; public transport is also an issue, with 51% of remote and rural areas stating public transport services are convenient compared with 88% in large urban areas and 79% in accessible small towns [114]. Distance from specialists and specialist facilities - for example, cardiac catheterization facilities - is inversely proportional to the likelihood of patients receiving specialist investigation [120]. In some cases (e.g. chronic heart failure) this can result in poorer outcomes for patients in remote areas [121]. However, there is an increasing awareness that technology may help overcome barriers to healthcare delivery and equity of access particularly in remote areas [122].

Teleclinics and communication technology use may be of most benefit to rural patients in that they greatly reduce or abolish travel times. Instead of patients requiring days off work to attend a clinic, they can potentially receive specialist clinical review via videoconferencing at a local venue. Furthermore, between formal clinic appointment times, it may be difficult to reach a physician at a convenient time for both physician and patient. The use of e-mail or web-based communication may better enable communication between the patient and the doctor, obviating the necessity of both parties being available at the same time. This potentially could increase adherence to treatment plans and thus improve overall health. In one study in patients with congestive cardiac failure, the introduction of telemedicine increased medication compliance and improved physical and mental well-being at a relatively low cost [123]. In cases where individuals do not have Internet access the nearest healthcare centers may facilitate contact between the patient and hospital doctors, and where there is a lack of patient skills this could be facilitated by local healthcare professionals (e.g. community nurses).

Acceptability of new services will be important in their subsequent implementation. The use of text messages to remind patients for hospital appointments in pediatric services in Hull resulted in a decrease in nonattendance. It is important that over 90% of these patients were happy to receive a text reminder of appointments. With regard to videoconsultations a review of patients using psychiatric services noted that some patients preferred videoconferencing compared with traveling to the clinic [124]. Although

some individuals perceive such appointments as impersonal, they felt that an adequate doctor-patient relationship was established. This was particularly the case in patients who were previously known to the doctor where a relationship had been formed prior to the videoconferencing appointment, suggesting that videoconferencing may be more appropriate for review patients rather than new patient consultations. Nevertheless, the benefits of more frequent or convenient communication with a doctor appeared to be more important than potential subtle reductions in quality of consultation (i.e. the patients were less affected by the way in which doctors interacted with them but rather whether this interaction occurred at all) [125]. However, telecommunication is obviously less robust in terms of physical examination, and there is a danger that physical signs will be missed [126]. Furthermore, patients and physicians share some concerns regarding communication technology [127]. Security and confidentiality are a particular concern with e-mail and web. Physicians report concern regarding developing a good rapport with patients and the danger of missing nonverbal queues, although case selection should reduce these risks. In some instances physicians feared that the introduction of telemedicine would increase clinical workload. This concern was surrounding the hours that would be spent reviewing e-mails and responding to individuals [128]. Good documentation of working practice and appropriately supported service redesign should address these issues.

This current study has demonstrated that older patients are less likely to use communication technologies in day-to-day living, and there is a concern that

older patients will be denied the benefits from using healthcare technologies. However, studies have shown that if support is given, new technologies can be successfully used even in very elderly populations [126]. Another concept is that of "teleassistance" where individuals in the community facilitate communication between physicians via telemedicine; these individuals could be community nurses, general practitioners, or others, although using non-healthcare professionals may be ethically challenging, and confidentiality of patients should always be protected.

The use of technology in daily life is likely to continue to increase and with it the proportion of the population with specific communication technology skills. Already there are many examples of the use of communication technology in the delivery of healthcare (e.g. mobile phones, online booking, NHS repeat prescriptions service). However, many of these technologies have been used in younger patients (e.g. young diabetic patients) and therefore may not be so easily implemented in a general clinic population: indeed, in our cohort age was the dominant influence on the use of communication technologies. Furthermore, high-tech monitoring of patients both in the hospital and in a non-hospital setting will continue to be developed [129, 130].

This was a single center study, and therefore the results may not be applicable to other centers. Nevertheless this was a study of all comers and therefore represented a breadth of patients in terms of age and geographical locations. Indeed, although other studies have investigated the use of technologies in

younger patients, our cohort was unselected and therefore represents a more general cardiac clinic population.

A.5 CONCLUSION

This study has identified that the use of communication technologies is not widespread within the cardiology outpatient community. Age is the strongest predictor of use of communication technology, with younger patients more likely to use e-mail, the web, mobile phone, and texting. This study has highlighted that there may be several potential barriers to the widespread implementation of communication technology to a general cardiology clinic population. Cognizance should be taken of these findings when attempting service redesign.

MATHEMATICAL FORMULATION OF REGRESSIONS AND GRADIENT DESCENT

Regression analysis is a statistical method to determine the relationship between a dependent variable and a number of independent variables. Details about functional forms, cost functions and a minimization algorithm are reported in the following sections.

