Toward Intelligent Optimization of Brace Treatment for Adolescent Idiopathic Scoliosis

by

David Eric Chalmers

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Biomedical Engineering

Department of Electrical and Computer Engineering
University of Alberta

Abstract

Electronic decision support systems have the potential to improve healthcare practices in many domains. This thesis investigates the use of data-driven decision support to help optimize brace treatment for children who have Adolescent Idiopathic Scoliosis (AIS).

AIS is a spinal deformity affecting 2-3% of adolescents. If left untreated, AIS may progress (worsen), negatively affecting the adolescent's emotional, social, and physical wellbeing and eventually necessitating surgical intervention. Brace treatment is the most common non-surgical treatment for AIS; in brace treatment a back brace applies corrective pressure to the torso, with the goal of preventing progression. Patients' faithfulness in wearing the brace as long and as tightly as prescribed affects treatment outcome. But the outcome also depends on patient characteristics, the nature of the deformity, and many other factors in addition to compliance. The relationships between these factors and treatment outcome are complex and not perfectly understood; as a result, brace treatment outcome is difficult to predict. As technology improves our ability to predict treatment outcome, the ability to optimize treatment protocols for individual AIS patients should improve as well.

This research envisions a complete system for collecting patient data and using it to generate treatment recommendations for new patients. In this system, electronic sensors collect information about patients' brace-wear habits, machine-learning techniques use sensor and other data to train prediction models, and these models' predictions of new patients' outcomes are used to customize treatment protocols to those patients. This work developed the components of this system and implemented them in a scalable hardware/software platform. Data from 31 patients was collected and processed by the system. Simulations were used to provide an initial assessment of the system's treatment recommendations.

Preface

The research described in this thesis received ethics approval from the Health Research Ethics Board – Health Panel, University of Alberta, project name: "Understanding biomechanical action of a brace for the treatment of scoliosis", reference number: Pro00003495.

Portions of the material in this thesis have been published in the following papers:

- E Chalmers, W Pedrycz, and E Lou, "Predicting the Outcome of Brace Treatment for Scoliosis using Conditional Fuzzy Clustering," Joint Congress of the International Fuzzy Systems Association and North American Fuzzy Information Processing Society, June 24-28 2013

  Some material from this conference paper appears in chapter 3 of the thesis. I conceived of and executed the work described in the paper. Dr. Pedrycz provided technical advice and direction, and assisted with manuscript preparation. Dr. Lou assisted with manuscript preparation.

- E Chalmers, W Pedrycz, and E Lou, "Human experts' and a fuzzy model's predictions of outcomes of scoliosis treatment: A comparative analysis", IEEE Transactions on Biomedical Engineering, 62(3): 1001-1007, 2015

  Material from this paper appears in chapter 3 of the thesis. I conceived of and executed the work described in this paper. I composed the manuscript with assistance from Dr. Predrycz and Dr. Lou.

- E Chalmers, D Hill, H Zhao, and E Lou, "Prescriptive analytics applied to brace treatment for AIS: a pilot demonstration", Scoliosis, 10(Suppl 2): S13, 2015

  Material from this paper appears in chapter 3 of the thesis. I conceived of and executed the work described in this paper. Mr. Hill, Dr. Zhao, and Dr. Lou assisted with manuscript preparation.

- E Chalmers, E Lou E, D Hill, V Zhao, "An Advanced Compliance Monitor for Patients Undergoing Brace Treatment for Idiopathic Scoliosis", Medical Engineering & Physics, 37(2):203-209, 2015

  Material from this paper appears in chapters 4 and 5. Dr. Lou conceived of the experiment described in the paper. I was responsible for designing the electronic device described and collecting and analyzing data. I composed the manuscript with assistance from all co-authors.

- Chalmers E, Mizianty M, Parent E, Yuan Y, Lou E, “Toward maximum-predictive-value classification”, Pattern Recognition, 47(12): 3949-3958, 2014

  Brief excerpts from this paper appear in chapter 6 of the thesis. I conceived of and executed the work described in this paper, with technical advice from Dr. Mizianty, Dr. Parent, and Dr. Yuan. I composed the manuscript with assistance from all co-authors.

References to these papers appear in footnotes in the appropriate sections of this thesis.

Acknowledgements

I would like to gratefully acknowledge the financial support I've received from the Natural Sciences and Engineering Research Council of Canada (NSERC), TD Bank Financial Group, the Women and Children's Health Research Institute (WCHRI), and Alberta Innovates. Without this support I would not have been able to pursue the work described in this thesis.

Dr. Lou has been the perfect supervisor. His mentorship has provided the guidance I needed as well as the freedom to pursue my own ideas and interests. My co-supervisor Dr. Zhao and committee member Dr. Duke have also provided needed encouragement, support, and advice. My gratitude goes out to the long list of people whom I have approached looking for advice or mentorship during my PhD: Doug Hill, Dr. Pedrycz, Dr. Kurgan, Dr. Parent, Dr. Mizianty, and Dr. Yuan, to name a few. I have always felt comfortable approaching these people and have always come away enriched.

Finally, I would like to thank my loving wife, Lena. Her love, support, and patience have been unlimited.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIS: Adolescent Idiopathic Scoliosis

ANN: Artificial Neural Network

CART: Classification and Regression Tree

CDSS: Clinical Decision Support System

CFCM: Conditional Fuzzy C-Means

CTS: Clinical Trial Simulation

FCM: Fuzzy C-Means

KDDM: Knowledge Discovery and Data Mining

MCC: Mathew's Correlation Coefficient

NPV: Negative Predictive Value

PCA: Principle Component Analysis

PDF: Probability Density Function

PMF: Probability Mass Function

PODSS: Prediction and Optimization-based Decision Support System

PPV: Positive Predictive Value

SQL: Structured Query Language

SVM: Support Vector Machine

SOSORT: Society on Scoliosis Orthopaedic and Rehabilitation Treatment

SRS: Scoliosis Research Society

TLSO: Thoraco-Lumbo-Sacral Orthosis

XML: eXtensible Markup Language

1. Introduction

This chapter provides a brief overview of brace treatment for children with Adolescent Idiopathic Scoliosis, and a formulation of the problem: optimization of brace treatment. It discusses how an electronic decision support system might be used to optimize brace treatment. The objectives of the work are stated, and the approach taken is summarized. Finally an outline of the remainder of this thesis is provided.

1.1. Brace treatment for Adolescent Idiopathic Scoliosis

Adolescent Idiopathic Scoliosis (AIS) is a three-dimensional deformity of the spine, affecting 2-3% of adolescents. Scoliosis affects adolescents' emotional and social wellbeing, and can cause physiological problems in severe cases. If left untreated the deformity might progress in severity until surgical correction is required.

Brace treatment is the most common non-surgical treatment for AIS. In brace treatment, a custom brace counteracts the Scoliotic curve by applying targeted pressure to the torso. Figure 1 shows a radiograph of a scoliotic spine, and one example of a brace. There are a myriad of brace types – worn for different portions of the day, constructed with different or different amounts of material, and using slightly different theories of curve correction – but all have a corrective

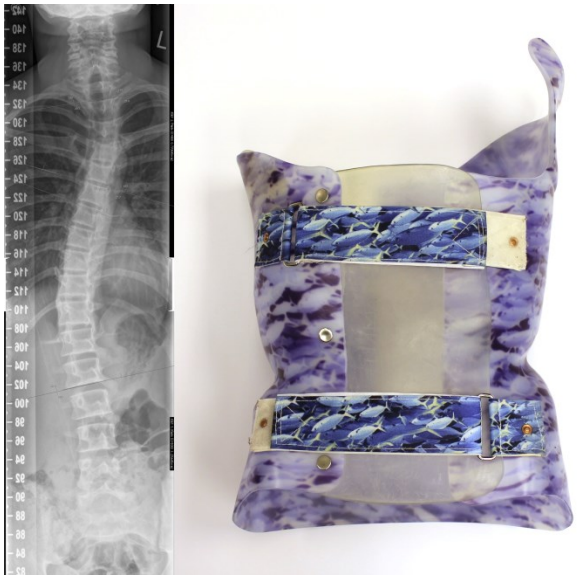


Figure 1: A radiographic image of a Scoliotic spine (left), and a brace used to counteract progression (right).

effect on the spinal curve while they are worn. The goal of the brace is simply to prevent progression of the deformity as the adolescent grows. The brace is usually worn until skeletal maturity.

Recent research suggests that a well-built and properly worn brace can lower the risk of Scoliosis progression, reducing the incidence of surgical intervention. However the factors influencing treatment success and their precise relationship to treatment outcome are not clear. As a result, predicting which individual cases will progress is a difficult task. It seems that treatment outcome is influenced by the patient's characteristics, and the nature, location, and severity of their Scoliosis. It is also apparent that the amount of pressure applied by the brace is positively correlated with the brace's immediate corrective effect [1], [2] and that this corrective effect is positively correlated with treatment success [3]. There is a further correlation between duration of brace-wear per day and treatment success [4], [5]. These relationships are documented in literature, but somewhat rudimentary: a more refined mathematical model of brace treatment would be needed to predict treatment outcome with much accuracy.

Since the relationships between patient characteristics, treatment parameters, and treatment outcome are not clear, existing brace treatment guidelines are general in nature. For example, the Scoliosis Research Society (SRS) Brace Manual's guidelines recommend that bracing should be used when the patient falls into certain ranges of Risser Grade (a measure of skeletal maturity) and Cobb angle (a measure of Scoliosis severity) [6]. These guidelines are shown in Table 1.

Risser Grade	Cobb Angle	Action
0-1	0-20°	Observe
0-1	20-40°	Brace
2-3	0-30°	Observe
2-3	30-40°	Brace
0-3	40-50°	Gray
0-4	≥50°	Surgery

Table 1: The Scoliosis Research Society Brace Manual's indications for observation, brace treatment, and surgical treatment.

Guidelines published by the Society on Scoliosis Orthopaedic and Rehabilitation Treatment (SOSORT) in 2011 include a table which is similar if more fine-grained – it includes each Risser grade as a separate category and ten categories of Cobb angle [7].

1.2.Motivation

With the large number of available brace types and brace-wear protocols, and the fact that "AIS" encompasses a wide range of three-dimensional deformities [8], it seems clear that general guidelines such as those published by the SRS and SOSORT do not adequately reflect the complexity of brace treatment or the diversity of patients. This is not a failing of the SRS or SOSORT; rather it is indicative of the current state of Scoliosis research. Existing Scoliosis literature is probably not yet informative enough to support the development of more sophisticated guidelines. Thus, evidence-based guidelines may not actually describe the brace treatment which is optimal for a given adolescent with AIS.

In a perfect world, brace treatment guidelines might consider more than just Risser grade and Cobb angle. Guidelines which considered (for example) patients' characteristics, the nature of their deformity, or their tolerance for treatment might better reflect the complexity of brace treatment. With a good enough ability to estimate case-by-case probabilities of treatment success, we might be able to replace *general* guidelines with *patient-specific* treatment plans.

This idea is, of course, not new. Individual healthcare providers already try to fine-tune guidelines to meet their patients' needs. Given two patients with similar Risser grades and Cobb angles, a physician might prescribe two different brace treatments: Based on past experience, intuition, and his/her interaction with the patient, the physician makes a judgement on how the general guidelines should be customized to a specific case.

The contrast in dispositions toward patient-specific treatment protocols is interesting. High-level organizations such as the SRS realize, perhaps, that the current body of Scoliosis research can only support the development of quite general guidelines. Meanwhile, physicians at the point of care realize the need for more patient-specific treatment, and routinely develop treatment plans which are individualized to varying degrees – and evidence-based to varying degrees.

Optimizing brace treatment probably means finding a middle ground: the choice of treatment should be justifiable by virtue of being based on solid data. But it should also acknowledge that different patients have different needs, and have some ability to accommodate them.

There may be opportunity to begin optimizing brace treatment at the clinic level. A system could be implemented which supports individualized treatment planning by providing evidence-based

recommendations. The system would provide information that general guidelines cannot, and help inform physicians' decision-making at the point of care. The goal would be to help physicians optimize treatment protocols to individual patients.

In this thesis, the system takes the form of an electronic decision support system.

1.3. Electronic Decision Support for Brace Treatment Planning

So how is the system to generate its evidence-based recommendations? If brace treatment guidelines are very general due to the current state of Scoliosis research, then the recommendations should not be drawn from research publications. Moreover, Trisha Greenhalgh et al. point out the negative consequences when "evidence-based" means "based on published research" [9]. This thesis will instead take "evidence-based" to mean "data-based", and assert that the recommendations should be based on an analysis of data collected from past patients. That is, the system will use a database of records of past patients to try to infer the optimal treatment for a given new patient.

The role of the system would be to enhance the physician's decision making. Humans have limited ability to process information: the large number of patient records and academic publications available to physicians represent an overwhelming volume of data. It is very difficult for the human brain to process all this information, determining which facts are relevant and how to use them in decision-making. A computer-based decision support system could programmatically select the most important information and use it to produce a sound recommendation. The system would accomplish this using machine learning techniques. A high-level representation of the envisioned system is shown in Figure 2.

The system accepts patient data from various sources. This data includes patients' clinical measurements, brace treatment details, and information about their brace-wear habits. This data is stored in a database for analysis. Machine learning techniques are used to (automatically) analyse the data and extract the key relationships: those that allow progression to be predicted for an individual AIS patient at the start of their brace treatment. These relationships are supplied to a "decision support engine" – a computer program which uses them to estimate optimal treatment parameters for new patients. Thus the system's treatment recommendations are individualized but still evidence (data) based – and what's more, the data originates from the same clinic at which the recommendations are being used. Published

research, on the other hand, involves diverse patient populations which may be different from the one seen at the local clinic.

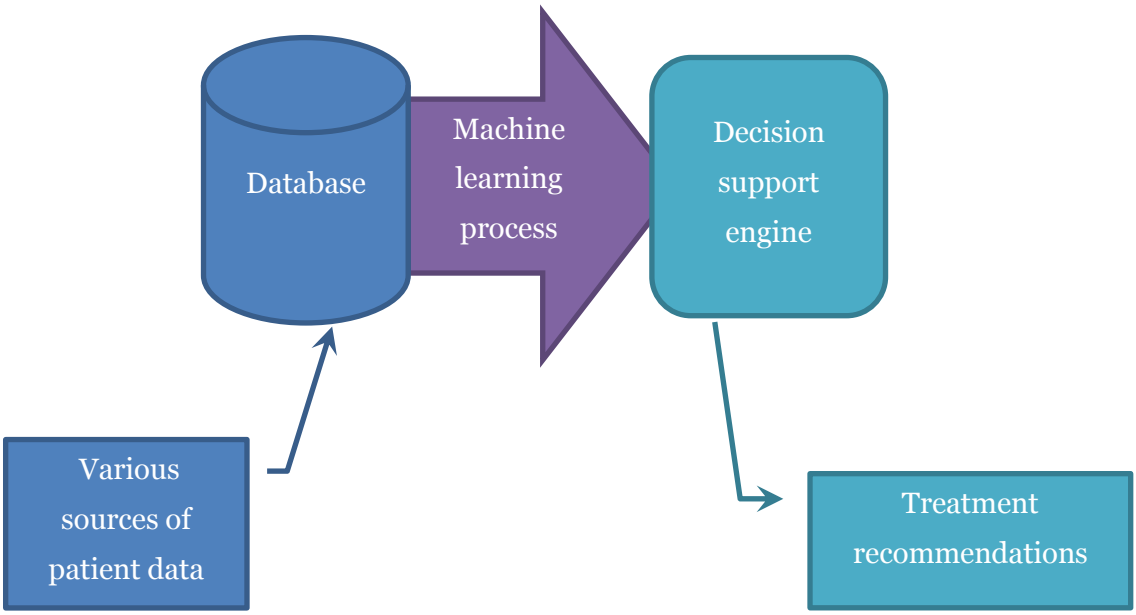


Figure 2: An illustration of the envisioned decision support system. Patient data accumulates in a database for analysis. A machine learning process distills from the data the key relationships that will be useful in recommending optimal brace treatment protocols.

The treatment recommendations would serve to advise and inform the physician in treatment planning. They would not be perfect, being limited by the amount and quality of the available data, and the quality of the machine learning process. Thus the physician may and should decide to reject them in some cases. But as the amount of stored data increases the recommendations would improve. The system could partially relieve clinic staff of the task of interpreting past patient data, and enhance their decision-making ability.

1.4.Objectives of This Work

This work investigates the feasibility of a decision support system as described in section 1.3. The system has been developed following that general description, and validated using actual patient data. The validation determines whether the system functions as intended, and whether it can be expected to produce valuable recommendations. The specific objectives are to:

1. Perform a preliminary validation of the concept of machine-learning based decision support for brace treatment using retrospective data from past patients.

2. Design and validate an electronic sensor which can be embedded into a brace and used to measure how long and how tightly a patient wears their brace.
3. Design and validate software which accepts data from the electronic sensor and other sources, stores it in a database, and allows users to perform simple statistical analyses and visualize the data.
4. Implement and test machine-learning software which "learns" from the data how to predict at the start of treatment whether a braced patient will progress (i.e. whether their treatment will succeed).
5. Design software which uses the prediction models to calculate recommended treatment parameters. Test the efficacy of these recommendations in a pilot simulation study.
6. Validate the overall system using data from patients undergoing brace treatment at the local Scoliosis clinic.

1.5. Proposed Approach

This work implements the decision support system as a scalable hardware/software platform which integrates the tasks of collecting patient data, managing the database, analysing the data, modelling the data, and using the models for decision support. Figure 3 illustrates the overall system.

The hardware portion of the platform includes an electronic device which can be embedded in a brace to monitor how patients wear their braces. The software portion of the platform includes a user interface for retrieving logged data from these devices. The data is stored in a relational database along with patients' other clinical measurements.

Analysis of the accumulated data is done from within the software platform, through the use of external "modules". Two types of modules are accommodated: visualization modules (used for simple statistical analysis and data visualization) and modelling modules (which apply machine learning techniques to model the data). As the users' needs change or the database grows, the platform can be expanded by adding new modules. This work also develops one modelling module and a small contingent of visualization modules.

The modelling modules interact with a decision support engine, which converts the knowledge contained in the models into case-specific recommendations. The decision support engine

accepts the details of a particular case, and then determines what combination of the models' non-static parameters produce a desired outcome. For example, suppose a model predicts treatment outcome based on patient Cobb angle, in-brace correction, and compliance. The decision support engine would calculate the in-brace correction and compliance required to obtain a desirable prediction, given the patient's Cobb angle. Chi et al call this style of decision support "Prediction and Optimization based Decision Support" [10], [11].

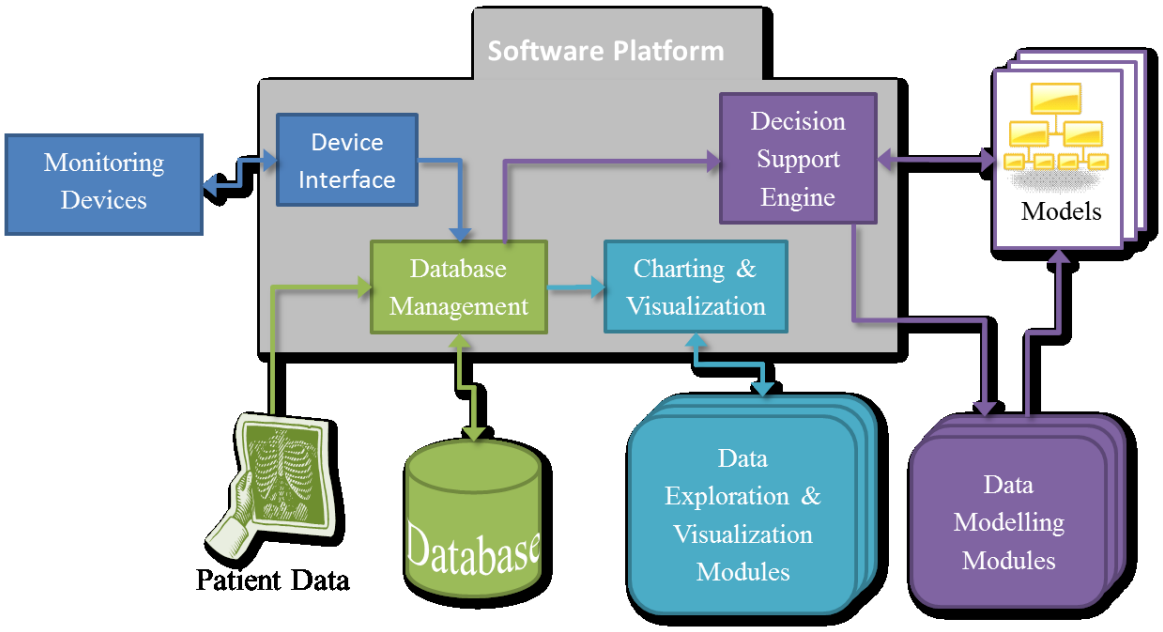


Figure 3: Illustration of the overall decision support system.

Validation of the software platform has been performed at the lab testing level (as identified by Wyatt and Spiegelhalter [12]). Field testing is out of the scope of this thesis. Thirty-one patients undergoing brace treatment had their braces instrumented with the electronic sensors developed during this work. Data from these patients has been collected and processed by the software platform. The modelling module was developed and cross validated using this data. A clinical trial simulation [13] estimates the efficacy of the treatment recommendations made by the decision support engine, in conjunction with the prediction model.

1.6.Thesis Overview

This thesis contains 6 chapters. After an introduction and review of related literature, the thesis describes a preliminary validation of the decision support concept. It then details the

development and testing of each component of the overall decision support system. The validation of the overall system is then described and test results are presented. Finally, conclusions are stated and recommendations for future work are put forward. Overviews of the following chapters are as follows:

Chapter 2 contains a review of literature on AIS, brace treatment, electronic decision support and predictive modelling, and applications of machine learning in Scoliosis. This review of published literature attempts to situate this thesis among the other research being conducted.

Chapter 3 describes a preliminary test of the concept of machine-learning-based decision support for brace treatment. This test used retrospective patient data to create a prediction model, and used this model to create treatment recommendations. The effect of the recommendations was estimated in a clinical trial simulation. This preliminary test was limited because the retrospective data did not include any information about patients' brace wear habits – this information is an important part of the proposed decision support system.

Chapter 4 discusses the design of each component of the overall decision support system: the electronic brace-wear sensing device, visualization and modelling modules, and the decision support engine. A software platform integrates these components into a complete system which accommodates future expansion.

Chapter 5 describes the validation of the overall decision support system. Test procedures and results for the electronic device are explained. Patient data is used to test the system's modelling module. The same data is used to validate the decision support engine in a clinical trial simulation. Some commentary is included to discuss the test results and what they reveal about the road to eventual implementation of a system like the one proposed here.

Chapter 6 concludes the thesis with a summary of the work. It then provides a discussion of the limitations of this work and makes recommendations for working toward optimal brace treatment protocols in the future.

2. Background

This chapter reviews published research on the topic of Adolescent Idiopathic Scoliosis: progression of AIS, predicting progression, and principles of brace treatment. It also overviews electronic clinical decision support, and general predictive modelling and machine learning techniques.

2.1. Scoliosis, Etiology, and Mechanisms of Progression

Adolescent Idiopathic Scoliosis (AIS) is a spinal deformity affecting 2-3% of adolescents [14]. Although it appears on typical radiographs as a "side-to-side" bending of the spine (see Figure 4), it is actually a complex three-dimensional deformity involving abnormal coronal and sagittal plane bending as well as vertebral rotation [15]–[17]. The coronal plane component can cause visible asymmetry of the shoulders or waist, while the rotational component can cause a visible rib hump or (in girls) breast asymmetry. AIS can involve deformation of the intervertebral discs [18], the vertebrae themselves [17], and the ribcage [19].

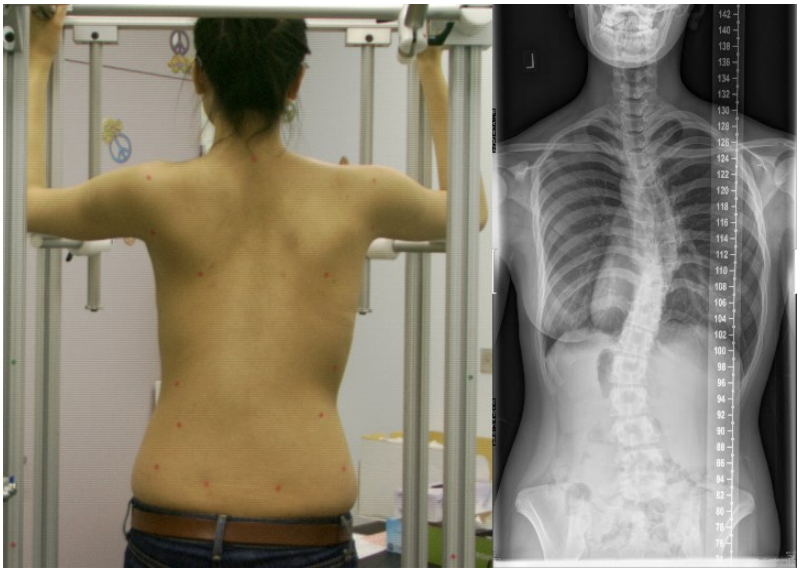


Figure 4: An adolescent with AIS (left), and a radiograph of the same adolescent (right).

The standard measurement for quantifying Scoliosis severity is the Cobb angle, measured from a posterior-anterior (or anterior-posterior) radiograph. It is the angle formed by the endplates of the two most tilted vertebrae along the spinal curve (Figure 5). Unfortunately, the two-dimensional nature of the Cobb angle makes it only an approximate measurement of Scoliosis

severity. The Cobb angle measurement is also very noisy; it suffers from measurement error with a 95% confidence interval of up to 7° for a single observer [20] and up to 9° for multiple observers [21]. However the Cobb angle remains the gold standard despite its drawbacks.

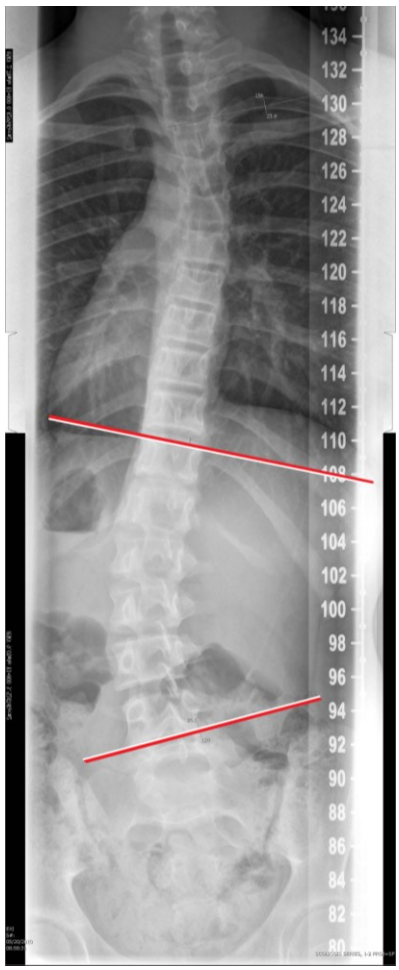


Figure 5: Cobb angle measurement. Lines are drawn along the endplates of the two most tilted vertebrae in the curve. The Cobb angle is the angle formed by these lines. The Cobb angle shown in this figure is roughly 26°.

Scoliosis can be life-threatening in very severe cases [22], but in general does not increase mortality rate. Usually severe curves are surgically corrected before becoming physiologically dangerous. However adolescents with AIS can have higher pain prevalence, lower social function, and lower self-image than healthy individuals [23].

The etiology of AIS is unclear, but some potential contributing mechanisms have been discussed in literature [24]. Millner and Dickson observed that scoliosis is preceded by a flattening of the spine in the sagittal plane [16]; they pointed out that a spine which is straighter than normal in

the sagittal plane would be more prone to bending in the coronal plane when pushed laterally – by the aorta, for example [25]. This idea of a Scoliotic spine being initially straighter and more slender seems supported by the fact that Scoliotic adolescents are often tall for their weight [26], [27]. Van Loon et al. made the related observation that forcibly exaggerating the spine's normal sagittal curves can correct Scoliotic deformities in the coronal plane [28]. The lordosing effect observed by Millner occurs around the adolescent growth spurt, which is also when AIS tends to progress the most. Adolescents' are also skeletally immature at the start of the growth spurt, making their spines more flexible [29]. This flexibility could make the spine more susceptible to deformation. Some deficiency of the intervertebral discs may also contribute to AIS: Yu et al. found an abnormal lack of an elastic fiber network in discs from Scoliotic individuals [30]. While it is unclear whether this was a cause or a result of Scoliosis, the weaker elastic fibers mean poorer structural properties for the disc.

Veldhuizen *et al.* suggested that AIS could originate due to postural instability caused by a small vestibular, visual, or proprioceptive defect [31]. The body's attempt to compensate for this defect could lead to asymmetrical muscle development, which ultimately unbalances forces on the spine. These asymmetrical loads on the spine can cause curve progression, as shown by Stokes [32],[33]. Stokes further suggests that the muscle activation strategies of the adolescent may be the difference between progressive and non-progressive Scoliosis [33].

The Heuter-Volkmann Principle seems to be a commonly accepted mechanism of Scoliosis progression [34]. In this context the principle simply states that the growth of the vertebral growth plates is impeded when the plates are loaded, and stimulated when they are unloaded. Thus if there is some curve in the spine, the uneven loading of the growth plates causes uneven growth. The uneven growth exacerbates the curve, which worsens the uneven growth plate loading. This cyclic mechanism – termed the "vicious cycle" by Stokes et al. – explains Scoliosis progression, and seems to be at the core of many conservative treatments.

There is likely some genetic basis for Scoliosis [25][35][36][37]. If literature in this area is not being oversold, genetic screening offers the potential for very early detection and possible prevention of AIS cases [36]. Developments in this area could prove very valuable in the future. At present though, it seems AIS is still correctly called "idiopathic".

2.2. Prognostic Factors and Methods for Predicting Progression

The natural history of AIS involves progression of the deformity [23], but the incidence of progressive versus non-progressive curves is unclear. For example Rogala et al. reported that 79% of curves between 20-30° progressed [38], while Lonstein and Carlson reported only 37% progression for the same range of Cobb angles [39]. The inconsistent reports in literature may be partly due to a historical lack of a standardized definition of "progression" – some researchers have defined progression as a certain magnitude of Cobb angle increase, while others have defined it as an increase past a certain threshold. The SRS has addressed this problem by recommending "progression" be defined as an increase in Cobb angle of 6° or more [40].

Professionals who treat AIS cases estimate their patient's risk of progression to make the best treatment decisions for their patients [7]. Literature has been flooded with statistical analyses of prognostic features, and sometimes contradictions appear. Table 2 shows a list of features which appear in literature, along with works which claim the feature has or does not have prognostic value.

Most works cited in Table 2 have investigated linear correlations between progression and the various features. The overall outcome of these researchers' efforts could be summarized by listing a few correlations that seem to have become generally accepted by clinicians. It is generally accepted that younger (especially in terms of skeletal maturity) patients are at higher risk of progression. So are deformities involving large (in terms of Cobb angle) curves, curves located high on the back, or multiple curves. The risk of progression is considered greatest during the adolescent growth spurt.

These correlations provide a general picture of what factors increase or decrease a patient's risk of progression. However we should note that most existing research has investigated simple linear relationships between the factors and progression, while the true nature of these relationships could be inherently non-linear. It has been shown that non-linear models outperform linear ones in describing scoliosis severity [41], [42]: the same could be true in modelling progression.

Some researchers have gone beyond simple descriptive statistics and proposed methods for predicting progression or non-progression in specific patients. Lonstein and Carlson proposed a "progression factor" – a function of patient's age, Risser sign, and Cobb angle – which related

Factor	Works which suggest prognostic value	Works which suggest little prognostic value
Age	[39][43][44][45]	[46][47]
Apex *	[45][46]	
Height or BMI	[25][48][49]	[43][48]
Cobb / curve magnitude	[25][39][43][44][47][48][50][51]	[45]
Curve type or direction †	[43][44][51]	
Family history/ genetics	[25][35][36][37][52]	[39][43]
Flexibility	[34]	
Growth velocity/ growth spurt	[25][44][46][50][53][54]	
Menarche	[44][46][48][50]	[45][47]
Risser or Bone Age	[39][43][44][45]	[50]
Rotation	[51]	[39][45]
Sagittal plane features	[16][44][55]	[43][45]
Sex	[25][43]	[39][47][48]
Trunk balance ‡	[45][51]	[43]

* Apex refers to the particular vertebra which falls at the "apex" of the spinal curve. It is an indication of the curve's vertical position in the spine.

† Various systems have been proposed to classify AIS cases by curve type; usually based on the number, location, and severity of curves in the spine. Curve direction refers to the direction (left or right) of a curve's deviation from the normal vertical.

‡ Measured as the horizontal deviation between the top of the thoracic spine from a vertical line passing through the center of the sacrum.

Table 2: Features which appear in literature related to an AIS patient's risk of progression. A few features appear prognostic in some studies but not others.

non-linearly to a percentage risk of progression [39]. Although Lonstein's progression factor has been widely used in clinical practice [56], it seems to have never been validated as an accurate predictor of progression. Because of this SOSORT has discontinued its use [7].

Peterson proposed a logistic regression model to predict progression using Risser sign, curve apex level, imbalance, and age [45]. Peterson claimed this model was 81% accurate in training (81% accurate on the dataset used to create it), but it was never validated on separate data. Like

Lonstein's progression factor, there does not seem to be any evidence that Peterson's risk of progression makes accurate predictions [57].

Ajemba *et al.* developed a support vector machine classifier [58] to predict progression using a small dataset of 44 patients, 14 of whom were braced [59]. The model used radiographic features as inputs, including several non-standard measurements not normally measured in routine practice. They reported a cross-validated prediction accuracy of 78% on the 14 braced patients using sagittal balance and wrist x-ray as model inputs. While Ajemba's method may have merit, his dataset was so small that a generalizable model could not be found – as evidenced by the gap between training and test accuracies (100% and 78% respectively).

Wu *et al.* have used fuzzy clustering and artificial neural networks [60], and a non-uniform rational B-spline technique [61] to predict a patient's Cobb angle at their fourth follow-up visit based on the previous three Cobb angle measurements. The absolute error in predicting the Cobb angle was 4.1 ± 3.3° in the later study: impressively close to the measurement error of the Cobb angle itself. Thus this method may be able to reduce patient's radiation exposure by replacing some x-ray measurements with Cobb angle estimates. However, Wu's dataset included 56 radiographs from only 11 AIS patients, and apparently has not been validated on additional data to ensure generalizability. Also note that Wu's method is applicable only after three visits (6-12 months apart) and so is not intended to achieve early identification of progressive curves.

Lee *et al.* used Classification and Regression Tree (CART) [62] analysis to develop a decision tree for classifying patients' risk of progression [48]. The tree used Cobb angle, age, and menarche (or height if menarche was not available) to place patients in one of four risk categories. But Lee was interested only in the hazard ratios for the patients in each category – his decision tree was not developed to generate actual predictions of progression or non-progression for particular patients. In fact, a glance at the leaf nodes of the resulting tree suggests the model may predict progressive curves quite inaccurately.

Lou *et* al. developed the only prediction model specifically for AIS patients treated with braces [3]. The model is a linear regression using Peterson's risk of progression, flexibility, quality of brace wear, quantity of brace wear, and the product of quality and quantity as inputs. Lou used force sensors embedded in the brace to measure the quality (related to the amount of force applied) and quantity (related to brace wear-time per day). Initial results showed prediction of progression within 3° on a test set of six subjects. Lou's method needs validation on a larger set

of patients, but looks very promising. Using this model, it is unclear how early in the treatment process a prediction could be made.

To summarize, numerous statistical analyses have been performed to identify factors with prognostic value. However as Lee's work points out:

> *"A factor being judged as prognostic does not necessarily imply that it is useful in classifying risk of curve progression." [48]*

One factor contributing to the truth of this statement – and possibly to the discrepancies observed in Table 2 – is that many of the statistical analyses in Scoliosis literature are explanatory rather than predictive in nature. Galit Shmueli explains that the traditional statistical analysis techniques practiced by most Scoliosis researchers are suitable for exploring causal relationships, but unsuited to developing good prediction models [63]. As a result many papers claim their findings support prediction of future events, when in reality they may be better suited to explaining past ones [64]. Moreover several methods of predicting Scoliosis progression have been proposed, but many have significant drawbacks. It seems safe to say that work on modelling Scoliosis progression is immature.

2.3. Brace Treatment

Brace treatment is the most common non-surgical treatment for AIS [14]. In general a brace is an orthotic which is usually (but not necessarily [65]) rigid, and custom fitted to each patient. The brace is designed to counteract the spinal curve while worn (Figure 6). Brace treatment is generally considered successful if the curve does not progress [40], although improvement of the curve can occur in some cases. Ultimately the brace is intended to prevent progression and thus avoid the need for surgical correction. There are a variety of brace types. Zaina et al. provide a thorough review of common brace types, and point out that there is no standard for designing braces or treatment protocols [66]. Here the focus will be on a few common brace types and the general principles of brace treatment.

In general bracing is recommended for patients with 20-40° Cobb angles which are predicted to be progressive [6], [7]. The prognostic factors discussed previously are usually considered by individual care providers prescribing brace treatment. If curves progress to 45-50°, surgical intervention is usually recommended.

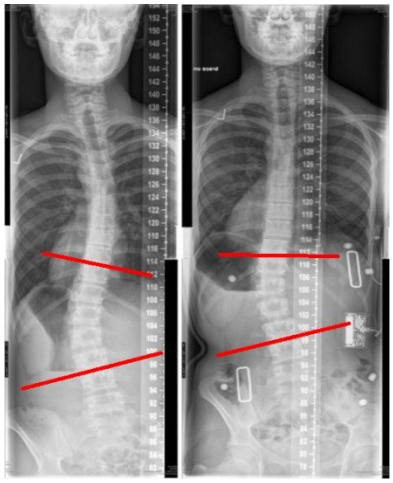


Figure 6: X-ray images of a patient normally (left) and wearing a brace (right). Lines defining the Cobb angle are shown.

One of the oldest types of braces is the Milwaukee brace or cervico-thoraco-lumbo-sacral orthosis (CTLSO). It involves an orthosis which rests on the pelvis, and three metal uprights supporting a neck ring which contacts the patient's throat and head. The brace provides a traction force to elongate the spine while additional pads counteract the spinal curve. The Milwaukee brace may have a higher success rate than other brace types [67], but is often unacceptable to adolescents because the neck support apparatus is highly obtrusive.

The Boston Brace is a thoraco-lumbo-sacral orthosis (TLSO) invented as a less-obtrusive alternative to the Milwaukee brace [68]. It uses a standardized prefabricated 'blank', which is cut and padded by the orthotist to achieve the desired shape. The brace is shaped to provide lumbar and pelvic flexion, reliefs opposite every area of force (to encourage movement of the torso in the desired direction), and force couples for de-rotation. These mechanisms allow the Boston brace to treat all aspects of the 3-D deformity. As with any brace though, good correction of the 3-D deformity by the Boston brace depends on the orthotist's skill [69].

The Charleston Bending Brace is a brace type which applies 3-point pressure to unbend a single Scoliotic curve. It can often achieve overcorrection of the curve (in-brace) and is intended for nocturnal wear only. The manual for the Charleston brace claims it can treat double curves, but calls this an "advanced technique" [70]. In practice this is uncommon, as this brace can actually worsen secondary curves [71].

The Cheneau brace has, historically, been more commonly used in Europe. It emphasises relief areas into which the body is expected to move. This led to the introduction of the Cheneau Light

brace: which removes brace material from relief areas, giving a lighter, airier feeling brace which has shown some success [72]

With the exception of the Boston brace, the traditional approach to brace construction involves a casting procedure. The orthotist wraps the patient in a wet casting material, and then manipulates their torso into the desired (corrected) position while the cast hardens. The hardened cast serves as a negative mold of the desired torso shape. Plaster can then be poured into the mold to create a positive model. Any additional corrections to the torso shape can be made on the plaster model. The brace itself is built by draping a sheet of hot plastic over the model.

A Providence brace uses a specialized system designed to facilitate effective casting. The Providence System consists of a large pegboard and adjustable pads that can be 'plugged in' anywhere on the board. After applying the cast, the patient lies on the board and pads are added to provide corrective force. The adjustable pads essentially simulate the brace itself. The makers of the Providence system advertise that it can achieve nearly 100% in-brace correction of the Cobb angle [73].

A more modern alternative to casting uses a 3-D scanner to capture a digital model of the patient's torso. The model is manipulated in the desired ways using specialized software, and then exported to a computerized milling machine which fabricates the physical model used in the final brace construction. This approach to brace construction is referred to as CAD/CAM (computer assisted design / computer assisted machining).

The Hueter-Volkmann principle and vicious cycle phenomenon are core concepts in brace treatment, just as they are central in explaining curve progression. By counteracting the Scoliotic curve, a brace unloads the normally compressed side of the vertebral growth plates. A brace which overcorrects the curve not only unloads the compressed side but also increases load on the normally uncompressed side of the growth plates; further equalizing the average forces on the plates. The Hueter-Volkmann principle in conjunction with the vicious cycle theory says that, where an untreated spinal curve would progress due to the cycle of uneven loading and uneven growth, in a treated spine this effect should be reduced. However Castro suggests that the principle may only apply to adolescents with particularly flexible spines [34].

A study of the Cheneau brace by Kotwicki et al. identified some passive and active mechanisms by which the brace counteracts the spinal curve [74]. The passive mechanisms included bending,

elongation, de-rotation, and tissue transfer. The bending effect described used a 3-point pressure system: one pressure point at the apex of the spinal curve and two opposing counter-pressure points above and below. The three-point system is a mechanism for directly counteracting the spinal curve, and is accompanied by an elongation of the torso. De-rotation of the torso counteracts abnormal vertebral rotation, and tissue transfer helps restore overall balance of the torso.

The active mechanisms identified by Kotwicki included the brace's guidance of vertebral growth (as explained by the Hueter-Volkmann principle). The brace also asymmetrically guided respiration and re-arranged trunk muscles to employ the patient's own muscular system in counteracting the curve. Kotwicki also noticed an anti-gravitational effect, whereby the body's proprioceptive system will activate muscles to counteract a perceived shift out of balance. A well-designed brace could use this effect to activate muscles which counteract the spinal curve. Grivas and Kaspiris studied common European braces and reiterated many of the mechanisms mentioned by Kotwicki [75]. But there is some debate on the particulars of these mechanisms' implementation. For example most experts seem to agree with the 3-point pressure concept, but disagree on where the main pressure point should be located [76].

In fact it may be difficult to generalize about exact implementation of corrective mechanisms, as each brace is custom made for a specific patient's deformity. As no two patients are identical, the effect of the brace on the spinal curve is highly variable [77]. In general the three-point pressure must be applied using a combination of brace-body contact and relief: a brace which is everywhere in contact with the body interferes with the guided respiration and possibly the tissue transfer mechanisms, and can reduce lung capacity [78],[79]. The brace must achieve good correction of both the Cobb angle and the vertebral rotation [80]. Greater than 25% Cobb angle correction has been suggested as a guideline for large curves [81], but overall the optimal amount of correction for a given patient is unclear. Moreover the maximum correction attainable is determined partly by the stiffness of the individual's spine. Less mature spines are more flexible and correct more easily than stiffer spines [82],[83]; as a result they require more correction in-brace for effective treatment [84].

The efficacy of brace treatment has been unclear in past literature [85]. A review by Rowe et al. claimed bracing was an effective treatment for AIS [86], but a later review by Dolan and Weinstein found literature on the subject to be "troublingly inconsistent" [87]. A 2010 Cochrane Review on brace treatment effectiveness concluded there was "very low quality evidence in favor

of using braces, making generalization very difficult" [88]. However, historically the literature on brace effectiveness has not considered patients' brace-wear habits, or has considered prescribed but not actual brace wear-time (inappropriate as actual compliance is usually unknown and often overestimated [89]). This neglect may have contributed to the confusion in literature.

Recent studies which have measured patient compliance indicate that brace treatment is effective when the brace is worn as prescribed. Katz et al. reported an 82% success rate among compliant brace-wearers (12 hours or more of brace wear per day) but only 31% among those with poor compliance (7 hours or less) [4]. Weinstein et al. reported 75% success among braced patients versus 42% among non-braced patients [5]. This difference in success rate became significant partway through Weinstein's study - maintaining a non-braced group became unethical and the study stopped early.

In seems the construction of the brace and the patient's brace-wear habits both contribute heavily to treatment success [90]. It has been suggested that successful treatment requires at least 12 hours per day of brace wear [4] or 18 hours for large curves [81], and Lou's prediction model indicates a negative correlation between wear-time and progression [3]. Lou's model also indicates a negative correlation between applied force and progression. Besides being commonsensical, the importance of wearing the brace tightly was confirmed by studies which show in-brace Cobb angle correction is correlated with brace strap tension [1][2]. Sanders et al. agree with the importance of compliance, but show that there is much room to optimize brace treatment protocols themselves [91].

However it should be noted that patient compliance, while an important factor in treatment outcome, is not the whole story. The Katz and Weinstein studies show 10-18% failure rates among the most compliant patients, and 30-40% success rates among those with very poor compliance (Figure 7). Clearly treatment outcome depends on other factors in addition to patient compliance: some cases may simply be more or less difficult to treat.

To summarize, a variety of brace types exist having slightly different philosophies of how to counteract the Scoliotic spinal curve. The brace's design allows it to utilize various corrective mechanisms, and to ensure a good design we (appropriately) rely on orthotist's experience as

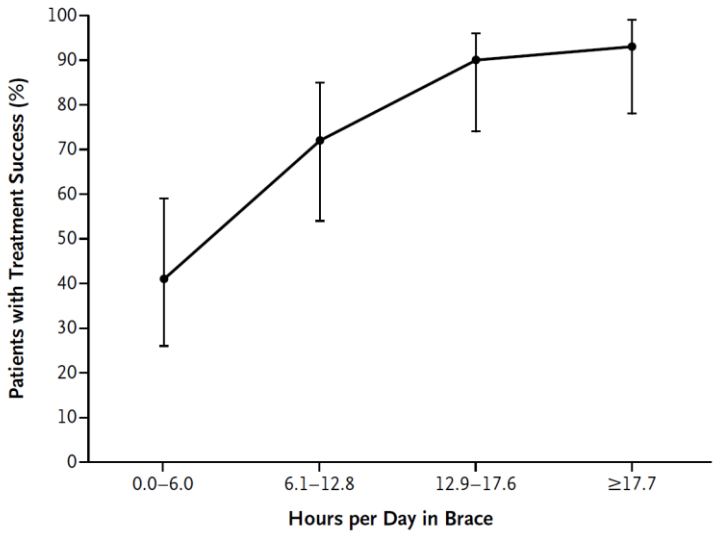


Figure 7: This chart from a study published by Weinstein et al. [5] shows a clear compliance-success association. But interestingly, when we look only at compliance, success rate plateaus around 90%, and 40% of the least compliant patients still achieve success.

much as scientific research. Brace treatment effectively reduces overall progression rates in compliant patients, though success for an individual is not purely a matter of compliance. Ideally all aspects of an individual's treatment would be optimized for maximum chance of success.

2.4. Efforts to Improve Brace Treatment

Efforts to improve brace treatment have generally fallen into two categories: improving the brace's ability to counteract the deformity, and improving our understanding of how patients are wearing their braces.

2.4.1. Improving Brace Effectiveness

Finite element analysis (FEA) is being used to investigate how braces might more effectively counteract spinal deformity. A finite element model simulates a real-world object or event by modelling it as a collection of small elements. In this case, a numerical model of a spine and ribcage can be created, and the effect of various forces can be simulated and observed. FEA could be a convenient way to model the effect of a brace on a patient's deformity, to optimize the brace's design before building it [71][92][93][94].

Gignac *et al.* developed patient-specific spine and ribcage models, then demonstrated that a computer program could determine the optimal forces for counteracting the deformity [95]. Cheng et al. modelled a single curve deformity and calculated the magnitude and location of the optimal correcting force [96]. Clin et al. used FEA to investigate the effects of altering various brace features. They determined that the placement of the opening, strap tension, rigid shell shape, lordosis design and design of the trochanter extension side were the most influential brace design factors [92]. FEA looks promising: in the past this technology was able to produce braces as effective as conventional methods [97]. It now produces more efficient and lighter braces [98].

2.4.2. Sensing Patients' Compliance with Prescribed Brace Wear

Investigations into how patients wear their braces have used various sensors embedded in the brace. These electronic compliance monitors generally consist of a sensor, clock, and a battery-powered data logger to record sensor readings. They generally fall into two categories: temperature sensing and force/pressure sensing. Vandal et al. used a device which measured tension in the brace straps. A threshold of 7.8 N was used to identify when the brace was worn [99]. Havey et al. placed four pressure switches inside the brace, and considered the brace to be worn when at least two switches were activated [100]. Lou et al. designed a compliance monitor which recorded force applied to the patient's body [101], [102]. Lou discretized the raw force readings to show time spent above, below, and within a reference force range. He found that patients typically wear their brace at 50-70% of the force level recommended by the orthotist – possibly decreasing the brace's effectiveness [102].

Temperature sensing is the most popular method of compliance monitoring. These compliance monitors set a threshold on temperature in the brace (between 28° and 32° Celsius [4], [5], [103]–[107]), to differentiate ambient temperature from the skin temperature of the patient. Studies using temperature-based compliance monitors have found that younger patients are more compliant than older ones [105], [107], patients are generally more compliant at night [104], and the knowledge of being monitored itself improves compliance [108]. There is some discrepancy on whether patient compliance should be expected to change over time [102], [106].

There are limitations to both force-based and temperature-based compliance monitoring. Temperature sensing detects when the brace is worn, but not how well it is worn: the temperature readings cannot differentiate between a patient who is wearing their brace properly, and one who is wearing their brace too loosely. Conversely, force sensing can detect

how well the brace is worn, but cannot differentiate between a patient who wears their brace too loosely and one who does not wear it at all: in both cases the sensor records zero force. The distinction is clinically relevant – if the brace is too loose the patient may simply need instruction on proper brace-wear, while not wearing the brace is more serious. Force monitors can also be sensitive to their placement in the brace, and could potentially lose contact with the patient's body in some positions. This would result in erroneous compliance readings.

In 2005 Lou *et al.* developed an active system which used an inflatable air bladder to regulate pressure at one part of the brace-body interface [109]. This system was tested on 5 subjects, and increased the pressure from 53% of the desired level to 68% on average. It seems reasonable to assume that this effect should improve treatment outcome, since brace pressure normally varies with posture [110]. Chalmers, et al. developed an improved version of the system which regulated pressure at up to 4 locations [111]. In laboratory tests this system increased pressure from 31% of the desired level to 62%.

2.5. Electronic Decision Support in Medicine

> *"Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?" – T.S. Eliot*

Almost all research activities generate information on some subject. Some of this information is disseminated in publications, becoming part of the body of knowledge on the subject. But often the process ends there – the knowledge is never translated into actual practice [112]. The knowledge translation process is necessary if the research is to create lasting value. Straus *et al.* explain knowledge translation:

> *"[Knowledge translation] is a move beyond the simple dissemination of knowledge into actual use of knowledge. Knowledge creation (i.e., primary research), knowledge distillation (i.e., the creation of systematic reviews and guidelines) and knowledge dissemination (i.e., appearances in journals and presentations) are not enough on their own to ensure the use of knowledge in decision-making." [113]*

A Clinical Decision Support System (CDSS) is a software tool which can facilitate knowledge translation by promoting use of best practices, condition-specific guidelines, and population-based management [114]. This promotion takes the form of case-specific advice, which the CDSS generates using details of the case [115] combined with knowledge about the domain in general. The CDSS's advice is meant to enhance users' (physicians, clinicians, nurses, etc.) natural decision making ability.

CDSSs can significantly affect the quality of medical decision making [116]. Garg et al. reviewed literature on CDSSs and found that practitioner performance was improved by 64% of CDSSs related to diagnosis, reminder, disease management, or prescription tasks [117]. Kuwamoto et al. performed a similar review with similar results (68% of CDSSs improved performance) [118]. Another review by Lobach et al. agreed that CDSSs improve practitioner performance in offering preventive services, ordering tests, or prescribing [119]. Strangely, Garg and Lobach both found that while practitioner performance improved, patient outcomes were unaffected. Lisboa and Taktak reviewed applications of artificial neural networks for decision support in cancer. They found that 78% of these systems improved healthcare provision, and none decreased it [120].

CDSSs can generally be grouped in two categories: knowledge-based and non-knowledge-based CDSS [121]. Knowledge based CDSS (also referred to as "Expert systems") consist of a knowledge base containing information about the domain, an inference engine which draws on the knowledge base to generate recommendations for a particular case, and a user interface which communicates the recommendation to the practitioner. The knowledge base is constructed by eliciting domain knowledge from domain experts and/or literature, and codifying them in a systematic way – often in the form of IF-THEN rules. This appeal of this approach is that the CDSS emulates a panel of experts. An early example of a knowledge-based CDSS is MYCIN [122]: MYCIN recommends antimicrobial therapies for patients with bacterial infection using hundreds of IF-THEN rules which were derived during discussion with physicians.

Non-knowledge-based CDSSs use no base of expert knowledge. Instead they employ machine learning algorithms which attempt to extract domain knowledge directly from retrospective data. In essence this type of CDSS replaces experts' knowledge of the domain, with a machine-learning-derived model of the domain itself. The advantage of this approach is that it avoids the process of eliciting a knowledge base from experts (this process can be complex [123]) and avoids introducing the experts' biases or preconceptions (if any) into the system's recommendations. An example of a non-knowledge-based CDSS is PAPNET [124], which uses a trained artificial neural network to assist in diagnosing cervical cancer. Dilsizian and Siegel predict great potential for non-knowledge-based CDSS:

> *"In the near future, artificial intelligence/machine learning will likely assist physicians with differential diagnosis of disease, treatment option suggestions, and recommendations… based on empirical data." [125]*

Chi *et al.* coined the term "Prediction and Optimization based Decision Support System" (PODSS) to refer to a particular type of non-knowledge-based CDSS, which uses machine learning coupled with optimization to generate recommendations [126]. The PODSS uses machine learning to generate a model which uses domain inputs to predict some domain output. Each of the inputs must be classified as "static" (variables which cannot be changed by the user; patient age and gender for example) or "non-static" (variables which can be influenced by the user, such as treatment procedures) [127]. An optimization process then determines what combination of non-static inputs will produce the desired model output. The PODSS concept of using advanced data analysis to recommend optimal decisions is known in the business field as "prescriptive analytics" [128]. Haas et al. state that the goal of prescriptive analytics is to "identify optimal business, policy, investment, and engineering decisions in the face of uncertainty" [129].

The PODSS concept was used by Liau et al, who modelled the output of a crude oil distillation unit using an artificial neural network, and then used the model in an optimization problem to find the operating conditions which would maximize production [130]. Song and Kusiak created a model of boiler operation and used evolutionary computation to find the control settings which would optimize performance [131]. Chi et al. designed a support vector machine classifier to predict hospital mortality based on several inputs related to the hospital and the patient. This model was used in a PODSS which recommended an optimal hospital choice for a particular patient [11]. Chi et al. also designed a k-nearest-neighbor classifier to predict whether a patient would develop heart disease, based on their characteristics and lifestyle. This model was used in a PODSS which found the optimal change in lifestyle to lower a person's risk of heart disease [10]. Chi et al. also used PODSS to optimize individualized warfarin treatment [13]. Buchan et al. proposed that similar model optimization techniques could be used to improve clinical treatment plans [132].

Buchan acknowledged that gaining medical knowledge through machine learning is a departure from the traditional hypothesis-based experimentation. He explained:

> *"Hypothesis-driven research and reductionist approaches to causality have served health science well… But they do not reflect the complexity of health" [132]*

Machine learning methods can be powerful knowledge discovery tools, but represent a somewhat different paradigm than some medical practitioners are used to. This has made the acceptance of machine learning methods – and CDSSs in general – an issue [133],[134].

Shortliffe claimed that one cannot expect clinicians to adopt a CDSS, even if it has been proven to function at the level of experts [135]. Feblowitz et al. found acceptance of CDSS by health care providers to be "highly variable" and called acceptance a "complex and individual phenomenon" [136].

Several literature reviews have identified a long list of attributes which contribute to a CDSS's successful implementation. Shortliffe states that transparency of the system is important [135]. Kuwamoto *et al.* found that successful systems integrate easily into a clinic's existing workflow, provide recommendations rather than mere assessments, and provide support at the time and place of decision making [118]. Lobach et al. confirmed Kuwamoto's findings and added three more important features: the CDSS should promote action rather than inaction, local users should be involved in its development, and it should provide information to patients as well as care providers [119]. Hsiao et al. and Lu et al. found that the quality of information provided by the CDSS (i.e. accuracy, completeness, and legibility) and the system's ease of use were important factors [137],[138]. Bates et al. mentioned the integration, promotion of action, and ease of use factors. He also said that a CDSS should have good service quality (technical support and maintenance of the system should be provided), it should work quickly and require as little user input as possible, its recommendations should be simple, and its implementation should be monitored to ensure user adoption [134]. A similar study by Castillo and Kelemen reiterated many of the same features [139]. Between these authors, the most commonly named features were: integration into the clinic's existing workflow, and ease of use. This long list of features illustrates the difficulty involved in designing and implementing a successful CDSS. However a review by Miller points out that, while it is true most CDSSs reported in literature do not achieve clinical implementation, there are some very successful CDSSs [140]. So while persuading medical professionals to adopt a CDSS may be difficult, it is certainly not impossible.

Wyatt and Spiegelhalter discussed the process of designing and validating a CDSS [12]. They divided the process into three stages: definition, prototyping and lab testing, and field testing. The definition stage identifies the CDSS's main objectives, often by building a rough version of the system and asking for user's comments. In the prototyping and lab testing stage, a working (though perhaps not final) version of the CDSS is built and tested. This testing involves the obvious tests of the system's accuracy and performance, but could also include testing of others' reactions to the system. For example, system users might be asked: "Is this system wanted?", "Is it easy to use?", and "Do its recommendations make sense?" Domain experts could be asked: "Is the system of good quality?", "Does it reason appropriately?", and "Are its recommendations

valuable?" Once the system has been finalized it must be field tested. Field testing can be a long and complex process [115], and involves investigating the system's level of integration with the clinic, its effects on the healthcare process, and its effects on patient outcomes.

2.6. Data Analysis and Predictive Modelling

The quality of a PODSS's recommendations depends on an accurate model of the domain. The process of creating and using such a model is sometimes called knowledge discovery and data mining (KDDM). Several models for the KDDM process itself have been proposed. For example, Cios et al. described a six-step KDDM process model [141] which has been used in research. Industry collaborators developed a similar model called the Cross-Industry Standard Process for Data Mining (CRISP-DM) [142]. Kurgan and Musilek reviewed these and several other models, and laid out a generic six-step model for the KDDM process [143] (Figure 8).

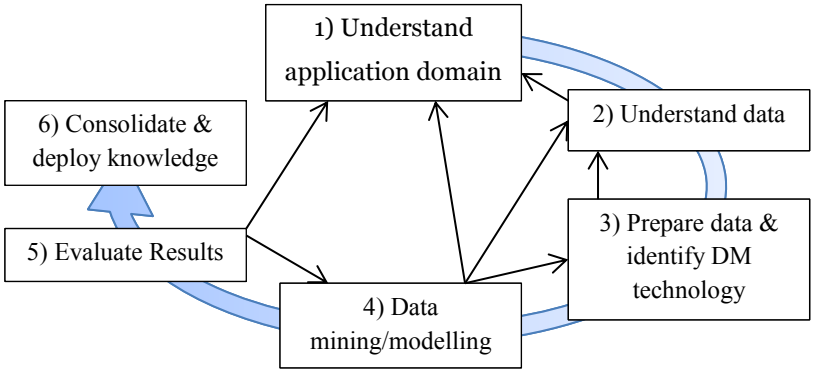


Figure 8: Knowledge discovery and data mining process model. The process follows steps 1-6, with possible feedback paths shown by black arrows.

Step 1 involves defining the goals of the KDDM venture, and understanding the current solutions and terminology in the domain. Step 2 is to investigate the quality of the available data with respect to the goals, and identify problems such as missing values, redundancy, etc. Step 3 involves cleaning, imputing, and reducing dimensionality of the data, and choosing an appropriate modelling technique. Step 4 applies this technique, and step 5 tests the resulting model's ability to meet the goal. Step 6 involves applying the resulting knowledge to achieve the goal. In practice the KDDM process is iterative, so feedback loops exist between some steps. For example, if step 5 discovers the domain model is unsatisfactory, it may be necessary to return to step 4 to try a different modelling technique, or to step 1 to modify the goal. As it relates to PODSS, this six-step model describes the entire process of collecting and investigating domain

information (data), and using machine learning to reduce it to a model which is then deployed as a component of a CDSS.

Data preparation (step 3) includes reducing the number of variables for use in a model. This is an important step in producing a generalizable model, and can be achieved using various transformation or selection methods. One example of a transformation for reducing data dimensionality is Principle Component Analysis (PCA) [144], which performs an orthogonal transformation to remove correlations in data (assuming the data follows a multivariate normal distribution). After the transformation, some of the new variables contribute little to the overall variance of the data and can be removed.

Data dimensionality could also be reduced by selecting a subset of variables for use in modelling, and discarding the rest. Selection methods can generally be classified as "filter" or "wrapper" methods. Filter methods use some heuristic criteria for selecting or ranking variables, where the heuristic criterion is independent of the machine learning technique used. One example of a filter method is Correlation-based Feature Selection [145], which searches for a subset of features having high correlation with the output variable but low correlation with each other.

Wrapper methods integrate the machine learning technique into the variable selection process. A search is conducted through the space of possible feature subsets, and the machine learning technique is applied to each. The value of each subset is tested by measuring the accuracy of the corresponding models using cross-validation. Different wrapper methods are distinguished primarily based on the search strategy used to explore feature subsets. Kohavi and John offer a thorough discussion of wrapper methods [146].

The machine learning technology used in step 4 could be one of many existing options. Shouval et al. explain that the technology would probably be machine learning based, not based on conventional statistical methods [147]. The goal of such technologies is to generate a model – mathematical or otherwise – which describes the relationship between inputs and outputs of the system or process of interest. This relationship may be valuable as a predictor of future outputs, or may be valuable in itself as a description of input-output interaction (the "prediction model" / "explanatory model" distinction [63]). The selection of a particular learning method depends on the problem and the data involved. If the goal involves predicting future outputs it is important to ensure generalizability of the resulting model, whatever machine learning method is used.

Using cross-validation and setting aside some data for model testing are important practices in this regard.

Support vector machines (SVM) are a popular technique which has already been mentioned. A support vector machine classifier finds a hyperplane to separate two classes of data in an n-dimensional space. The optimal hyperplane is the one which maximizes the margin between itself and the closest points on either side. To accommodate data in which the classes are not linearly separable, the SVM is given a degree of tolerance for misclassified data. A Kernel function can also be used to non-linearly transform the n-dimensional data into a space with higher dimension, which may make the classes more separable. Since the main objective of the SVM algorithm is maximizing the margin around the separating boundary, SVMs can often produce good, generalizable models using small training datasets. Xue provides a complete description of SVM [58].

Decision trees were also mentioned previously: the CART algorithm is one example [62], the C4.5 algorithm is another [148, p. 5]. Decision trees are an attractive method of performing classification tasks, as they are usually completely human-readable. Algorithms which generate decision trees (such as CART, C4.5, etc.) build the tree node-by-node, deciding at each step the variable and split giving the best value. The method for calculating this value varies by algorithm, as do methods of pruning the tree (removing less-valuable nodes).

Logistic regression is another modelling technique which produces human-readable models; it may be described as the analog of linear regression for classification problems. Logistic regression fits the logistic function to n-dimensional data with two classes, represented numerically as 1 (the “case” class) and 0 (the “non-case” class). The weights of each feature in the data are adjusted to minimize the error between the logistic function output (which has a range of 0-1) and the class values of each point in the training dataset. Logistic regression models are common in medical research, where they are often used to analyse correlations between an output of interest and several inputs. When used for prediction, the logistic function output loosely represents the probability that a set of inputs corresponds to a “case”.

Fuzzy logic has seen some application to Scoliosis. Where conventional logic uses absolute concepts of “true” (or “1”) and “false” (or “0”), fuzzy logic accommodates truth with a degree of membership ranging from 0 to 1. This allows for approximate reasoning, which is often better suited to tackling real-world problems. Some specific methods include fuzzy clustering (which partitions data into overlapping or “fuzzy” subsets based on similarity), and fuzzy rule bases (a

model consisting of fuzzy IF-THEN rules, where the activation of the consequent is proportional to the truth value of the antecedent). A description of these and other fuzzy processing techniques is given by Pedrycz and Gomide [149].

Another method used in medical research is artificial neural network (ANN) – a machine learning algorithm inspired by the interaction of biological neurons in the brain. An ANN consists of one or more layers of simple functional blocks called "neurons", each implementing a simple activation function which produces one output from one or more inputs. Weighted connections are made between neurons in adjacent. The first layer accepts the input variables, and the last layer produces the model's output(s). A training process tunes the interconnection weights until the ANN can emulate relationships seen in the training data. ANNs can be extremely powerful, and can model complex non-linear relationships. However this also means that it is difficult to create a generalizable model with limited training data. A thorough discussion of ANN is given by Engelbrecht [150].

Many more modelling techniques have appeared in medical literature, and more are being invented all the time. The feedback loops in the KDDM process allow several different approaches to be tried before finding one which suits the data.

As a final note, it is important to realize there are different ways to analyse data: Shmueli identifies two paradigms with unique goals: explanatory and predictive analyses [63]. Explanatory modelling tries to identify causal relationships in the data, or in other words, to identify the mechanism that generated the data. Explanatory analysis techniques concentrate on finding a model which *describes* the available data – results are usually reported in terms of R^2 and p values (often emphasizing p=0.05 as the threshold for "significance" – a practice which statisticians like Regina Nuzzo would like to stamp out [151]). In contrast, predictive modelling finds patterns in data for the sole purpose of predicting future data – the results of a predictive modelling endeavour would be reported in terms of predictive (rather than descriptive) power, accuracy, etc. Techniques used for explanatory modelling are useful in many respects, but as Leo Breiman points out, often unsuitable for predicting future events [152]. Breiman laments:

> *"Hundreds, perhaps thousands of articles were published claiming proof of something or other because the coefficient was significant at the 5% level." [152]*

A myriad of papers have investigated associations between Scoliosis progression and various patient characteristics. Most of these analyses have been explanatory in nature, meaning their results may have limited use in predicting progression in new patients. While some exceptions

are described in section 2.7, there is an opportunity to apply predictive modelling to problems in Scoliosis.

2.7. Machine Learning and Decision Support in Scoliosis

Machine learning and CDSS have been applied in the Scoliosis field; most of these applications were discussed in a review by Phan et al. [153]. Those which relate to predicting Scoliosis progression are described in section 2.2. Most of the others relate to assessment of Scoliosis severity or classification of curve type.

Jaremko et al. used an ANN to estimate Cobb angle using surface topography [42]. The surface topographic measurements were taken from digitized 3-D models of the patient's torso, and described various deformations of the torso surface. A genetic algorithm selected measurements for use in the ANN. The final ANN's Cobb angle estimates had a 66th percentile error of 5°, and 85th percentile error of 10°.

A similar work by Ramirez et al. used a SVM to classify patients' Cobb angle as "mild" (<30°), "moderate" (30°-44°), or severe (>44°) using surface topography and clinical measurements [154]. They reported 85% accuracy in classifying patients as "mild", or "non-mild", and 69% accuracy for the three-class problem. Ramirez' intended use of the system was to assist physicians in planning treatment. The 44° threshold corresponds to the range where surgical treatment is recommended, but it is unclear whether the 30° threshold is meaningful.

Tang *et al.* designed a system to automatically measure a Cobb-angle-like index from radiographic images. The system used fuzzy processing to identify the center of each vertebra, and then measured the discrepancy between the center points and a best-fit straight line. Tang reported good correlation (0.92) between this measurement and Cobb angles above 10°, making the system a potentially valuable support tool for assessment of Scoliosis severity [155].

Adankon et al. used surface topographic measurements to predict Scoliotic curve type (thoracic, double, or lumbar/thoracolumbar) using an SVM [156]. A large number of measurements were reduced to 53 variables using PCA, and a custom kernel was devised for SVM classification. Adankon reported classification accuracies of 84-97% for the various curve types. Such a system could potentially replace x-rays used to determine curve type.

Several researchers have applied machine learning to the problem of classifying types of Scoliotic deformity. Mezghani et al. used Kohonen Self-Organizing Maps (an extension of artificial neural networks) to automatically classify patients into “normal”, “moderate Scoliosis”, and “severe Scoliosis” categories using biplane X-rays [157]. These categories were defined with the goal of progressing away from Cobb angle toward a more 3-D representation of Scoliosis severity. The use of a computer method automated and standardized the classification. Sangole et al. and Stokes, et al. used clustering methods to identify new Scoliosis sub classifications based on several measurements taken from a 3-D reconstruction of the spine [158], [159]. Duong et al. performed a similar endeavour using fuzzy clustering [160]. These works indicated that 3-D descriptions of Scoliosis are more useful than the traditional Cobb angle alone.

Phan et al. used a self-organizing map to learn an AIS classification scheme from a database of radiographs [161]. The resulting map was compared to the Lenke classification system (which intends to classify patients based on how they should be surgically treated). The map better described physicians’ surgical choices than the Lenke system. Speaking of this work, Kang et al. encouraged future development of computational data analysis in the Scoliosis field, but stressed that further validation and humanized front-end software was required [162].

Nault *et al.* developed a CDSS to assist with surgical planning for patients with main thoracic curves [163]. The system used a fuzzy rule base to determine whether surgical correction should include the high thoracic and/or lumbar curves. The system’s recommendations on 30 test cases were compared to recommendations from 5 surgeons; Nault reported “good agreement” between the two (kappa values were 0.71 for the high thoracic curves and 0.64 for the lumbar curves).

In summary, some excellent work has been done applying computer methods to prediction, diagnosis, classification, and planning tasks in the Scoliosis domain. However this thesis represents the first step towards an electronic decision support system for brace treatment planning.

3. Brace Treatment Decision Support: A Preliminary Validation

In this chapter, retrospective AIS patient data is used in a preliminary validation of the concept of brace treatment decision support. The patient data is extracted from the clinic's records. Two models are developed to predict progression in braced patients based on start-of-treatment measurements. The best performing model is validated by comparing its prediction accuracy with that of Scoliosis experts. It is then used to generate hypothetical treatment recommendations for a set of test patients. A clinical trial simulation estimates that these recommendations may have reduced overall progression rates as well as the aggressiveness of treatment in some cases. The limitations of this preliminary validation are discussed.

3.1. Predicting Progression in Braced AIS Patients*

3.1.1. Patient Data

Patient data used in this preliminary validation was obtained retrospectively from the local scoliosis clinic's database, with approval from the local ethics board. Records were extracted for patients aged 9-16 years, who had been diagnosed with idiopathic scoliosis and had pre-brace Cobb angles between 20° and 45°. All patients had been treated in 2006 or later, and had finished brace treatment at the time of this study. The start date of 2006 was chosen because the primary orthotist had received new training and revised his treatment protocol at that time. One hundred two patient records met the criteria; 12 were excluded because key radiographic measures had not been recorded. Sixty two records were used as a training set for developing prediction models, twenty eight were used as a test set to evaluate prediction performance. Each record contained the treatment outcome – a Cobb angle progression in degrees (measured at the time of discharge from the brace), and 14 clinical measurements taken at the start of treatment. These clinical measurements included: age at brace fitting, sex, diagnosis, brace type, height,

* Material in this section has been published in the following papers:
- E Chalmers, W Pedrycz, and E Lou, "Predicting the Outcome of Brace Treatment for Scoliosis using Conditional Fuzzy Clustering," Joint Congress of the International Fuzzy Systems Association and North American Fuzzy Information Processing Society, June 24-28 2013
- E Chalmers, W Pedrycz, and E Lou, "Human experts' and a fuzzy model's predictions of outcomes of scoliosis treatment: A comparative analysis", IEEE Transactions on Biomedical Engineering (in press)
- E Chalmers, L Westover, J Jacob, A Donauer, H Zhao, E Parent, M Moreau, J Mahood, D Hedden, and E Lou, "Predicting success or failure of brace treatment for adolescents with idiopathic scoliosis", submitted to Medical & Biological Engineering & Computing, currently under review after revision

weight, Scoliotic curve type, Cobb angle, curve direction, vertebra number of curve apex, in-brace correction of major Cobb angle, Scoliometer measurement, height velocity, and weight velocity. Table 3 shows the distributions of each measurement.

Variable Number	Variable	Distribution
1	Age at fitting	13.5 ± 1.7 years
2	Sex	75 girls, 15 boys
3	Diagnosis	84 AIS, 6 JIS
4	Brace type	48 TLSO, 30 Charleston, 12 Boston
5	Height	159.9 ± 10.8 cm
6	Weight	49.0 ± 10.5 kg
7	Curve type	63 single, 27 double
8	Major Cobb angle	30 ± 7°
9	Major curve direction	56 Right, 34 Left
10	Major curve apex*	T11 ± 3 vertebrae
11	In-brace correction†	TLSO: 43 ± 29% Charleston: 112 ± 42% Boston: 23 ± 18%
12	Scoliometer measurement‡	9 ± 4°
13	Height velocity	5.6 ± 4.5 cm/year
14	Weight velocity	6.6 ± 6.4 kg/year

* Apex refers to the particular vertebra which falls at the "apex" of the spinal curve. It is an indication of the curve's vertical position in the spine.

† the instantaneous Cobb angle reduction achieved by the brace, expressed as a percentage of the pre-brace Cobb angle

‡ a measurement of vertebral rotation, taken by placing a specialized inclinometer on the patient's back while they are in a forward bending position

Table 3: Clinical measurements considered as predictor variables in predictive modelling.

One limitation of this data is that it does not include Risser sign – a measurement of skeletal maturity taken from a radiographic view of the pelvis which is believed to be a useful prognostic indicator in predicting progression. For many of the patients in this sample, Risser sign could not be measured due to cropping of the radiographs. Height velocity was included in the data instead; Little et al. suggest that height velocity is a better indicator of both growth and curve progression than Risser sign [50].

The 90 patients included 75 girls and 15 boys, aged 13.5 ± 1.7 years. Patients had been prescribed full-time thoraco-lumbar-sacral orthoses (60 cases) or night-time only braces (30 cases) by the attending surgeon or nurse practitioner, according to established guidelines. Pre-brace Cobb angles were 30 ± 7°. Forty one patients (46%) experienced >5° progression of their deformity by the end of brace treatment. The overall mean Cobb angle change was 5 ± 10°. Each record was labelled "progressed" if the patient's major Cobb angle increased more than 5° by the end of treatment, and "non-progressed" otherwise.

Missing values are a common problem in medical datasets. This dataset contained two missing values for height, three for weight, six for in-brace correction, fourteen for Scoliometer measurement, and six each for height velocity and weight velocity. These missing values were imputed using nearest neighbor imputation [164]: The missing value was imputed as the weighted average of the values from the 4 subjects who were most similar (based on Euclidean distance using all predictor variables).

3.1.2. Logistic Regression Model

A logistic regression model was trained to predict whether a braced patient's Cobb angle would progress by the end of treatment, based on measurements taken at the start-of-treatment. Logistic regression is an attractive technique for this task because it is familiar to most medical researchers – who use it as a statistical analysis tool for investigating relative strengths of association between inputs and an outcome. It is also a "white box" model – which is an advantage because clinical professionals are more likely to accept a prediction model when they can see (and agree with) the way it makes predictions [135].

A logistic regression models two-class data supplied as a training data set S = {[$\mathbf{X_1}$,Y_1], [$\mathbf{X_2}$,Y_2]… [$\mathbf{X_n}$,Y_n]}, where each instance $\mathbf{X_i} \in \mathbb{R}^d$ has a corresponding outcome $Y_i \in \{0,1\}$. Yi represents the class assignment for each instance: in this case, "progressed" (0) or "non-progressed" (1). In its basic form it fits the logistic function:

$$f(\mathbf{X}) = \frac{1}{1 + e^{-(\beta_0 + \boldsymbol{\beta_1}\mathbf{X})}} \qquad \text{Equation 1}$$

to the data by tuning the $\boldsymbol{\beta}$ parameters to maximize the log-likelihood function $l(\boldsymbol{\beta})$:

$$l(\boldsymbol{\beta}) = \sum_{i=0}^{n} \left[Y_i \log\left(f(\mathbf{X}_i)\right) + (1 - Y_i) \log\left(1 - f(\mathbf{X}_i)\right) \right] \qquad \text{Equation 2}$$

After the training process, $f(\mathbf{X})$ represents an estimate of the probability $P(Y=1|\mathbf{X})$. This estimate will be very good in the ideal case where the two classes follow normal distributions with similar covariance matrices but different means. In general though, the estimate can still be fairly good even in non-ideal situations [165].

A threshold of 0.5 was set on the model's output $f(\mathbf{X})$ to classify new patients as progressive ($f(\mathbf{X}) \geq 0.5$) or non-progressive ($f(\mathbf{X}) < 0.5$). This could be considered the "default" threshold, because $f(\mathbf{X}) = 0.5$ represents equal chances of progression and non-progression.

3.1.2.1. Predictor Variable Selection

Reducing the number of variables was necessary to ensure an interpretable model, and to remove variables which are redundant (containing roughly the same information as some other variable) or irrelevant (providing no information useful for predicting progression). Building a model using redundant or irrelevant variables can lead to overfitting; a condition in which the model describes idiosyncrasies in the data rather than underlying trends.

The number of variables was reduced using selection rather than transformation (e.g. principle component analysis) to preserve interpretability of the final model. A subset of useful features were selected from the list in Table 3, using a wrapper approach and a tiered cross-validation as described by Kohavi [146]. Cross-validation is an intuitive, low-bias method for validating a model by sequentially training and testing it on different portions of the same dataset [166]. It allows evaluation of a model's predictive power, rather than merely how well it fits the training data [152].

The variable selection was performed within a five-fold cross-validation scheme as illustrated in Figure 9. That is, five separate variable selection processes were performed on unique but overlapping subsets of the full training data. The entire procedure was repeated five times (the division into five folds was re-randomized each time) for a total of 25 selection processes performed on different subsets of the training data.

The variable selection process itself exhaustively searched through variable subsets of increasing size. For each variable subset, a logistic regression model was tested in a leave-one-out cross-validation. The Matthew's Correlation between each model's predictions and the true outcomes was used to select the best model within each search, and a running count tracked the number of times (out of 25) each variable appeared in a best subset.

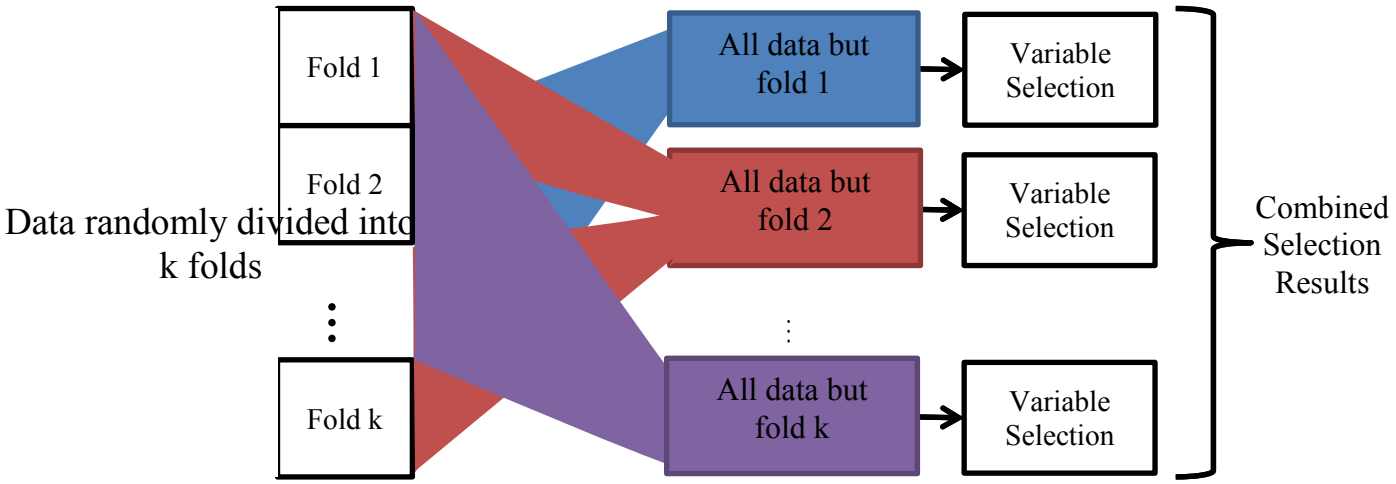


Figure 9: Illustration of variable selection using a k-fold cross-validation scheme.

The tiered use of cross validation (leave-one-out inside of five-fold cross validation) helps prevent overfitting [156]; Some overfitting may occur within each search, however these effects are averaged out across the multiple searches. Conversely, the variables describing underlying trends are selected often, and these selections accumulate across multiple searches. The number of selections is shown in Figure 10.

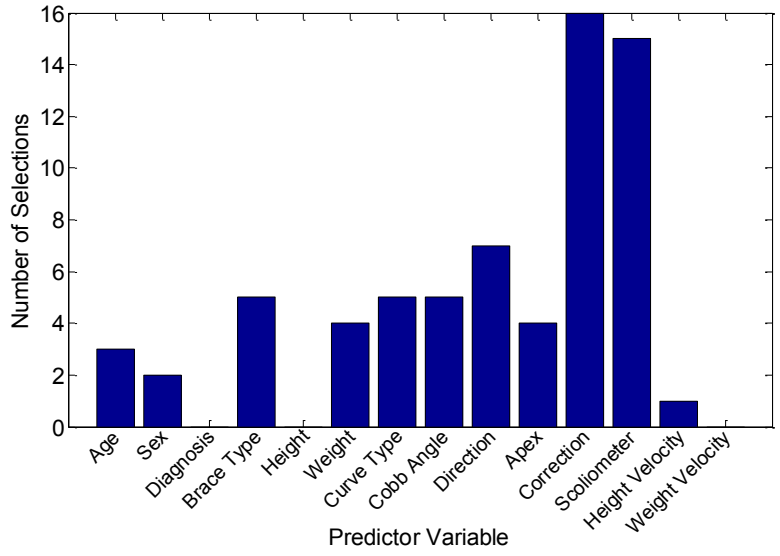


Figure 10: Number of times each variable was selected during the predictor variable selection process. Variable numbers correspond to Table 3. Variables 11 (in-brace correction) and 12 (Scoliometer measurement) were clearly preferred.

In-brace correction and Scoliometer measurement were selected 16 and 15 times respectively, while all other variables were selected 7 times or less. Based on these results in-brace correction and Scoliometer measurement were chosen as predictor variables for the model.

3.1.2.2. Training the Final Model

In a small dataset even the simple logistic regression model can be influenced by random noise and outliers. To mitigate this effect we padded the dataset with randomly generated data – a form of dithering which can improve a model`s robustness [167]. The mean and covariance of the available data were measured and used to generate additional data in a multivariate normal distribution (Figure 11). These additional data points can be thought of as "hypothetical" subjects, postulated following the distributions observed in the real subject data. Adding this noise improves the model's noise immunity by "training" it to expect some noise. The final padded dataset contained 692 points (62 original subjects, and 630 randomly generated subjects).

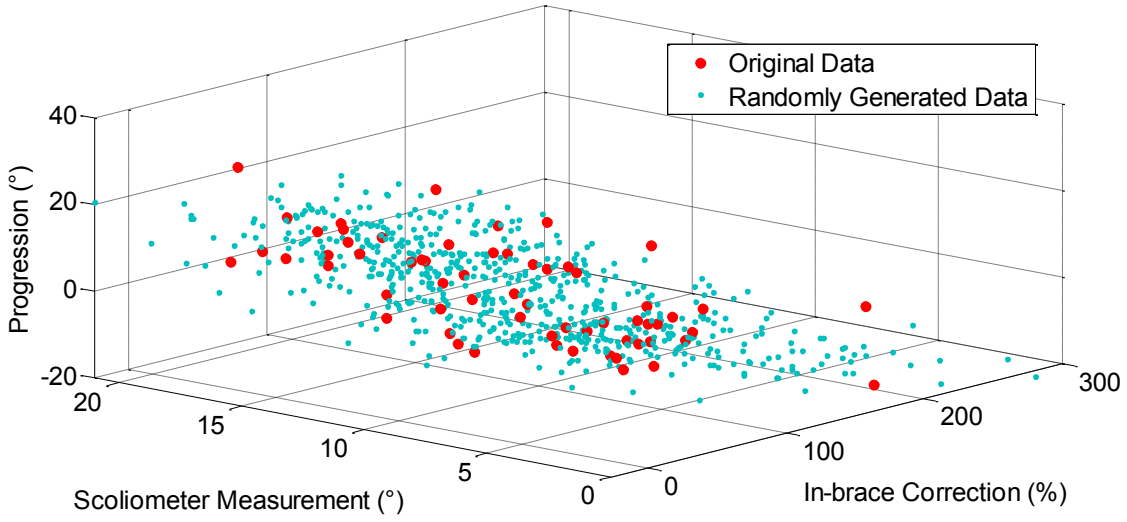


Figure 11: In-brace correction, Scoliometer measurement, and progression for original subjects and randomly generated data.