B.1 LINEAR REGRESSION

For linear regression, the relationship between input vector X and output is determined by the function f defined as follows:

$$f(B, X) = B^T X \quad (\text{B.1})$$

where X is the vector of predictor variables (with 1 as first element).

Given a training set with M data points $\{(X^{(1)}, y^{(1)}), (X^{(2)}, y^{(2)}) \dots (X^{(M)}, y^{(M)})\}$ with $X^{(m)} \in \mathbb{R}^{n+1}$, $x_0^{(m)} = 1$ and $y^{(m)} \in \mathbb{R}$, the cost function is set as follows:

$$J(B) = -\frac{1}{2M} \sum_{m=1}^M (B^T X^{(m)} - y^{(m)})^2 \quad (\text{B.2})$$

B.2 LOGISTIC REGRESSION

The model output is the probability $f(B, X)$ of a data point belonging to a certain class given known values of X (vector containing predictors). The general form of the functional dependence is expressed by formula 2.1 and could be rewritten as:

$$f(B, X) = \frac{1}{1 + e^{-B^T X}} \quad (\text{B.3})$$

where X is the vector of predictor variables (with 1 as first element) and B is the vector of coefficients to be determined by the logistic regression algorithm.

Given a training set with M data points $\{(X^{(1)}, y^{(1)}), (X^{(2)}, y^{(2)}) \dots (X^{(M)}, y^{(M)})\}$ with $X^{(m)} \in \mathbb{R}^{n+1}$, $x_0^{(m)} = 1$ and $y^{(m)} \in \{0, 1\}$, the cost function $J(B)$ to be minimized in order to find the coefficient vector B is:

$$J(B) = -\frac{1}{M} \left[\sum_{m=1}^M y^{(m)} \log f(B, X^{(m)}) + (1 - y^{(m)}) \log (1 - f(B, X^{(m)})) \right] \quad (\text{B.4})$$

B.3 GRADIENT DESCENT

In order to minimize the cost function an iterative methods is used: for example, the gradient descent algorithm could be used to carry out such task and few details are here reported.

Given a first solution guess $B^{[0]}$ and a learning rate α , coefficients at the $k + 1$ iteration are computed as follows:

$$b_j^{[k+1]} = b_j^{[k]} - \alpha \frac{\partial}{\partial b_j} J(B^{[k]}) \quad (\text{B.5})$$

That is:

$$\mathbf{B}^{[k+1]} = \mathbf{B}^{[k]} - \alpha \nabla J(\mathbf{B}^{[k]}) \quad (\text{B.6})$$

Coefficients vector \mathbf{B} for the model will be the one satisfying the following:

$$J(\mathbf{B}) = \min_{\theta} J(\theta) \quad (\text{B.7})$$

BIBLIOGRAPHY

- [1] T. Mazzocco, A. Hussain, S. Hussain, and A. A. Shah, "A novel mortality model for acute alcoholic hepatitis including variables recorded after admission to hospital," *Computers in biology and medicine*, vol. 44, pp. 132–135, 2014.
- [2] T. Mazzocco and A. Hussain, "Novel logistic regression models to aid the diagnosis of dementia," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3356–3361, 2012.
- [3] T. Mazzocco and A. Hussain, "A side-effects mapping model in patients with lung, colorectal and breast cancer receiving chemotherapy," in *e-Health Networking Applications and Services (Healthcom), 2011 13th IEEE International Conference on*, pp. 34–39, IEEE, 2011.
- [4] T. Gandiya, A. Dua, G. King, T. Mazzocco, A. Hussain, and S. J. Leslie, "Self-reported "communication technology" usage in patients attending a cardiology outpatient clinic in a remote regional hospital," *Telemedicine and e-Health*, vol. 18, no. 3, pp. 219–224, 2012.
- [5] G. Goertzel, "Clinical decision support system," *Annals of the New York Academy of Sciences*, vol. 161, no. 2, pp. 689–693, 1969.

- [6] V. Sintchenko and H. Garsden, "Clinical decision support: new approaches to usability study," *HIC 2002: Proceedings: Improving Quality by Lowering Barriers*, p. 32, 2002.
- [7] H. Kaur and S. K. Wasan, "Empirical study on applications of data mining techniques in healthcare," *Journal of Computer Science*, vol. 2, no. 2, p. 194, 2006.
- [8] D. Avison and T. Young, "Time to rethink health care and ICT?," *Communications of the ACM*, vol. 50, pp. 69–74, 2007.
- [9] M. Ashworth, P. Seed, D. Armstrong, S. Durbaba, and R. Jones, "The relationship between social deprivation and the quality of primary care: a national survey using indicators from the UK quality and outcomes framework," *British Journal of General Practice*, vol. 57, pp. 441–448, 2007.
- [10] J. Gammon, M. Heulwen, and D. Gould, "A review of the evidence for suboptimal compliance of healthcare practitioners to standard/universal infection control precautions," *Journal of Clinical Nursing*, vol. 17, pp. 157–167, 2007.
- [11] D. F. Lobach, "Electronically distributed, computer-generated, individualized feedback enhances the use of a computerized practice guideline," in *Proceedings of the AMIA Annual Fall Symposium*, p. 493, American Medical Informatics Association, 1996.
- [12] D. K. Litzelman, R. S. Dittus, M. E. Miller, and W. M. Tierney, "Requiring physicians to respond to computerized reminders improves their