Logistic regression was applied to the training data after variable selection and generation of additional data. The model's "training" accuracy was estimated in a leave-one-out cross validation of the training data. Although the model was trained using all training data (the original as well as randomly generated data) only the predictions on the original 62 subjects were used in cross-validated performance estimation. A final model was trained on all available

training data and tested on the 28-patient test set to evaluate its prediction accuracy. The final model's output $f(\mathbf{X})$ is depicted as a function of patients' scoliometer and in-brace correction in Figure 12. The logistic regression coefficients for in-brace correction and Scoliometer were -0.0099 and 0.1445 respectively, with a constant term of -0.7607.

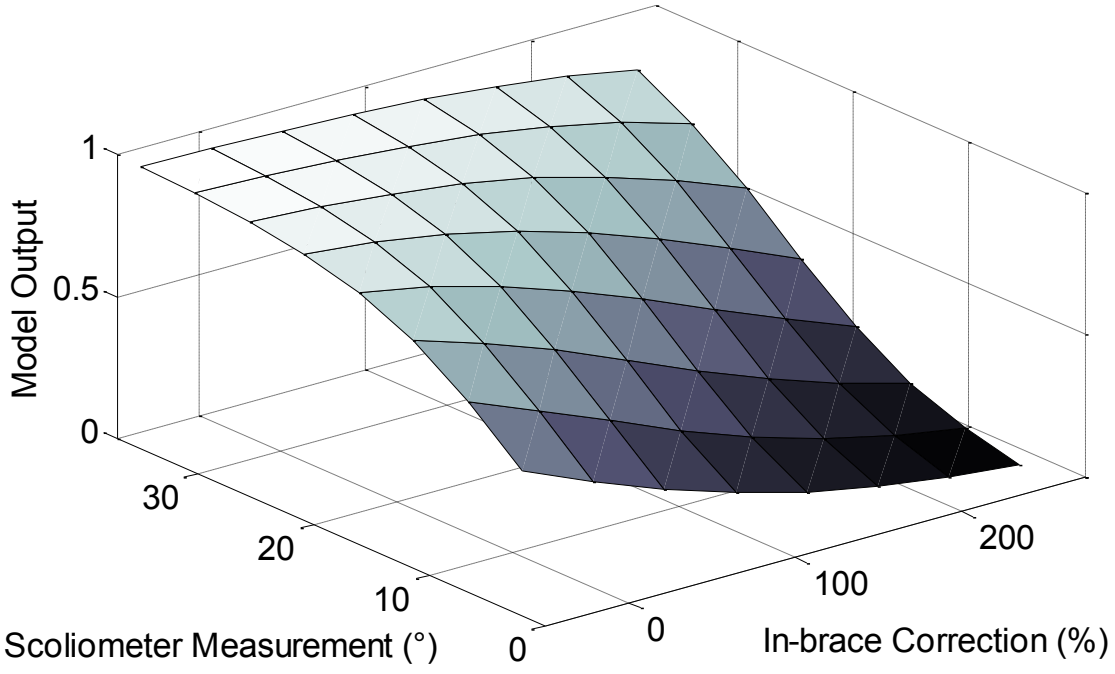


Figure 12: Logistic function output shown for the range of in-brace corrections and Scoliometer measurements. The model output loosely represents the probability that the subject will progress.

Applying the threshold of $f(\mathbf{X}) = 0.5$ created the decision boundary for classifying patients into the predicted progression or predicted non-progression categories. The decision boundary is shown in Figure 13 along with markers representing the 90 AIS patients in the training and test datasets. Note that some of the patient's braces overcorrected their deformities – thus the in-brace correction percentage can exceed 100%.

The model's operation seems reasonable – the logistic regression coefficients indicate that progression correlates positively with Scoliometer measurement and negatively with in-brace correction (as expected). The selection of in-brace correction as a predictor makes sense, as does the selection of Scoliometer given recent research which has found association between transverse-plane deformities and progression [77], [168], [169].

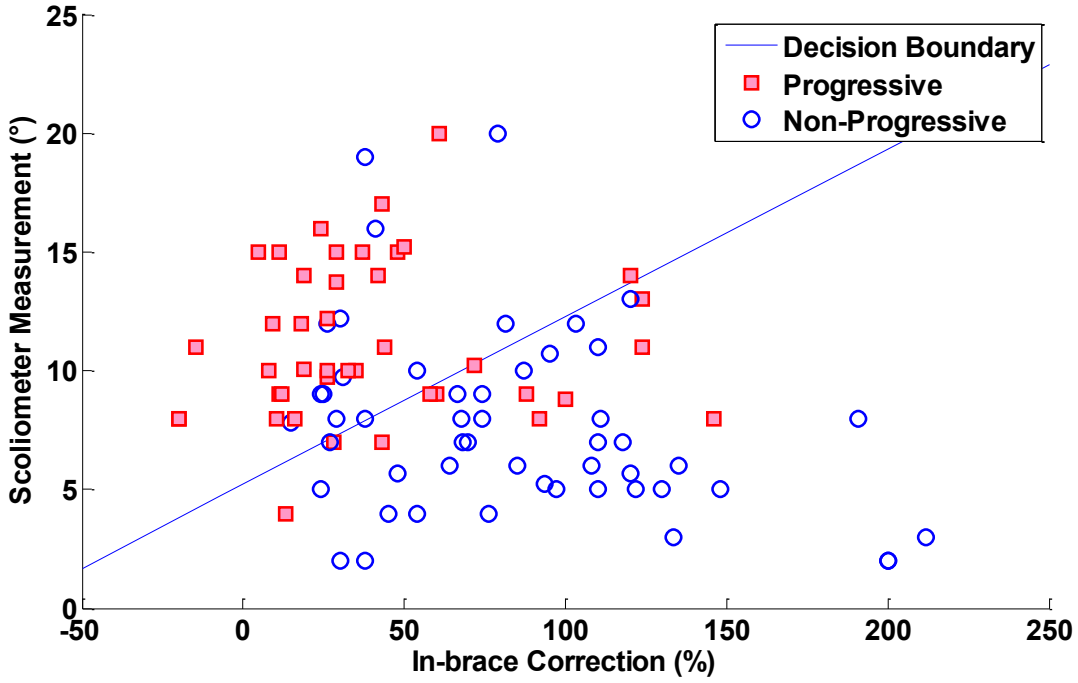


Figure 13: In-brace correction and Scoliometer measurement for all 90 subjects. The straight line shows the decision boundary separating predicted progression (above the line) from predicted non-progression (below).

3.1.3. Fuzzy Model

The logistic regression model described in section 3.1.2 is simple, interpretable, and reasonably effective. A second model was also developed to see if better performance could be obtained using a more sophisticated approach. Generally speaking, a more sophisticated model is a less interpretable one, with a greater risk of rejection by users. Kruppa et al. point out:

> *"The fundamental question is whether a clinician will trust the findings obtained with a fancy non-interpretable machine and use this in clinical routine" [165]*

Fuzzy processing offers an attractive compromise between complexity and interpretability, by allowing complex mathematical constructs to be represented linguistically.

Fuzzy processing makes use of fuzzy sets. In conventional set theory, membership (belongingness) is Boolean: elements either do, or do not belong to a set. By contrast, a fuzzy set's members exhibit degrees of membership in the range [0, 1] [149]. An element's membership in a fuzzy set is given by a membership function. For example, a fuzzy set defined by the triangular membership function tri(2x-1), includes the element x=1.5 with a membership

value of 0.5. The element x=1 has membership of 1 (full membership) and x= -1 has membership of 0 (no membership).

Fuzzy set theory allows sets to more accurately reflect the "fuzzy" nature of real-world concepts. For example, a Cobb angle increase of 6° or more is often labelled "progression", while an increase of 5° or less is labelled "non-progression". This dichotomy is quite artificial (is a 5° change fundamentally that different from a 6° change?) and does not reflect the true continuous nature of progression. However the *concepts* of "progression" and "non-progression" are an important part of clinical practice. We could get more realistic definitions of progression/non-progression and still retain the concepts by defining "progression" and "non-progression" as overlapping fuzzy sets of Cobb angle changes. The fuzzy sets could be defined such that a, say, 0° Cobb angle change would have high membership in the "non-progression" set, 10° would have high membership in the "progression" set, and a 5.5° change would have intermediate membership in both. The fuzzy transition between "progression" and "non-progression" concepts would reflect the continuous nature of progression more accurately than the conventional 5° threshold.

3.1.3.1. Fuzzy C-Means Clustering (FCM) and Conditional FCM

The second model for predicting progression in braced patients used Conditional Fuzzy C-Means Clustering, a fuzzy processing technique which is an extension of FCM.

FCM is a clustering algorithm which partitions data into overlapping (fuzzy) subsets, based on similarity. Conventional clustering methods (K-Means [170] for example) partition groups of similar observations into clusters, with each observation belonging exclusively to one cluster. By contrast, fuzzy clusters include each observation to some degree – in other words, each observation has some degree of membership in each cluster, based on its proximity to the cluster's representative or "prototype" point. The fuzzy cluster assignments characterize the uncertainty faced in real-world classification problems.

The FCM algorithm discovers *c* clusters in a data set S = $\{\mathbf{X_1}, \mathbf{X_2}, \ldots \mathbf{X_n}\}$, where each instance $\mathbf{X_i} \in \mathbb{R}^d$. A partition matrix $\mathbf{U}=[u_{ik}]$ describes the cluster memberships: u_{ik} is the membership of the k^{th} observation in the i^{th} cluster. The i^{th} cluster is represented by a prototype $\mathbf{v_i} \in \mathbb{R}n$. The algorithm iteratively positions these prototypes and updates the partition matrix U to minimize the sum of square distances, which is the objective function used in the clustering:

$$Q = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik}^{m} \|\mathbf{X_k} - \mathbf{v_i}\|^2 \qquad \text{Equation 3}$$

The parameter m is the fuzzification coefficient, which controls the "fuzziness" of the transitions between clusters. It is often set to 2. The use of Euclidean distance allows a closed-form calculation of the optimal prototypes as the following weighted average:

$$\mathbf{v_i} = \frac{\sum_{k=1}^{n} {u_{ik}}^{m} \mathbf{X_k}}{\sum_{k=1}^{n} {u_{ik}}^{m}} \qquad \text{Equation 4}$$

Cluster memberships are then re-assigned given the new prototype locations:

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|\mathbf{X_k} - \mathbf{v_i}\|}{\|\mathbf{X_k} - \mathbf{v_j}\|} \right)^{2/(m-1)}} \qquad \text{Equation 5}$$

The algorithm iterates until the changes in membership assignments are sufficiently small. Equation 5 – in conjunction with the final prototypes – acts as a membership function for each fuzzy cluster.

Being an unsupervised learning algorithm, FCM finds clusters which are generic: they have no meaning beyond the similarity of the points they contain. The Conditional FCM algorithm (CFCM) [171] transforms FCM into a semi-supervised learning method where the clustering process is guided by additional information – namely the observations' memberships in several "contexts". Each context is a fuzzy set defined on some variable of interest, usually an outcome or predicted variable.

CFCM gives the partition matrix an extra dimension to represent context, so u_{ikj} denotes the membership of the kth observation in the ith of several (let us say c_j) clusters corresponding to the jth context. Each observation $\mathbf{X}$ comes with a corresponding row in the table w, where w_{kj} indicates the kth observation's membership in the jth context. The clustering is performed for each context separately. Partition matrix entries are calculated:

$$u_{ikj} = \frac{w_{jk}}{\sum_{h=1}^{c_j} \left(\frac{\|\mathbf{X_k} - \mathbf{v_i}\|}{\|\mathbf{X_k} - \mathbf{v_h}\|} \right)^{2/(m-1)}} \qquad \text{Equation 6}$$

Thus the clustering process within each context is constrained by observations' memberships in that context. The upshot of this is that CFCM discovers clusters of observations which belong to particular contexts. That is, it discovers clusters *of a particular type*, rather than FCM's generic and data-driven-only clusters (see Figure 14). For this application to brace treatment, CFCM is

used to discover clusters of patients who all have similar brace treatment outcomes. Predicting a new patient's outcome is then a matter of measuring their membership in these clusters.

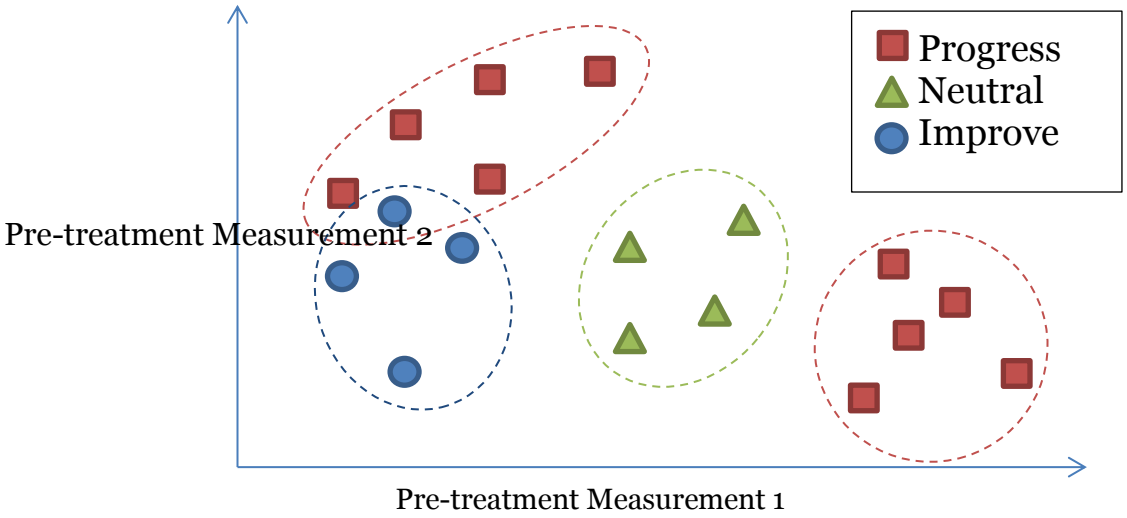


Figure 14: CFCM discovers clusters of observations belonging to particular contexts. This (hypothetical) example shows clusters of patients who improve, remain neutral, or progress during brace treatment (three unique contexts). Without the context information, a conventional clustering algorithm might merge the "improved" cluster with a "progressed" cluster.

3.1.3.2. CFCM-based Prediction Model

Three meaningful brace-treatment outcomes or *contexts* were identified: "improvement", "neutral", and "progression". Each context was defined as a fuzzy set of Cobb angle progressions measured after discharge from brace treatment. These fuzzy sets' membership functions are shown in Figure 15. Each of these fuzzy sets represents a unique treatment outcome, and together they reflect the nature of our modeling problem. The "neutral" context was designed to intersect the "improved" context at -0.5° progression, and the "progressed" context at 5.5°. This is convenient because the Scoliosis Research Society's standardized definition of progression is a Cobb angle increase greater than 5° [40]. The fuzzy sets were based on Gaussian membership functions with a variance of 1° - which is in the range of estimated values for variance of the Cobb angle measurement [21]. Thus, for example, the "progressed" fuzzy set suggests the degree to which a measured progression represents an actual progression greater than 5°. Using these fuzzy sets as contexts in the CFCM clustering resulted in clusters of patients who tend to progress, improve, or remain neutral after brace treatment. Thus a Cobb angle improvement is considered to be an inherently different thing than simply not progressing.

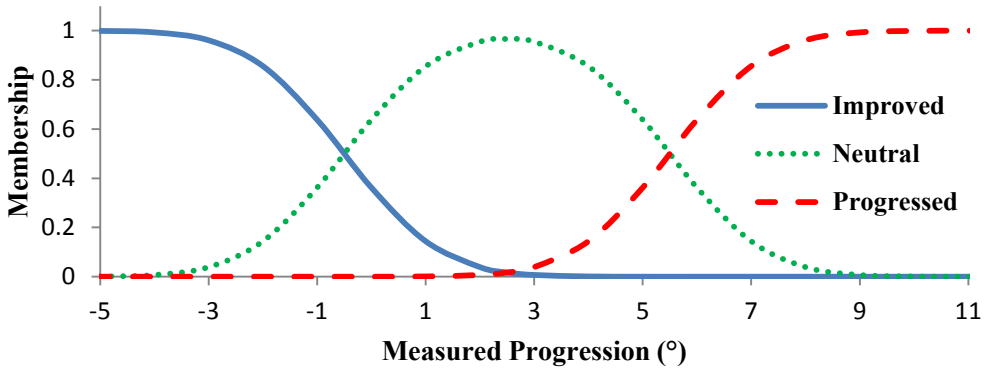


Figure 15: Fuzzy sets defining the three treatment outcome contexts used in CFCM clustering. A measured Cobb angle progression is mapped to three context memberships using these membership functions.

The 62 patient records in the training set were each assigned degrees of membership in the "progressed", "improved", and "neutral" contexts based on their measured progression at the end of brace treatment. The CFCM clustering grouped the 62 patients into 8 fuzzy clusters: 2 from the "neutral" context, and 3 from each of the other contexts. The number of clusters (prototypes) was chosen empirically, by plotting the fuzzy clusters' goodness of fit as a function of the number of prototypes (see Figure 16). Points of diminishing returns in these plots indicated suitable numbers of prototypes. This was done for each of the three outcomes separately.

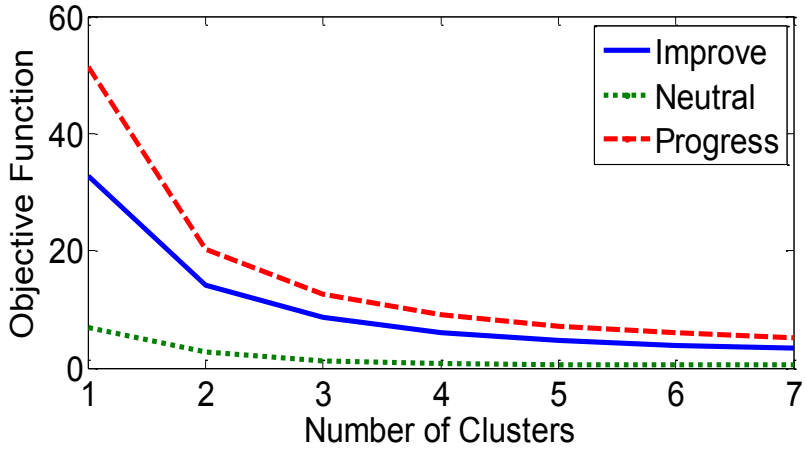


Figure 16: Objective function Q versus number of clusters in each context. Points of diminishing returns suggest appropriate numbers of clusters for each context.

The clustering used a custom distance function instead of conventional Euclidean distance. Distance between two points was calculated as the angle between the vectors connecting each

point to the global mean $\bar{\mathbf{X}}$ of the training data. That is, the distance between any two points $\mathbf{X_a}$ and $\mathbf{X_b}$ is:

$$\|\mathbf{X_a} - \mathbf{X_b}\| = \cos^{-1}\left(\frac{\acute{\mathbf{X}}_\mathbf{a} \cdot \acute{\mathbf{X}}_\mathbf{b}}{|\acute{\mathbf{X}}_\mathbf{a}||\acute{\mathbf{X}}_\mathbf{b}|}\right) \qquad \text{Equation 7}$$

where

$$\acute{\mathbf{X}}_\mathbf{a} = \mathbf{X_a} - \bar{\mathbf{X}}, \qquad \acute{\mathbf{X}}_\mathbf{b} = \mathbf{X_b} - \bar{\mathbf{X}} \qquad \text{Equation 8}$$

This distance function compares patients on the basis of how they differ from average. Ultimately this means that the prototypes reveal how patients who tend to progress or improve differ from average. The concept is illustrated in Figure 17. Each prototype was placed by a downhill simplex algorithm (Equation 4 is derived assuming the use of Euclidean distance).

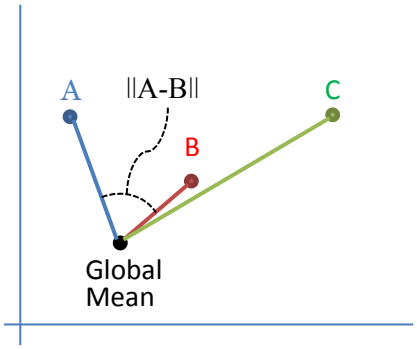


Figure 17: Illustration of how distance between two points is represented. In this example $\|$B-C$\|$ < $\|$A-B$\|$, because B and C lie in similar 'directions' from the global mean.

Treatment outcome for new patients was predicted by first calculating the patient's membership in every cluster using the conventional FCM membership calculation. That is, the membership u_g of a new patient (described by feature vector $\mathbf{X}$) in prototype $\mathbf{v_g}$'s cluster is:

$$u_g = \frac{1}{\sum_{h=1}^{c}\left(\frac{\|\mathbf{X} - \mathbf{v_g}\|}{\|\mathbf{X} - \mathbf{v_h}\|}\right)^{2/(m-1)}} \qquad \text{Equation 9}$$

where c is the total number of clusters in all contexts. The memberships for all clusters in a given context were then summed to give the membership in that context. This produced predicted memberships in the "improved", "progressed", and "neutral" outcome categories. The model provided these three membership values along with a dichotomous progressed/non-progressed label (per the Scoliosis Research Society standard for reporting brace treatment

outcome [40]). The label was “progressed” if the “progressed” membership was larger than the “neutral” and “improved” memberships and “non-progressed” otherwise.

3.1.3.3. Predictor Variable Selection

Reducing the number of variables was necessary to ensure an interpretable model, and because the 62-patient training set was too small to accommodate eight, 14-dimensional prototypes. To preserve interpretability of the model, feature reduction was achieved through selection rather than transformation (i.e. principle component analysis, etc.).

The variable selection procedure was similar to that described in section 3.1.2.1. It exhaustively searched through feature subsets of size 5 or less. For each feature subset, the modeling process described above was tested in a 10-fold cross validation. The feature subset which maximized correlation between the model’s predictions and the true outcome was selected. This procedure was performed inside of a 5-fold cross validation, and repeated 5 times (with the 5 folds re-randomized each time) for a total of 25 searches. Features were ranked on the number of times they were selected by a search. Selection results are shown in Figure 18.

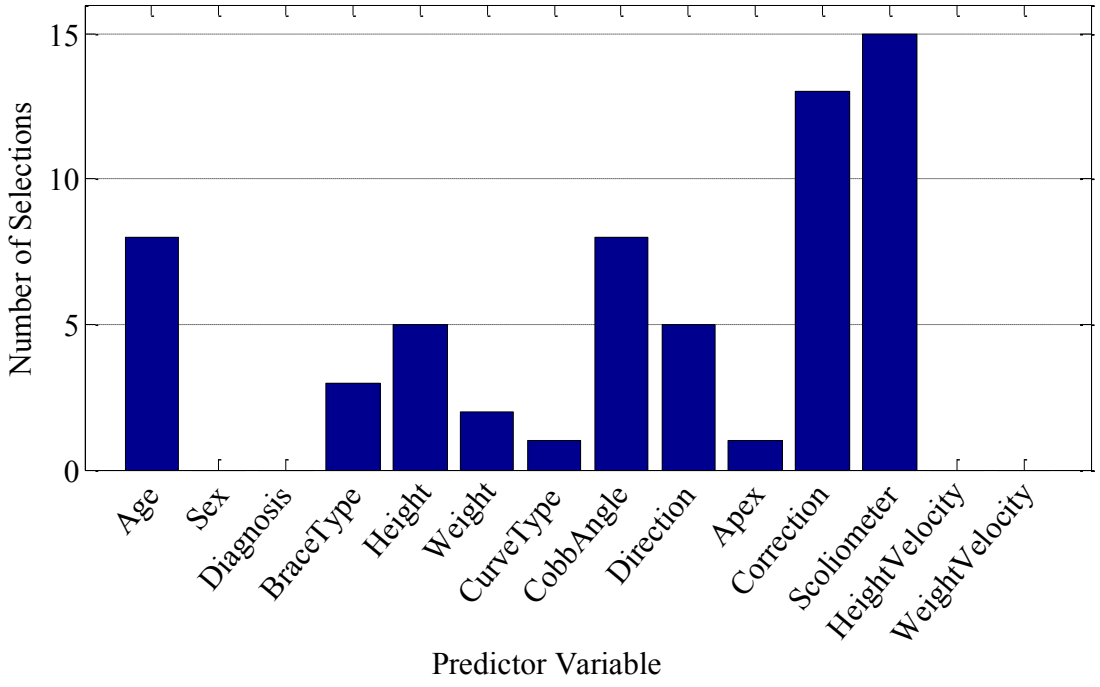


Figure 18: Number of selections (out of 25) for each variable during the selection process.

3.1.3.4. Training the Final Model

The final model was built by adding features one-at-a-time in order of rank and testing in a 10-fold cross validation. The addition of features stopped when adding a feature did not improve accuracy. The final model retained four features: Scoliometer measurement, in-brace correction, Cobb angle, and patient age. Thus the feature selection process retained two features describing the scoliosis, one describing the patient, and one describing the treatment. The final CFCM-derived prototypes are shown in Table 4. The final model was tested on the 28 patient test dataset.

Prototype #	Context	Scoliometer (°)	In-brace Correction (%)	Cobb Angle (°)	Age (yrs)
	Global Mean	9	66	30	13.6
1	Improve	7	61	27	10.8
2		5	142	25	12.9
3		7	87	22	14.6
4	Neutral	4	60	27	13.6
5		16	60	41	15.4
6	Progress	10	101	26	15.3
7		10	20	38	13.6
8		14	38	31	12.3

Table 4: Feature values for each prototype, and the global mean.

Most prototypes in Table 4 seem consistent with existing literature: prototypes 1-4 have mild deformities and/or good in-brace correction, so we would not expect them to progress. Prototype 5 has a more severe deformity but is quite old (near skeletal maturity) – so the fact that they are "neutral" is reasonable. Prototypes 7 and 8 have severe deformities and poor correction, so they would be expected to progress.

However, prototype 6 does not seem to belong to the "progress" context when compared to prototype 5. Prototype 6 had a more mild deformity and better correction, but apparently progressed more than prototype 5. This is an interesting discrepancy; the most likely explanation is that prototype 6 describes a cluster of patients with good in-brace correction but poor compliance, i.e. they were not faithful in their brace-wear.

3.1.4. Comparison of Logistic Regression & Fuzzy Model Performance

Table 5 is a confusion matrix showing the logistic regression model's prediction performance. It shows the number of progressive/non-progressive cases which were correctly or incorrectly classified during training and testing. The model achieved 73% accuracy in the leave-one-out cross validation on the training data. The final model included the two most-selected variables: it can be noted that adding the third-ranked variable (curve direction) had no effect on cross-validated accuracy, and using only the top-ranked variable (in-brace correction) resulted in slightly lower accuracy of 68%. The model's predictions on the test data were 75% accurate. Table 6 shows the confusion matrix for the fuzzy model. The fuzzy model's cross validated accuracy in training was 77%, and it's accuracy on the test data was 82%.

		Predicted Outcome	
		Progress	Non-Progress
True Outcome	Progress	25 [10]	9 [2]
	Non-Progress	8 [5]	20 [11]

Table 5: Confusion matrix showing the numbers of correct and incorrect predictions made by the logistic regression model on the training [test] data. Training performance was measured in cross validation.

		Predicted Outcome	
		Progress	Non-Progress
True Outcome	Progress	24 [12]	5 [0]
	Non-Progress	9 [5]	24 [11]

Table 6: Confusion matrix showing the numbers of correct and incorrect predictions made by the fuzzy model on the training [test] data. Training performance was measured in cross validation.

Figure 19 compares the prediction performance of the logistic regression model to that of the fuzzy model in terms of Mathew's Correlation Coefficient (MCC), sensitivity, specificity, and F1 and F2 measures. Mathew's Correlation measures the correlation between the treatment outcomes predicted by the model and the true outcomes. An MCC of 1 indicates perfect agreement, an MCC of 0 indicates no correlation, and an MCC of -1 indicates perfect

disagreement. Sensitivity is the fraction of progressive cases which were correctly identified, while specificity is the number of non-progressive cases which were correctly identified. The F1 measure is a balanced measurement of performance which considers both sensitivity and precision – the fraction of the "progress" predictions which are actually correct. The F2 measure is similar to the F1 measure but places a greater emphasis on sensitivity. The statistics shown in Figure 19 were calculated by pooling the training and test predictions shown in Table 5 and Table 6.

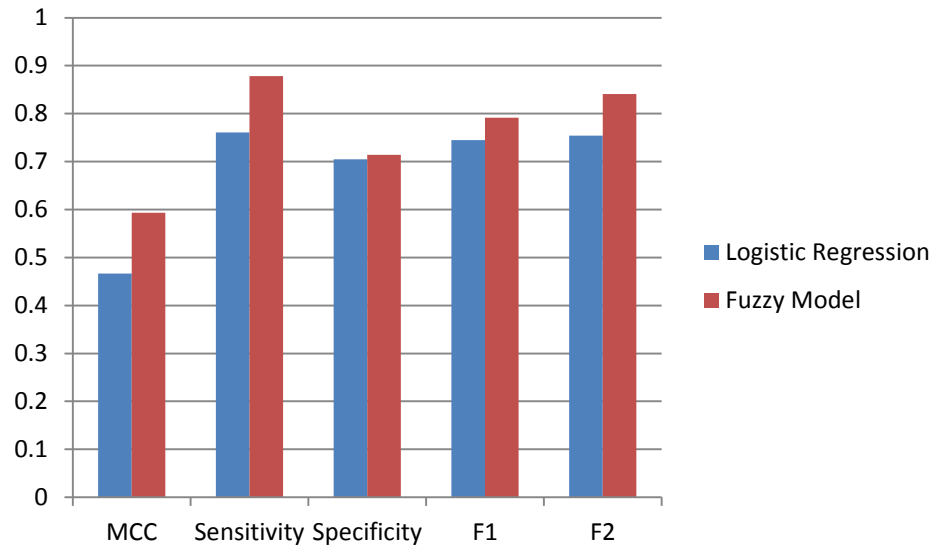


Figure 19: Comparison of selected performance measures for the logistic regression model and the fuzzy model.

The fuzzy model's predictions on the pooled training and test data are significantly different from the logistic regression model's predictions: McNemar's test gives a one-tailed p value of 0.049. The fuzzy model shows a slight advantage in each of the metrics shown in Figure 19. Of particular interest are the sensitivity and F2 measure. This is because for the task of predicting progression, a "false negative" is more costly than a "false positive". That is, sensitivity (the ability to identify progressive individuals) is more important than specificity (the ability to identify non-progressive individuals) in this application.

In conclusion, the fuzzy model has a modest advantage over the logistic regression in terms of predictive power. It has a convenient advantage in terms of sensitivity, in particular. The fuzzy model has a human-readable interpretation (via its prototypes, shown in Table 4), but the logistic regression retains an advantage in interpretability, being a much simpler model. Its simplicity would also make it easier to implement. Since this application has no pressing need

for model simplicity, the fuzzy model is preferable to the logistic regression due to its superior prediction performance.

3.2. Fuzzy Model's Predictions Compared to Experts'[†]

A physician makes Scoliosis treatment decisions using their own experience to predict risk of progression. If a computer model can improve these predictions, it will likely improve the resulting treatment decisions. Comparing computer models to human experts is a common benchmark. Nault et al. [163] designed a fuzzy system to recommend the number of vertebrae to fuse during surgical correction of AIS. The system's recommendations on a set of test cases were compared to surgeons' selections on the same cases; the system's recommendations had "good agreement" with the surgeons. A comparison of this nature is intended to show that the computer model operates at the level of experts, or that it emulates the operation of experts.

Alternatively, a model's prediction performance can be compared to experts' performance vis-à-vis test cases. For example, Uyar et al. [172] designed a model to predict implantation of embryos in in-vitro fertilization. The model's predictions were compared to embryologists' predictions on new cases over a period of two months, with the result that the model's predictions were more accurate than the embryologists'. Rodriquez-Gonzalez et al. [116] compared the diagnoses of a computerized CDSS to those of physicians and found the CDSS's diagnoses to be more accurate. Studies like these attempt to show a computer model's ability to "out-predict" experts. The intent is to show the model's potential to enhance an expert's natural prediction ability.

A computer model is certainly not guaranteed to have better predictive power than experts. Farion et al. [173] developed a model to diagnose asthma exacerbations, but found that it performed more poorly than physicians. Cornu et al. [174] evaluated a CDSS intended to identify drug-drug interactions, and found it identified fewer interactions than pharmacists with more false-positives. Demonstrating a prediction model's superior performance over experts is an important step in demonstrating its usefulness. If it makes better predictions, it could potentially be used to enhance the expert's decision making process. On the other hand a

[†] Material in this section has been published in the paper: E Chalmers, W Pedrycz, and E Lou, "Human experts' and a fuzzy model's predictions of outcomes of scoliosis treatment: A comparative analysis", IEEE Transactions on Biomedical Engineering (in press)

physician would have no use for a model's predictions if they could produce better ones themselves.

This work compared the fuzzy model described in section 3.1.3 to AIS experts, on the basis of ability to predict Scoliosis progression during brace treatment.

3.2.1. Panel of Experts

The experts participating in the comparison included:

- Two orthopaedic surgeons, each with approximately 30 years of experience treating Adolescent Scoliosis.
- One nurse practitioner with 3 years of experience working with Scoliosis cases.
- Three orthotists with approximately 10, 20, and 7 years of experience constructing and adjusting braces for treating AIS.
- One research scientist with 22 years of experience in Scoliosis research, and over 40 refereed publications related to brace treatment.
- One clinical engineer with 25 years of experience working in Scoliosis clinics, and over 50 refereed publications related to Scoliosis and brace treatment.

Almost all brace-treated AIS cases in the local health region are overseen by the nurse practitioner or one of the two orthopaedic surgeons, though a small number are seen by a third surgeon who was unable to participate in the study. All braces in the region are designed and built by the three orthotists. The orthotists and surgeon or nurse practitioner work together to devise a brace treatment predicted to minimize a patient's risk of progression. Predicting the risk of progression is an inherent part of this process. The scientist and clinical engineer research scoliosis progression and risk factors, and both have multiple publications on the topic of predicting progression. These eight experts represent the most knowledgeable personnel on brace treatment in the region.

3.2.2. Comparison Procedure

Records of the 28 patients in the test dataset were supplied to each of the experts. Each record contained the following measurements taken at the start of the patients' brace treatment: age, gender, time since menarche (where applicable), type of brace used and in-brace correction applied, height, weight, Cobb angle, location of scoliotic curve apex, Scoliometer measurement,

height velocity, weight velocity, and observed progression prior to bracing. The true outcomes (the Cobb angle progression of each patient by the end of brace treatment) were not shown to the experts. Each expert was asked to predict whether each of the 28 patients would progress by the end of brace treatment. Experts did not collaborate with each other, as they typically make decisions on their own in clinical settings. No time limit was imposed on the exercise.

The fuzzy model's predictions (the dichotomous progressed/non-progressed labels) for these 28 patients were also obtained, based on their scoliometer measurements, in-brace corrections, Cobb angles, and ages. The model's and each expert's predictions were evaluated for agreement with the actual treatment outcomes using Mathew's Correlation Coefficient (MCC) [33]. The experts were also considered as a panel by taking their mode prediction. Sensitivity, specificity, and positive and negative predictive values (PPV and NPV) were calculated for the panel and the fuzzy model. McNemar's test was used to check statistical significance of the differences between predictors. The multi-rater kappa was also calculated to measure agreement between the expert's predictions.

In addition to comparing with experts' predictions, the fuzzy model was also compared to alternatives from literature: prediction methods proposed by Lonstein et al. [39] and Peterson et al. [45] were implemented for comparison with the fuzzy model. These methods have seen perhaps the most clinical use; Lonstein's was used in guidelines published by the Society on Scoliosis Orthopaedic and Rehabilitation Treatment (SOSORT) until 2011 [56]. SOSORT's 2011 guidelines avoided sanctioning any method, however [7]. Peterson's method has been cited numerous times and its predictions are one of the inputs to Lou's model [3].

Lonstein's method uses a patient's Cobb angle, Risser sign, and age as inputs. Peterson's uses Risser sign, age, imbalance of the spine, and apical level of the curve. Risser sign information did not exist for 6 of the 28 test patients (due to X-ray cropping). Thus Lonstein's and Peterson's methods were applied to the remaining 22 patients. Mathew's Correlation between the methods' predictions and actual outcome was calculated for these 22 patients, and compared to the correlation for the fuzzy model.

3.2.3. Results

The multi-rater kappa measure of agreement among experts was 0.47, showing only moderate agreement between experts. Some experts tended to favor predictions of "progression", while others favored "non-progression". There was no noticeable correlation between the expert's level

of experience and prediction performance. This seems to illustrate the difficulty of predicting brace treatment outcome.

The model outperformed (in terms of MCC) all but one expert, and performed similarly to the panel of experts. Figure 20 shows MCC for each expert's predictions, as well as for the panel and the model. The expert's MCC scores had mean 0.55 and standard deviation 0.08 – if experts' scores were drawn from a corresponding normal distribution then the model's improvement in MCC is statistically significant at p=0.03.

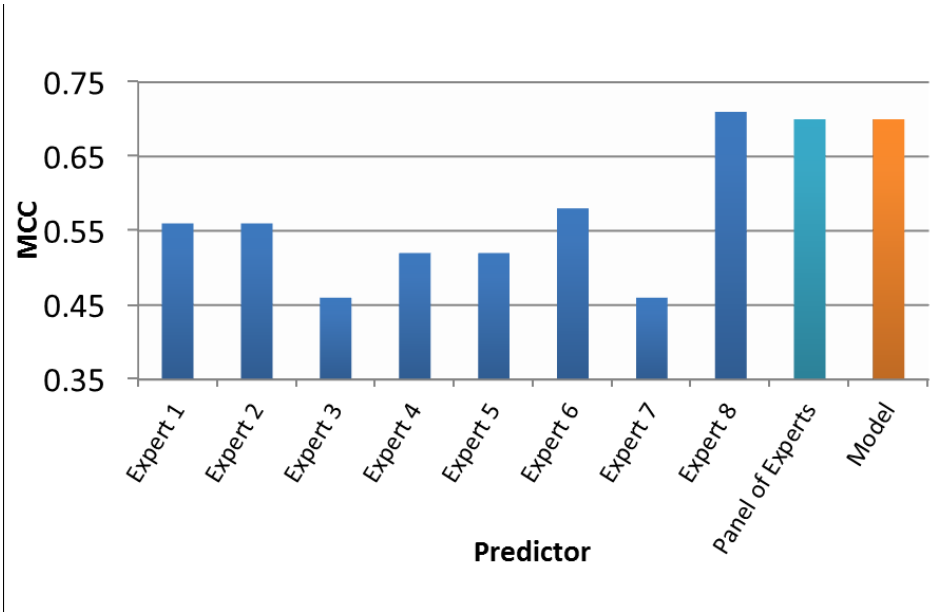


Figure 20: Mathew's correlation between predicted and actual outcomes for each expert's predictions, the panel of experts (the mode prediction across experts), and the model.

A detailed performance comparison between the panel of experts and the fuzzy model is shown in Figure 21. Sensitivities for the panel and model were 91% and 100% respectively, specificities were 80% and 69%, PPVs were 77% and 71%, and NPVs were 92% and 100%.

McNemar's test showed p-values less than 0.1 when comparing the model to experts 1, 2, 6, and 8, and when comparing expert 1 to experts 3, 4, and 7. Interestingly, one of these cases of significantly different predictions (the model compared to expert 8) was the case in which the expert out-predicted the model.

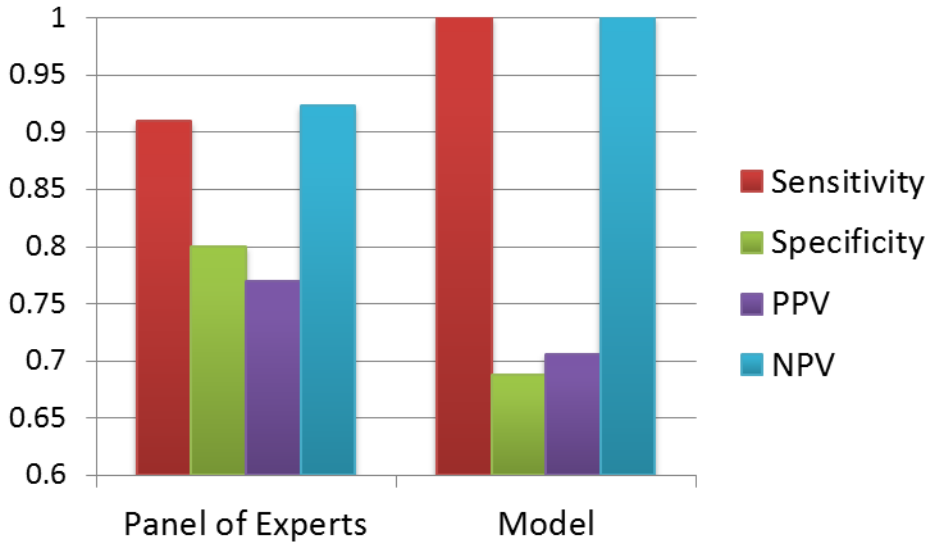


Figure 21: Comparison of statistics describing the prediction performances of the fuzzy model and the panel of experts.

Figure 22 shows the comparison between Lonstein and Peterson's prediction methods and the model. Predictions generated using Lonstein's method showed an insignificant (negative) correlation with actual outcomes.