- compliance with preventive care protocols," *Journal of general internal Medicine*, vol. 8, no. 6, pp. 311–317, 1993.
- [13] D. Hunt, R. Haynes, S. Hanna, and K. Smith, "Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review," *The Journal of the American Medical Association*, vol. 280, pp. 1339–1346, 1998.
- [14] A. X. Garg, N. K. Adhikari, H. McDonald, M. P. Rosas-Arellano, P. Devereaux, J. Beyene, J. Sam, and R. B. Haynes, "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes," *JAMA: the journal of the American Medical Association*, vol. 293, no. 10, pp. 1223–1238, 2005.
- [15] E. S. Berner, *Clinical Decision Support Systems*. Springer, 2007.
- [16] W. B. Kannel, D. McGee, and T. Gordon, "A general cardiovascular risk profile: the Framingham study," *The American journal of cardiology*, vol. 38, no. 1, pp. 46–51, 1976.
- [17] C. Hug, *Detecting hazardous intensive care patient episodes using real-time mortality models*. PhD thesis, Massachusetts Institute of Technology, 2009.
- [18] E. Coiera, "Clinical decision support systems," *Guide to health informatics*, vol. 2, no. 1, 2003.
- [19] M. J. Rantz, M. Skubic, R. J. Koopman, L. Phillips, G. L. Alexander, S. J. Miller, and R. D. Guevara, "Using sensor networks to detect urinary tract

- infections in older adults," in *e-Health Networking Applications and Services (Healthcom)*, 2011 13th IEEE International Conference on, pp. 142–149, IEEE, 2011.
- [20] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success," *Bmj*, vol. 330, no. 7494, p. 765, 2005.
- [21] P. S. Roshanov, N. Fernandes, J. M. Wilczynski, B. J. Hemens, J. J. You, S. M. Handler, R. Nieuwlaat, N. M. Souza, J. Beyene, H. G. C. V. Spall, A. X. Garg, and R. B. Haynes, "Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials," *BMJ*, vol. 346, 2013.
- [22] C. M. Clancy and K. Cronin, "Evidence-based decision making: global evidence, local decisions," *Health affairs*, vol. 24, no. 1, pp. 151–162, 2005.
- [23] V. Patkar, D. Acosta, T. Davidson, A. Jones, J. Fox, and M. Keshtgar, "Cancer multidisciplinary team meetings: evidence, challenges, and the role of clinical decision support technology," *International journal of breast cancer*, vol. 2011, 2011.
- [24] A. Wright, D. W. Bates, B. Middleton, T. Hongsermeier, V. Kashyap, S. M. Thomas, and D. F. Sittig, "Creating and sharing clinical decision support content with web 2.0: issues and examples," *Journal of biomedical informatics*, vol. 42, no. 2, pp. 334–346, 2009.

- [25] R. Scales and M. Embrechts, "Computational intelligence techniques for medical diagnostics," in *Proceedings of Walter Lincoln Hawkins, Graduate Research Conference from the World Wide Web: <http://www.cs.rpi.edu/~bivenj/MRC/proceedings/papers/researchpaper.pdf>*, 2002.
- [26] K. C. Desouza, "Knowledge management in hospitals: a process oriented view and staged look at managerial issues," *International journal of healthcare technology and management*, vol. 4, no. 6, pp. 478–497, 2002.
- [27] S. K. Wasan, V. Bhatnagar, and H. Kaur, "The impact of data mining techniques on medical diagnostics," *Data Science Journal*, vol. 5, pp. 119–126, 2006.
- [28] K. J. Ottenbacher, H. R. Ottenbacher, L. Tooth, and G. V. Ostir, "A review of two journals found that articles using multivariable logistic regression frequently did not report commonly recommended assumptions," *Journal of clinical epidemiology*, vol. 57, no. 11, pp. 1147–1152, 2004.
- [29] J. Concato, A. R. Feinstein, and T. R. Holford, "The risk of determining risk with multivariable models," *Annals of internal medicine*, vol. 118, no. 3, pp. 201–210, 1993.
- [30] S. C. Bagley, H. White, and B. A. Golomb, "Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain," *Journal of clinical epidemiology*, vol. 54, no. 10, pp. 979–985, 2001.

- [31] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Data preprocessing for supervised learning," *International Journal of Computer Science*, vol. 1, no. 2, pp. 111–117, 2006.
- [32] D. B. Suits, "Use of dummy variables in regression equations," *Journal of the American Statistical Association*, vol. 52, no. 280, pp. 548–551, 1957.
- [33] C. Dougherty, *Introduction to econometrics*, vol. 2. Oxford University Press Oxford, 2002.
- [34] A. Field, *Discovering Statistics Using SPSS 3th (third) edition*. Sage Publications Ltd, 2010.
- [35] A. R. Feinstein, *Multivariable analysis: an introduction*. Yale University Press, 1996.
- [36] R. L. Ott, M. Longnecker, and L. Ott, *A first course in statistical methods*. Thomson-Brooks/Cole, 2004.
- [37] S. Menard, *Applied logistic regression analysis*. No. 106, Sage, 2002.
- [38] R. H. Myers, *Classical and modern regression with applications*, vol. 2. Duxbury Press Belmont, CA, 1990.
- [39] G. H. John, R. Kohavi, K. Pfleger, *et al.*, "Irrelevant features and the subset selection problem," in *ICML*, vol. 94, pp. 121–129, 1994.
- [40] D. P. MacKinnon, J. L. Krull, and C. M. Lockwood, "Equivalence of the mediation, confounding and suppression effect," *Prevention Science*, vol. 1, no. 4, pp. 173–181, 2000.

- [41] M. A. Babyak, "What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models," *Psychosomatic medicine*, vol. 66, no. 3, pp. 411–421, 2004.
- [42] J. Concato and A. R. Feinstein, "Monte Carlo methods in clinical research: applications in multivariable analysis," *Journal of investigative medicine: the official publication of the American Federation for Clinical Research*, vol. 45, no. 6, p. 394, 1997.
- [43] F. E. Harrell Jr, K. L. Lee, D. B. Matchar, T. A. Reichert, *et al.*, "Regression models for prognostic prediction: advantages, problems, and suggested solutions," *Cancer treatment reports*, vol. 69, no. 10, pp. 1071–1077, 1985.
- [44] P. Peduzzi, J. Concato, E. Kemper, T. R. Holford, and A. R. Feinstein, "A simulation study of the number of events per variable in logistic regression analysis," *Journal of clinical epidemiology*, vol. 49, no. 12, pp. 1373–1379, 1996.
- [45] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *IJCAI*, vol. 14, pp. 1137–1145, 1995.
- [46] B. Efron, "Bootstrap methods: another look at the jackknife," *The annals of Statistics*, pp. 1–26, 1979.
- [47] J.-H. Kim, "Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap," *Computational Statistics & Data Analysis*, vol. 53, no. 11, pp. 3735–3745, 2009.

- [48] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [49] E. Lidbrink, J. Elfving, J. Frisell, and E. Jonsson, "Neglected aspects of false positive findings of mammography in breast cancer screening: analysis of false positive cases from the Stockholm trial," *BMJ*, vol. 312, pp. 273–276, 1996.
- [50] M. Nano, J. Kollias, G. Farshid, P. Gill, and M. Bochner, "Clinical impact of false-negative sentinel node biopsy in primary breast cancer," *British Journal of Surgery*, vol. 11, pp. 1430–1434, 2002.
- [51] I. Gram, E. Lund, and S. Slenker, "Quality of life following a false positive mammogram," *British Journal of Cancer*, vol. 62, pp. 1018–1022, 1990.
- [52] M. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine," *Clinical Chemistry*, vol. 39, pp. 561–577, 1993.
- [53] S. Sougioultzis, E. Dalakas, P. C. Hayes, and J. N. Plevris, "Alcoholic hepatitis: from pathogenesis to treatment," *Current Medical Research and Opinion*, vol. 21, no. 9, pp. 1337–1346, 2005. PMID: 16197651.
- [54] M. Fox, J. Fox, and M. Davies, "Diagnosis and management of chronic liver disease in older people," *Reviews in Clinical Gerontology*, vol. 21, pp. 1–15, 2 2011.