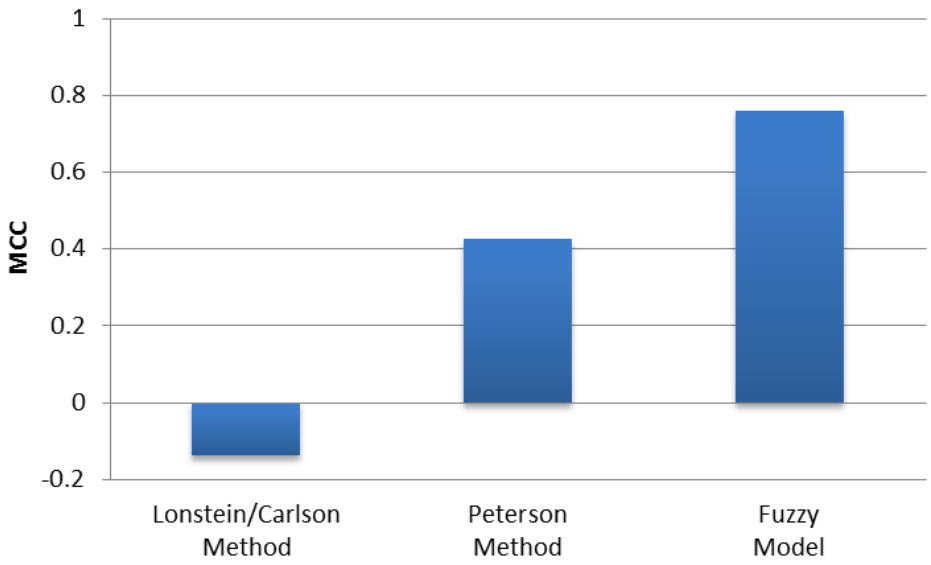


Figure 22: Mathew's correlation between predicted and actual outcomes for Lonstein and Peterson's prediction methods, and the model.

3.3. Fuzzy-Model-Based Treatment Recommendations[‡]

This work envisions a decision support system based on machine-learning-derived prediction model(s), in the style described by Chi [126]. The model's function is to predict treatment outcome based on parameters measureable at the start of treatment. These predictions are fed into a decision support engine, which communicates to an end user how a particular patient's predicted outcome changes as a function of these parameters. Essentially the system enables a "what-if" analysis; providing some insight into what might happen in the future, given particular actions or conditions in the present.

Using a computer model's predictions of future events to identify optimal decisions in the present is sometimes called "prescriptive analytics" [129], and is a logical next step if the prediction model is trustworthy. In fact the prescriptive analytics concept describes how we make most important decisions: when faced with a decision between several alternatives, we often mentally forecast the result of each action before choosing the one which seems most desirable.

Figure 23 shows the fuzzy model's predictions of a particular patient's memberships in the "improve", "neutral", and "progress" outcome contexts. The memberships are shown as functions of the patients' in-brace correction and scoliometer measurement. This type of plot shows graphically how the predictions vary in response to these parameters.

In this preliminary validation, a decision support engine was implemented which used the fuzzy model's predictions to provide treatment recommendations. The model predicts treatment outcome based on three uncontrollable features (Scoliometer, Cobb angle, and age) and one controllable one (in-brace correction). Thus, the decision support engine used the model's predictions to recommend optimum in-brace correction given a patient's Scoliometer, Cobb angle, and age.

First the model generated a complete set of predictions for every possible combination of all four

---

[‡] Material in this section has been published in the following papers:
- E Chalmers, D Hill, H Zhao, and E Lou, "Prescriptive analytics applied to brace treatment for AIS: a pilot demonstration", Scoliosis (in press)
- E Chalmers, W Pedrycz, and E Lou, "Human experts' and a fuzzy model's predictions of outcomes of scoliosis treatment: A comparative analysis", IEEE Transactions on Biomedical Engineering (in press)

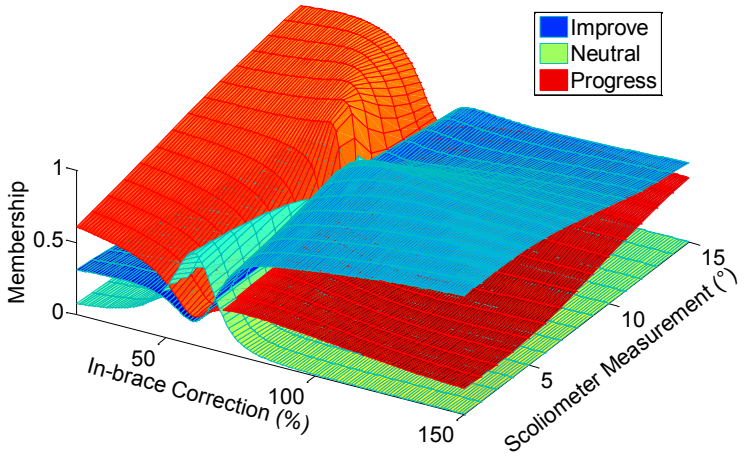


Figure 23: Predictions for a 13-year-old patient with 30° Cobb angle. "Improve", "Neutral", and "Progress" memberships are shown as a function of both Scoliometer measurement and in-brace correction.

features (given a specified resolution), creating a multi-dimensional lookup table of "improve", "neutral", and "progress" membership predictions for the range of possible cases. The decision support engine allowed a user to enter a patient's Scoliometer measurement, Cobb angle, and age. It then retrieved the predictions corresponding to these characteristics and the range of in-brace corrections. The results were plotted, allowing the user to visualize the patient's predicted memberships in the "improve", "neutral", and "progress" contexts as a function of the in-brace correction applied. For an example of this plot see Figure 24, which is actually a cross section of the plot shown in Figure 23.

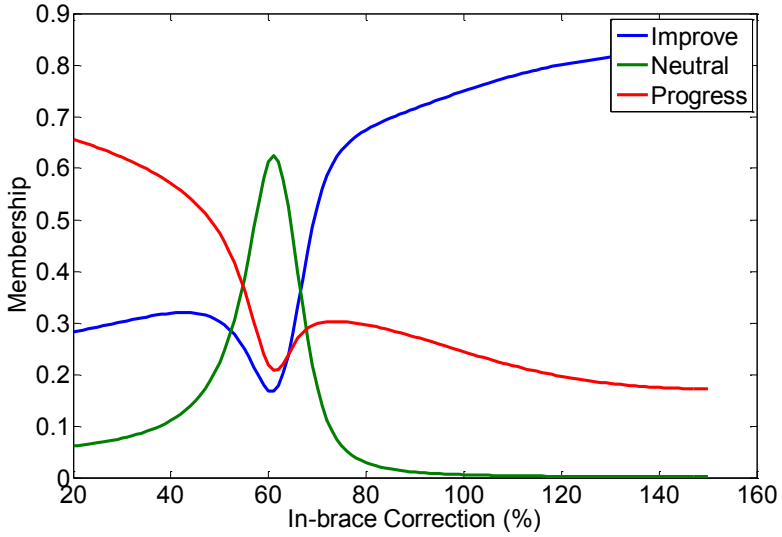


Figure 24: Predicted memberships in the "Improve", "Neutral", and "Progress" categories for a 13-year-old patient with 3° Scoliometer measurement and 30° Cobb angle.

Considering Figure 24, observe that "progress" is the dominant membership for in-brace corrections below 55%, so correction greater than 55% is indicated for this patient. But we also notice the "improve" membership shows a point of diminishing returns at about 75% correction. This is valuable information, as applying too much correction might result in an uncomfortable brace and low compliance.

This work chose a recommended in-brace correction based on the model's predictions, by identifying the correction which made one of the "neutral" or "improve" memberships the dominant membership. An illustrative example is shown in Figure 25. For this patient, 60% in-brace correction gives maximum "neutral" context membership and a relatively low "progress" context membership. Thus 60% correction may be a good target correction for this patient as predictions indicate their curve would likely not progress. A brace intended for nocturnal wear might attempt more correction, but the chart indicates a point of diminishing returns in the improve membership at about 75% correction; thus 75% may a good target correction if this patient is to receive a night brace.

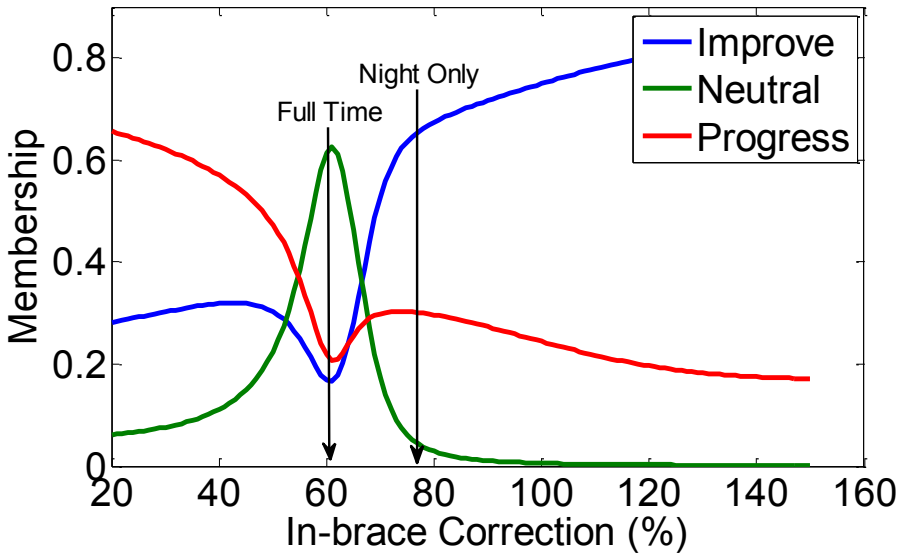


Figure 25: Using the fuzzy model's predictions of treatment outcome to identify suitable in-brace correction recommendations.

3.4. How Good are The Recommendations?[§]

3.4.1. Clinical Trial Simulation

Section 3.3 described how treatment recommendations were produced using computer-generated predictions of treatment outcome. The obvious question is: what would be the effect of these recommendations if they were applied to actual patients? Ultimately, the efficacy of the recommendations would need to be determined in a formal clinical trial, with the goal of seeing whether using the CDSS lead to improved treatment outcomes. However computer simulation can provide an initial estimate of the recommendations' effect based on retrospective data.

A clinical trial simulation (CTS) technique proposed by Chi [126] was used to estimate the efficacy of the in-brace correction recommendations. Chi et al used the procedure to estimate the effect of lifestyle change recommendations on risk of heart disease [10], and the effect of individualized Warfarin treatment protocols [13]. The CTS randomly divided available patient data into two equally-sized groups: A and B. Separate prediction models were trained using the data from each group. Model A was then used to recommend in-brace corrections for the patients in group B, using the procedure described in section 3.3. Model B then predicted the new treatment outcomes for the group B patients given Model A's in-brace correction recommendations. Thus, the CTS simulated a case-control study: the group B patients with model A's recommendations applied are the "cases", while the original group B patients serve as the matched controls. Since the two models used to create the recommendations and predict outcomes are trained and used separately, the CTS provides an unbiased estimate of the recommendations' effect. The CTS procedure is illustrated in Figure 26.

The procedure described in section 3.3 was used to select two target corrections for each patient in Group B: one for a full-time and one for a night-time brace. The CTS used the target corresponding to the brace type that had actually been prescribed to the patient, as recorded in their charts. However if no suitable full-time correction could be found (i.e. the minimum recommended correction was too high to be achieved by a full time brace) then the patient was switched to a night-brace in the simulation. Sixty percent correction or higher was considered unreasonably high for a full-time brace.

[§] Material in this section has been published in the paper: E Chalmers, D Hill, H Zhao, and E Lou, "Prescriptive analytics applied to brace treatment for AIS: a pilot demonstration", Scoliosis (in press)

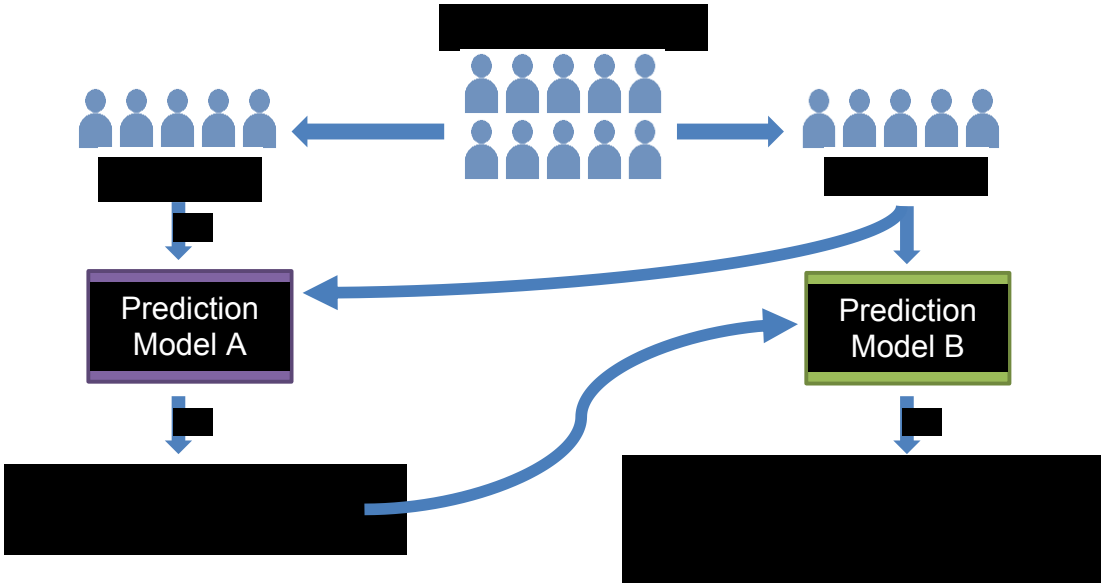


Figure 26: Diagram illustrating the clinical trial simulation procedure. Model A recommends in-brace corrections for patients in group B, with Model B predicting the recommendations' effect. Predicted outcomes given the recommendations were compared to outcomes in the group B patients' charts.

Overall progression rates from the patient charts were compared to (predicted) progression rates under the recommended in-brace corrections; the difference in progression rate was measured. Progression was defined as a >5° increase in Cobb angle by the end of treatment [10].

As the progression rate under the recommended corrections is predicted by the fuzzy model, and the model's predictions are not perfect, there is some uncertainty associated with the progression rate estimated in the CTS. The progression rate can be treated as a random variable. The distribution of the progression rate was estimated by first measuring model B's negative predictive value (NPV) and positive predictive value (PPV) in a five-fold cross validation on the group B patients. The probability mass function (PMF) for the number of progressions given the recommended in-brace corrections was calculated based on these values. That is, each of the model's "progress" predictions represents a progressive case with probability equal to the model's PPV. Each "non-progress" prediction represents a progressive case with probability equal to 1-NPV. The PMF for progressive cases in the entire group B was then calculated given that the PMF of the sum of two independent discrete random variables is:

$$f_3(j) = \sum_k f_1(k) \cdot f_2(j-k) \qquad \text{Equation 10}$$

where $f_1(x)$ and $f_2(x)$ are the mass functions for the two variables, and $f_3(x)$ is the mass function for the sum.

3.4.2. Simulation Results

The group B patients included 23 patients who had progressed by the end of treatment, and 22 who had not. Fourteen patients had been treated with night braces, while 31 had received full time braces.

Eight group B patients who had originally been prescribed fulltime braces were switched to night braces in the CTS. The predicted outcome memberships for one such patient is shown in Figure 27. Eighty percent correction was selected as a suitable night brace correction for this patient, but the plot reveals no suitable full-time brace correction less than this.

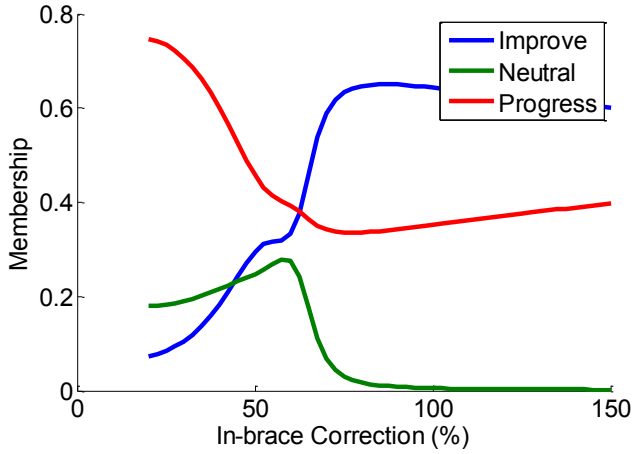


Figure 27: Predicted "Improve", "Neutral", and "Progress" memberships for one of the patients in the CTS. For this patient, 80% was identified as a suitable correction – too large a target for a full-time brace but suitable for a night brace.

Recommended corrections for full-time and night braces are shown in histograms in Figure 28 and Figure 29 respectively. Recommended corrections ranged from 20%-58% for full-time braces (mean 37%) and from 65%-130% for night braces (mean 91%). In 17 of the 45 cases (38%), the recommended correction was less than the correction that had actually been applied clinically, as recorded in the patients' charts. Figure 30 shows a scatterplot of the recommended

in-brace corrections versus the actual corrections applied clinically, and indicates which patients were predicted to progress in the CTS. The group of patients who received lower corrections in the CTS did not suffer an increased progression rate.

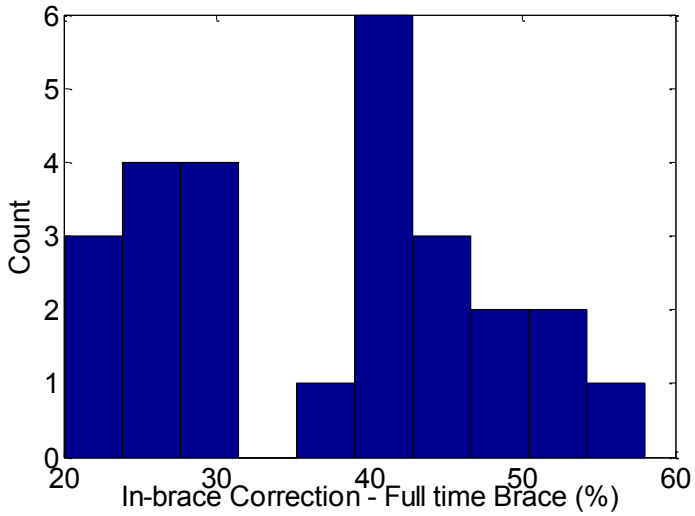


Figure 28: Histogram showing the distribution of recommended in-brace corrections for full time braces in the CTS.

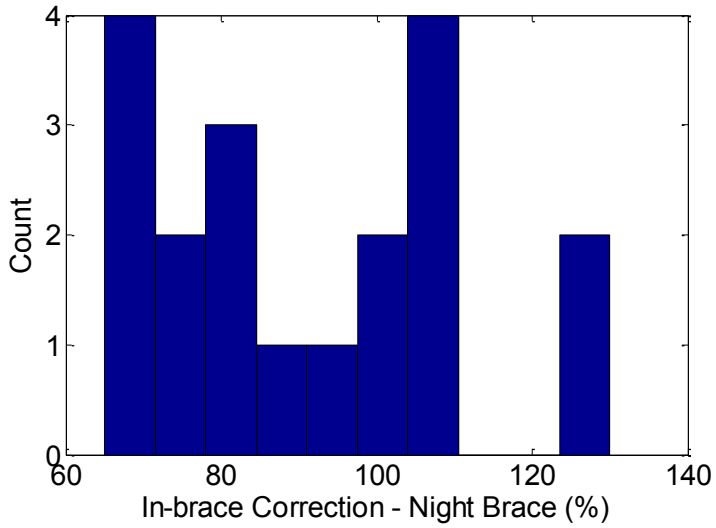


Figure 29: Histogram showing the distribution of recommended in-brace corrections for night braces in the CTS.

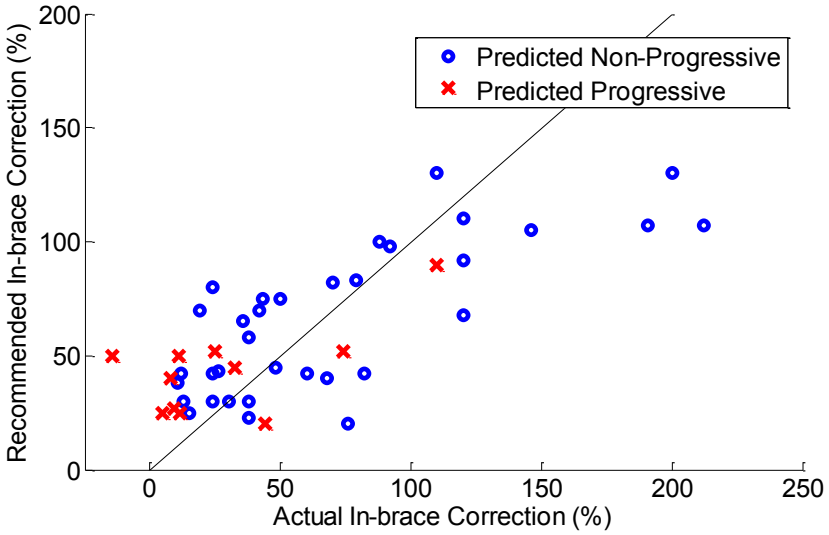


Figure 30: Recommended and actual in-brace corrections for group B patients in the CTS. Those above the diagonal line received higher corrections in the CTS than they had received clinically. Those below the line received lower corrections.

Model B's positive predictive value was 0.73. Its negative predictive value was 0.70. In the CTS, predicted outcomes for group B included 11 progressions – a 52% reduction from the original 23. Combining 11 progression predictions and 34 non-progress predictions as per Equation 10 produced a PMF for progressive cases, shown in Figure 31.

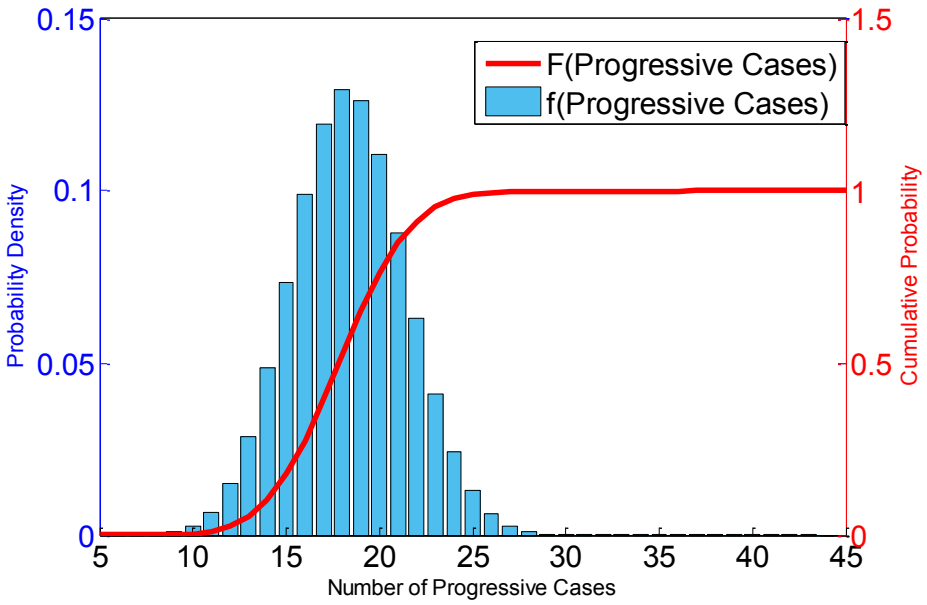


Figure 31: Estimated probability distribution of progressive cases in group B, given the recommended in-brace corrections. Bars show the probability mass function. The solid line shows the cumulative distribution function.

The mean of this distribution is 18 progressions (a 22% improvement over the original 23). The corresponding CDF (also illustrated in Figure 31) indicated a 4.7% chance of there being more than 23 progressions. Thus the improvement in progression rate can be considered statistically significant at the p=0.05 level.

3.5. Limitations and Concluding Remarks

The CTS estimated that the model-recommended in-brace corrections could significantly reduce progression rates among patients undergoing brace treatment. However there are several limitations to the work presented in this chapter.

First, the method of choosing recommended in-brace corrections described in section 3.3 was performed manually in a rather time-consuming process. It also required a small degree of subjective assessment. It would be preferable if this process were automated to save the user's time, and to ensure objective and reproducible recommendations.

Second, while the model-recommended in-brace corrections were estimated to reduce progression rate by 22%, it is unclear whether the recommended corrections would actually be achievable in practice. Overall the observed ranges of recommendations for full-time and night braces agree with literature and corrections observed at our clinic, but some individual patients with stiff curves may not be capable of large corrections. Thus, what is perhaps most interesting is that about one-third of the recommended corrections were *lower* than that actually applied. This may suggest some potential to build less aggressive (more comfortable) braces without compromising treatment outcome.

Another limitation is that the data used in this work included no measurements of patient compliance with prescribed brace-wear. The prediction model can only be used to recommend parameters which it uses in making predictions. The fuzzy model was used to recommend in-brace correction (its other inputs – age, Cobb angle, and Scoliometer measurement – cannot be controlled), but recommendations of brace wear-time may be more valuable. This is because in practice we may have limited control over the in-brace correction. Moreover the orthotist has limited time with the patient and often doesn't know the correction achieved by the brace until an in-brace radiograph is taken some time later. Thus fine-tuning a brace to achieve a specific

target correction is impractical. Health care providers have more control over the amount of brace-wear that is prescribed than they do over how much correction a brace applies.

There is another aspect of the work in this chapter which may be seen as a limitation. The prediction model(s), the comparison with experts, and the CTS all cast predicting progression as a classification problem: "progress" and "non-progress" were the only predictions allowed. This paradigm might engender a deterministic view of curve progression: that progression has definite causes and that we could predict it perfectly if only we had enough information. The pitfalls of taking this viewpoint when predicting future events have been discussed by Gigerenzer et al. [175] and Nate Silver [176]. While it is important to retain the concepts of "progression" and "non-progression" as distinct events, it may be best if predictions of progression were probabilistic. For example, instead of predicting a patient to "progress" (and calculating the PPV corresponding to this prediction), the prediction could give the patient a "75% probability of progression". This probabilistic viewpoint would better reflect our imperfect ability to predict progression, and better communicate the uncertainty in our predictions.

To conclude, chapter 3 has described a complete validation of the concept of electronic decision support for brace treatment planning. Section 3.4.2 reported encouraging results, while section 3.5 has pointed out some limitations. Chapter 4 will discuss the development of a complete decision support platform. This platform will allow ongoing collection of patient data (including compliance data), and conversion of this data into treatment recommendations through prediction models.

4. A Hardware & Software Platform for Providing Brace Treatment Decision Support

This chapter describes the development of a complete decision support system which builds on the general approach described in chapter 3. The system includes a hardware device used for collecting brace-wear data, and software components which use patient data to generate prediction models and decision support. Development of each component is described. A software platform integrates the components in a complete, scalable system.

4.1.Overview

Chapter 3 detailed a preliminary validation of the concept of electronic decision support for brace treatment. First, prediction models were developed which predict treatment outcome using start-of-treatment measurements. Second, optimal in-brace corrections were estimated using the predictions. Finally, a clinical trial simulation estimated that the recommended parameters (if achievable) would reduce overall progression rates.

The dataset used in the preliminary validation was limited in that it contained no brace-wear compliance information. Thus no measurements of compliance were used by the prediction model, meaning the decision support stage could not recommend patients' optimal brace wear time. This chapter describes a complete decision support system which overcomes this limitation. The system is implemented as a platform which accommodates the following tasks:

- Collecting patients' brace-wear (compliance) information
- Storage of compliance data and other routine clinical measurements
- Visualization and simple analysis of the stored data
- Automatic training of prediction models
- Generation of treatment recommendations

A diagram of the overall system is shown in Figure 32.

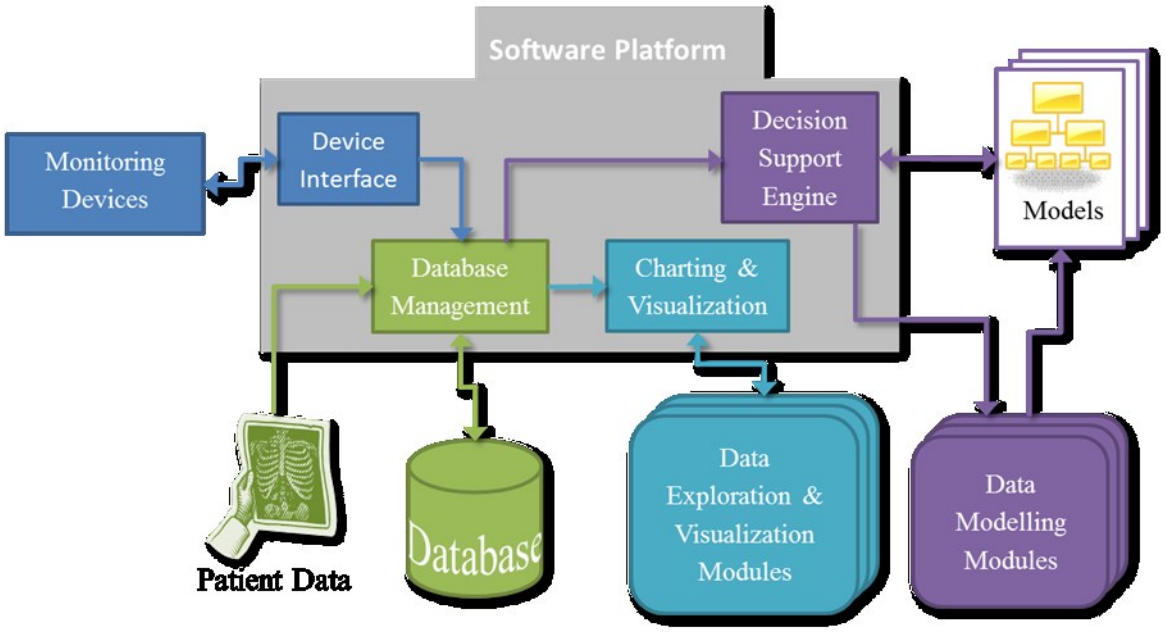


Figure 32: Illustration of the overall decision support system.

The platform includes an electronic monitoring device which records the force applied by the brace and the duration of patients' brace-wear. This device communicates with the software portion of the platform, which comprises the data storage, model training, and decision support components.

Arrows in Figure 32 illustrate exchange of information between the components. The software platform interfaces with the hardware devices to download compliance data. This data is stored in a relational database along with additional clinical information entered by the user. A "charting and visualization" component draws from the database to perform various visualization and simple analysis tasks. A decision support engine component trains prediction models which predict the outcome of brace treatment using the measurements stored in the database. The decision support engine uses these models to predict treatment outcomes for a particular patient given a range of treatment options. It then uses these predictions to estimate the optimal treatment for the patient, with the specific nature of the recommendation depending on the model used.

The following sections describe the development of each component in more detail, starting with components external to the software platform, and concluding with the platform itself. Section 4.2 describes the hardware device designed to collect compliance information. Section 4.3 describes the relational database used to store patient data, the visualization modules, and

the modelling module. Section 4.4 describes the software platform which integrates these pieces into a complete system.

4.2. Compliance Monitoring Device[**]

4.2.1. First Generation Compliance Monitor

An electronic compliance monitor was developed to measure patients' brace-wear habits during long-term treatment. This information is used by the software platform in making predictions and recommendations.

Most compliance monitors for use in AIS are temperature-based – detecting when the brace is worn or unworn based on body heat. An alternative is to detect brace-wear by sensing force or pressure at the brace-body interface, and a preliminary study by Lou et al. [3] has suggested that such force readings can be a predictor of treatment outcome. For this reason the device described here was originally designed to be a force-sensing compliance monitor – a new design based on a previous generation force-sensing device used by Lou et al. [102].

The monitor consisted of a force sensor, data logger, and battery. The data logger sampled the output of the force sensor using an analog-to-digital converter, and stored the digital samples in on-board memory as illustrated in Figure 33. The data logger communicated wirelessly, for convenience in downloading logged data and performing other operations.

The monitoring device was designed for minimal size and power consumption, so that it could be embedded in a brace for long-term monitoring. The key component of the data logger is the CC2530F256 system-on-chip (Texas Instruments Inc. Texas, USA). It includes an 8051 microcontroller, an onboard radio transceiver, and a sigma-delta analog to digital converter. Low-power design was accommodated by the 8051 microcontroller, which uses a simple, 8-bit processor with low power requirements and features several low-power operating modes. The included radio transceiver followed the IEEE 802.15.4 standard; a standard including the physical layer and media access control for low-power wireless communication. The analog-to-digital converter had 11 bits effective resolution, allowing sufficiently precise digitization of the

[**] Material for this section has been published in the paper: E Chalmers, E Lou E, D Hill, V Zhao, "An Advanced Compliance Monitor for Patients Undergoing Brace Treatment for Idiopathic Scoliosis", Medical Engineering & Physics (in press)

force sensor output. The CC2530F256 also contains 256 KB of on-board flash memory which was used for storing data; this reduced the size of the device by avoiding the need for separate memory.

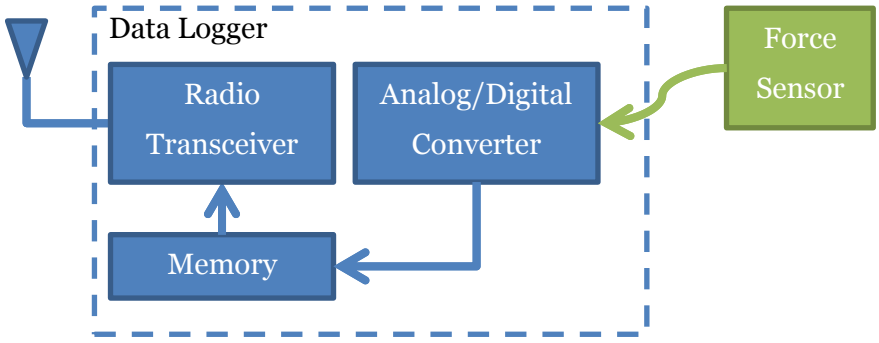


Figure 33: High-level illustration of the compliance monitoring device.

The key component of the force sensing module was the FS1500 force sensor (Honeywell International, Inc. New Jersey, USA) with a sensing range of 14.7 N. The force sensor was covered by a rubber housing, which helps to direct force onto the force-sensitive area. The design and reliability testing of the force sensor module was performed previously [177]. The force readings were electronically amplified and then digitized to have a range of 10 N and a resolution of 0.05 N. The force of 10 N was considered to be the maximum force expected in brace [109]. The 0.05 N resolution meant 201 discrete force levels, so that a force reading could be stored using a single byte of memory.

A two-point calibration procedure was used to calibrate each sensor individually. This procedure applied 1 N and 3 N forces to the force sensor and recorded the corresponding sensor outputs. The slope and offset of the sensor response was calculated from these values. The data logging device performed the calibration calculations, and stored the resulting slope and offset values.

During long-term operation the device sampled the output of the force sensor at a user-defined interval. These samples were stored in the CC2530F256's onboard flash memory. Between samples, power to the force sensor was turned off and the microcontroller entered a low-power state in which only a low-power oscillator was left running. Use of this low-power state allowed battery life to be extended beyond six months (the typical time span between patients' clinic visits).

4.2.2. Second Generation Compliance Monitor

As discussed in section 2.4.2, there are limitations to both force and temperature-based compliance monitoring. A second generation of the compliance monitoring device featured new firmware, which allowed temperature sensing using the CC2530F256's internal temperature sensor. By employing both temperature and force sensing, the new device provides a more complete picture of brace-wear compliance. After sampling the force sensor output, the new device samples the CC2530F256's onboard temperature sensor 4 times at a rate of 50 Hz, and averages these readings to obtain a single temperature measurement. Oversampling in this way improves the temperature reading's precision [178].

The device uses a memory-efficient data logging scheme. Every three sampling periods, the device logs force and temperature information in a 4-byte packet. The packet contains a two bit header for communication purposes, and the three force readings (one byte each). The remaining 6 bits encode the difference in temperature between the current packet and the previous one. If the difference is too large to be encoded in six bits, any remainder is carried over to the next interval. The difference between successive temperature samples is usually small, making temperature difference a more memory-efficient metric than the actual temperature. When logged data is retrieved the (relative) temperature curve can be reconstructed by integrating temperature changes over all samples.

The CC2530F256 includes 256 KB of flash memory, with 22 KB being reserved for program code. The data logging scheme requires 4 bytes for three samples, meaning the device can accommodate (234 KB) x (1024 bytes/KB) / (4 bytes) x (3 samples) = 179712 samples. Since the brace is usually donned for at least several hours at a time, and major changes in brace force are infrequent while the brace is worn, a two or three-minute sampling interval is assumed to provide sufficient data. The device's memory capacity accommodates 250 days of sampling at a 2 minute sampling interval, or 374 days at a 3 minute interval.

Since the compliance monitor spends as much time as possible in a low-power state, it consumes an average current of only 36 μA while sampling at a 2 minute interval. With a 400 mAh battery, the battery life is (400 mAh) / (0.036 mA) = 11,111 hours = 15.4 months (12 months if we apply a 20% de-rating factor to the battery).

The monitor's low-power and memory-efficient design allow battery life and memory capacity greater than the typical 6-month window between clinic visits. It is also small enough to allow easy and unobtrusive installation into the brace for long-term use: The data logger measures 5.2 x 2.5 x 0.8 cm, and the force sensor measures 2.4 x 2.4 x 0.8 cm. The compliance monitor is

shown in Figure 34. Its modular design in which the force sensor is detachable from the data logger is convenient because the force sensor is usually embedded in a brace pressure pad during installation, and is not easily removable. When the patient arrives for a clinic visit, the data logger can be quickly detached and replaced by a logger with a fresh battery.

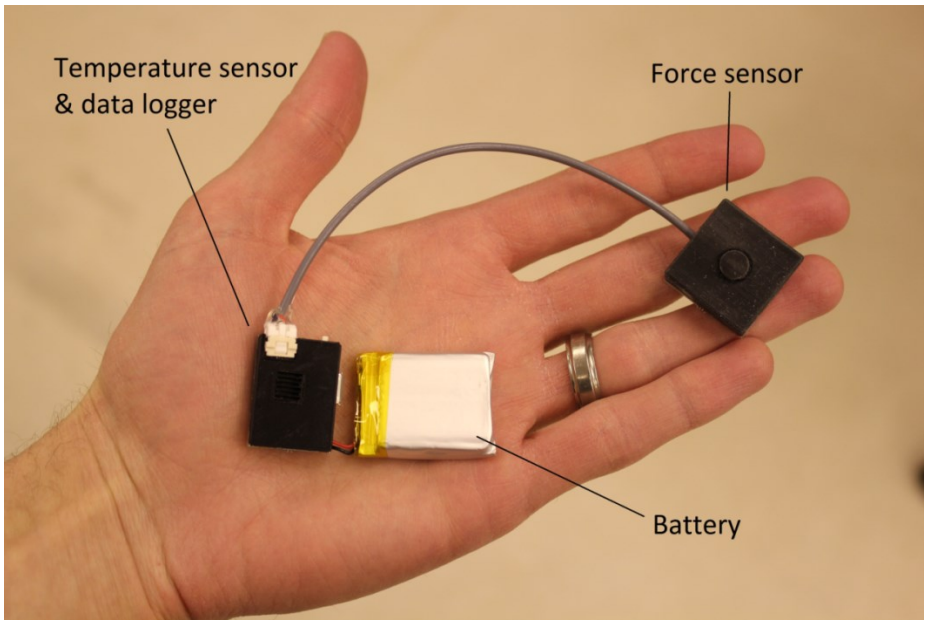


Figure 34: The passive brace-wear monitoring device.

4.3. Software Components

4.3.1. Database

A Microsoft® Access database was created to store patient data, including clinical measurements collected during routine examinations, and data collected from compliance monitoring devices. The tables and relationships between them are shown in Figure 35.

The database consists of six tables:

- DemographicsTable: Contains patients' name, sex, and birthdate.
- ExamTable: Contains fields describing various clinical measurements. Currently the only fields utilized are Cobb angle(s), Scoliometer measurement, height, weight, age, and a boolean field indicating whether the Cobb angle was measured while the patient was in-brace.

- DeviceAssignmentsTable: Records which hardware device ID numbers have been assigned to particular patients.
- DeviceTypesTable: Lists the types of available hardware devices (currently limited to the passive monitoring device described in section 4.2, or an active pressure control device designed outside the scope of this thesis).
- BraceWearTable: Stores measurements and supporting data collected from the monitoring devices.
- TargetValuesTable: This table allows target forces/pressures to be associated with specific monitoring devices. The targets could be desired or intended levels of force chosen by the orthotist.

Relationships between the tables allow data from several tables to be combined in a single query using appropriate SQL (Structured Query Language) "JOIN" statements.

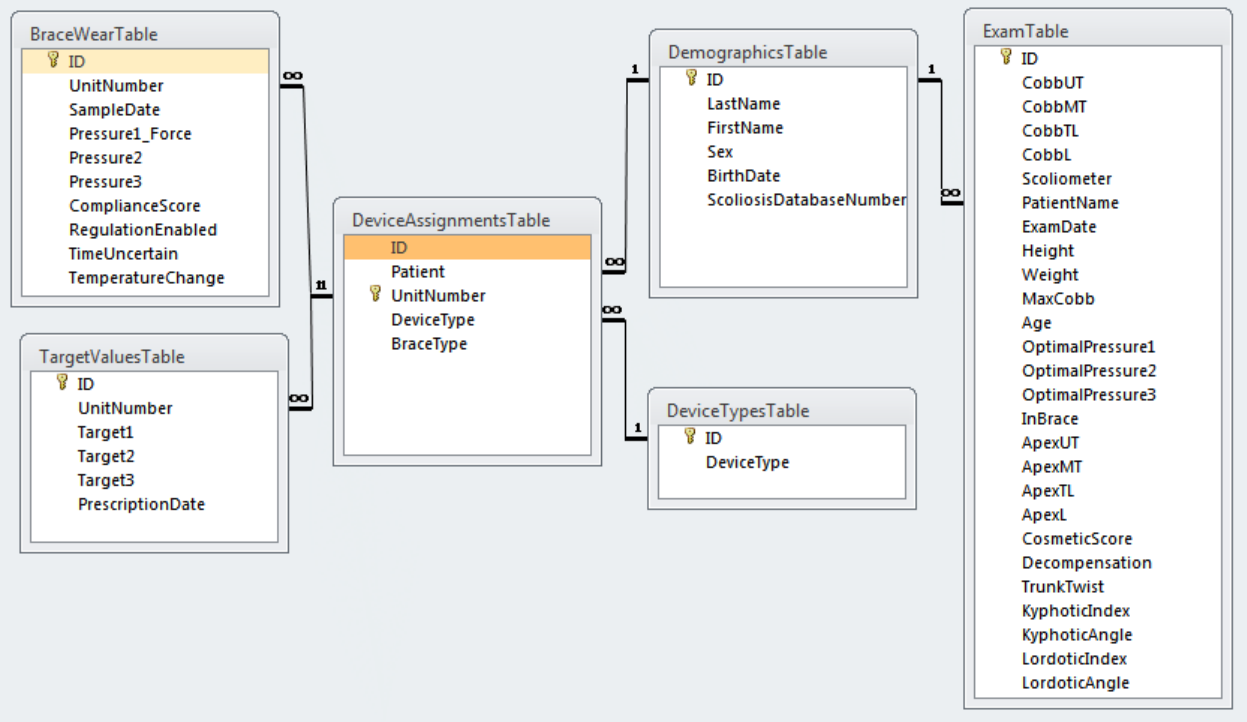


Figure 35: Layout of the tables in the relational database used to store patient data.