- [55] R. Bruha, K. Dvorak, and J. Petrtyl, "Alcoholic liver disease," *World journal of hepatology*, vol. 4, no. 3, p. 81, 2012.
- [56] W. C. Maddrey, J. K. Boitnott, M. S. Bedine, F. L. Weber Jr, E. Mezey, R. I. White Jr, *et al.*, "Corticosteroid therapy of alcoholic hepatitis," *Gastroenterology*, vol. 75, no. 2, pp. 193–199, 1978.
- [57] R. N. H. Pugh, I. M. Murray-Lyon, J. L. Dawson, M. C. Pietroni, and R. Williams, "Transection of the oesophagus for bleeding oesophageal varices," *British Journal of Surgery*, vol. 60, no. 8, pp. 646–649, 1973.
- [58] C. Child, "Ill: The liver and portal hypertension," *Philadelphia, WB Saunders Co*, p. 23, 1964.
- [59] E. H. Forrest, C. D. J. Evans, S. Stewart, M. Phillips, Y. H. Oo, N. C. McAvoy, N. C. Fisher, S. Singhal, A. Brind, G. Haydon, J. O'Grady, C. P. Day, P. C. Hayes, L. S. Murray, and A. J. Morris, "Analysis of factors predictive of mortality in alcoholic hepatitis and derivation and validation of the Glasgow alcoholic hepatitis score," *Gut*, vol. 54, no. 8, pp. 1174–1179, 2005.
- [60] W. Dunn, L. H. Jamil, L. S. Brown, R. H. Wiesner, W. R. Kim, K. V. N. Menon, M. Malinchoc, P. S. Kamath, and V. Shah, "MELD accurately predicts mortality in patients with alcoholic hepatitis," *Hepatology*, vol. 41, no. 2, pp. 353–358, 2005.
- [61] A. Duseja, N. S. Choudhary, S. Gupta, R. K. Dhiman, and Y. Chawla, "Apache ii score is superior to sofa, ctp and meld in predicting the short-

- term mortality in patients with acute-on-chronic liver failure (ACLF)," *Journal of Digestive Diseases*, pp. 484–490, 2013.
- [62] M. Sheth, M. Riggs, and T. Patel, "Utility of the Mayo End-Stage Liver Disease (MELD) score in assessing prognosis of patients with alcoholic hepatitis," *BMC gastroenterology*, vol. 2, no. 1, p. 2, 2002.
- [63] A. Chedid, C. Mendenhall, P. Gartside, S. French, T. Chen, and L. Rabin, "Prognostic factors in alcoholic liver disease. A cooperative study group," *The American journal of gastroenterology*, vol. 86, pp. 210–216, 02 1991.
- [64] S. Masson, I. Emmerson, E. Henderson, E. Fletcher, A. Burt, C. Day, and S. Stewart, "PWE-285 Clinical but not histological factors predict long-term prognosis in patients with biopsy proven advanced alcoholic liver disease," *Gut*, vol. 61, no. Suppl 2, pp. A413–A414, 2012.
- [65] L. Castera, "Noninvasive methods to assess liver disease in patients with hepatitis B or C," *Gastroenterology*, vol. 142, pp. 1293–1302.e4, 05 2012.
- [66] W. Srikureja, N. L. Kyulo, B. A. Runyon, and K.-Q. Hu, "MELD score is a better prognostic model than Child-Turcotte-Pugh score or Discriminant Function score in patients with alcoholic hepatitis," *Journal of Hepatology*, vol. 42, no. 5, pp. 700 – 706, 2005.
- [67] W. Maddrey, "Alcoholic liver disease," *Current hepatology*, vol. 1, pp. 71–85, 1995.

- [68] P. Mathurin, C. L. Mendenhall, R. L. C. Jr, M.-J. Ramond, W. C. Maddrey, P. Garstide, B. Rueff, S. Naveau, J.-C. Chaput, and T. Poynard, "Corticosteroids improve short-term survival in patients with severe alcoholic hepatitis (AH): individual data analysis of the last three randomized placebo controlled double blind trials of corticosteroids in severe AH," *Journal of Hepatology*, vol. 36, no. 4, pp. 480 – 487, 2002.
- [69] T. F. Imperiale and A. J. McCullough, "Do corticosteroids reduce mortality from alcoholic hepatitis? A meta-analysis of the randomized trials," *Annals of Internal Medicine*, vol. 113, no. 4, pp. 299–307, 1990.
- [70] "Management of patients with dementia: a national clinical guideline," tech. rep., Scottish Intercollegiate Guidelines Network, 2006.
- [71] S. Turner, S. Iliffe, M. Downs, J. Wilcock, M. Bryans, E. Levin, J. Keady, and R. O'Carroll, "General practitioners' knowledge, confidence and attitudes in the diagnosis and management of dementia," *Age and ageing*, vol. 33, no. 5, pp. 461–467, 2004.
- [72] S. Cahill, M. Clark, H. O'Connell, B. Lawlor, R. Coen, and C. Walsh, "The attitudes and practices of general practitioners regarding dementia diagnosis in Ireland," *International journal of geriatric psychiatry*, vol. 23, no. 7, pp. 663–669, 2008.
- [73] E. C. Hansen, C. Hughes, G. Routley, and A. L. Robinson, "General practitioners' experiences and understandings of diagnosing dementia:

- factors impacting on early diagnosis," *Social science & medicine*, vol. 67, no. 11, pp. 1776–1783, 2008.
- [74] S. Klöppel, C. M. Stonnington, J. Barnes, F. Chen, C. Chu, C. D. Good, I. Mader, L. A. Mitchell, A. C. Patel, C. C. Roberts, *et al.*, "Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method," *Brain*, vol. 131, no. 11, pp. 2969–2974, 2008.
- [75] M. J. Prince, J. L. de Rodriguez, L. Noriega, A. Lopez, D. Acosta, E. Albanese, R. Arizaga, J. R. Copeland, M. Dewey, C. P. Ferri, *et al.*, "The 10/66 dementia research group's fully operationalised DSM-IV dementia computerized diagnostic algorithm, compared with the 10/66 dementia algorithm and a clinician diagnosis: a population validation study," *BMC Public Health*, vol. 8, no. 1, p. 219, 2008.
- [76] M. Prince, C. P. Ferri, D. Acosta, E. Albanese, R. Arizaga, M. Dewey, S. I. Gavrilova, M. Guerra, Y. Huang, K. Jacob, *et al.*, "The protocols for the 10/66 dementia research group population-based research programme," *BMC Public Health*, vol. 7, no. 1, p. 165, 2007.
- [77] L. Oteniya, R. Coles, and J. Cowie, "DemNet: a clinical decision support system to aid the diagnosis of dementia," in *Proceedings of the 22ndHealthCare Computing Conference*, pp. 289–297, Citeseer, 2005.
- [78] F. V. Jensen and T. D. Nielsen, *Bayesian networks and decision graphs*. Springer, 2007.

- [79] J. Cowie, L. Oteniya, and R. Coles, "Diagnosis of dementia and its pathologies using Bayesian belief networks," in *ICEIS (2)*, pp. 291–295, 2006.
- [80] L. Oteniya, *Bayesian belief networks for dementia diagnosis and other applications: a comparison of hand-crafting and construction using a novel data driven technique*. PhD thesis, University of Stirling, 2008.
- [81] R. Petersen, J. Stevens, M. Ganguli, E. Tangalos, J. Cummings, and S. DeKosky, "Practice parameter: early detection of dementia. Mild cognitive impairment (an evidence-based review). Report of the quality standards subcommittee of the American Academy of Neurology," *Neurology*, vol. 56, no. 9, pp. 1133–1142, 2001.
- [82] H. Brodaty and C. M. Moore, "The clock drawing test for dementia of the Alzheimer's type: a comparison of three scoring methods in a memory disorders clinic," *International journal of geriatric psychiatry*, vol. 12, no. 6, pp. 619–627, 1997.
- [83] J. Moroney, E. Bagiella, D. Desmond, V. C. Hachinski, P. Mölsä, L. Gustafson, A. Brun, P. Fischer, T. Erkinjuntti, W. Rosen, *et al.*, "Meta-analysis of the Hachinski Ischemic Score in pathologically verified dementias," *Neurology*, vol. 49, no. 4, pp. 1096–1105, 1997.
- [84] S. Daskalaki, I. Kopanas, and N. Avouris, "Evaluation of classifiers for an uneven class distribution problem," *Applied artificial intelligence*, vol. 20, no. 5, pp. 381–417, 2006.

- [85] D. D. R. Cox, *The analysis of binary data*, vol. 32. CRC Press, 1989.
- [86] N. J. Nagelkerke, "A note on a general definition of the coefficient of determination," *Biometrika*, vol. 78, no. 3, pp. 691–692, 1991.
- [87] *Cancer statistics: registrations, England, 2007*. Office for National Statistics, 2009.
- [88] *Cancer Incidence*. ISD Scotland (NHS), 2010.
- [89] *Cancer Incidence in Wales, 2003-2007*. Public Health Wales, 2009.
- [90] *Cancer Incidence and Mortality*. Northern Ireland Cancer Registry, 2009.
- [91] *The Diagnosis and Treatment of Lung Cancer*. National Collaborating Centre for Cancer, Cardiff (Wales), 2005.
- [92] *Early and locally advanced breast cancer: diagnosis and treatment*. National Collaborating Centre for Cancer, Cardiff (Wales), 2009.
- [93] A. Molassiotis, C. Stricker, B. Eaby, L. Velders, and P. Coventry, "Understanding the concept of chemotherapy-related nausea: the patient experience," *European Journal of Cancer Care*, vol. 17, pp. 444–453, 2008.
- [94] N. Kuderer, D. Dale, J. Crawford, L. Coser, and G. Lyman, "Mortality, morbidity, and cost associated with febrile neutropenia in adult cancer patients," *Cancer*, vol. 106, pp. 2258–2266, 2006.
- [95] G. Bonadonna, P. Valaguassa, A. Moliterni, and C. Brambilla, "Adjuvant cyclophosphamide, methotrexate, and fluorouracil in node positive