4.3.2. Visualization Modules

The visualization modules allow visualization of the data in various ways, or calculation of various descriptive statistics. For example, given a set of compliance data retrieved from the passive monitoring device, we may wish to calculate the patient's overall compliance percentage or plot their compliance over time. Figure 36 shows an example of such a plot – in this figure a patient's brace-wear compliance (as a percentage of 24 hours) is plotted over time. This particular patient's family went on vacation during July 2013, and apparently the patient did not wear their brace during this time – though their brace-wear had also been declining during the weeks previous.

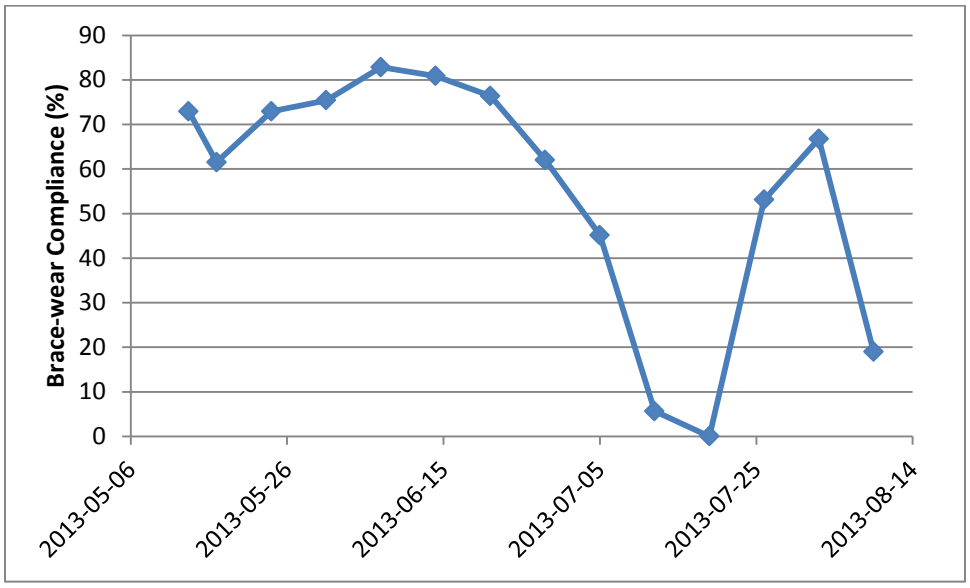


Figure 36: Sample plot of one patient's brace-wear compliance (as a percentage of 24 hours) over time.

Each visualization module is an XML (eXtensible Markup Language) document which encodes all information needed to perform the visualization or analysis task. The document includes the SQL statement used to retrieve the relevant data from the database, identifies any input required by the user (patient numbers, date ranges, etc.), and specifies how the results should be displayed. It also includes user-friendly descriptions of the visualization and instructions.

At present, seven visualization modules have been developed to perform the following tasks:

- Plot brace-wear compliance by time of day (see Figure 48 on page 89 for an example).
- Plot compliance over time (see Figure 36 for an example). Readings can be averaged by day, week, or month.

- Plot a histogram of forces recorded by a monitoring device.
- Calculate average brace-wear compliance using force measurements. Forces greater than the 4th centile reading are considered to represent brace-wear.
- Plot force readings over time. Readings can be averaged by day, week, or month.
- Plot force readings over time, adjusted for compliance. Only readings which are inferred to represent brace-wear are included in the calculation of daily/weekly/monthly averages.
- Create a generic scatterplot of Cobb angle progression per year versus a user-specified field from the ExamTable table. All patients who have the relevant data recorded in the database are used. Progression per year is calculated by taking the overall Cobb angle progression observed in the database and normalizing to one year. An example scatterplot showing progression per year versus Scoliometer measurement is shown in Figure 37. The XML code for the module used to create this plot is given in Appendix A.

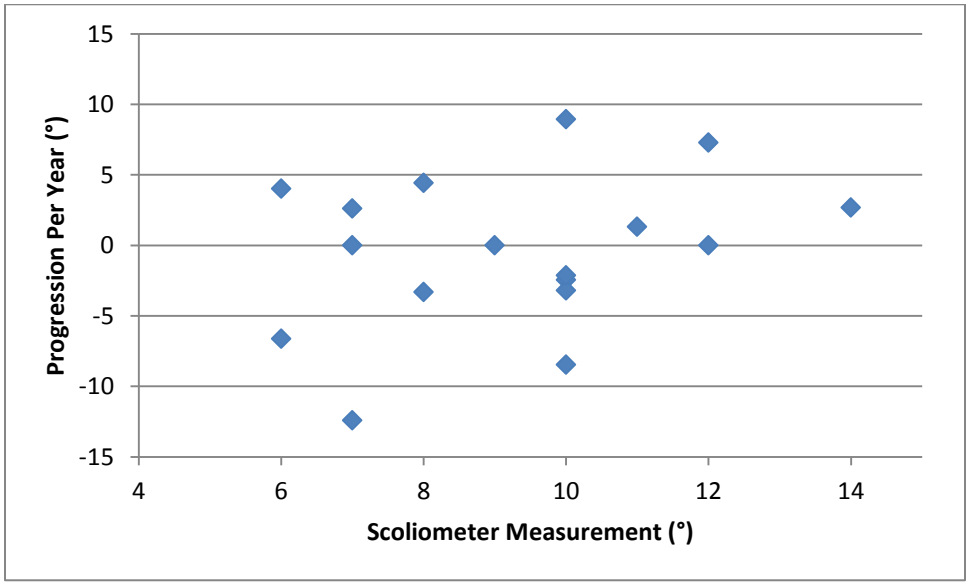


Figure 37: Example scatterplot created using a visualization module. Each point represents one patient from the database.

4.3.3. Modelling Module

Just as a visualization module contains all the information necessary to calculate a descriptive statistic or create a plot, a modelling module contains everything needed to train a prediction model. The training process creates a model while predicts treatment outcome using the

measurements recorded in the database. The model can later be used by the software platform's decision support engine as it identifies optimal brace treatment parameters for new patients.

A modelling module consists of two executables and an XML document. The first executable accepts a data set consisting of predictor variables and corresponding outcomes for a number of patients. It uses this data to train and save a prediction model, and reports the cross-validated prediction performance of the model. No limits are placed on the type of prediction model created. The second executable accepts a set of predictor variables and uses the saved prediction model to predict the outcome. The XML document tells the software platform how to call each executable and contains human-readable descriptions of the model and its expected performance.

One modelling module was created for this thesis. It was written in MATLAB (The MathWorks, Natick, MA). The module trains a logistic regression model to predict curve progression after one year of brace treatment. The remainder of section 4.3.3 describes the methodology used in developing the module. Specific details of the model's implementation and testing are given in chapter 5.

4.3.3.1. General predictive modelling approach

Section 3.5 noted that classifying progressive non-progressive cases might best be seen as a probabilistic problem. That is, we should seek a probability rather than a predicted class assignment. Producing predictions of the form: "there is a 66% chance that patient X will progress" allows us to continue framing treatment outcome as a "progress"/ "non-progress" dichotomy, but also clearly communicates the uncertainty in the prediction. Communicating the inherent uncertainty in a prediction is important if the prediction is to benefit the decision-making process [175], [176]. Logistic regression, naïve Bayes classifiers, and neural networks are examples of standard modelling algorithms which can give probabilistic predictions.

Because of the small population in the local scoliosis clinic, the dataset used to train the model for this thesis would be small. This imposes certain constraints on the choice of modelling method. In particular the model must be kept simple, as a more powerful model is prone to overfitting a small dataset. This is intuitive but has also been demonstrated empirically by Chan et al. [179], who found that with small training set, a linear model outperformed a quadratic one on a classification problem involving two classes of unequal covariance (in such a problem the quadratic model is theoretically the optimal choice). Overfitting can be further mitigated by

minimizing the number of decisions made by the designer. Bouesteix and Schmid call this number of decisions the "degrees of freedom", and note that "Increasing the degrees of freedom of the analyst might also increase the risk of conscious or subconscious overoptimism and 'fishing for significance'" [64]. Sometimes design decisions are explicit, such as choosing the number of neurons in a neural network or setting kernel parameters in a support vector machine. Other decisions are hidden inside algorithms, such as a wrapper feature selector which trains multiple models and chooses the one which performs "best" on the training data. The modelling module developed for this thesis employs logistic regression, which gives a simple and interpretable model. Refer to section 3.1.2 for a description of logistic regression.

The model predicted treatment outcome in terms of Cobb angle progression after one year of brace treatment. However, Cobb angle measurements are known to have some inherent measurement error, so there is some uncertainty associated with these progression measurements. In keeping with the probabilistic philosophy of this modelling module, the raw progression measurements in degrees were converted to a probability that the Cobb angle had increased. This was done by assuming the Cobb angle measurement error is normally distributed with a 95% confidence interval of 5° (a standard deviation of roughly 1.25°) [20]. Since measuring progression involves "before" and "after" Cobb angle measurements, the error is compounded: the sum of two normally distributed random variables with variances σ^2_a and σ^2_b is a normally distributed random variable with variance $\sigma^2_a + \sigma^2_b$. Thus the progression measurement error has standard deviation $\sigma = 1.77°$ or 95% confidence interval of roughly 7.1°. The probability of a measured progression corresponding to an actual Cobb angle increase can then be calculated and is shown in Figure 38. The logistic regression was performed on these probabilistic outcomes.

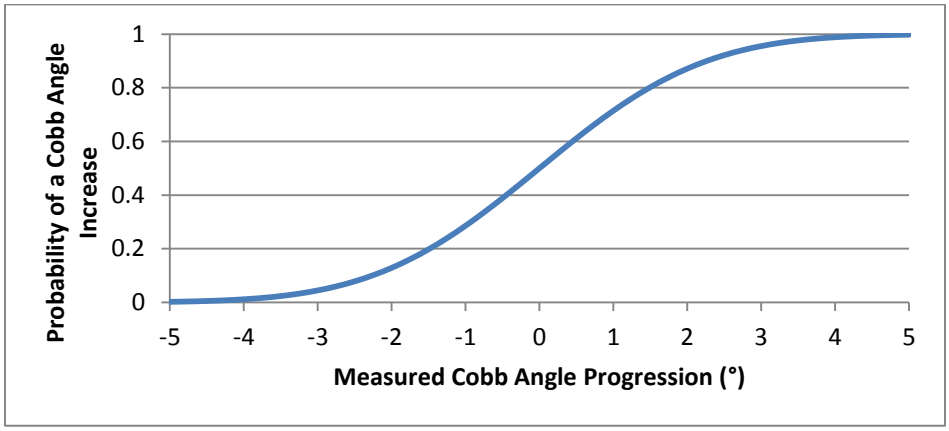


Figure 38: Probability of a Cobb angle increase given a measured Cobb angle progression, assuming measurement error in a single Cobb angle measurement is normally distributed with 95% confidence interval of 5°.

4.3.3.2. Predictor Variables and Variable Selection

Since the training dataset was expected to be small, the number of features included was small compared to the dataset described in section 3.1.1. To increase the modelling module's ability to create a generalizable model, seven features were selected for inclusion as candidate predictor variables. These features were chosen based on existing knowledge and past experience including that described in section 3. Steyerberg et al. described the need to rely on existing knowledge when a dataset is too small to allow completely automated knowledge discovery:

> *"Sensible modeling should find a balance between external knowledge from outside the data versus what can be learned from the data. The smaller the data set available, the more we have to rely on external information."[180]*

The seven features were patient age, Cobb angle, Scoliometer measurement, in-brace correction, average force applied by the brace (as recorded by the compliance monitor), brace-wear compliance, and height velocity. Height velocity was used as a proxy for Risser sign – which is usually considered a good predictor but was not available for all patients at our center. Little et al. [50] suggest that height velocity provides a better indication than Risser sign of the period of maximum Cobb angle progression. Average force and brace-wear compliance were calculated using the 60 days of force sensor data closest to the start-of-treatment date. The 4th centile force reading was calculated in a 24-hour sliding window, and compliance was calculated as the percentage of readings greater than the 4th centile threshold. The average force was calculated after first normalizing the force readings to the 95th centile value. Sixty days were used because there are typically two months between the examination where the brace is prescribed and the

next follow-up visit. If the model can make predictions using only these first two months of data, it can be used to predict a patient's risk of progression at their follow-up visit.

A filter rather than a wrapper-based selection scheme was used to select a feature subset for use in the logistic regression. This helps mitigate the potential for overfitting. The modelling module implemented the Correlation-based Feature Selection (CFS) scheme proposed by Mark Hall [145]. Given a full set of features, this method tries to select a subset of features which have high correlation with the output, but low correlation with each other. This is accomplished by maximizing the correlation estimate:

$$r_{zc} = \frac{k\overline{r_{zi}}}{\sqrt{k + k(k-1)\overline{r_{ii}}}} \qquad \text{Equation 11}$$

where $\overline{r_{zi}}$ is the mean correlation between features in the subset and the output, $\overline{r_{ii}}$ is the mean correlation of the features with each other, and *k* is the number of features in the subset.

The modelling module searched through feature subsets of size 1 and 2, choosing the subset giving the best value of r_{zc}. This feature selection scheme is attractive for use on small datasets because it is completely automated – having few "degrees of freedom". Being a filter method, it executes independently of the prediction model itself. Thus potential for overfitting is minimized.

4.4. Software Platform

The software platform integrates the hardware devices, database, visualization modules, and modelling modules into a complete system. It was written in the Microsoft C# programming language and has four main components used for device interface, database management, charting and visualization, and decision support. The following sections describe each of these components.

4.4.1. Device Interface

The device interface portion of the software can communicate with the monitoring device described in section 4.2. Communication is wireless via an IEEE 802.15.4 radio dongle (Adaptive Modules Inc. Hove, East Sussex). A screenshot of the device interface screen is shown in Figure 39.

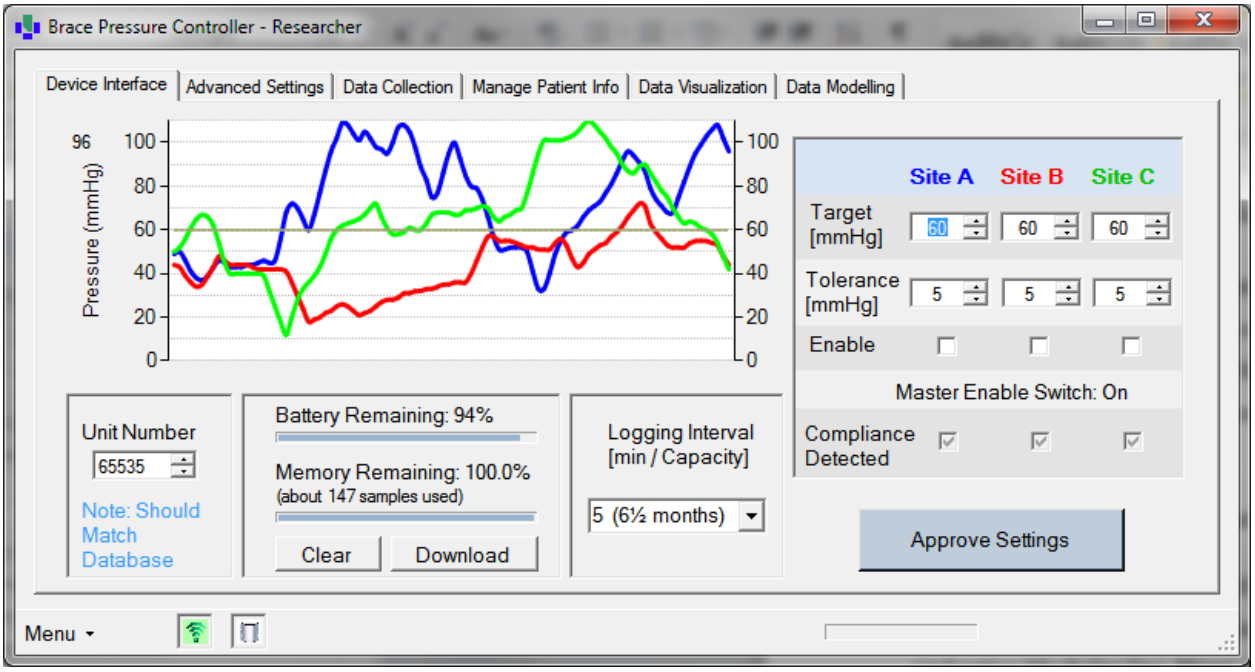


Figure 39: Screenshot of the device interface screen in the software platform.

The device interface allows the user to view and change the devices' settings, download logged data, and perform a sensor calibration procedure for the monitoring device. It also allows calibration settings to be transferred from one device to another. This is a necessary feature because the data logger's battery would not last through a complete two-to-three year brace treatment. Thus the logger must be replaced during the patient's clinic visits (or the patient must be burdened with the task of recharging its battery). The force sensor is usually embedded in the brace's pressure pad and cannot easily be removed with the logger, so it is left in place and the sensor-specific calibration settings are transferred from the old data logger to the new one.

4.4.2. Database Management

The database management portion of the software platform provides a front-end for the database and encompasses all tasks related to data storage. User interfaces allow the user to view as well as create patient records, and enter measurements collected during routine clinic visits. An illustration of one of these interfaces is shown in Figure 40.

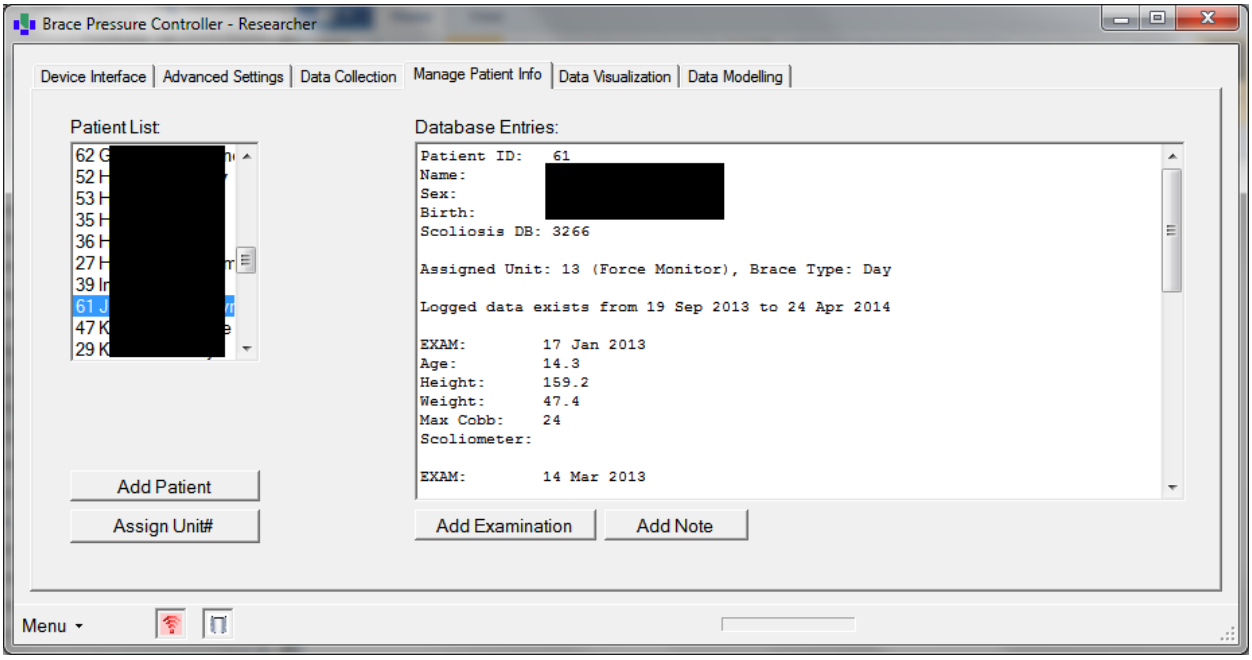


Figure 40: Screenshot of the database management screen in the software platform. Patient names and information have been blocked in this image.

The database management component receives downloaded data from the device interface, and stores it in the BraceWearTable of the database. It also retrieves data and provides it as needed to both the charting and visualization component and the decision support engine.

4.4.3. Charting and Visualization

The charting and visualization component displays plots and statistics as described by the individual visualization modules. The software allows the user to select from the list of available modules, and then automatically updates in response to the information encoded in the module's XML. After parsing the XML, the interface displays the visualization's description and provides fields for the user to enter any required input. When the visualization is run the SQL commands are parsed from the XML document and passed to the database management component, which returns the requested data. This data is then plotted according to the module's specifications. A screenshot of the visualization screen is shown in Figure 41. This screenshot shows a particular patient's brace wear (as a percentage of 24 hours) over time, averaged by week. This patients' average brace wear generally fluctuates between 40% and 60% (9.6 - 14.4 hours per day).

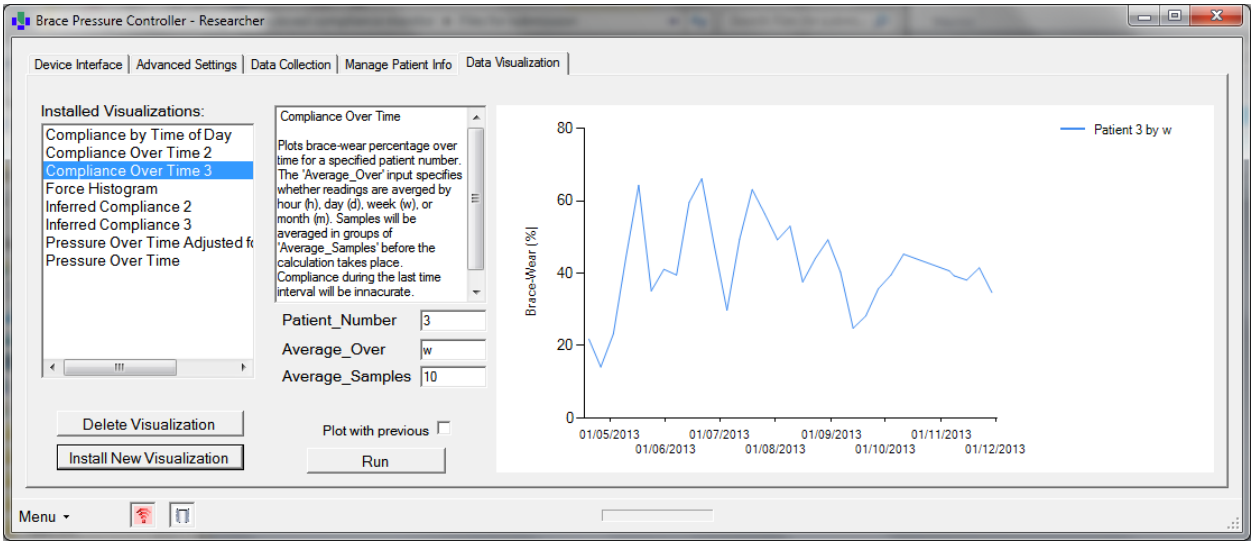


Figure 41: Screenshot of the data visualization screen in the software platform.

4.4.4. Decision Support Engine

The decision support engine is the capstone of the system: Ultimately this thesis is about translating data into treatment recommendations, and the decision support engine is where this translation happens. The decision support engine handles two main tasks: training prediction models, and using the trained models to generate treatment recommendations. A screenshot of the interface used by the decision support engine is shown in Figure 42. This interface allows the user to train a prediction model of one of the available types, and use the trained model(s) to generate recommendations for specific patients.

As part of the model training process (but a separate step from the actual training), the decision support engine retrieves data for all patients in the database and calculates features used in predictive modelling. The resulting dataset is exported as a separate file which is used during by modelling modules to train their prediction models.

The software allows the user to select from a list of available modelling modules. When the user chooses to train a prediction model, the decision support engine parses the XML file of the corresponding modelling module and runs the executable used for model training. This executable reads the training data exported by the software and uses it to train the actual prediction model. The executable then saves the model and a new XML file containing information about the model and it's cross validated performance which can be displayed to the user. This XML file also specifies which of the features in the dataset are actually used by the

model, and which ones are considered "controllable" (i.e. we have some control over them in practice and values for them can be recommended by the decision support engine) versus "uncontrollable" (i.e. features like patient age and Cobb angle, which may be valuable predictors but which we cannot control).

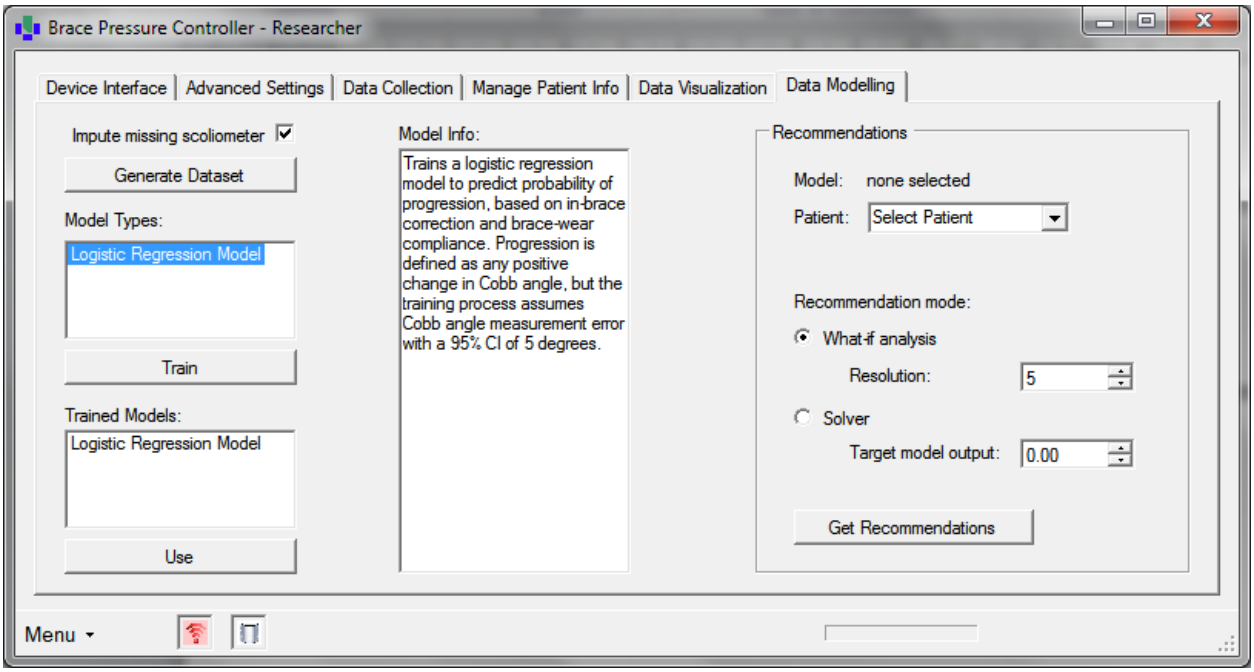


Figure 42: Screenshot of the screen used for training models and generating recommendations.

The decision support engine uses trained models to recommend optimal treatment parameters for specific patients from the database. The engine features two recommendation modes: "what-if analysis", and "solver". The "what-if analysis" uses a model to predict treatment outcomes for every possible combination of its input features (given a user-specified resolution), creating a multi-dimensional lookup table outcome predictions for the range of possible cases. The results are displayed graphically; the user is allowed to manipulate values of the controllable features and observe the effect on predicted outcome.

An example of this display is shown in Figure 43. In this example the prediction model used two controllable features as inputs: in-brace correction and brace-wear compliance. The "what-if" analysis screen allows the user to select one of these features to appear on the horizontal axis of a chart which plots predicted outcome as a function of this feature. The other feature appears as a slider beside the chart – the user can manipulate this feature's value and observe the change in the predicted outcome curve. Figure 43 shows the predicted outcome (probability of

progression) as a function of the "Compliance" controllable variable, when the "InBrace Correction" controllable variable is set to 0.83 (83% curve correction).

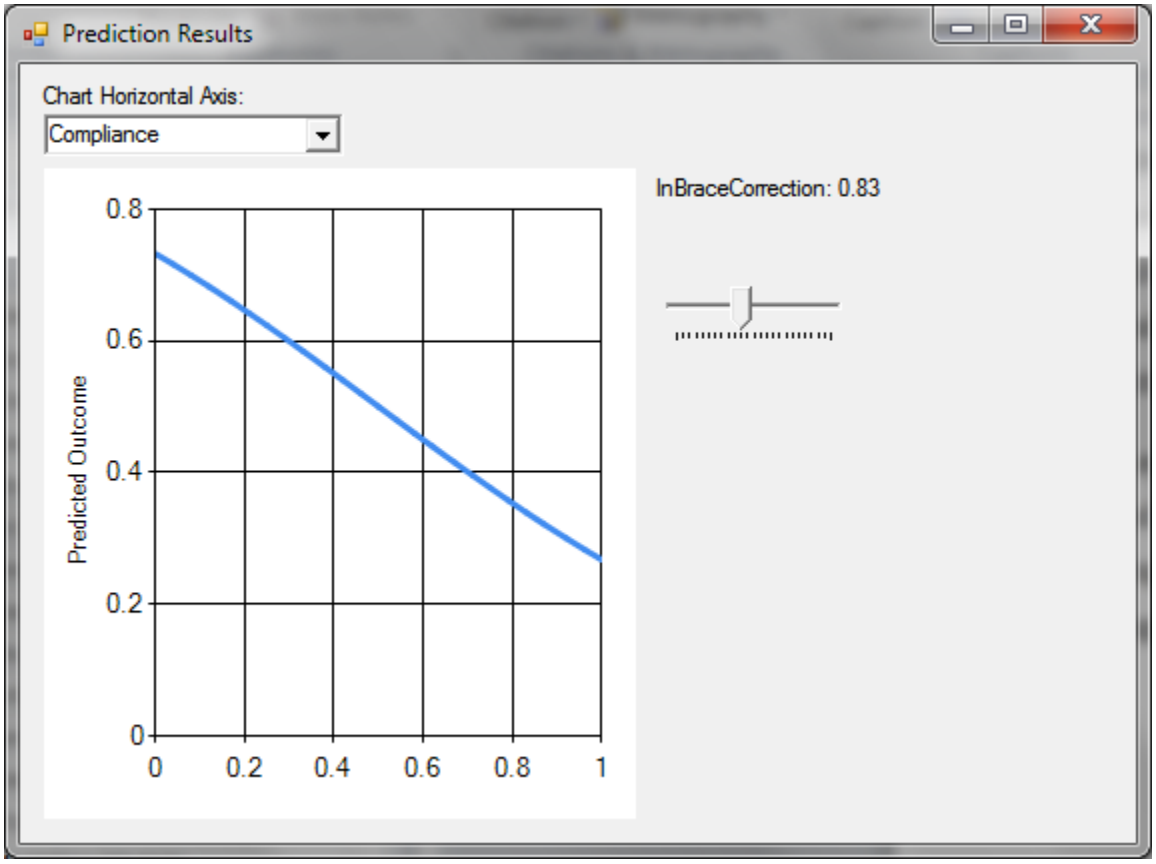


Figure 43: Screenshot of the "what-if" analysis screen.

The second recommendation mode – "solver" – attempts to find a single optimal set of treatment parameters for a given patient. It does this using a prediction model in the style of Chi's PODSS [126]. The prediction model essentially becomes an objective function, and an optimization algorithm searches for treatment parameter combinations which optimize the patient's predicted outcome.

The most efficient method of solving this optimization problem would depend on the type of model involved. In the case of logistic regression, a closed-form solution exists which defines all feature combinations which achieve a specified outcome prediction. However for many other types of models a closed-form solution would not exist, so the decision support engine uses a genetic algorithm as a general optimization technique which could be applied to any arbitrary modelling module.

A genetic algorithm (GA) is an optimization technique inspired by the concepts of natural selection and evolution. Given an optimization problem, a GA randomly generates a number ("population") of candidate solutions called "genomes". None of these solutions will be very good (they are random), but some will be better than others: A "fitness function" quantifies how good each genome is. The fittest are crossed and mutated in various ways to produce a second generation of genomes, and the fittest of the parents and offspring form the new population. The process iterates until (hopefully) a genome with sufficient fitness is created. Engelbrecht gives a comprehensive description of GA [150].

The decision support engine uses a custom GA written in C#. The GA maintains a population of individual sets of controllable feature values. In each iteration, it uses the modelling module to predict treatment outcome for each set. The fitness function simply measures the absolute difference between the prediction and a user-specified target. After five iterations the best controllable feature combination is presented to the user.

The GA uses a population size of 1000 sets of feature values, which are initially randomly generated. It uses a roulette selection operation to select the fittest sets for crossover and mutation. Roulette selection gives each individual a probability of selection which is proportional to its fitness. Pairs of selected sets are crossed using scattered crossover, which randomly selects features whose values are swapped between the pair. Sets are mutated by adding a random quantity to a randomly selected feature value.

4.5. System Scalability

Hardware devices communicate with the software using a set of custom commands. Any new sensor or device which could be made to communicate over an IEEE 802.15.4 channel using this command set could be used with the interface, with minimal modification to the software platform's source code. Thus the system has some limited potential for scalability in terms of adding new sensing devices.

The system is designed for much greater scalability in terms of visualization and modelling tasks. Since the visualization modules are external to the software platform, new visualization and analysis tasks can be added without access to the software's source code. Any data which can be retrieved using an SQL query can be displayed by the software platform. The relevant instructions need simply be encoded in a new XML document.

The practice of passing information between the software platform and the modelling module by reading and writing files was inspired by Tan et al. [181]. Tan et al. created a platform system for health analysis and simulation, and used a similar data-passing approach which they called "loose coupling". The advantage of this approach is that it makes the development of modelling modules very flexible: the executables can be programmed in any language. For example they could make use of R, Python, or MATLAB libraries which support modelling, or use Weka's Java application programming interfaces. This facilitates the addition of new modelling modules. The only requirements for developing new modules are that they include the required XML document and pass data to and from the software in the designated way. Any modelling module can be used by the software platform to generate recommendations.

5. System Validation and Testing

This chapter describes testing and validation of the decision support system components described in chapter 4. Laboratory testing and tuning is performed on the passive monitoring device. It is then used on several AIS patients to assess the value of its combined force-and-temperature-based compliance monitoring. The quality of the prediction model's predictions is tested in cross-validation on data from thirty-one patients. The same data is used in a clinical trial simulation to estimate the efficacy of treatment recommendations made using the model.

5.1. Compliance Monitor Testing[††]

5.1.1. In-lab Testing and Tuning

The compliance monitor records both force and temperature data. The force data is perhaps most informative as a continuous variable, but the temperature change readings must be translated into a categorical variable: "brace worn" or "brace unworn". This requires a threshold to be set on the temperature change data. A temperature change which exceeds the threshold indicates that the sensor is warming up due to the patient donning the brace. A negative change with magnitude greater than the threshold indicates a cooling down after the brace is taken off. Once these events are flagged, the brace-wear pattern of the patient can be reconstructed.

The optimal thresholds were determined using a test apparatus and test data from five healthy volunteers. The test apparatus (shown in Figure 44) was a section of plastic brace material padded like an actual brace. The compliance monitor was installed in a recessed section of the plastic, exactly as it would be in a real brace. A belt secured the apparatus to the wearer's body, such that the apparatus simulated actual brace wear. Each volunteer donned and removed the apparatus two or three times during a four hour period, keeping a log of when the apparatus was worn. The experiment was carried out in an environment of approximately 21°C. The temperature threshold was then tuned to achieve the best possible match between the logs and the temperature-based estimates of brace wear.

[††] Some material for this section has been published in the paper: E Chalmers, E Lou E, D Hill, V Zhao, "An Advanced Compliance Monitor for Patients Undergoing Brace Treatment for Idiopathic Scoliosis", Medical Engineering & Physics (in press)

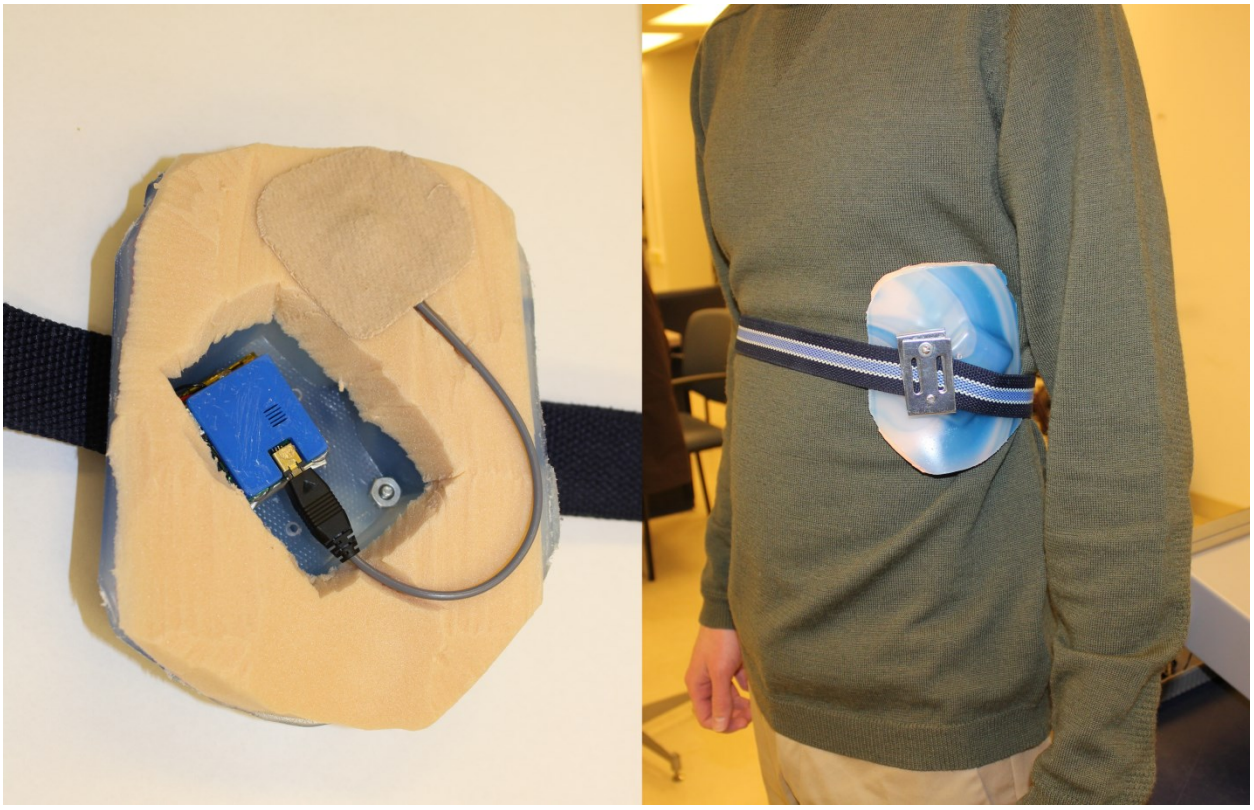


Figure 44: Test apparatus used to test the compliance monitor on healthy volunteers. The apparatus simulates the force sensor's placement in an actual brace.

The chosen threshold allowed the volunteer's brace-wear to be calculated with an average error of 1%: comparable to the 0.1 - 3% range of errors for temperature-based compliance monitors in literature [4], [89], [105], [182]. A cross validation was used to estimate roughly the expected error when the compliance monitor is applied to new patients. The cross validation was conducted in 5 steps: in the i^{th} step of 5, temperature thresholds were selected using all data except the i^{th} volunteer's data, and then these thresholds were tested on the i^{th} volunteer's data. The average of the resulting 5 errors was 4.8%. This cross validation procedure gives a somewhat more realistic accuracy estimate, compared to the practice of setting and testing thresholds on the same data.

When using temperature to monitor compliance, there is some risk that the brace will simply be placed – unworn – in a warm place. The rise in temperature would cause the temperature-based compliance monitor to erroneously begin recording compliance. To avoid this situation the temperature-based compliance state was reset to "unworn" after 24 hours of consecutive "worn" readings.

The force readings can also be converted into the categorical variable: “brace worn” or “brace unworn”. The brace was considered to be worn when the force reading exceeded the fourth centile force reading. The fourth centile was used because the maximum prescribed brace wear time is usually 23 hours per day, meaning the brace is expected to be off at least 4% of the time. The fourth centile seemed more appropriate than the minimum, which could sometimes represent a random spurious sensor reading. The average error of the force-based estimate on the healthy volunteers was 1%.

5.1.2. Patient Testing

Temperature sensing is the de facto standard method of sensing patient’s compliance with brace-wear. To assess the value of the new monitor’s combined temperature/force approach, seven AIS patients wore the monitor in a pilot study. The temperature-based and force-based compliance estimates were compared.

The monitors were installed in the braces of all consenting AIS patients who received a thoraco-lumbo-sacral orthosis (TLSO) brace beginning August 2013. Each of these patients had been prescribed the TLSO (a full-time brace prescribed to be worn up to 23 hours per day) by the orthopaedic surgeon attending the local scoliosis clinic. The patients included 6 girls and 1 boy, with mean age 13.7 ± 1.7 years, and mean major Cobb angle 35 ± 6°. Five patients had right thoracic curves, one was left thoracic, and one was left lumbar. Ethics approval was granted by the local ethics board.

Force sensors were installed by the orthotist at the major correction pad area (Figure 45). A section of the main pressure pad in each brace was cut away to accommodate the force sensor, and the sensor was covered with an adhesive covering. The main pressure pad placement is an important aspect of brace design, and so the main pad was considered the optimal location for a single force sensor. The monitors were installed during the patients’ clinic visits or brace fittings, and retrieved at the following visit – producing an average of 93 days of data per patient (range 41-218 days). Force sensors were confirmed to be working properly when the patients returned to the clinic.

Temperature-based and force-based brace wear-time estimates were calculated for each day and as an overall average. A Wilcoxon test checked for significant differences between daily force-based and temperature-based estimates for each patient. The fourth centile threshold force used in the force-based estimate was calculated for each twenty-four hour period separately; this

accounts for changes in the baseline force over time, due to gradual shifting of the sensor or wearing of the brace padding.

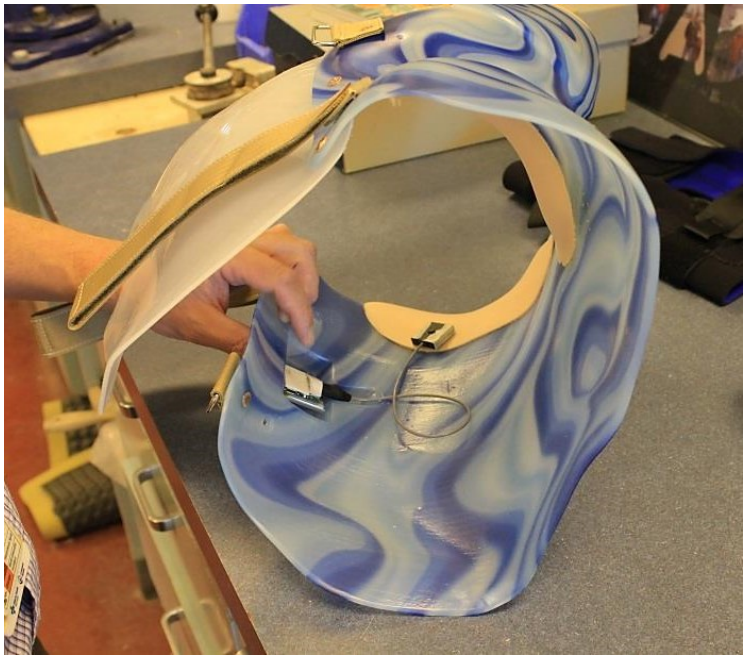


Figure 45: Installing the monitor in a brace.

The average force reading while the brace was worn (per the force-based estimate) was calculated. Force-based compliance by time of day was also measured for each patient and normalized to their overall compliance, to show the distribution of brace wear by time of day.

5.1.3. Patient Test Results

Figure 46 shows the compliance estimates for each patient. Three patients (patients 3, 5, and 6) showed close agreement between temperature and force-based compliance estimates. However, the other six patients showed statistically significant ($p<0.01$) differences between the two estimates. In three of these six cases the force sensor estimated significantly lower compliance than the temperature sensor.

The average forces measured during brace wear are shown in Figure 47. These values show the force measured during periods when the force sensor deemed the brace to be worn. The values seem reasonable given force measurements taken by Lou et al. [102] in a clinic setting (where the patient's brace-wear is at its best). There is a loose correlation (Pearson $r=0.65$, $p=0.06$) between these values and the force-based compliance estimates in Figure 46, indicating that

patients who wore their braces tightly generally tended to receive higher force-based compliance estimates.

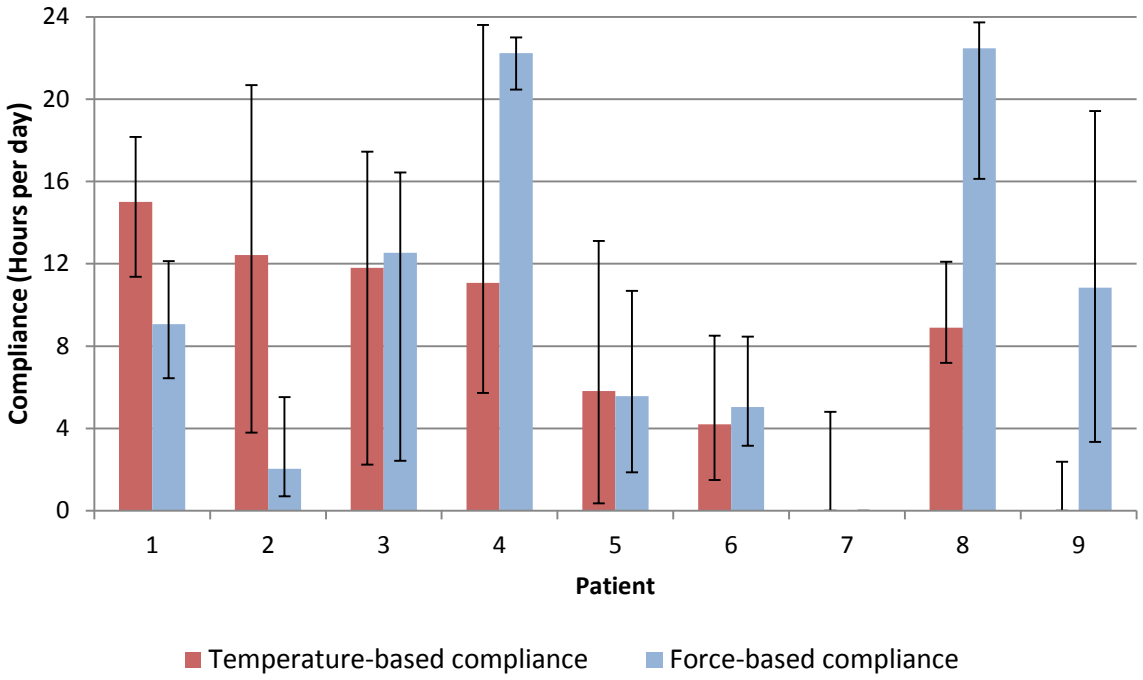


Figure 46: Median daily force-based and temperature-based compliance estimates for each patient. Error bars show 16th and 84th centiles. Only patients 3, 5, and 6 show good agreement between force and temperature-based estimates.

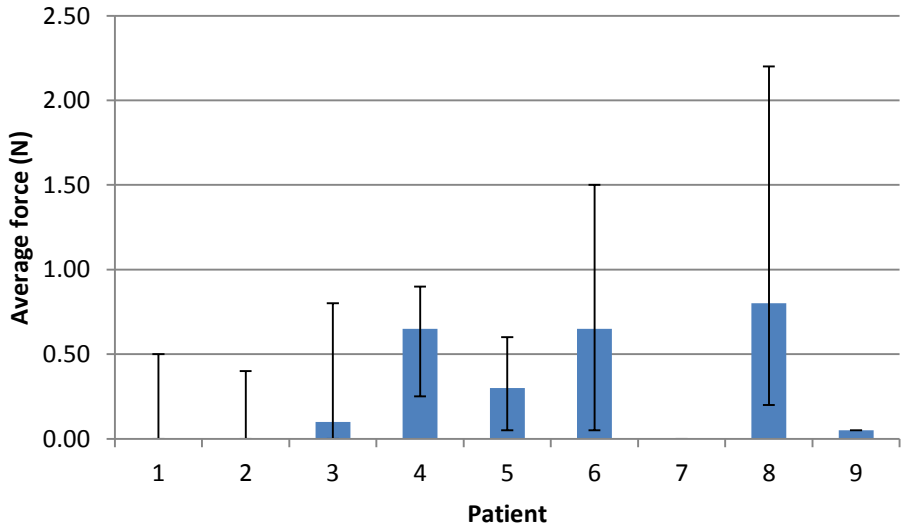


Figure 47: Median forces during brace-wear measured for each patient. Error bars show 16th and 84th centiles.

It should be noted that the force sensors were calibrated before being installed in the braces. The installation involved covering the sensor with foam padding of varying thickness, so the sensor readings may not actually be accurate measurements of force. Fortunately, their value as measurements of force is not relevant within the scope of this thesis; ultimately, they need only be useful for predicting treatment outcome.

Figure 48 shows each patient's measured probability distribution of brace wear by time of day; patient 7 was excluded from this figure because their monitor recorded no force. This chart shows the portion of each patient's daily brace wear occurring during intervals centered at 00:00, 06:00, 12:00, and 18:00; thus it shows patients' brace-wear patterns vis-à-vis the time of day. For example, consider the line representing patient 6: roughly 55% of the brace-wear recorded during the test occurred at 18:00, 40% occurred at 12:00, and almost none occurred between 00:00 – 06:00. Thus this patient tends to wear their brace in the afternoon and early evening, and generally sleeps without it. In contrast, patient 5 wears the brace most frequently at night, and infrequently during the mid-day. This figure seems to agree with previous studies, which have found patients to be generally more compliant at night than during the day [104], [106].

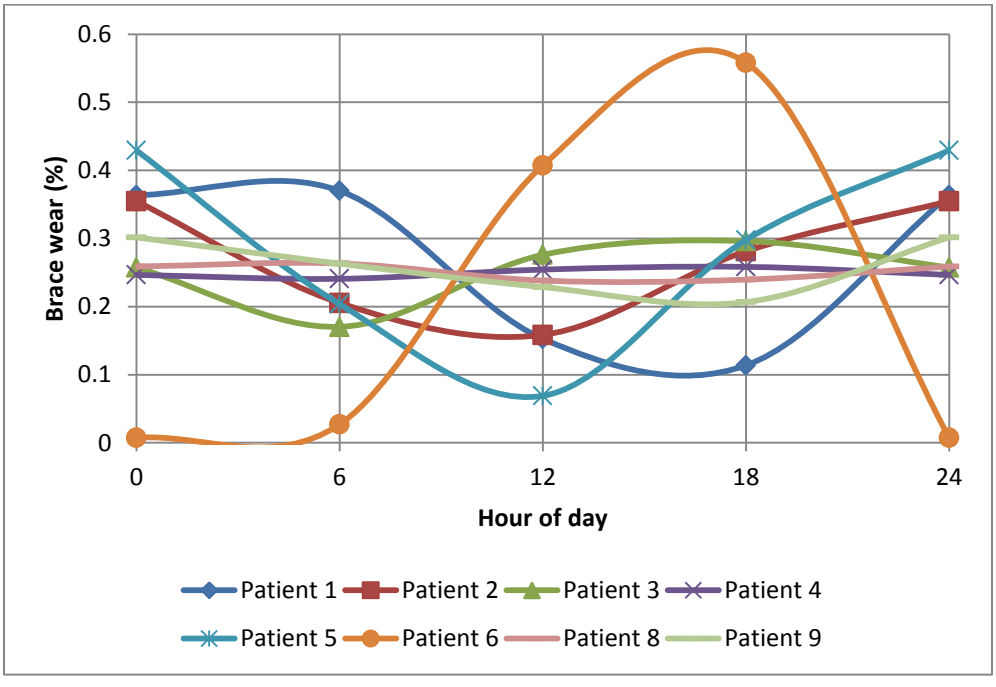


Figure 48: Distribution of each patient's force-based brace-wear estimate by time of day.

The test results suggest that temperature-based compliance estimates can differ significantly from force-based estimates. Previous work has shown that patients wear their braces at only 39%-78% of the prescribed tightness [102], suggesting that temperature-based compliance estimates should be higher than force-based estimates (a loose brace would not activate the force sensor but should still activate the temperature sensor). For example, patients 1 and 2 probably don their braces loosely – resulting in intermittent force readings (as illustrated in Figure 47).

However, Figure 46 shows force-based estimates being *higher* than temperature-based estimates just as often as not. There is no definitive explanation for this anomaly, but these patients' force readings seem more realistic than their temperature readings: the force readings show reasonable periods where the brace was unworn, usually just before and after school time or in the evenings. The rest of the time there is significant force with mild fluctuations, just as we expect from in-brace force readings. The temperature readings, on the other hand, often switch inexplicably between 'worn' and 'unworn' states, with no corresponding activity in the force readings. These unexpected events usually occurred at regular times in the morning or evening, presumably during patients' morning or nighttime routines. If these patients wore insulating shirts under their braces, or if the sensor was installed such that it did not have good thermal contact with the body, the temperature readings would have been biased by ambient air. The compliance monitor's temperature sensing function worked well in a laboratory setting, but it appears that in practical application its reliability may not be perfect. The discrepancies between force-based and temperature-based compliance monitoring could have implications for practitioners who use temperature-based monitors. In future testing patients should be asked to keep a brace-wear diary for comparison to the sensor estimates. Though ultimately the true accuracy of the diary would be unknown, it may shed some light on which sensor's brace-wear estimate is more accurate. Note though, that a brace-wear estimate need not necessarily be an accurate measure of wear time to be useful for prediction.

Even when each sensing modality functions properly, there are limitations to temperature-only and force-only compliance monitoring. Temperature-based monitoring could give a good indication of *when* the brace is worn, but not *how well* it is worn. Temperature sensing also becomes ineffective if ambient temperature is similar to skin temperature or if the sensor has a poor thermal connection to the patient. Conversely, force-based monitoring can indicate *how well* the brace is worn, but gives a poor estimate of *how often*. The more complete information provided by a force-and-temperature combination could enable some clinically relevant

distinctions. For example, Figure 46 shows poor force-based and temperature-based compliance estimates for patient 7: this patient simply does not wear their brace as prescribed (and openly admitted this during their clinic visit). Patients 1 and 2 wore their braces relatively often (according to the temperature-based estimate), but force readings suggest it was worn loosely. Patient 3 has similar temperature and force-based compliance estimates: when they wore their brace it was likely worn at the appropriate tightness. A health-care provider will give these patients different recommendations in response to this brace wear information.

5.2. Validation of Prediction Model and Decision Support Engine

5.2.1. Patient Data

Data from thirty-one AIS patients who had compliance monitors installed in their newly prescribed braces was stored and processed using the software platform. Fourteen night braces and seventeen full-time braces were included in this sample. These patients had been recruited to wear the compliance monitors, starting in 2008. An additional five patients had monitors installed in their braces but dropped out of the study thereafter. One of the patients was excluded from the dataset because of erroneous force readings recorded by their monitor. Distributions of the seven variables listed in section 4.3.3.2, and the outcome (Cobb angle progression per year), are shown in Table 7.

Variable	**Distribution**
Age	12.7 ± 1.8 years
Major Cobb Angle	29 ± 7°
Height Velocity	2.0 ± 1.3 cm/year
Scoliometer Measurement	10 ± 3°
In-brace Correction of Cobb Angle	Full-time brace: 38 ± 20% Night-time brace: 100 ± 34%
Average Force*	0.27 ± 0.2
Brace-wear compliance	Full-time brace: 9.2 ± 6 hrs/day Night-time brace: 6.6 ± 2.7 hrs/day
Cobb Angle Progression per Year	1 ± 7°

* After normalizing all force readings to the 95th centile value.

Table 7: Description of the patient data used for model and decision support validation.

Few patients had a clinic visit that fell exactly one year after beginning brace treatment. As a result, Cobb angle progression was measured over durations ranging from 8 to 16 months. All

measured progressions were converted to "per year" quantities for consistency. Not all the patients in this group had been given the compliance monitor described in section 4.2; many had used a previous device which did not include a temperature sensor. Thus all brace-wear estimates in this dataset are force-based as described in section 4.3.3.2.

Height velocity could not be calculated for 10 of the patients because the necessary height measurements were not available. The missing height velocity entries were imputed using the mean of the available height velocity instances. Scoliometer measurements did not exist for 5 of the patients. The missing scoliometer measurements were imputed using a linear regression model. This model was created by querying the local clinic's database for age, height, weight, and Cobb angle measurements from all examinations of idiopathic scoliosis patients aged 9-16. This returned 1367 examination records from 648 patients. A wrapper feature selection process selected Cobb angle as the single best predictor of Scoliometer measurement out of the four possibilities: a linear regression model using Cobb angle predicted Scoliometer measurement with a mean absolute error of 2.9° in 10-fold cross-validation (the default prediction method of assigning all cases the mean Scoliometer value gave a mean absolute error of 3.9°). The requirements for age, height, and weight were then removed from the database query, resulting in 1892 examination records from 783 patients. The linear regression model was build using this data, with the final model being: Scoliometer = 0.2166 * Cobb Angle + 3.5165.

The pooling of both night and full-time braces in the data may seem unusual to those more familiar with traditional hypothesis-based research. When the research goal involves a conventional hypothesis test or comparison, it is often advisable to subdivide the data into homogeneous subgroups for analysis. But when the goal is simply to achieve good predictions, it is appropriate – even preferable – to use more heterogeneous training data. Not only will the resulting model be generally applicable (rather than being applicable to only one brace type), but the model may actually be made stronger when heterogeneous training data are used. Justin Washtell explains this concept in the context of advertising:

> *"The central premise of predictive modeling is precisely that one size does not fit all - otherwise we would just assign the same outcome to all cases and be done with it. The intention is that to whatever extent customer group A is different to customer group B, our algorithms should recognize this [and] the resultant model should treat the two groups differently. At the same time, to whatever extent customer group A is similar to customer group B, we would still like our algorithms to identify those similarities and discover general rules. Our joint model - having benefited from the entire data - will be much stronger overall." [183]*

5.2.2. Cross-validation of Prediction Model

The correlation-based feature selection was applied as a pre-filter to the dataset described in section 5.2.1. The results of the correlation based feature selection are shown in Figure 49. The compliance feature produced a higher value of r_{zc} (as per Equation 1Equation 11; see page 76) than any other single feature or pair of features. However, combinations of average force, compliance, and in-brace correction all performed similarly.

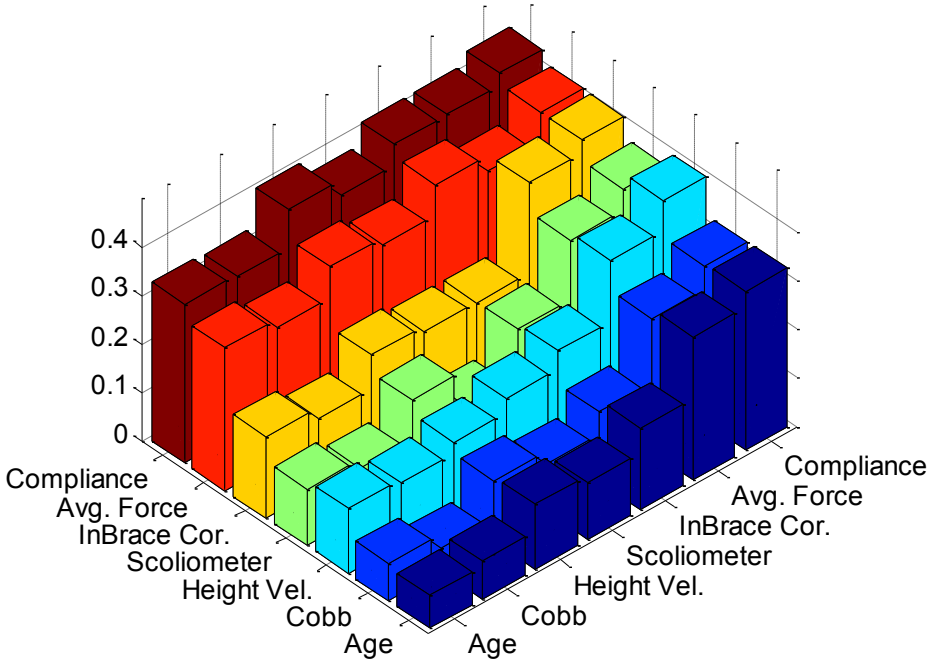


Figure 49: Results of correlation-based feature selection algorithm for each feature and pair of features. The compliance feature produced the highest score (by a small margin).

The logistic regression model was tested in a leave-one-out cross-validation on the filtered data. For comparison, "default" predictions were also generated by predicting each test patient to have the same probability of progression observed in the training patients. That is, given the set of training data S = {[X1,Y1], [X2,Y2]... [Xn,Yn]}, where each patient instance Xi ϵ $\mathbb{R}^d$ has a corresponding outcome Yi ϵ [0,1], the default method ignores the features Xi and generates a predicted probability of progression $\hat{Y}$ by simply averaging the probabilistic outcomes Yi:

$$\hat{Y} = \frac{1}{n}\sum_{i}^{n} Y_i \qquad \text{Equation 12}$$

The Brier Score and median prediction error were calculated for both the logistic regression and the default method. Brier Score is the mean square error between a model's probability prediction and the corresponding true outcome. The Brier Scores were 0.16 and 0.18 for the logistic regression and default predictions respectively. The median errors were 0.35 and 0.50 for the logistic regression and default predictions respectively. This performance improvement over the default method shows that the logistic regression is utilizing information contained in the patient data.

The distribution of errors for each method is shown in histograms in Figure 50. The distribution for the logistic regression is subjectively more desirable, having more bulk in the lower half of the error range. However the difference between the two sets of predictions is not statistically significant on the small validation dataset (p=0.40 by the Wilcoxon test).

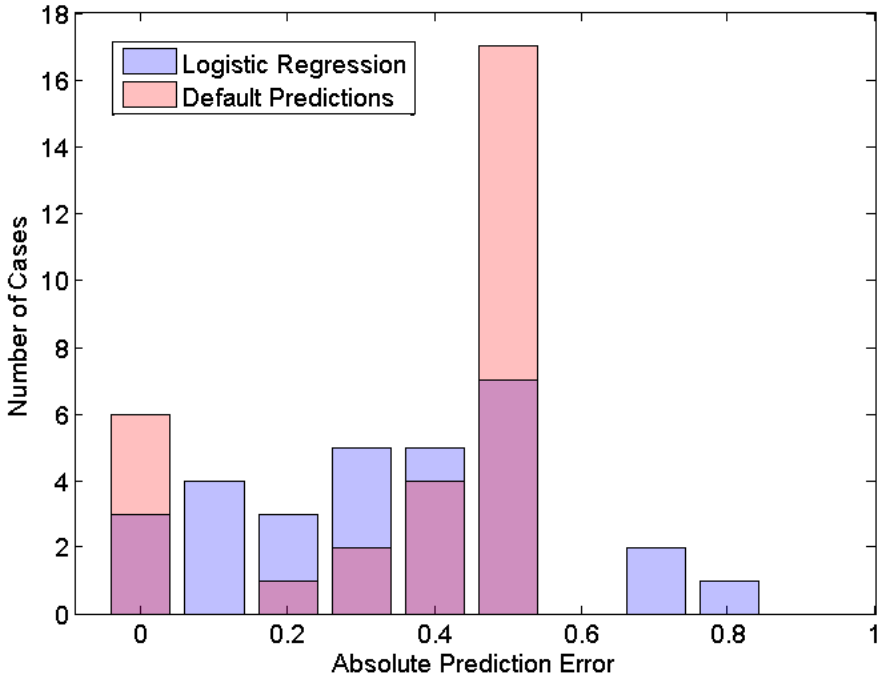


Figure 50: Overlaid histograms showing the distribution of errors obtained using the logistic regression (blue) model, and the default method (red). (Areas of overlap appear purple).

While the logistic regression model demonstrated an improvement over the default method there are two main factors which limit the predictive power achievable at this stage. The first is

the small training data set; as the database grows, it will come to support more meaningful and powerful models. The second factor is the outcome being taken after one year rather than after the entire treatment; “progression” may be a stochastic process which is easier to predict in the long term than the short term. This logistic regression model may not yet be suitable for clinical use – unless an experiment like that described in section 3.2 were to demonstrate clinical value. However the model is valuable as a demonstration of how the software platform may be used to generate predictive models. Recommendations based on this model may also have value as described in the following section.

Figure 51 shows a calibration plot for the model’s cross-validated predictions. “Calibration” describes how *correct* a probabilistic prediction is, vis-à-vis the corresponding actual outcomes for instances which received the prediction. For example, suppose ten patients are each predicted to have an 80% chance of progression. If eight of them actually progress, the predictions were well-calibrated. The calibration plot in Figure 51 was calculated by first applying Gaussian smoothing (using a normal distribution with $\sigma = 0.05$) to the set of predictions to compensate for the small dataset. At each point, the vertical deviation between the calibration curve and the ideal diagonal line was scaled by the relative weight of predictions at that point; this compensates for large deviations which are the result of an under-represented prediction.

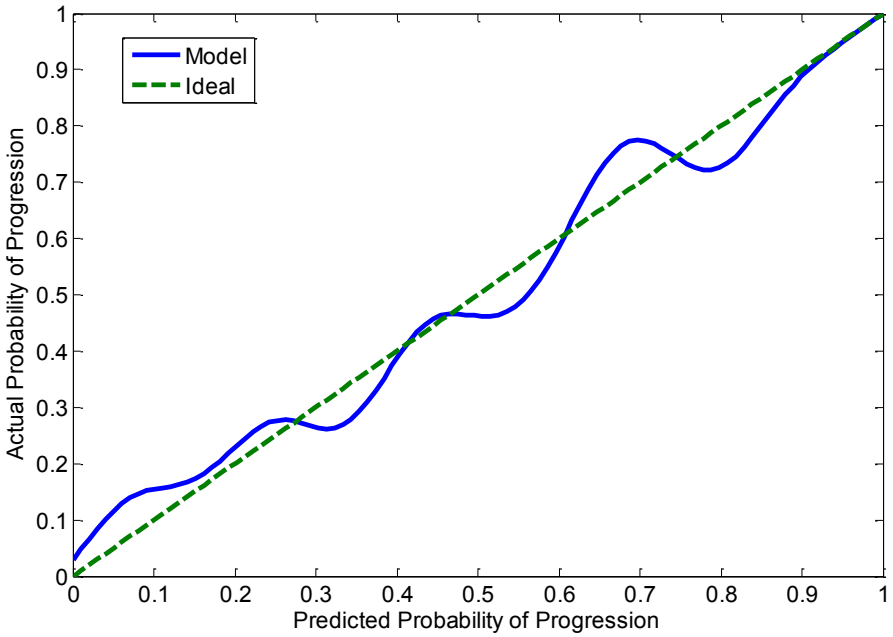


Figure 51: Calibration plot for the logistic regression model's cross-validated predictions.

The logistic regression model's output is plotted in Figure 52, along with dots representing each patient. With brace-wear expressed as a fraction of 24 hours, the logistic regression coefficient for brace-wear compliance was -3.25, with a constant term of 1.07. The probability of progression can be calculated using these coefficients and the brace wear as per Equation 1.

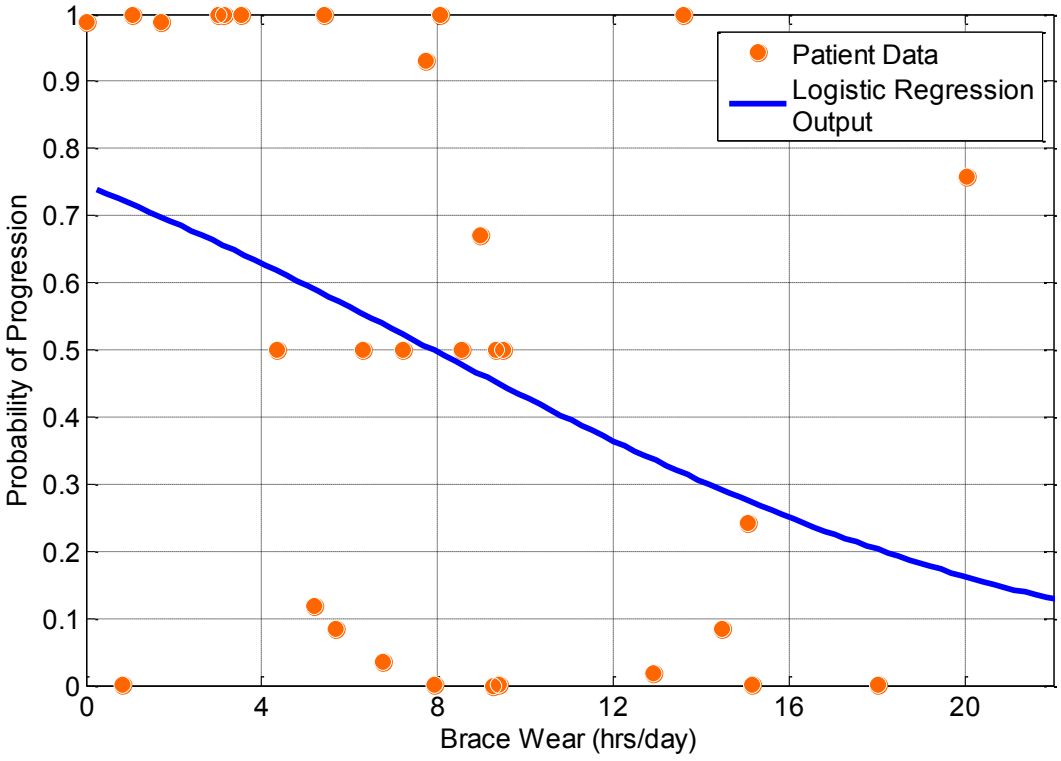


Figure 52: Logistic-regression model predictions as a function of brace-wear time. Each dot represents one of the patients in the training dataset.

5.2.3. Validation of Decision Support Engine

The primary purpose of the prediction model(s) is to be used by the decision support engine to generate treatment recommendations, as described in section 4.4.4. The logistic regression model described in sections 4.3.3 and 5.2.2 was used to generate recommended brace wear-times. These recommendations were calculated by solving Equation 1 (on page 34) for brace-wear time given a specified target probability of progression:

$$Recommended\ brace\ wear = \frac{-\ln\left(\frac{1}{p} - 1\right) - \beta_0}{\beta_1} \qquad \text{Equation 13}$$

where p is the target probability of progression, β_0 is the logistic regression constant term, and β_1 is the logistic regression coefficient for brace wear. The recommended wear-times produced by Equation 13 were capped at 8 hours/day for night braces and 23 hours/day for full-time braces. The target probability of progression was set to 10% lower than the overall probability

observed in the data. The average probability of progression in the data described in section 5.2.1 was 0.50, so the target probability was set to 0.40.

The effect of the recommended brace wear was estimated in a clinical trial simulation. The CTS randomly divided available patient data into two equally-sized groups: A and B. Separate prediction models were trained using the data from each group. Model A was then used to calculate recommended wear-times for the patients in group B, as per Equation 13. Model B then predicted the new treatment outcomes for the group B patients given Model A's recommendations. Refer to Figure 26 (page 58) for an illustration of the general CTS procedure.

This CTS procedure was repeated 1000 times, with the group A / group B assignments re-randomized each time. The 1000 repetitions allowed various distributions to be estimated. Figure 53 shows a histogram of the wear-time recommendations produced during all runs of the CTS. The average recommendation was 12.6 hours/day, though the recommendations were limited to 8 hours/day for patients who had night braces. Figure 54 shows the distribution of the difference between the wear-time recommendation and the actual measured wear-time. In 34% of cases the recommendation was less than or equal to the actual wear-time.

The CTS-estimated effect of the recommendations is illustrated in Figure 55 and Figure 56. Figure 55 shows the distribution of individual patients' change in progression rate during all runs of the CTS. Approximately 67% of patients experienced a decrease in risk of progression. The mean change in probability of progression was -0.13. Figure 56 shows the distribution of overall group B risk of progression. The mean group probability of progression was 0.37; slightly better than the 0.40 target used in the simulations. Of the 1000 CTS runs, 3.6% showed a group probability of progression greater than the measured average of 0.50. Thus the improvement in group probability of progression could be considered statistically significant.

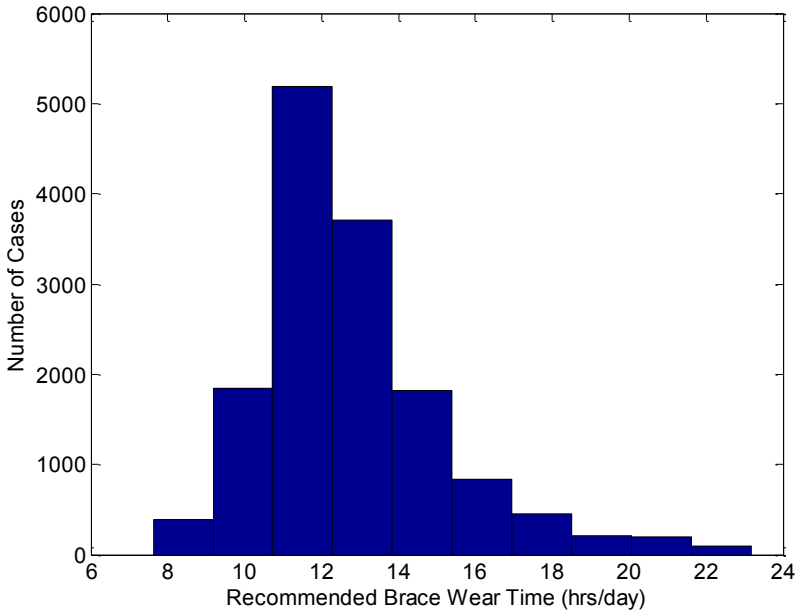


Figure 53: Histogram showing the distribution of brace wear-time recommendations produced during the clinical trial simulations.

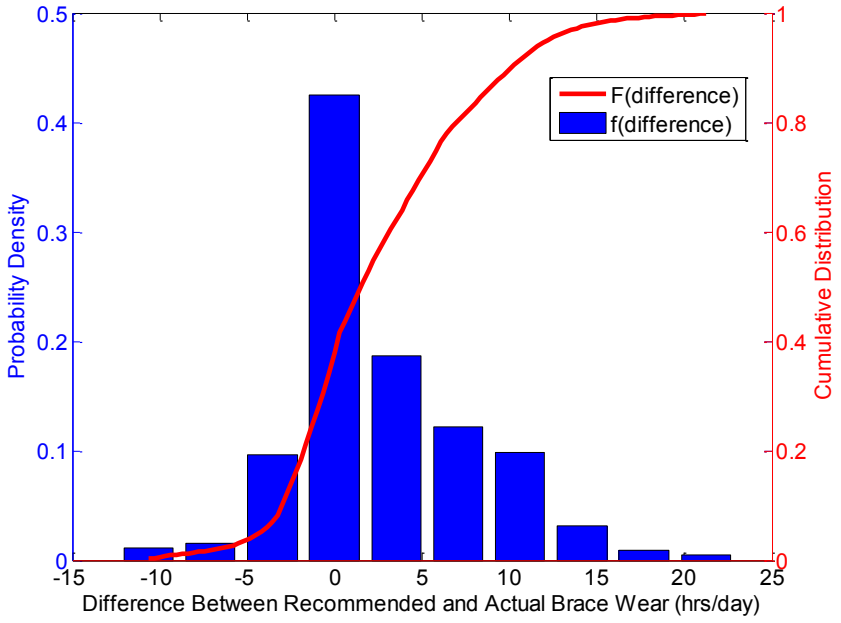


Figure 54: Empirically measured probability distribution and cumulative distribution of the difference between recommended wear-times and patients' actual measured wear times.

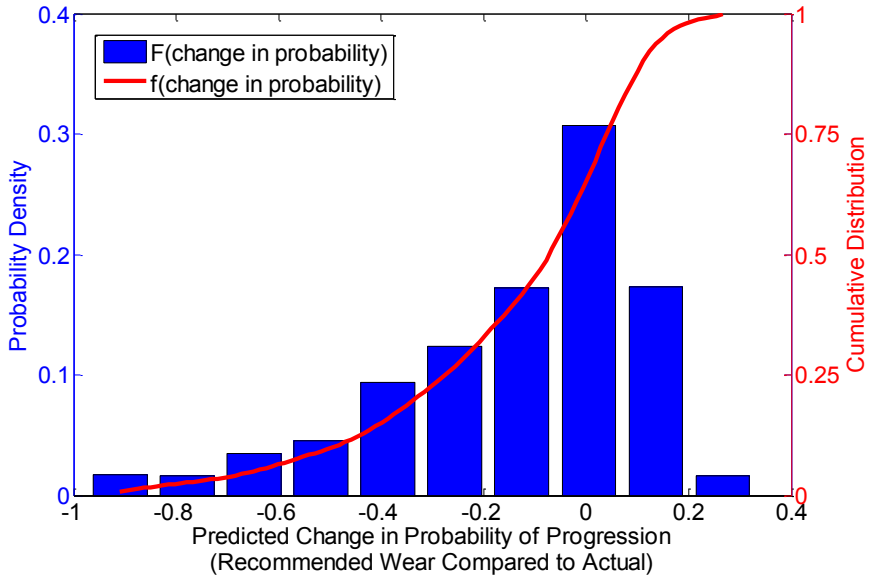


Figure 55: Empirically measured probability distribution and cumulative distribution of the estimated change in probability of progression across all individual group B patients in all runs.

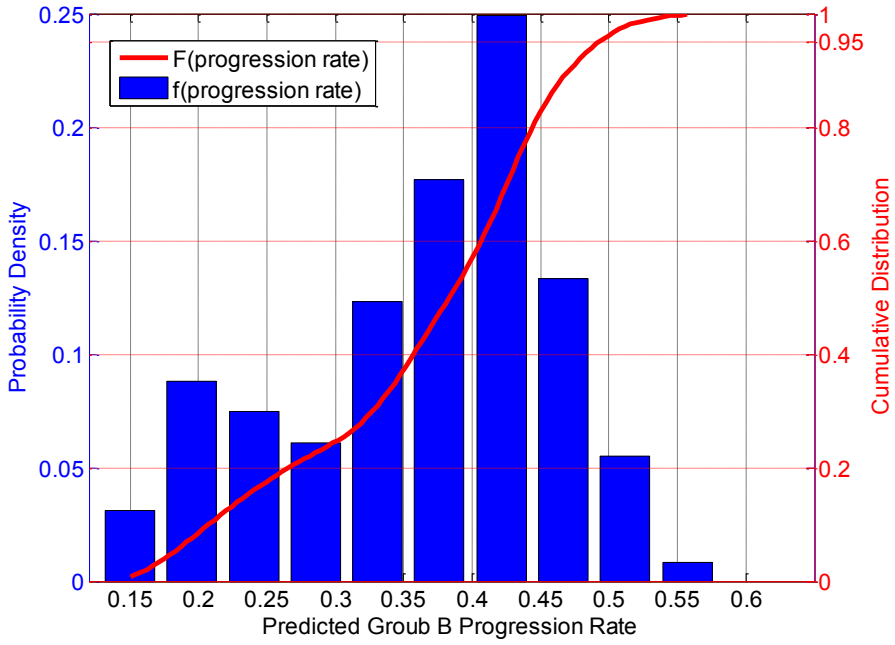


Figure 56: Empirically measured probability distribution and cumulative distribution of overall group B risk of progression across all runs of the clinical trial simulation.

The three main points to be gleaned from the CTS results are:

- Average recommended wear-time was 12.6 hours/day (reduced to 8 for night braces)

- The recommended wear times produced a significant reduction in *overall* progression rate, but
- Predicted risk of progression for *individual* patients increased significantly in some cases.

The (average) 12.6 hours/day of brace wear meant an increase for many of the patients in the simulation; and this caused the reduction in overall progression rates – a success vis-à-vis the goals and design of the simulation. However, some patients had in practice worn their braces much more than this, and the reduction to 12.6 hours produced a corresponding increase in their individual probability of progression. If there is a message of clinical significance, it may be that encouraging non-compliant patients to wear a full-time brace for the prescribed 18-23 hours/day (a daunting prospect for some adolescents) may be unnecessary. It may be enough to help these patients work toward 12-13 hours/day.

As the software platform is used to accumulate more patient data, more sophisticated prediction models will be possible. Having more accurate and more individualized predictions would give a narrower distribution of possible group progression rates (as compared to Figure 56) and a lower incidence of increased individual risk of progression (as compared to Figure 55).

6. Conclusions and Recommendations

A brief recap of the work is provided, and the major contributions of this thesis are listed. Limitations of the work in this thesis are discussed, and some suggestions for future research in this area are put forward.

6.1. Summary of Work

After a literature review, this thesis explored the idea of data-driven decision support for brace treatment of AIS; specifically the idea that a prediction model could be used to generate optimum in-brace correction targets for each patient. Two prediction models were developed using a retrospective dataset of 90 patients – both predicted whether or not a braced patient's Cobb angle would progress by the end of treatment. The first model was a logistic regression model which made predictions on the basis of patients' Scoliometer measurement and in-brace correction. The second model was a fuzzy model based on conditional fuzzy clustering, which made predictions on the basis of patient's Scoliometer measurement, in-brace correction, Cobb angle, and age. Both models were cross-validated on the training portion of the dataset and further validated on the test portion to ensure generalizability.

The fuzzy model's predictive power was found to be superior to that of the logistic regression. It was subjected to additional testing by comparison with human AIS experts. In this test the model predicted more accurately than seven out of eight experts. This comparison provided some evidence that the model had the ability to enhance the experts' natural predictive ability.

A method of using the model to generate patient-specific in-brace correction recommendations was described: the model was used to predict treatment outcome for a range of possible corrections, and the correction giving the most desirable outcome was selected. Larger corrections were not always more desirable, and in one third of cases the recommended correction was lower than what had actually been applied clinically. A clinical trial simulation (CTS) was used to estimate the effect of applying the recommended in-brace corrections. The CTS estimated that the recommendations (if achievable) would lower overall progression rates. Ultimately a real clinical trial would be necessary to prove that recommendations generated in this way could change outcomes. However the CTS is an inexpensive way to estimate what the outcome of the real clinical trial would be, and thus provide some justification for moving forward.

The thesis then described a complete decision support system. The system was designed following the general approach used in chapter 3: patient data is collected and used to train prediction models, which are used to generate treatment recommendations. However several methodological improvements were made to the approach in chapter 3. The decision support system includes a means of collecting brace-wear information for use in prediction and recommendation, where the approach in chapter 3 relied purely on measurements taken in the clinic. The system treats the prediction task as a probabilistic rather than a classification problem, and the system includes an algorithmic method of selecting treatment recommendations rather than the manual method used in chapter 3.

The system includes an electronic monitoring device which can be embedded in a patient's brace during long-term treatment. The device monitors patients' brace-wear by measuring temperature and force applied at the main brace pressure pad. A software platform stores patient data and allows it to be visualized or modelled using an expandable set of visualization and modelling modules. This thesis includes the development of several visualization modules and one modelling module, which trains a simple logistic regression model to predict progression after one year. The platform also includes a decision support engine which can facilitate "what-if" analysis or search for a given patients' optimal treatment parameters using a genetic algorithm. The complete system facilitates collection and storage of brace-wear data and other clinical data, visualization of the data, training of prediction models, and generation of treatment recommendations using the models.