- breast cancer patients: the results of 20 years of follow-up," *The New England Journal of Medicine*, vol. 332, pp. 901–906, 1995.
- [96] D. Carr, L. Goudas, D. Lawrence, W. Pirl, J. Lau, D. DeVine, B. Kulpelnick, and K. Miller, "Management of cancer symptoms: pain, depression, and fatigue," tech. rep., Agency for Healthcare Research and Quality, 2002.
- [97] M. Groenvold, P. Fayers, M. Petersen, M. Sprangers, N. Aaronson, and H. Mouridsen, "Breast cancer patients on adjuvant chemotherapy report a wide range of problems not identified by health-care staff," *Breast Cancer Research and Treatment*, vol. 103, pp. 185–195, 2007.
- [98] M. Jefford and M. H. Tattersall, "Informing and involving cancer patients in their own care," *The Lancet Oncology*, vol. 3, pp. 629–637, 2002.
- [99] E. L. Poleshuck, J. Katz, C. H. Andrus, L. A. Hogan, B. F. Jung, D. I. Kulick, and R. H. Dworkin, "Risk factors for chronic pain following breast cancer surgery: a prospective study," *The Journal of Pain*, vol. 7, pp. 626–634, 2006.
- [100] J. Armer, M. Radina, D. Porock, and S. Culbertson, "Predicting breast cancer-related lymphedema using self-reported symptoms," *Nursing Research*, vol. 52, pp. 370–379, 2003.
- [101] J. A. Talcott, J. Manola, J. A. Clark, I. Kaplan, C. J. Beard, S. P. Mitchell, R. C. Chen, M. P. O'Leary, P. W. Kantoff, and A. V. D'Amico, "Time course and predictors of symptoms after primary prostate cancer therapy," *Journal of Clinical Oncology*, vol. 21, pp. 3979–3986, 2003.

- [102] G. Dranitsaris, D. Rayson, M. Vincent, J. Chang, K. Gelmon, D. Sandor, and G. Reardon, "The development of a predictive model to estimate cardiotoxic risk for patients with metastatic breast cancer receiving anthracyclines," *Breast Cancer Research and Treatment*, vol. 107, pp. 443–450, 2008.
- [103] A. Louis, T. Turner, M. Gretton, A. Baksh, and J. Cleland, "A systematic review of telemonitoring for the management of heart failure," *European Journal of Heart Failure*, vol. 5, pp. 583–590, 2003.
- [104] N. Kearney, L. McCann, J. Norrie, L. Taylor, P. Gray, M. McGee-Lennon, M. Sage, M. Miller, and R. Maguire, "Evaluation of a mobile phone-based, advanced symptom management system (ASyMS©) in the management of chemotherapy-related toxicity," *Supportive Care in Cancer*, vol. 17, pp. 437–444, 2009.
- [105] N. Kearney, L. Muir, M. Miller, I. Hargan, and P. Gray, "Using handheld computers to support patients receiving outpatient chemotherapy," *European Journal of Cancer Supplements*, vol. 1, p. S368, 2003.
- [106] R. Maguire, L. McCann, M. Miller, and N. Kearney, "Nurse's perceptions and experiences of using of a mobile-phone-based Advanced Symptom Management System (ASyMS©) to monitor and manage chemotherapy-related toxicity," *European Journal of Oncology Nursing*, vol. 12, pp. 380–386, 2008.

- [107] N. Kearney, L. Kidd, M. Miller, M. Sage, J. Khorrami, M. McGee, J. Cassidy, K. Niven, and P. Gray, "Utilising handheld computers to monitor and support patients receiving chemotherapy: results of a UK-based feasibility study," *Supportive Care in Cancer*, vol. 14, pp. 742–752, 2006.
- [108] R. Maguire, M. Miller, M. Sage, J. Norrie, L. McCann, L. Taylor, and N. Kearney, "Results of a UK based pilot study of a mobile phone based advanced symptom management system (ASyMS©) in the remote monitoring of chemotherapy related toxicity," *Clinical Effectiveness in Nursing*, vol. 9, pp. 202–210, 2005.
- [109] L. Forbat, R. Maguire, L. McCann, N. Illingworth, and N. Kearney, "The use of technology in cancer care: applying Foucault's ideas to explore the changing dynamics of power in health care," *Journal of Advanced Nursing*, vol. 65, pp. 306–315, 2009.
- [110] R. Maguire, J. Cowie, C. Leadbetter, K. McCall, K. Swingler, L. McCann, and N. Kearney, "The development of a side effect risk assessment tool (ASyMS©-SERAT) for use in patients with breast cancer undergoing adjuvant chemotherapy," *Journal of Research in Nursing*, vol. 14, pp. 27–40, 2009.
- [111] L. McCann, R. Maguire, M. Miller, and N. Kearney, "Patients' perceptions and experiences of using a mobile phone based Advanced Symptom