In testing, the monitoring device achieved less than 5% error in laboratory testing, and was used in seven patients' braces where it provided more complete information than conventional temperature-based monitors. Compliance monitor and other clinical data for thirty-one patients was processed and stored by the software platform. Brace wear-time was selected by a feature selection algorithm as the single most valuable feature in this dataset, and was used to train a logistic regression model to predict one-year progression based on brace-wear time. This model achieved a Brier score of 0.16 in cross-validation, and presented an improvement in predictive power over a default prediction method. The model's probabilistic predictions also appear to be fairly well calibrated.

The prediction model was used to calculate recommended brace wear-times. These recommendations were calculated to reduce the patient group's overall probability of progression by ten percent. A clinical trial simulation estimated these recommendations to

provide a statistically significant reduction in overall probability of progression; the mean value of this reduction was 13%. The average brace-wear recommendation in the clinical trial simulation was 12.6 hours per day.

6.2. Summary of Contributions

The main contributions of this thesis are:

- An exploration of the concept of CDSS for use in brace treatment planning. Current practice relies to a large extent on established guidelines and algorithms. While these guidelines are often based on some simple statistical analysis of patient data, the computerized use of locally collected data to support decision making represents a paradigm shift. This thesis provides a thorough description of a system that leverages patient data to provide a source of information and recommendations during treatment planning. In contrast to established guidelines, these recommendations would be based on locally-collected data describing the types of patients and treatments seen at that clinic itself.
- A preliminary demonstration of the concept of electronic decision support for brace treatment planning. A CTS in Chapter 3 estimated that computer-recommended in-brace corrections, if achievable, could reduce progression rates. A similar CTS in chapter 5 estimated that computer-recommended wear-times, if adhered to, could reduce a patient groups' overall probability of progression to a desired level.
- A template for predictive modelling and data-driven decision support in this domain. This thesis employed appropriate predictive modelling techniques: models were tested in cross-validation and further validated on separate data where possible. Models were evaluated on the basis of their predictive power, rather than their statistical goodness-of-fit. The amount of data available was considered when choosing techniques and "degrees of freedom". Clinical trial simulations estimated the distribution of possible outcomes, should the models be used in practice.

Additional contributions include:

- The fuzzy model described in section 3.1.3, which provides valuable predictions of treatment outcome based on pre-treatment clinical measurements. The modelling technique is novel, and seems to have good aptitude for this prediction task.

- Description and validation of a compliance monitoring device described in section 4.2, which combines both force and temperature-based monitoring to provide more complete information than other existing temperature-only compliance monitors. The force data collected by this monitor has been used in this thesis to create a prediction model for brace treatment outcome.

6.3. Limitations and Recommendations for Future Work

6.3.1. Patient Data

The given timeframe allowed a limited amount of data to be collected for use in chapter 5's validation of the decision support system. As a result the predictive model developed as described in section 4.3.3 was kept simple to avoid over-fitting the data. Its testing as described in section 5.2 was limited to cross-validation, with additional testing on new data reserved for a future time. As the platform developed and described in this thesis is used, more and more patient data will accumulate in the database. Additional model testing, development of new and more sophisticated models, and more precise CTS results will be possible given enough data.

It is difficult to say how much data is "enough". In fact there may be no way of knowing *a priori* how many observations a dataset should contain in order to support a given predictive modelling task. Some crude rules of thumb have been proposed in terms of events-per-variable – the ratio of observations in the dataset to predictor variables considered by the machine learning algorithm. A simulation study by Peduzzi et al. [184] suggested that logistic regression analysis requires 10 events-per-variable. A later study by Vittinghoff and McCulloch [185] suggested this rule of thumb is too conservative, and could be relaxed. A similar study by Sahiner et al. [186] suggested 5 events-per-variable as an appropriate minimum. However, all these studies have considered only linear or generalized linear models. Moreover, events-per-variable rules of thumb are rough guidelines at best, and are routinely violated in fields like gene microarray analysis [187] and econometrics [188], where the events-per-variable ratio can be much less than one. Conversely, some prediction goals may require many more than 10 events-per-variable.

Rather than try to estimate how much data will be required for a prediction task, it may be better to focus on the prediction performance. Raudys and Jain [189] point out that a measurement of a model's prediction performance is a random variable, as it depends on the

particular training and test samples used for evaluation. Cross-validation as used in this thesis can be a useful tool to estimate the distribution of possible prediction performance [190]. As the amount of available data increases, this measured distribution should approach its "true" form. For example, consider the group probability of progression shown in Figure 56. This distribution is already acceptable given the goals of the CTS, but additional data could cause the confidence interval of this distribution to shrink further. If additional data no longer affects the distribution, then it is likely that the existing data already gives a sufficient representation of the patient population.

To speed the collection of patient data, a clinic should consider implementing compliance monitoring as a standard procedure. This would drastically increase the amount of available compliance monitor data. Data-sharing agreements could be also established between multiple clinics, as has been done in other fields [191]. However such agreements could potentially detract from the clinic-specific nature of the system's recommendations. As a minor note, data collection could also be facilitated by modifying the software platform to automatically pull patients' clinical data from the clinic's database, rather than requiring this data to be manually entered. In the future the concept of a stand-alone platform could be discarded altogether; it may be possible to directly integrate the decision support system with existing software used by the clinic.

Section 5.2.1 noted that several patients who received compliance monitors dropped out of the study soon after. While this could potentially create bias in a conventional analysis, it may not be a concern for the work in this thesis. The prediction model predicts probability of progression given a patient's brace-wear (measured by the monitoring device). If there is a "type" of patient who is likely to refuse a monitoring device, the model will not apply to them and they should be removed from the dataset. However, caution should then be exercised in applying the model-based recommendations to this type of patient.

There are several limitations to the data used for modelling. For example, only a snapshot (60 days) of brace-wear data was used in calculating the average force and brace-wear variables shown in Table 7. Thus it is assumed that brace-wear habits are relatively consistent throughout treatment (or at least that these 60 days can be used to predict outcome). Orthotists have suggested in private conversation that this assumption is reasonable. Moreover, assumptions such as this are often unavoidable; if the predictions required brace wear measurements taken throughout treatment, the prediction model would be inapplicable to current patients in

practice (i.e. it would not be able to deliver a prediction until the end of treatment – when the prediction was no longer needed). Another such assumption is that the single in-brace correction measurement used to quantify the brace's effectiveness is a good indication of the correction experienced during daily brace-wear. In practice, the in-brace radiograph may tell us more about what correction is *possible* than what correction is typical. Still, future work could develop a prediction model which uses brace-wear measurements taken throughout treatment, and is used purely for creating recommendations.

Another limitation of the compliance monitor data is the fact that it represents force at a single location in the brace. Each monitor was installed by the orthotist at the brace's main pressure pad – the location deemed by the orthotist to be of primary interest in terms of applied force. Thus, force readings are comparable between patients in the sense that they all represent primary pressure pad forces. However braces involve several applied forces acting in concert, so a single force sensor will never completely describe the effect of the brace. Future work should strive to incorporate multiple sensors in a non-obtrusive way. This would allow better measurement of the force distribution inside the brace (though it likely would not affect the force-based compliance estimates).

The discrepancy between force and temperature-based brace wear estimates should be investigated. The force-based estimates are most likely the more useful of the two; they seemed more realistic during patient testing, and were a stronger predictor of progression than the other variables listed in Table 7. Still, a comparison with patients' brace-wear diaries or with another temperature-based monitor could be beneficial.

Future work should make use of new clinical measurements as they become available. For example, this thesis has considered Cobb angle measurements taken using conventional 2-D radiographs. 3-D radiography is becoming more commonplace and may provide measurements with better predictive value. Work by Courvoisier et al. [77], [192] has demonstrated that the EOS 3-D radiography system (EOS Imaging) can be used to assess the efficacy of a brace – measuring correction of rotational deformity in addition to Cobb angle. Courvoisier et al. [193] and Nault et al. [168], [169] have used EOS to show that measures of transverse-plane deformity correlate strongly with progression. Other examples of potentially useful predictors which may

be more readily available in the future include measurements of spinal stiffness, or the 3-D orientation of the force applied by the brace's pressure pad.

6.3.2. Prediction Model[‡‡]

Castillo and Kellemen point out a risk that users may come to distrust a CDSS if it frequently provides erroneous information [139]. As more data becomes available it will be important to conduct additional testing of prediction models to ensure their prediction performance in acceptable. If possible, it would be beneficial to test the model(s) by comparison with experts as described in section 3.2. Such testing can be a good demonstration of a prediction model's value (or a revelation that the lack thereof); this is important when the model is a component of a non-knowledge based decision support system, which derives its knowledge from data. Knowledge-based decision support systems have less need of such validation, as they simply recite the codified knowledge the experts have programmed them with.

The patient data described in section 5.2.1 included force-based brace wear information only. Temperature information was discarded as it was not available for all patients. As discussed in section 5.1.3, the combination of force and temperature data provides a more complete picture of brace wear than either force or temperature-based monitoring alone. Future work could include both kinds of measurements in the set of potential predictor variables. Alternatively, future work could seek to combine the two measurements into a single indicator in a meaningful way.

The accumulation of patient data over time will allow experimentation with a variety of modelling techniques. The simple logistic regression model described in this thesis is designed to provide a probabilistic prediction of progression, which is useful in understanding any patient's risk of progression and in producing treatment recommendations. Most standard modelling techniques will result in similar probabilistic predictions or classifications, which help us understand who is *likely to* progress. In the context of brace treatment, it may also be useful to know who is almost *certain to* progress or not progress. Patients who have an extremely low (or extremely high) likelihood of progression may not be good candidates for brace treatment at all.

[‡‡] Some material in this section has been published in the paper: Chalmers E, Mizianty M, Parent E, Yuan Y, Lou E, "Toward maximum-predictive-value classification", Pattern Recognition, 47(12): 3949-3958, 2014

Identifying cases where the outcome is almost certain is a difficult task: in statistical terms the predictor must have nearly 100% positive predictive value (PPV) with respect to the outcome of interest. One could try to generate such predictions using a standard probabilistic prediction technique, and identifying cases where the predicted probability is extreme. For example, a logistic regression model could be trained to predict progression. Patients who are predicted by the model to have a very high probability of progression could be flagged as being almost certain to progress. However such an approach would not give optimal high-predictive-value predictions. The reason for this is illustrated in Figure 57 using data from the "ecoli" dataset [194]. The logistic regression creates a decision boundary by considering the entire dataset: the natural decision boundary created by setting a threshold of 0.5 on the model output can be shifted by increasing this threshold. The shifted boundary would produce predictions with higher PPV, but may unnecessarily decrease the true-positive rate (TPR). The optimal decision boundary may require a completely different decision boundary.

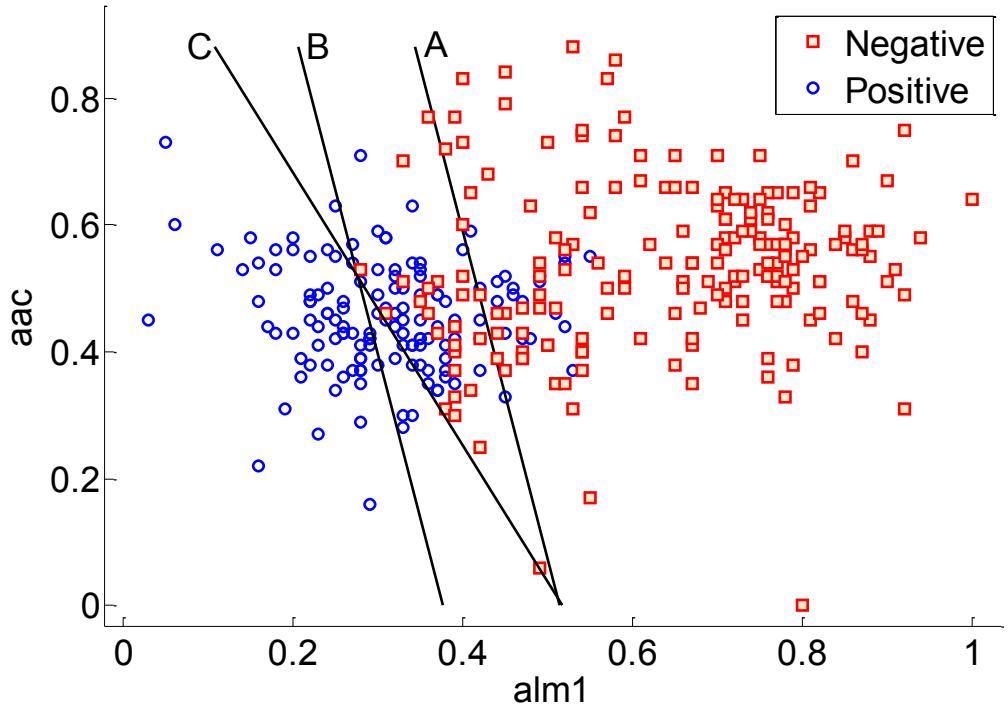


Figure 57: A sample two-class classification problem taken from the "ecoli" dataset. A standard logistic regression produces decision boundary A. Increasing the decision threshold on class probability can maximize PPV by shifting the boundary to B, but can never produce optimal linear boundary C (the boundary with maximum TPR at 100% PPV).

Some work on maximizing the predictive value of predictions has been done by Chalmers et al. [195]. However further developments would be required to confidently identify patients who are very likely to progress despite brace treatment, or very unlikely to progress even without it.

6.3.3. Decision Support Engine

The clinical trial simulation described in section 5.2.3 measured the effect of model-generated brace wear recommendations. The average brace wear recommendation was 12.6 hours per day, which meant an increase in wear time for most patients. The CTS results showed that if patients adhered to the model-generated brace wear recommendations, there would be a decrease in overall risk of progression. These results are, perhaps, not surprising: it stands to reason that an overall increase in brace wear would produce an overall decrease in probability of progression. In general a machine-learning-based decision support system will not give results which seem particularly unusual. A main advantage of these systems may be that they quantify relationships which may have been vague previously.

In the CTS the model's recommendations were designed to produce a 10% reduction in overall probability of progression. This was achieved successfully, but at the expense of some individual's probabilities of progression. CTS is a useful and important technique for evaluating decision support systems like the one described in this thesis. However in the future the recommendation protocol could be made more sophisticated, to decrease group probability of progression with minimal detriment to any individual probability of progression.

An interesting theoretical limitation of the CTS has to do with the notions of causality and confounders. It is possible that the predictor variables which give the best predictions of an event are not actually causes of that event. As a silly example, suppose that patients who are predisposed to be poor brace wearers also tend to have diets high in Progressium – an as yet undiscovered mineral which weakens the human spine. The prediction model used in the CTS predicts probability of progression based on brace wear time, and indicates that lower brace-wear raises the probability of progression. But in fact, the real cause of progression in poor brace wearers is not the lack of brace wear, but rather the intake of so much Progressium. It is important to note that whether or not brace-wear has a causal relationship with progression is irrelevant to the quality of the predictions themselves. However it is relevant to recommendations made using the model, which assume that changes in brace wear can actually change the probability of progression.

Another way of explaining the difficulty is by contrasting the population level with the individual level. The prediction model was trained using data from a population of patients, and describes a relationship between brace wear and progression which was observed in that population. There is no question that this relationship is useful for *predicting* progression in future patients. However we have no guarantee that the population level observation still applies at the individual level. We observe that – in the population – better brace wearers have lower risk of progression. But this does not necessarily guarantee that an *individual* can *change* their risk of progression by changing their brace wear. That is an assumption.

This assumption is implicit in much of our activity. For example, if a patient's brace achieves little correction, their health care provider might recommend that the brace be adjusted to attempt more correction. The health care provider has observed in the population of past patients that higher in-brace correction correlates with treatment success, and so assumes that this patient's chance of success can be improved with a more corrective brace. This is a perfectly reasonable assumption – and adjusting the brace is probably the correct action given the information available. But still there is some chance that the observed success of well-corrected brace patients has nothing to do with the in-brace correction, but rather some other factor such as the spinal curve's flexibility or predisposition to correct in a brace. The upshot of this is that a CTS provides a good initial estimate of the efficacy of computer-generated recommendations, but does not prove anything. Ultimately a real clinical trial would be necessary to evaluate a decision support system's effect on patient outcomes. Moreover, an electronic decision support system's recommendations should always be delivered to a human expert – not trusted blindly.

One final limitation of the decision support engine is in the way it searches for "optimal" treatment parameters. In the case of the logistic regression model described in section 4.3.3, calculating brace-wear recommendations is trivial: the brace-wear giving a specified probability of progression can be easily calculated, and the model's predictions are a function of a single predictor variable and can be easily visualized. However, the platform will accommodate prediction models of any complexity. The genetic algorithm described in section 4.4.4 searches for the treatment parameters which produce a user-specified output from an arbitrary model.

One problem with this approach is that the genetic algorithm finds *one* "optimal" set of treatment parameters – the set that produces a model output closest to the user's specification. But for a given model, there may be a variety of treatment options which would all give the same predicted outcome. In practice it would be valuable to know about all these options. As a simple

example, Figure 58 shows the output of a logistic regression model trained using the brace-wear and in-brace correction variables of the dataset described in section 5.2.1. For any given probability of progression, an infinite number of brace-wear / in-brace correction combinations can be found which produce the probability prediction. These solutions extend along a line whose orientation is defined by the logistic surface.

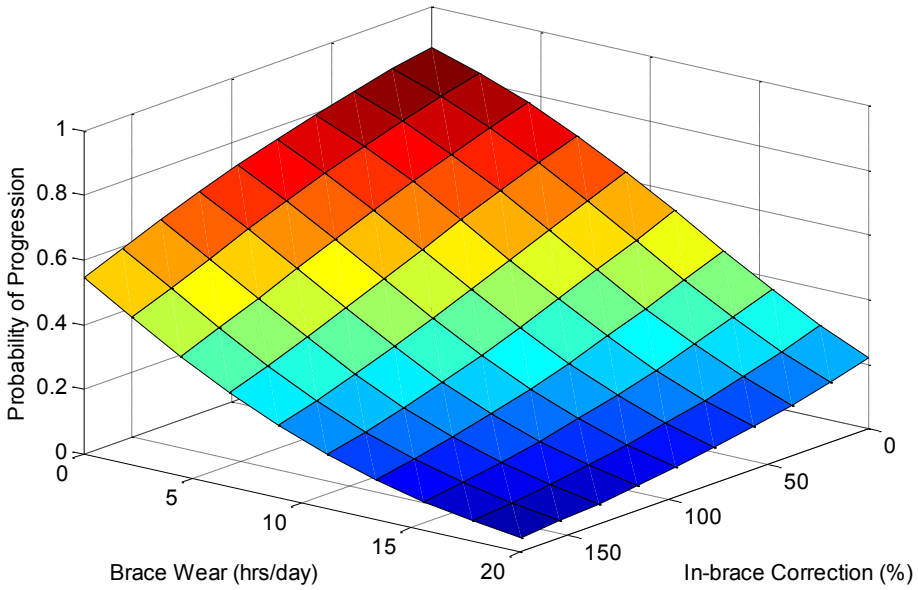


Figure 58: Output of a sample logistic regression model trained using the Brace Wear and In-brace Correction variables from the dataset described in section 5.2.1.

Most optimization techniques are designed to seek a single, global optimum to an objective function. To identify multiple treatment parameter combinations for recommendation would require a novel optimization technique. Inventing such a technique would ultimately make recommendation systems like the one described in this thesis more useful in practice.

6.3.4. A Final Note on "Optimal" Treatment Recommendations

This thesis has discussed the concept of "optimizing" brace treatment for AIS. The grand vision underlying this thesis is that an electronic decision support system could recommend treatment parameters which are optimal in the strongest sense: the recommended parameters would consider all aspects of brace treatment relevant to a specific patient, and give that patient their individual best chance of success. The work presented in this thesis is an important step toward

this ideal, but in its present state does not achieve optimal recommendations in this strong sense.

To truly optimize treatment, the decision support system would need a very rich database of features. If the feature set captured all information relevant to progression, it would allow the best possible predictions of progression and the most meaningful recommendations. As mentioned in section 6.3.1, one important way to improve the database would be to add three-dimensional measurements of Scoliosis. This work has used the two-dimensional Cobb angle as the quantification of Scoliosis severity (as is the standard). The prediction models are affected by the crudeness of Cobb angle as a measurement of severity. The decision support engine tries to optimize Cobb angle progression, unaware of the effect this optimization might have on the holistic, three-dimensional deformity.

Furthermore, the decision support engine cannot account for factors which are not described in the database. The engine optimizes prediction model outputs by tuning the inputs to the prediction model(s). But if there are important treatment parameters which are not represented in the database, they will never be used by the prediction models or considered by the decision support engine's optimization process. For example, aspects of brace design likely play a major role in treatment outcome, but cannot be optimized by the current system because they are not described by the available data.

In the future, investments should be made to create richer, more descriptive, and higher-quality databases. This thesis is a first step toward a system for optimizing brace treatment on an individual level, and its results provide a taste of what might be possible with further development. Creating *truly* optimized treatment protocols will be a difficult but worthwhile challenge.

References

[1] J. Clin, C.-E. Aubin, S. Parent, and H. Labelle, “Biomechanical Modeling of Brace Treatment of Scoliosis: Effects of Gravitational Loads,” *Med. Biol. Eng. Comput.*, vol. 49, no. 7, pp. 743–753, 2011.

[2] M. Wong, A. Mak, K. Luk, J. Evans, and B. Brown, “Effectiveness and biomechanics of spinal orthoses in the treatment of adolescent idiopathic scoliosis (AIS),” *Prosthet. Orthot. Int.*, vol. 24, no. 2, pp. 148–162, Aug. 2000.

[3] E. Lou, D. Hill, E. Parent, J. Raso, M. Moreau, J. Mahood, and D. Hedden, “Prediction of curve progression for AIS patients treated with a TLSO Brace,” *Scoliosis*, vol. 4, no. Suppl+2, pp. O38–O38, 2009.

[4] D. Katz, J. Herring, R. Browne, D. Kelly, and J. Birch, “Brace Wear Control of Curve Progression in Adolescent Idiopathic Scoliosis,” *J. Bone Jt. Surg.*, vol. 92, no. 1, pp. 1343–1352, 2010.

[5] S. L. Weinstein, L. A. Dolan, J. G. Wright, and M. B. Dobbs, “Effects of Bracing in Adolescents with Idiopathic Scoliosis,” *N. Engl. J. Med.*, vol. 369, pp. 1512–1521, Sep. 2013.

[6] D. Rowe, “Idiopathic Scoliosis,” in *Manual of Brace Treament for Idiopathic Scoliosis*, 2003.

[7] S. Negrini, A. Aulisa, L. Aulisa, and A. Circo, “2011 SOSORT guidelines: Orthopaedic and Rehabilitation treatment of idiopathic scoliosis during growth,” *Scoliosis*, vol. 7, no. 1, pp. 3–37, 2012.

[8] J. Y. Margulies, “Reviewer’s comment concerning ‘3D analysis of brace treatment in idiopathic scoliosis’ (ESJO-D-12-00300R" by A. Courvoisier, X. Drevelle, R. Vialle, J. Dubousset and W. Skalli),” *Eur. Spine J.*, vol. 22, no. 11, pp. 2456–2457, Nov. 2013.

[9] T. Greenhalgh, J. Howick, N. Maskrey, and for the Evidence Based Medicine Renaissance Group, “Evidence based medicine: a movement in crisis?,” *BMJ*, vol. 348, no. jun13 4, pp. g3725–g3725, Jun. 2014.

[10] C.-L. Chi, N. W. Street, J. G. Robinson, and M. A. Crawford, “Individualized patient-centered lifestyle recommendations: An expert system for communicating patient specific cardiovascular risk information and prioritizing lifestyle options,” *J. Biomed. Inform.*, vol. 42, no. 1, pp. 1164–1174, Jul. 2012.

[11] C.-L. Chi, N. W. Street, and M. M. Ward, “Building a hospital referral expert system with a Prediction and Optimization-Based Decision Support System algorithm,” *J. Biomed. Inform.*, vol. 41, no. 2, pp. 371–386, 2008.

[12] J. Wyatt and D. Spiegelhalter, “Evaluating medical expert systems: what to test and how?,” *Med. Inform. (Lond.)*, vol. 15, no. 3, pp. 205–217, Jul. 1990.

[13] C.-L. Chi, V. A. Fusaro, P. Patil, M. A. Crawford, C. F. Contant, and P. J. Tonellato, “An approach to optimal individualized warfarin treatment through clinical trial simulations,” in *Biomedical Engineering Conference (CIBEC), 2010 5th Cairo International*, 2010, pp. 116–120.

[14] E. Lou, D. Hill, and J. Raso, “Brace treatment for adolescent idiopathic scoliosis,” *Stud. Health Technol. Inform.*, vol. 135, no. 1, pp. 265–273, 2008.

[15] A. White and M. Panjabi, “The Clinical Biomechanics of Scoliosis,” *Clin. Orthop.*, vol. 118, pp. 100–112, Jul. 1976.

[16] P. Millner and R. Dickson, “Idiopathic Scoliosis: Biomechanics and Biology,” *Eur. Spine J.*, vol. 5, no. 6, pp. 362–73, Dec. 1996.

[17] D. Birchall, D. Hughes, and B. Williamson, “Demonstration of vertebral and disc mechanical torsion in adolescent idiopathic scoliosis using three-dimensional magnetic resonance imaging,” *Scoliosis*, vol. 2, no. Suppl+1, p. S27, 2007.

[18] T. P. C. Schlösser, M. van Stralen, R. C. Brink, W. C. W. Chu, T.-P. Lam, K. L. Vincken, R. M. Castelein, and J. C. Y. Cheng, "Three-Dimensional Characterization of Torsion and Asymmetry of the Intervertebral Discs versus Vertebral Bodies in Adolescent Idiopathic Scoliosis," *Spine*, Jun. 2014.
[19] D. Wever, A. Veldhuizen, J. Klein, P. Webb, G. Nijenbanning, J. Cool, and J. von Horn, "A biomechanical analysis of the vertebral and rib deformities in structural scoliosis," *Eur. Spine J.*, vol. 8, no. 4, pp. 252–260, Aug. 1999.
[20] R. Morrissy, G. Goldsmith, E. Hall, D. Kehl, and G. Cowie, "Measurement of the Cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error," *J. Bone Joint Surg. Am.*, vol. 72, no. 3, pp. 320–327, Mar. 1990.
[21] M. Gstoettner, K. Sekyra, N. Walochnik, P. Winter, R. Wachter, and C. Bach, "Inter- and intraobserver reliability assessment of the Cobb angle: manual versus digital measurement tools," *Eur. Spine J.*, vol. 16, no. 10, pp. 1587–1592, 2007.
[22] S. Swank, R. Winter, and J. Moe, "Scoliosis and cor pulmonale," *Spine*, vol. 7, no. 4, pp. 343–354, Aug. 1982.
[23] M. A. Asher and D. C. Burton, "Adolescent idiopathic scoliosis: natural history and long term treatment effects," *Scoliosis*, vol. 1, pp. 2–10, 2006.
[24] R. G. Burwell, "Aetiology of idiopathic scoliosis: current concepts," *Pediatr. Rehabil.*, vol. 6, no. 3–4, pp. 137–170, Dec. 2003.
[25] D. Katz, "The etiology and natural history of idiopathic scoliosis," *J. Prosthet. Orthot.*, vol. 15, no. 4, pp. S3–S10, Oct. 2003.
[26] F. Zaina, S. Donzelli, M. Lusini, L. Vismara, P. Capodaglio, L. Neri, and S. Negrini, "Adolescent idiopathic scoliosis and eating disorders: Is there a relation? Results of a cross-sectional study," *Res. Dev. Disabil.*, vol. 34, no. 4, pp. 1119–1124, Jan. 2013.
[27] M. Ramirez, J. Martinez-Llorens, J. Sanchez, J. Bago, A. Molina, J. Gea, and E. Caceres, "Body composition in adolescent idiopathic scoliosis," *Eur. Spine J.*, vol. 22, no. 2, pp. 324–329, Feb. 2013.
[28] P. J. M. van Loon, B. A. G. Kühbauch, and F. B. Thunnissen, "Forced lordosis on the thoracolumbar junction can correct coronal plane deformity in adolescents with double major curve pattern idiopathic scoliosis," *Spine*, vol. 33, no. 7, pp. 797–801, Apr. 2008.
[29] Z. Chen, Y. Zhao, Y. Bai, K. Liu, S. He, C. Wang, J. Zhang, and M. Li, "Factors affecting curve flexibility in skeletally immature and mature idiopathic scoliosis," *J. Orthop. Sci.*, vol. 16, no. 2, pp. 133–138, Mar. 2011.
[30] J. Yu, J. Fairbank, S. Roberts, and J. Urban, "The elastic fiber network of the anulus fibrosus of the normal and scoliotic human intervertebral disc," *Spine*, vol. 30, no. 16, pp. 1815–1820, Aug. 2005.
[31] A. Veldhuizen, D. Wever, and P. Webb, "The aetiology of idiopathic scoliosis: biomechanical and neuromuscular factors," *Eur. Spine J.*, vol. 9, no. 3, pp. 178–184, Jun. 2000.
[32] I. Stokes, "Mechanical effects on skeletal growth," *J. Musculoskelet. Neuronal Interact.*, vol. 2, no. 3, pp. 277–280, Mar. 2002.
[33] I. Stokes, "Analysis and simulation of progressive adolescent scoliosis by biomechanical growth modulation," *Scoliosis*, vol. 2, no. Suppl+1, p. P12, 2007.
[34] F. Castro, "Clinical studies: Adolescent idiopathic scoliosis, bracing, and the Hueter-Volkmann principle," *Spine J.*, vol. 3, no. 3, pp. 180–185, 2002.
[35] M. Inoue, S. Minami, H. Kitahara, Y. Otsuka, Y. Nakata, M. Takaso, and H. Moriya, "Idiopathic scoliosis in twins studied by DNA fingerprinting: the incidence and type of scoliosis.," *J. Bone Joint Surg. Br.*, vol. 80, no. 2, pp. 212–217, Mar. 1998.
[36] C. Good, "The Genetic Basis of Adolescent Idiopathic Scoliosis," *J. Spinal Res. Found.*, vol. 4, no. 1, pp. 13–17, 2009.

[37] J. Ogilvie, “Adolescent idiopathic scoliosis and genetic testing,” *Curr. Opin. Pediatr.*, vol. 22, no. 1, pp. 67–70, Feb. 2010.
[38] E. Rogala, D. Drummond, and J. Gurr, “Scoliosis: incidence and natural history. A prospective epidemiological study.,” *J. Bone Jt. Surg.*, vol. 60, no. 2, pp. 173–176, Mar. 1978.
[39] J. Lonstein and M. Carlson, “The Prediction of Curve Progression in Untreated Idiopathic Scoliosis during Growth,” *J. Bone Jt. Surg.*, vol. 66, no. 7, pp. 1061–1071, Sep. 1984.
[40] S. Richards, R. Bernstein, C. D’Amato, and G. Thompson, “Standardization of Criteria for Adolescent Idiopathic Scoliosis Brace Studies,” *Spine*, vol. 30, no. 18, pp. 2068–2075, 2005.
[41] J. Jaremko, S. Delorme, J. Dansereau, H. Labelle, J. Ronsky, P. Poncet, J. A. Harder, R. Dewar, and R. F. Zernicke, “Use of Neural Networks to Correlate Spine and Rib Deformity in Scoliosis,” *Comput. Methods Biomech. Biomed. Engin.*, vol. 3, no. 3, pp. 203–213, 2000.
[42] J. Jaremko, P. Poncet, J. Ronsky, J. A. Harder, J. Dansereau, H. Labelle, and R. F. Zernicke, “Genetic Algorithm—Neural Network Estimation of Cobb Angle from Torso Asymmetry in Scoliosis,” *J. Biomech. Eng.*, vol. 124, no. 5, pp. 496–503, Oct. 2002.
[43] W. Bunnell, “The natural history of idiopathic scoliosis before skeletal maturity.,” *Spine*, vol. 11, no. 8, pp. 773–776, Oct. 1986.
[44] M. Ylikoski, “Growth and progression of adolescent idiopathic scoliosis in girls,” *J. Pediatr. Orthop.*, vol. 14, no. 5, pp. 320–324, Sep. 2005.
[45] L.-E. Peterson and A. Nachemson, “Prediction of Progression of the Curve in Girls Who Have Adolescent Idiopathic Scoliosis of Moderate Severity,” *J. Bone Jt. Surg.*, vol. 77, no. 6, pp. 823–827, Jun. 1995.
[46] M. Ylikoski, J. Peltonen, and M. Poussa, “Biological Factors and Predictability of Bracing in Adolescent Idiopathic Scoliosis,” *J. Pediatr. Orthop.*, vol. 9, no. 6, pp. 680–683, Dec. 1989.
[47] K.-J. Tan, M. M. Moe, R. Vaithinathan, and H.-K. Wong, “Curve Progression in Idiopathic Scoliosis,” *Spine*, vol. 34, no. 7, pp. 697–700, Nov. 2009.
[48] C. F. Lee, D. Fong, K. Cheung, J. Cheng, and B. Ng, “A new risk classification rule for curve progression in adolescent idiopathic scoliosis,” *Spine J.*, vol. 12, no. 11, pp. 989–995, Jun. 2012.
[49] E. M. Clark, H. J. Taylor, I. Harding, J. Hutchinson, I. Nelson, J. E. Deanfield, A. R. Ness, and J. H. Tobias, “Association between components of body composition and scoliosis: a prospective cohort study reporting differences identifiable before the onset of scoliosis,” *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.*, vol. 29, no. 8, pp. 1729–1736, Aug. 2014.
[50] D. Little, K. Song, D. Katz, and J. Herring, “Relationship of peak height velocity to other maturity indicators in idiopathic scoliosis in girls.,” *J. Bone Joint Surg. Am.*, vol. 82A, no. 5, pp. 685–693, May 2000.
[51] S. Weinstein and I. Ponseti, “Curve progression in idiopathic scoliosis,” *J. Bone Joint Surg. Am.*, vol. 65-A, no. 4, pp. 447–455, Apr. 1983.
[52] N. Miller, “Genetics of familial idiopathic scoliosis,” *Clin. Orthop.*, vol. 462, pp. 6–10, Sep. 2007.
[53] F. Escalada, E. Marco, E. Duarte, J. Ma Muniesca, R. Boza, M. Tejero, and E. Caceres, “Assessment of angle velocity in girls with adolescent idiopathic scoliosis,” *Scoliosis*, vol. 4, no. 1, p. 20, 2009.
[54] M. Chazono, S. Soshi, Y. Kida, K. Hashimoto, and T. Inoue, “Height velocity curves in female patients with idiopathic scoliosis,” *Stud. Health Technol. Inform.*, vol. 176, no. 1, pp. 202–205, 2012.
[55] B. Ran, G.-Y. Zhang, F. Shen, J.-Y. Chen, J.-B. Wu, F.-C. Zhao, K.-J. Guo, D.-Y. Qi, B. Zhou, X.-Y. Chen, X.-Z. Zhang, Y.-H. Qiao, and M. Li, “Comparison of the sagittal profiles among

thoracic idiopathic scoliosis patients with different Cobb angles and growth potentials," *J. Orthop. Surg.*, vol. 9, no. 1, p. 19, Mar. 2014.
[56] H.-R. Weiss, S. Negrini, M. Rigo, T. Kotwicki, M. Hawes, T. Grivas, T. Maruyama, and F. Landauer, "Indications for conservative management of scoliosis (guidelines)," *Scoliosis*, vol. 1, no. 1, p. 5.
[57] D. Uyttendaele, P. Dangerfield, E. Lou, D. Hill, J. Raso, J. Mahood, and M. Moreau, "Prediction of Brace Treatment Outcomes by Monitoring Brace Usage," *Stud. Health Technol. Inform.*, vol. 123, no. 1, pp. 239–244, Jun. 2006.
[58] H. Xue, "Support Vector Machines," in *The top ten algorithms in data mining*, Chapman and Hall, 2009, pp. 37–59.
[59] P. Ajemba, L. Ramirez, N. Durdle, D. Hill, and J. Raso, "A Support Vectors Classifier Approach to Predicting the Risk of Progression of Adolescent Idiopathic Scoliosis," *IEEE Trans. Inf. Technol. Biomed.*, vol. 9, no. 2, pp. 276–282, Jun. 2005.
[60] H. Wu, J. Ronsky, P. Poncet, F. Cheriet, D. Xue, J. A. Harder, and R. F. Zernicke, "Prediction of Scoliosis Progression in Time Series Using a Hybrid Learning Technique," in *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, Shanghai, China, 2005, pp. 6452–6455.
[61] H. Wu, J. Ronsky, F. Cheriet, J. Kupper, J. A. Harder, D. Xue, and R. F. Zernicke, "Prediction of scoliosis progression with serial three-dimensional spinal curves and the artificial progression surface technique," *Med. Biol. Eng. Comput.*, vol. 48, no. 1, pp. 1065–1075, Jul. 2010.
[62] D. Steinberg, "CART: Classification and Regression Trees," in *The Top Ten Algorithms in Data Mining*, Chapman and Hall, 2009, pp. 179–201.
[63] G. Shmueli, "To Explain or to Predict?," *Stat. Sci.*, vol. 25, no. 3, pp. 289–310, Aug. 2010.
[64] A.-L. Boulesteix and M. Schmid, "Machine learning versus statistical modeling," *Biom. J.*, vol. 56, no. 4, pp. 588–593, Jul. 2014.
[65] H.-R. Weiss and M. Werkmann, "Soft braces in the treatment of Adolescent Idiopathic Scoliosis (AIS) - Review of the literature and description of a new approach.," *Scoliosis*, vol. 7, no. 1, pp. 11–20, 2012.
[66] F. Zaina, J. C. De Mauroy, T. Grivas, M. T. Hresko, T. Kotwizki, T. Maruyama, N. Price, M. Rigo, L. Stikeleather, J. Wynne, and S. Negrini, "Bracing for scoliosis in 2014: state of the art," *Eur. J. Phys. Rehabil. Med.*, vol. 50, no. 1, pp. 93–110, Feb. 2014.
[67] B. Smith and T. Harrigan, "The Milwaukee Brace," in *Bracing Manual of the Scoliosis Research Society*, 1998.
[68] Boston Brace International, "Reference Manual for the Boston Scoliosis Brace." 2003.
[69] H. Labelle, J. Dansereau, C. Bellefleur, and B. Poitras, "Three-dimensional effect of the Boston brace on the thoracic spine and rib cage," *Spine*, vol. 21, no. 1, pp. 59–64, Jan. 1996.
[70] R. Hooper, F. Reed, and C. Price, "The Charleston Bending Brace," in *Bracing Manual of the Scoliosis Research Society*, 2003.
[71] J. Clin, C. Aubin, S. Parent, and H. Labelle, "A Biomechanical Study of the Charleston Brace for the Treatment of Scoliosis," *Spine*, vol. 35, no. 19, pp. E940–E947, Sep. 2010.
[72] H.-R. Weiss, M. Werkmann, and C. Stephan, "Correction effects of the ScoliOlogiC Chêneau light brace in patients with scoliosis," *Scoliosis*, vol. 2, no. 1, pp. 1–6, 2007.
[73] C. d' Amato, "The Providence Scoliosis System," in *Bracing Manual of the Scoliosis Research Society*, 2003.
[74] T. Kotwicki and J. Cheneau, "Biomechanical action of a corrective brace on thoracic idiopathic scoliosis: Cheneau 2000 orthosis," *Disabil. Rehabil.*, vol. 3, no. 3, pp. 146–153, May 2008.
[75] T. Grivas and A. Kaspiris, "European braces widely used for conservative scoliosis treatment," *Stud. Health Technol. Inform.*, vol. 158, pp. 157–166, 2010.

[76] M. Rigo, S. Negrini, H.-R. Weiss, T. Grivas, T. Maruyama, and T. Kotwicki, "SOSORT consensus paper on brace action: TLSO biomechanics of correction (investigating the rationale for force vector selection)," *Scoliosis*, vol. 1, no. 1, p. 11, 2006.
[77] A. Courvoisier, X. Drevelle, R. Vialle, J. Dubousset, and W. Skalli, "3D analysis of brace treatment in idiopathic scoliosis," *Eur. Spine J.*, vol. 22, no. 11, pp. 2449–2455, Nov. 2013.
[78] J. Kennedy, C. Robertson, A. Olinsky, D. Dickens, and P. Phelan, "Pulmonary restrictive effect of bracing in mild idiopathic scoliosis," *Thorax*, vol. 42, no. 12, pp. 959–961, Dec. 1987.
[79] G. Katsaris, A. Loukos, J. Valavanis, M. Vassiliou, and P. Behrakis, "The immediate effect of a Boston brace on lung volumes and pulmonary compliance in mild adolescent idiopathic scoliosis," *Eur. Spine J.*, vol. 8, no. 1, pp. 2–7, Feb. 1999.
[80] S. Upadhyay, I. Nelson, E. Ho, L. Hsu, and J. Leong, "New Prognostic Factors to Predict the Final Outcome of Brace Treatment in Adolescent Idiopathic Scoliosis," *Spine*, vol. 20, no. 5, pp. 537–545, 1995.
[81] D. Katz and A. Durrani, "Factors That Influence Outcome in Bracing Large Curves in Patients With Adolescent Idiopathic Scoliosis," *Spine*, vol. 26, no. 21, pp. 2354–2361, Nov. 2001.
[82] D. Perie, J. Sales De Gauzy, and M. Ho Ba Tho, "Biomechanical evaluation of Cheneau-Toulouse-Munster brace in the treatment of scoliosis using optimisation approach and finite element method," *Med. Biol. Eng. Comput.*, vol. 40, no. 3, pp. 296–301, May 2002.
[83] X. Drevelle, Y. Lafon, E. Ebermeyer, I. Courtois, J. Dubousset, and W. Skalli, "Analysis of Idiopathic Scoliosis Progression by Using Numerical Simulation," *Spine*, vol. 35, no. 10, pp. E407–E412, May 2010.
[84] J. Clin, C. Aubin, A. Sangole, H. Labelle, and S. Parent, "Correlation Between Immediate In-Brace Correction and Biomechanical Effectiveness of Brace Treatment in Adolescent Idiopathic Scoliosis," *Spine*, vol. 35, no. 18, pp. 1706–1713, Aug. 2010.
[85] M. Płaszewski and J. Bettany-Saltikov, "Non-surgical interventions for adolescents with idiopathic scoliosis: an overview of systematic reviews," *PloS One*, vol. 9, no. 10, p. e110254, 2014.
[86] D. Rowe, S. Bernstein, M. Riddick, F. Adler, J. Emans, and D. Gardner-Bonneau, "A Meta-Analysis of the Efficacy of Non-Operative Treatments for Idiopathic Scoliosis," *J. Bone Jt. Surg.*, vol. 80, no. 6, pp. 923–925, Jun. 1998.
[87] L. Dolan and S. Weinstein, "Surgical Rates After Observation and Bracing for Adolescent Idiopathic Scoliosis," *Spine*, vol. 32, no. 19S, pp. S91–S100, 2007.
[88] S. Negrini, S. Minozzi, J. Bettany-Saltikov, F. Zaina, N. Chockalingam, T. Grivas, T. Kotwicki, T. Maruyama, M. Romano, and E. Vasiliadis, "Braces for Idiopathic Scoliosis in Adolescents," *Spine*, vol. 35, no. 13, pp. 1285–1293, 2010.
[89] A. Morton, R. Riddle, R. Buchanan, D. Katz, and J. Birch, "Accuracy in the prediction and estimation of adherence to bracewear before and during treatment of adolescent idiopathic scoliosis.," *J. Pediatr. Orthop.*, vol. 28, no. 3, pp. 336–341, Apr. 2008.
[90] F. Landauer, C. Wimmer, and H. Behensky, "Estimating the final outcome of brace treatment for idiopathic thoracic scoliosis at 6-month follow-up," *Pediatr. Rehabil.*, vol. 6, no. 3–4, pp. 201–207, 2003.
[91] J. O. Sanders, P. O. Newton, R. H. Browne, D. E. Katz, J. G. Birch, and J. A. Herring, "Bracing for idiopathic scoliosis: how many patients require treatment to prevent one surgery?," *J. Bone Joint Surg. Am.*, vol. 96, no. 8, pp. 649–653, Apr. 2014.
[92] J. Clin, C. Aubin, S. Parent, A. Sangole, and H. Labelle, "Comparison of the biomechanical 3D efficiency of different brace designs for the treatment of scoliosis using a finite element model," *Eur. Spine J.*, vol. 19, no. 7, pp. 1169–1178, Jul. 2010.

[93] W. Nie, M. Ye, Z. Liu, and C. Wang, "The Patient-Specific Brace Design and Biomechanical Analysis of Adolescent Idiopathic Scoliosis," *J. Biomech. Eng.*, vol. 131, no. 4, p. 8, Apr. 2009.
[94] J. Berteau, M. Pithioux, S. Mesure, G. Bollini, and P. Chabrand, "Beyond the classic correction system: a numerical nonrigid approach to the scoliosis brace," *Spine J.*, vol. 11, no. 5, pp. 424–431, May 2011.
[95] D. Gignac, C. Aubin, J. Dansereau, and H. Labelle, "Optimization method for 3D bracing correction of scoliosis using a finite element model," *Eur. Spine J.*, vol. 9, no. 3, pp. 185–190, Jun. 2000.
[96] F.-H. Cheng, S.-L. Shih, W.-K. Chou, C.-L. Liu, W.-H. Sung, and C.-S. Chen, "Finite element analysis of the scoliotic spine under different loading conditions," *Biomed. Mater. Eng.*, vol. 20, no. 5, pp. 251–259, 2010.
[97] F. Desbiens-Blais, J. Clin, S. Parent, H. Labelle, and C.-E. Aubin, "New brace design combining CAD/CAM and biomechanical simulation for the treatment of adolescent idiopathic scoliosis," *Clin. Biomech.*, vol. 27, no. 10, pp. 995–1005, Dec. 2012.
[98] C.-E. Aubin, N. Cobetto, J. Clin, F. Desbiens-Blais, H. Labelle, S. Le May, and S. Parent, "Improved Brace Design Combining CAD/CAM and Finite Element Simulation for the Conservative Treatment of Adolescent Idiopathic Scoliosis (AIS): Preliminary Results of a Randomized Control Trial," in *Proceedings of The 10th Meeting of the International Research Society of Spinal Deformities*, Sapporo, Japan, 2014, p. 80.
[99] S. Vandal, C. Rivard, and R. Bradet, "Measuring the compliance behavior of adolescents wearing orthopedic braces," *Issues Compr. Pediatr. Nurs.*, vol. 22, no. 2–3, pp. 59–73, Apr. 1999.
[100] R. Havey, T. Gavin, A. Patwardhan, S. Pawelczak, K. Ibrahim, G. Andersson, and S. Lavender, "A reliable and accurate method for measuring orthosis wearing time," *Spine*, vol. 27, no. 2, pp. 211–214, Jan. 2002.
[101] E. Lou, J. Raso, D. Hill, J. Mahood, and M. Moreau, "Correlation between quantity and quality of orthosis wear and treatment outcomes in adolescent idiopathic scoliosis," *Prosthet. Orthot. Int.*, vol. 28, no. 1, pp. 49–54, Apr. 2004.
[102] E. Lou, D. Hill, D. Hedden, J. Mahood, M. Moreau, and J. Raso, "An objective measurement of brace usage for the treatment of adolescent idiopathic scoliosis," *Med. Eng. Phys.*, vol. 33, no. 3, pp. 290–294, 2011.
[103] T. Rahman, J. Bowen, M. Takemitsu, and C. Scott, "The Association Between Brace Compliance and Outcome for Patients With Idiopathic Scoliosis," *J. Pediatr. Orthop.*, vol. 25, no. 4, pp. 420–422, Aug. 2005.
[104] G. Nicholson, "The Objective Measurement of Spinal Orthosis Use for Treatment of AIS," *Spine*, vol. 28, no. 19, pp. 2243–2250, Oct. 2003.
[105] M. Takemitsu, J. Bowen, T. Rahman, J. Glutting, and C. Scott, "Compliance monitoring of brace treatment for patients with idiopathic scoliosis," *Spine*, vol. 29, no. 18, pp. 2070–2074, Sep. 2004.
[106] A. Helfenstein, M. Lankes, K. Ohlert, D. Varoga, H. Hahne, H. Ulrich, and J. Hassenpflung, "The objective determination of compliance in treatment of adolescent idiopathic scoliosis with spinal orthoses," *Spine*, vol. 31, no. 3, pp. 339–344, Feb. 2006.
[107] C. C. Hasler, S. Wietlisbach, and P. Büchler, "Objective compliance of adolescent girls with idiopathic scoliosis in a dynamic SpineCor brace," *J. Child. Orthop.*, vol. 4, no. 3, pp. 211–218, Jun. 2010.
[108] D. Miller, J. Franzone, H. Matsumoto, J. Gomez, and J. Avendano, "Electronic Monitoring Improves Brace-Wearing Compliance in Patients With Adolescent Idiopathic Scoliosis," *Spine*, vol. 37, no. 9, pp. 717–721, 2012.

[109] E. Lou, D. Hill, J. Raso, M. Moreau, and J. Mahood, “Smart orthosis for the treatment of adolescent idiopathic scoliosis,” *Med. Biol. Eng. Comput.*, vol. 43, no. 6, pp. 746–750, Nov. 2005.
[110] C. Aubin, H. Labelle, Y. Petit, D. Gignac, J. Joncas, and J. Dansereau, “Variability of strap tension in brace treatment for adolescent idiopathic scoliosis,” *Spine*, vol. 24, no. 4, pp. 349–354, Feb. 1999.
[111] E. Chalmers, E. Lou, D. Hill, V. Zhao, and W. MS, “Development of a Pressure Control System for Brace Treatment of Scoliosis,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 557–563, Jul. 2012.
[112] T. Nutley and H. W. Reynolds, “Improving the use of health data for health system strengthening,” *Glob. Health Action*, vol. 6, no. 0, p. 1, 2013.
[113] S. E. Straus, J. Tetroe, and I. Graham, “Defining knowledge translation,” *Can. Med. Assoc. J.*, vol. 181, no. 3/4, pp. 165–168, Aug. 2009.
[114] L. Perreault and J. Metzger, “A pragmatic framework for understanding clinical decision support,” *J. Healthc. Inf. Manag.*, vol. 13, no. 2, pp. 5–22, 1999.
[115] J. Wyatt and D. Spiegelhalter, “Field trials of medical decision-aids: potential problems and solutions.,” in *Proceedings of the Annual Symposium on Computer Application in Medical Care*, Washington, DC, 1991.
[116] A. Rodriguez-Gonzalez, J. Torres-Nino, M. Mayer, G. Alor-Hernandez, and M. Wilkinson, “Analysis of a multilevel diagnosis decision support system and its implications: a case study,” *Comput. Math. Methods Med.*, vol. 2012, pp. 1–9, 2012.
[117] A. Garg, N. Adhikari, H. McDonald, P. Rosas-Arellano, P. Devereaux, J. Beyene, J. Sam, and B. Haynes, “Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review,” *J. Am. Med. Assoc.*, vol. 293, no. 10, pp. 1223–1238, Mar. 2005.
[118] K. Kuwamoto, C. Houlihan, A. Balas, and D. Lobach, “Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success,” *Br. Med. J.*, vol. 330, no. 7494, pp. 765–768, Apr. 2005.
[119] D. Lobach, G. Sanders, T. Bright, A. Wong, and R. Dhurjati, “Enabling health care decisionmaking through clinical decision support and knowledge management.,” *Evid. ReportTechnology Assess.*, vol. 203, pp. 1–784, Apr. 2012.
[120] P. Lisboa and A. Taktak, “The use of artificial neural networks in decision support in cancer: A systematic review,” *Neural Netw.*, vol. 19, no. 4, pp. 408–415, May 2006.
[121] E. Berner and T. Lande, “Overview of Clinical Decision Support Systems,” in *Clinical Decision Support Systems: Theory and Practice*, 2nd ed., New York, NY: Springer Science + Business Media, 2007, pp. 3–22.
[122] E. Shortliffe, R. Davis, S. Axline, B. Buchanan, C. Green, and S. Cohen, “Computer-based consultations in clinical therapeutics: Explanation and rule acquition capabilities of the MYCIN system,” *Comput. Biomed. Res.*, vol. 8, no. 4, pp. 303–320, 1975.
[123] P. Clercq, J. Blom, H. Korsten, and A. Hasman, “Approaches for creating computer-interpretable guidelines that facilitate decision support,” *Artif. Intell. Med.*, vol. 31, no. 1, pp. 1–27, May 2004.
[124] H. Doornewaard, Y. van der Schouw, Y. van der Graaf, A. Bos, J. Habbema, and J. van den Tweel, “The diagnostic value of computer-assisted primary cervical smear screening: a longitudinal cohort study,” *Mod. Pathol.*, vol. 12, no. 11, pp. 995–1000, Nov. 1999.
[125] S. E. Dilsizian and E. L. Siegel, “Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment,” *Curr. Cardiol. Rep.*, vol. 16, no. 1, p. 441, Jan. 2014.
[126] C.-L. Chi, “Medical decision support systems based on machine learning,” PhD, University of Iowa, Iowa, USA, 2009.

[127] M. McNabb, Y. Cao, T. Devlin, B. Baxter, and A. Thornton, “Measuring MERCI: Exploring Data Mining Techniques for Examining the Neurologic Outcomes of Stroke Patients Undergoing Endo-vascular Therapy at Erlanger Southeast Stroke Center,” in *34th Annual International Conference of the IEEE EMBS*, San Diego, California, 2012, pp. 4704–4707.
[128] A. Basu, “Five pillars of prescriptive analytics success,” *Anal. Mag.*, pp. 8–12, 2013.
[129] P. Haas, P. P. Maglio, P. G. Selinger, and W.-C. Tan, “Data is Dead... Without What-if Models,” in *PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES*, Seattle, Washington, 2011, vol. 2, pp. 1486–1489.
[130] L. C.-K. Liau, T. C.-K. Yang, and M.-T. Tsai, “Expert system of a crude oil distillation unit for process optimization using neural networks,” *Expert Syst. Appl.*, vol. 26, no. 2, pp. 247–255, Feb. 2004.
[131] Z. Song and A. Kusiak, “Optimization of Temporal Processes: A Model Predictive Control Approach,” *IEEE Trans. Evol. Comput.*, vol. 13, no. 1, pp. 169–179, Feb. 2009.
[132] I. Buchan, J. Winn, and C. Bishop, “A unified modeling approach to data-intensive healthcare,” in *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, 2009, pp. 91–95.
[133] D. Hudson and M. Cohen, “Overcoming Barriers to Development of Cooperative Medical Decision Support Models,” in *34th Annual International Conference of the IEEE EMBS*, San Diego, California, 2012, pp. 2194–2197.
[134] D. Bates, G. Kuperman, S. Wang, T. Gandhi, and A. Kittler, “Ten Commandments for Effective Clinical Decision Support: Making the Practice of Evidence-based Medicine a Reality,” *J. Am. Med. Inform. Assoc.*, vol. 10, no. 6, pp. 523–530, Nov. 2003.
[135] T. Shortliffe, “Medical Thinking: What Should We Do?,” presented at the Medical Thinking: what do we know? A Review Meeting, London, 22-Jun-2006.
[136] J. Feblowitz, S. Henkin, J. Pang, H. Ramelson, L. Schneider, F. L. Maloney, A. R. Wilcox, D. W. Bates, and A. Wright, “Provider use of and attitudes towards an active clinical alert: a case study in decision support,” *Appl. Clin. Inform.*, vol. 4, no. 1, pp. 144–152, 2013.
[137] J. Hsiao, W. Wu, and R. Chen, “Factors of accepting pain management decision support systems by nurse anesthetists,” *BMC Med. Inform. Decis. Mak.*, vol. 13, no. 1, p. 16, Jan. 2013.
[138] C.-H. Lu, J. Hsiao, and R.-F. Chen, “Factors Determining Nurse Acceptance of Hospital Information Systems,” *Comput. Inform. Nurs.*, vol. 30, no. 5, pp. 257–264, May 2012.
[139] R. Castillo and A. Kelemen, “Considerations for a Successful Clinical Decision Support System,” *Comput. Inform. Nurs.*
[140] R. Miller, “Medical diagnostic decision support systems -- past, present, and future: a threaded bibliography and brief commentary,” *J. Am. Med. Inform. Assoc.*, vol. 1, no. 1, pp. 8–27, Jan. 1994.
[141] K. Cios, “A knowledge discovery approach to diagnosing myocardial perfusion,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 19, no. 4, pp. 17–25, Aug-2000.
[142] C. Shearer, “The CRISP-DM model: the new blueprint for data mining,” *J. Data Warehous.*, vol. 15, no. 4, pp. 13–19, 2000.
[143] L. Kurgan and P. Musilek, “A survey of knowledge discovery and data mining process models,” *Knowl. Eng. Rev.*, vol. 21, no. 1, pp. 1–24, Mar. 2006.
[144] J. Shlens, “A tutorial on principal component analysis,” *Syst. Neurobiol. Lab. Univ. Calif. San Diego*, vol. 82, 2005.
[145] M. Hall, “Correlation-based Feature Selection for Machine Learning,” Ph.D., University of Waikato, Waikato, 1999.
[146] R. Kohavi and G. John, “Wrappers for feature subset selection,” *Artif. Intell.*, vol. 97, no. 1–2, pp. 273–324, Dec. 1997.

[147] R. Shouval, O. Bondi, H. Mishan, A. Shimoni, R. Unger, and A. Nagler, "Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT," *Bone Marrow Transplant.*, vol. 49, no. 3, pp. 332–337, Mar. 2014.
[148] N. Ramakrishnan, "C4.5," in *The Top Ten Algorithms in Data Mining*, Chapman and Hall, 2009, pp. 1–19.
[149] W. Pedrycz and F. Gomide, *Fuzzy Systems Engineering. Toward Human-Centric Computing*, 1st ed. New Jersey: Wiley, 2007.
[150] A. Engelbrecht, *Computational Intelligence: An Introduction*, 2nd ed. John Wiley & Sons, 2007.
[151] R. Nuzzo, "Scientific method: Statistical errors," *Nature*, vol. 506, no. 7487, pp. 150–152, Feb. 2014.
[152] L. Breiman, "Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author)," *Stat. Sci.*, vol. 16, no. 3, pp. 199–231, Aug. 2001.
[153] P. Phan, N. Mezghani, C.-E. Aubin, J. Guise, and H. Labelle, "Computer algorithms and applications used to assist the evaluation and treatment of adolescent idiopathic scoliosis: a review of published articles 2000-2009," *Eur. Spine J.*, vol. 20, no. 7, pp. 1058–1068, Jul. 2011.
[154] L. Ramirez, N. Durdle, J. Raso, and D. Hill, "A Support Vector Machines Classifier to Assess the Severity of Idiopathic Scoliosis From Surface Topography," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 84–91, Jan. 2006.
[155] F. Tang, L. Chan, H. Lau, P. Tsui, and C. Cheung, "Computer-Generated Index for Evaluation of Idiopathic Scoliosis in Digital Chest Images: A Comparison with Digital Measurement," *J. Digit. Imaging*, vol. Supplement 1, pp. S113–S120, Nov. 2008.
[156] M. Adankon, "Non invasive classification system of scoliosis curve types using least-squares support vector machines," *Artif. Intell. Med.*, vol. 56, no. 2, pp. 99–107, Oct. 2012.
[157] N. Mezghani, R. Chav, L. Humbert, S. Parent, W. Skalli, and J. de Guise, "A computer-based classifier of three-dimensional spinal scoliosis severity," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 3, no. 1–2, pp. 55–60, Jun. 2008.
[158] A. Sangole, C.-E. Aubin, H. Labelle, I. Stokes, L. Lenke, R. Jackson, and P. Newton, "Three-dimensional classification of thoracic scoliotic curves," *Spine*, vol. 34, no. 1, pp. 91–99, Jan. 2009.
[159] I. Stokes, A. Sangole, and C.-E. Aubin, "Classification of scoliosis deformity three-dimensional spinal shape by cluster analysis," *Spine*, vol. 34, no. 6, pp. 584–590, Mar. 2009.
[160] L. Duong, F. Cheriet, and H. Labelle, "Three-Dimensional Classification of Spinal Deformities Using Fuzzy Clustering," *Spine*, vol. 31, no. 8, pp. 923–930, Apr. 2006.
[161] P. Phan, N. Mezghani, E. K. Wai, J. de Guise, and H. Labelle, "Artificial neural networks assessing adolescent idiopathic scoliosis: comparison with Lenke classification," *Spine J.*, vol. 13, no. 11, pp. 1527–1533, Nov. 2013.
[162] D. G. Kang, R. A. Lehman Jr., and L. G. Lenke, "Challenges in the classification of adolescent idiopathic scoliosis and the utility of artificial neural networks," *Spine J.*, vol. 13, no. 11, pp. 1534–1537, Nov. 2013.
[163] M.-L. Nault, H. Labelle, C.-E. Aubin, A. Sangole, and M. Balazinski, "Fuzzy-logic–assisted Surgical Planning in Adolescent Idiopathic Scoliosis," *J. Spinal Disord. Tech.*, vol. 22, no. 4, pp. 263–269, Jun. 2009.
[164] G. Batista and M. Monard, "An Analysis of Four Missing Data Treatment Methods for Supervised Learning," *Appl. Artif. Intell.*, vol. 17, no. 1, pp. 519–533, 2003.
[165] J. Kruppa, Y. Liu, H.-C. Diener, T. Holste, C. Weimar, I. R. König, and A. Ziegler, "Probability estimation with machine learning methods for dichotomous and multicategory outcome: Applications," *Biom. J.*, vol. 56, no. 4, pp. 564–583, Feb. 2014.

[166] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” presented at the International Joint Conference on Artificial Intelligence, 1995, vol. 14, pp. 1137–1145.
[167] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition*, 2nd ed. Morgan Kaufmann, 2005.
[168] M.-L. Nault, J. Mac-Thiong, M. Roy-Beaudry, J. Deguise, H. Labelle, and S. Parent, “Three-dimensional Spine Parameters Can Differentiate Between Progressive and Nonprogressive Patients With AIS at the Initial Visit: A Retrospective Analysis,” *J. Pediatr. Orthop.*, vol. 33, no. 6, pp. 618–623, Jun. 2013.
[169] M.-L. Nault, J.-M. Mac-Thiong, M. Roy-Beaudry, I. Turgeon, J. deGuise, H. Labelle, and S. Parent, “Three-Dimensional Spinal Morphology Can Differentiate Between Progressive and Nonprogressive Patients With Adolescent Idiopathic Scoliosis at the Initial Presentation: A Prospective Study. [Miscellaneous Article],” *Spine May 01 2014*, vol. 39, no. 10, 2014.
[170] J. Ghosh and A. Liu, “K-Means,” in *The Top Ten Algorithms in Data Mining*, Chapman and Hall, 2009, pp. 21–35.
[171] W. Pedrycz, “Conditional fuzzy C-Means,” *Pattern Recognit. Lett.*, vol. 17, no. 6, pp. 625–631, May 1996.
[172] A. Uyar, A. Bener, and H. N. Ciray, “Predictive Modeling of Implantation Outcome in an In Vitro Fertilization Setting: An Application of Machine Learning Methods,” *Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak.*, May 2014.
[173] K. J. Farion, S. Wilk, W. Michalowski, D. O’Sullivan, and J. Sayyad-Shirabad, “Comparing Predictions Made by a Prediction Model, Clinical Score, and Physicians: Pediatric asthma exacerbations in the emergency department.,” *Appl. Clin. Inform.*, vol. 4, no. 3, pp. 376–391, 2013.
[174] P. Cornu, S. Steurbaut, S. Soštarić, A. Mrhar, and A. G. Dupont, “Performance of a clinical decision support system and of clinical pharmacists in preventing drug-drug interactions on a geriatric ward,” *Int. J. Clin. Pharm.*, vol. 36, no. 3, pp. 519–525, Jun. 2014.
[175] G. Gigerenzer, W. Gaissmaier, E. Kurz-Milcke, L. M. Schwartz, and S. Woloshin, “Helping Doctors and Patients Make Sense of Health Statistics,” *Psychol. Sci. Public Interest Wiley-Blackwell*, vol. 8, no. 2, pp. 53–96, Nov. 2007.
[176] N. Silver, *The Signal and the Noise: Why So Many Predictions Fail-but Some Don’t*. Penguin, 2012.
[177] K. R. Evans, E. Lou, and G. Faulkner, “Optimization of a Low-Cost Force Sensor for Spinal Orthosis Applications,” *IEEE Trans. Instrum. Meas.*, vol. 62, no. 12, pp. 3243–3250, 2013.
[178] R. G. Lyons, *Understanding Digital Signal Processing*, 2nd ed. New Jersey: Prentice Hall, 2004.
[179] H.-P. Chan, B. Sahiner, and L. Hadjiiski, “Sample size and validation issues on the development of CAD systems,” *Int. Congr. Ser.*, vol. 1268, pp. 872–877, Jun. 2004.
[180] E. W. Steyerberg, T. van der Ploeg, and B. Van Calster, “Risk prediction with machine learning and regression methods,” *Biom. J.*, vol. 54, no. 4, pp. 601–606, Feb. 2014.
[181] W.-C. Tan, P. J. Haas, R. L. Mak, C. A. Kieliszewski, P. G. Selinger, P. P. Maglio, S. Glissman, M. Cefkin, and Y. Li, “Splash: A Platform for Analysis and Simulation of Health,” in *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, Miami, FL, 2012, pp. 543–552.
[182] T. Rahman, B. Borkhuu, A. Littleton, W. Sample, E. Moran, S. Campbell, K. Rogers, and J. Bowen, “Electronic monitoring of scoliosis brace wear compliance,” *J. Child. Orthop.*, vol. 4, no. 4, pp. 343–347, Aug. 2010.
[183] J. Washtell, “3 Must-Ask Questions Before Choosing That Machine Learning Algorithm!” [Online]. Available: http://www.analyticbridge.com/profiles/blogs/wait-why-are-you-using-that-algorithm?xg_source=activity. [Accessed: 05-Jan-2015].

[184] 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,” *J. Clin. Epidemiol.*, vol. 49, no. 12, pp. 1373–1379, 1996.
[185] E. Vittinghoff and C. E. McCulloch, “Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression,” *Am. J. Epidemiol.*, vol. 165, no. 6, pp. 710–718, Mar. 2007.
[186] B. Sahiner, H.-P. Chan, N. Petrick, R. F. Wagner, and L. Hadjiiski, “Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size,” *Med. Phys.*, vol. 27, no. 7, pp. 1509–1522, Jul. 2000.
[187]M. Molla, M. Waddell, D. Page, and J. Shavlik, “Using Machine Learning to Design and Interpret Gene-Expression Microarrays,” *AI Mag.*, vol. 25, no. 1, p. 23, Mar. 2004.
[188] J. Fan, J. Lv, and L. Qi, “Sparse high dimensional models in economics,” *Annu. Rev. Econ.*, vol. 3, p. 291, 2011.
[189] S. J. Raudys and A. K. Jain, “Small sample size effects in statistical pattern recognition: Recommendations for practitioners,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 3, pp. 252–264, 1991.
[190] R. Nisbet, J. F. Elder, and G. Miner, *Handbook of statistical analysis and data mining applications [electronic resource] / Robert Nisbet, John Elder, Gary Miner*. Burlington, Mass. : Academic Press, c2009 (Norwood, Mass. : Books24x7.com [generator]), 2009.
[191]E. Roelofs, A. Dekker, E. Meldolesi, R. G. P. M. van Stiphout, V. Valentini, and P. Lambin, “International data-sharing for radiotherapy research: an open-source based infrastructure for multicentric clinical data mining,” *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol.*, vol. 110, no. 2, pp. 370–374, Feb. 2014.
[192] A. Courvoisier, R. Vialle, and W. Skalli, “EOS 3D Imaging: assessing the impact of brace treatment in adolescent idiopathic scoliosis,” *Expert Rev. Med. Devices*, vol. 11, no. 1, pp. 1–3, Jan. 2014.
[193] A. Courvoisier, X. Drevelle, J. Dubousset, and W. Skalli, “Transverse plane 3D analysis of mild scoliosis,” *Eur. Spine J.*, vol. 22, no. 11, pp. 2427–2432, Nov. 2013.
[194] K. Bache and M. Lichman, “UCI Machine Learning Repository [http://archive.ics.uci.edu/ml].” University of California, School of Information and Computer Science.
[195]E. Chalmers, M. Mizianty, E. Parent, Y. Yuan, and E. Lou, “Toward maximum-predictive-value classification,” *Pattern Recognit.*, vol. 47, no. 12, pp. 3949–3958, Dec. 2014.

Appendix A: Creating Visualization Modules

Visualization modules are XML-style documents with a .vis file extension. The following table describes the elements allowed in the document. Mandatory elements are marked with an asterisk (*). Where several values are allowed for an element's content, the default value is enclosed in brackets.

Element	**Sub-Element**	**Description**
name*		The name of this module as displayed to the user. Ideally this will match the filename.
description*		A description to be displayed to the user. Should contain all information needed to understand and use the module.
input		The name of an input to be entered by the user. This input can be referred to elsewhere in the document. Up to 4 'input' elements can exist in a visualization module.
sql*		The SQL query used to extract the data for charting
series*		Each column returned by the SQL query will be plotted as a separate series. Use 'series' elements and corresponding sub-elements to specify how each should be displayed.
	seriesName*	The series name to be displayed in the chart legend
	chartType	One of the following options: [column], point, line, stackedColumn, polar, radar
	dataLabels	"on" to turn on data labels for this series
	xDataType	One of the following options: [double], or dateTime
	yDataType	One of the following options: [double], or dateTime
	xRange	Minimum and maximum x-axis values, separated by a comma
	yRange	Minimum and maximum y-axis values, separated by a comma
xLabel		X-axis label to be displayed on chart
yLabel		Y-axis label to be displayed on chart
xTick		One of the following options: off, or a positive value specifying the x-axis tick interval. If this element is unused the tick interval will be chosen automatically.
yTick		One of the following options: off, or a positive value specifying the y-axis tick interval. If this element is unused the tick interval will be chosen automatically.
grid		Specifies gridlines to be applied to the chart. One of the following options: [off], x, y, xy

Below is example code for a visualization module which creates a generic scatterplot of Cobb angle progression per year versus a user-specified field from the ExamTable table in the database (refer to section 4.3.2 and Figure 37).

```
<?xml version="1.0" encoding="utf-8"?>
<visualization>
	<name> Scatter with Progression </name>
	<description>
	Scatter plot of a user defined field with Cobb angle progression per year. The
	field can be any examTable field (e.g. 'MaxCobb', 'Scoliometer', or 'Age').
	</description>
	<input> userField </input>

	<sql>
		SELECT startExam.userField, (endExam.MaxCobb - startExam.MaxCobb)/
		((endExam.ExamDate - startExam.ExamDate)/365) as progressionPerYr
		FROM
		(
			(SELECT PatientName, MIN(ExamDate) as minDate,
				MAX(ExamDate) as maxDate
			FROM ExamTable WHERE InBrace = false
			GROUP BY PatientName
			) as dateRanges
			INNER JOIN ExamTable startExam
			ON dateRanges.PatientName = startExam.PatientName
				AND dateRanges.minDate = startExam.ExamDate
		)
		Inner JOIN ExamTable endExam
		ON dateRanges.PatientName = endExam.PatientName
			AND dateRanges.maxDate = endExam.ExamDate
		WHERE endExam.ExamDate - startExam.ExamDate > 182
			AND startExam.userField IS NOT NULL;
	</sql>
	<series>
		<seriesName> Progression vs userField </seriesName>
		<chartType> point </chartType>
	</series>
	<yLabel> Progression Per Year </yLabel>
	<xLabel> userField </xLabel>
</visualization>
```

Appendix B: Creating Modelling Modules

A modelling module consists of three files initially, with at least one more being generated by the module during the training process. The module should be located in the software platform's "AppData" folder. The three files supplied by the user are:

1. **The model training executable**

This executable should enact the model training process using the training data supplied by the software platform. This training data will be in a comma-separated-value file named "trainingData.csv".

When the training process is complete it should create (in the folder where it resides) an XML document with a .tnd file extension. It may save additional data as needed. The .tnd file represents the trained model to the software platform. It contains elements listed in the following table, with mandatory elements marked by an asterisk (*):

Element	Attribute	Description
name*		The name of the trained model, to be displayed to the user
predictExeName*		The name of the prediction executable (see the following section)
date*		The date on which the training process took place
description*		A description of the trained model, to be displayed to the user.
controllableVariable		The file should contain one controllableVariable element for each controllable variable used by the prediction model. This element's content is the variable's name.
	min	The minimum allowed value for this variable
	max	The maximum allowed value for this variable
uncontrollableVariable		The file should contain one uncontrollableVariable element for each uncontrollable variable used by the prediction model. This element's content is the variable's name.
outputType*		One of the following options: console, file. Specifies whether the prediction executable returns predictions in the console window or as a separate file (see the following section).

The following example .tnd file was generated by the logistic regression modelling module used in this thesis.

```
<?xml version="1.0" encoding="utf-8"?>
<trainedModel>
	<name>Logistic Regression Model</name>
	<predictExeName>LogisticRegressionModelPredict.exe</predictExeName>
	<date>03-Dec-2014</date>

	<description>
		This model predicts probability of progression (any Cobb angle increase)
		based on in-brace correction and compliance.

		Coefficients:
		Correction: -0.47314
		Compliance: -2.0153
		Intercept: 1.4016

		Brier score in leave-one-out cross validation is 0.16
	</description>

	<controllableVariable min="0" max="2">InBraceCorrection</controllableVariable>
	<controllableVariable min="0"max="1">Compliance</controllableVariable>
	<outputType>console</outputType>
</trainedModel>
```

2. The prediction executable

This executable reads patient data supplied by the software platform in a comma-separated-value file named “testData.csv”. This data will have the same format as the original training data. The prediction executable should make a prediction for each instance (patient) in the data set and deliver these predictions in a comma-separated list either in the console window or a separate file named “predictions.txt”.

3. The model description document

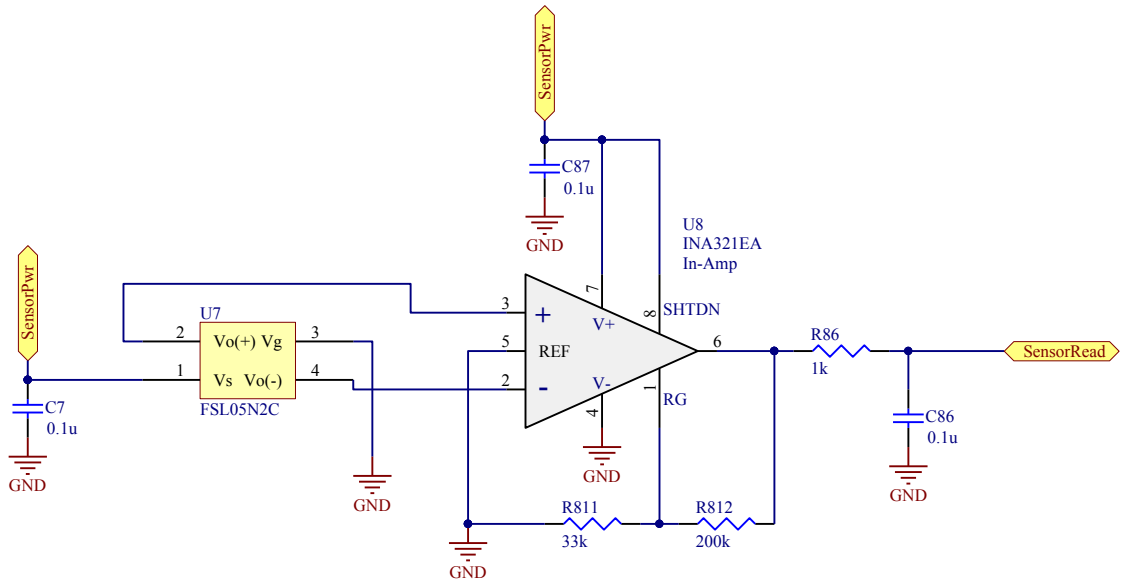
This document is an XML document with a .mdl file extension which represents a model type to the software platform. The document contains the three elements described in the following table:

Element	Description
name	The name of the model type, to be displayed to the user
description	A description of the model type, to be displayed to the user.
train	The name of the model training executable
predict	The name of the prediction executable

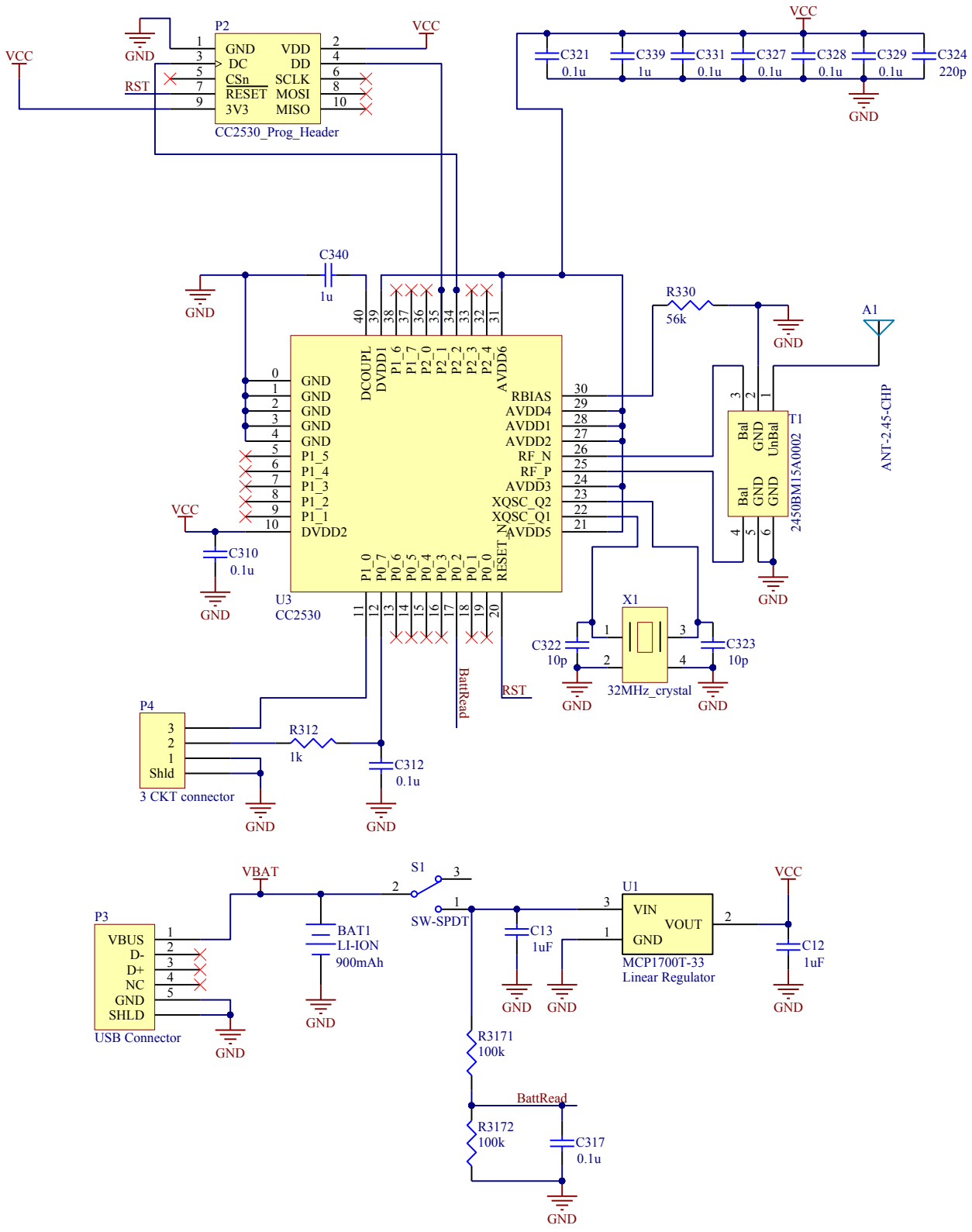
The following example shows the contents of the .tnd file for the logistic regression module used in this thesis.

```
<?xml version="1.0" encoding="utf-8"?>
<modelType>
	<name> Logistic Regression Model </name>

	<description>
	Trains a logistic regression model to predict probability of progression, based on
	in-brace correction and brace-wear compliance. Progression is defined as any
	positive change in Cobb angle, but the training process assumes Cobb angle
	measurement error with a 95% CI of 5 degrees.
	</description>

	<train>LogisticRegressionModelTrain.exe</train>
	<predict>LogisticRegressionModelPredict.exe</predict>
</modelType>
```

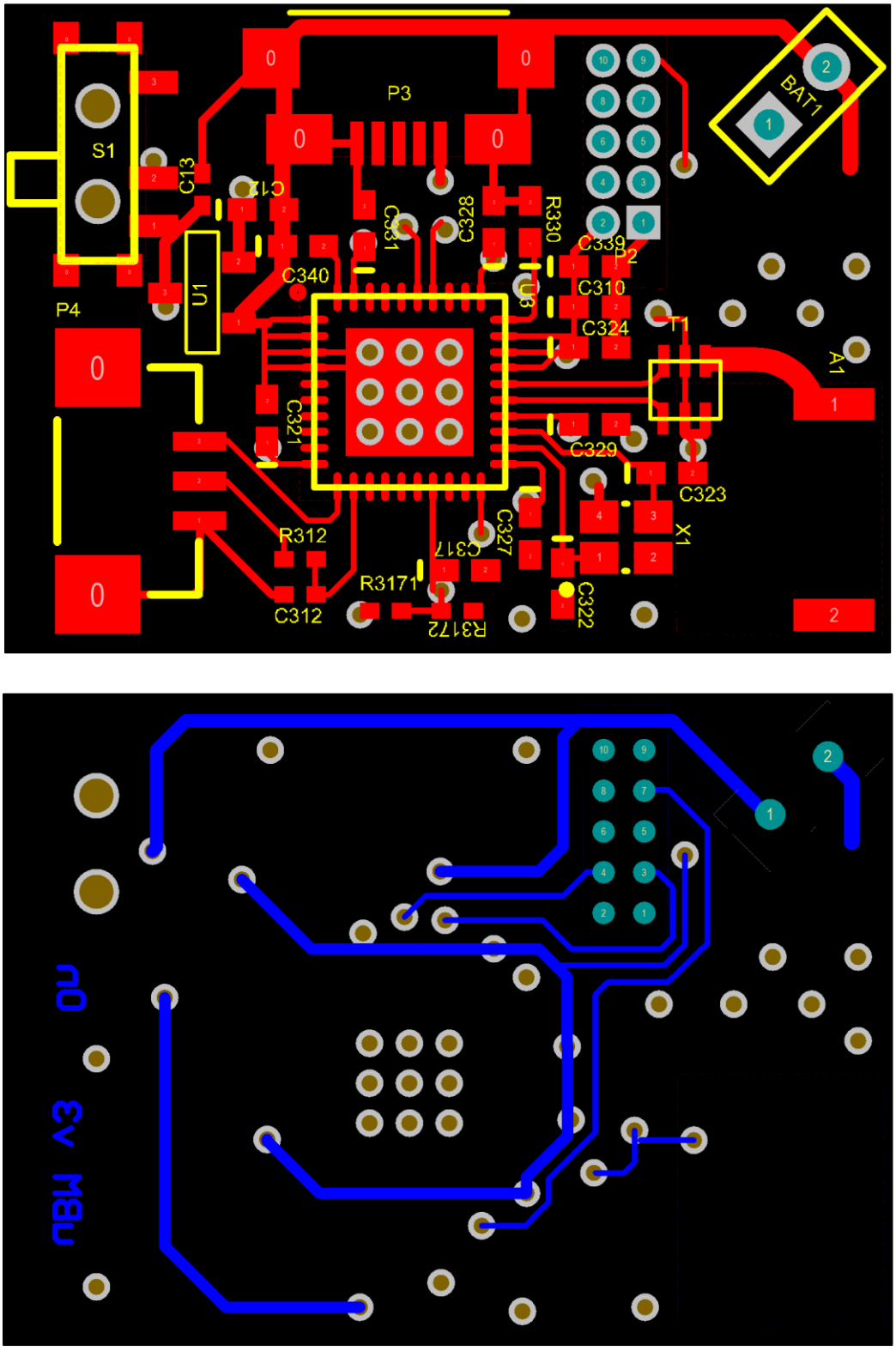
Appendix C: Compliance Monitor Schematic and Circuit Board Layout



Schematic for the force sensor portion of the monitoring device



Schematic for the data logger portion of the monitoring device



Top and bottom layer circuit board layouts