- Management System (ASyMS©) to monitor and manage chemotherapy related toxicity," *European Journal of Cancer Care*, vol. 18, pp. 156–164, 2009.
- [112] R. Harrison, W. Clayton, and P. Wallace, "Can telemedicine be used to improve communication between primary and secondary care?," *Bmj*, vol. 313, no. 7069, pp. 1377–1380, 1996.
- [113] J. L. DelliFraine and K. H. Dansky, "Home-based telehealth: a review and meta-analysis," *Journal of Telemedicine and Telecare*, vol. 14, no. 2, pp. 62–66, 2008.
- [114] J. Deaville, *The nature of rural general practice in the UK: preliminary research*. Institute of Rural Health Tregynon, Powys, 2001.
- [115] S. J. Katz, N. Nissan, and C. A. Moyer, "Crossing the digital divide: evaluating online communication between patients and their providers," *Am J Manag Care*, vol. 10, no. 9, pp. 593–598, 2004.
- [116] S. J. Leslie, M. Hartswood, C. Meurig, S. P. McKee, R. Slack, R. Procter, and M. A. Denvir, "Clinical decision support software for management of chronic heart failure: development and evaluation," *Computers in biology and medicine*, vol. 36, no. 5, pp. 495–506, 2006.
- [117] K. D. Mandl, I. S. Kohane, and A. M. Brandt, "Electronic patient-physician communication: problems and promise," *Annals of internal Medicine*, vol. 129, no. 6, pp. 495–500, 1998.

- [118] M. L. Hughes, S. J. Leslie, G. K. McInnes, K. McCormac, and N. R. Peden, "Can we see more outpatients without more doctors?," *Journal of the Royal Society of Medicine*, vol. 96, no. 7, pp. 333–337, 2003.
- [119] T. Raza, M. Joshi, R. M. Schapira, and Z. Agha, "Pulmonary telemedicine: a model to access the subspecialist services in underserved rural areas," *International Journal of Medical Informatics*, vol. 78, no. 1, pp. 53–59, 2009.
- [120] M. MacLeod, A. Finlayson, J. Pell, and I. Findlay, "Geographic, demographic, and socioeconomic variations in the investigation and management of coronary heart disease in Scotland," *Heart*, vol. 81, no. 3, pp. 252–256, 1999.
- [121] R. Clark, K. Eckert, S. Stewart, S. Phillips, J. Yallop, A. Tonkin, and H. Krum, "Rural and urban differentials in primary care management of chronic heart failure: new data from the CASE study," *Medical Journal of Australia*, vol. 186, no. 9, pp. 441–445, 2007.
- [122] E. Rygh and P. Hjortdahl, "Continuous and integrated health care services in rural areas. A literature study," *Rural and Remote Health*, vol. 7, no. 3, p. 766, 2007.
- [123] P. Whitten, A. Bergman, M. A. Meese, K. Bridwell, and K. Jule, "St. Vincent's home telehealth for congestive heart failure patients," *Telemedicine and e-Health*, vol. 15, no. 2, pp. 148–153, 2009.
- [124] S. Simpson, J. Knox, D. Mitchell, J. Ferguson, J. Brebner, and E. Brebner, "A multidisciplinary approach to the treatment of eating disorders via

- videoconferencing in north-east Scotland," *Journal of telemedicine and telecare*, vol. 9, no. suppl 1, pp. 37–38, 2003.
- [125] B. J. Wakefield, J. E. Holman, A. Ray, M. Scherubel, T. L. Burns, M. G. Kienzle, and G. E. Rosenthal, "Outcomes of a home telehealth intervention for patients with heart failure," *Journal of telemedicine and telecare*, vol. 15, no. 1, pp. 46–50, 2009.
- [126] C. Chandhanayingyong, B. Tangtrakulwanich, and T. Kiriratnikom, "Teleconsultation for emergency orthopaedic patients using the multimedia messaging service via mobile phones," *Journal of telemedicine and telecare*, vol. 13, no. 4, pp. 193–196, 2007.
- [127] G. King, H. Richards, and D. Godden, "Adoption of telemedicine in Scottish remote and rural general practices: a qualitative study," *Journal of telemedicine and telecare*, vol. 13, no. 8, pp. 382–386, 2007.
- [128] B. Bergeron, "E-mail: a realistic conduit for patient-doctor communications?," *The Journal of medical practice management: MPM*, vol. 15, no. 4, pp. 208–210, 1999.
- [129] M. S. Manikandan and S. Dandapat, "Wavelet-based ECG and PCG signals compression technique for mobile telemedicine," in *Advanced Computing and Communications, 2007. ADCOM 2007. International Conference on*, pp. 164–169, IEEE, 2007.
- [130] M. F. A. Rasid and B. Woodward, "Bluetooth telemedicine processor for multichannel biomedical signal transmission via mobile cellular

networks," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 9, no. 1, pp. 35-43, 2005.