TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABREVIATIONS

AIS	Adolescent idiopathic scoliosis
ATR	Angle of trunk rotation
AVL	Apical vertebra level
AVR	Apical vertebrae rotation
BFP	Best fit plane
C°	Cobb angle
COL	Center of lamina
CPM	Center of pedicle method
CT	Computed Tomography
FBT	Forwarded bending test
FH3DUS	Freehand 3D ultrasound system
GT	Geometric torsion
ICC	Intra correlation coefficient
MRI	Magnetic Resonance Imaging
MT	Main thoracic
PA	Posteroanterior plane
PMC	Plane of maximum curvature
PT	Proximal thoracic
SAP	Superior articular process
SP	Spinous process
SRS	Scoliosis research society
SS	Sanders Stage
TLL	Thoracolumbar/lumbar
TP	Transverse process
US	Ultrasound
2D	Two dimensions
3D	Three dimensions

INTRODUCTION

Adolescent Idiopathic Scoliosis (AIS) is a 3D deformation of the spine, mainly visible as a lateral curvature in the form of an elongated “S” or “C” shape from the posteroanterior plane. AIS is the most common type of scoliosis, and it is highly prevalent in adolescent between 10 and 18 years of age, or until skeletal maturity. Between 1% to 4% of the adolescent population−mainly girls−is affected by AIS. AIS starts in early puberty, a time when children are growing rapidly, and although not all the curves will be progressive, 1 over 1000 will require a surgical treatment.

The “idiopathic” part of AIS means that the cause is not known. However, genetic studies indicate that there exists an increased risk of developing AIS when there are first degree relatives with this condition. In addition, AIS can be related to other factors such as environmental, central nervous system abnormalities, skeletal and muscle growth, hormonal and metabolic, or other factors not yet identified.

AIS is usually diagnosed by a physical examination or postural screening exam at school. Common signs of AIS are asymmetry in shoulder height or shift of the trunk, where the hips look uneven, which cause that one leg appears to be longer than the other. In addition, a back hump can be visualized when the patient is bending forward.

Clinical assessment and classification of IAS rely on 2D radiographic observations of the spine in the posteroanterior and sagittal planes. These radiographs are taken in a standing position having a full view of the shape of the spine. A sense of abstraction is needed by clinicians who evaluate these projections to figure not only how the spine looks in the 3D space, but also how the curvature will progress.

Advances in technology are changing the current 2D description of AIS towards a 3D characterization. These 3D descriptors could be important to improve the understanding of AIS, as well as to improve assessment, prediction of progression and treatment.

0.1 Problem statement and motivation

The goal of studying scoliosis is to understand the medical condition, and to select the optimal treatment or surgical strategy for each patient. Since the spine is a 3D structure, experts who evaluate 2D images of the spine need experience, abstraction and visualization skills in order to avoid misinterpretation. Likewise, the evaluation is not deterministic and can change from expert to expert.

The current gold standard for the assessment of the magnitude of the curve is a 2D measurement used to evaluate a 3D structure. Nevertheless, two patients sharing the same profile in 2D will not necessarily share the same morphology in the 3D space. Hence, the treatments should be adapted specifically to the 3D shape of the spine (Labelle et al., 2011). Recent advances in technology allow to generate new techniques to characterize the spine in the 3D space. However, these descriptors are difficult to interpret and to measure in a clinical setting where only 2D radiographs are available.

Classification methods emerge to present a way to ease the visualization of common patterns in a dataset. In the case of spine deformities, the Lenke classification method helps to group similar curves. Nevertheless, although this system is widely used in clinical practice, it does not provide a full understanding of the deformation in the 3D space, since it depends on the analysis of 2D radiographs. The importance of a 3D classification system resides in improving the description and comprehension of AIS, in a way that can be reproduced with reliable outcomes.

Prediction of the development of the curves through time is also a relevant task. Knowing beforehand how the curve could change in the future, would help clinicians to improve treatments. Clinical indices such as chronological, skeletal, and menarcheal age, curve magnitude, and curve location have been studied (Cheng et al., 2015), however these are not robust enough to predict the deformations.

Computer-based approaches to describe, classify and predict the evolution of spine deformities would help to validate manual measurements, to decrease time for evaluating and treating patients, and to improve reproducibility.

Since patients with AIS are young, their tissues are still immature and sensitive to X-rays, which is the main imaging technology used to evaluate scoliosis. Patients with high risk of progression need to be evaluated frequently, every 4 to 6 months. Some studies (Doody et al., 2000; Hoffman, Lonstein, Morin, Visscher, & Harris III, 1989; Ronckers et al., 2010; Ronckers, Doody, Lonstein, Stovall, & Land, 2008) have shown that young women are especially sensitive to the exposure to ionizing radiation. Therefore, age, gender and recurrent exposure to radiation may increase the risk of developing breast or lung cancer (Levy, Goldberg, Mayo, Hanley, & Poitras, 1996). The development of radiation-free imaging technology to monitor spinal deformities progression would be of high interest for the management of AIS.

0.2 Research objectives and contributions

This thesis presents a framework designed to characterize and model the variation of the shape of the spine affected with AIS. This framework includes three contributions: 1) two measurement techniques for computing 3D descriptors of the spine, and a classification method to categorize spine deformations, 2) a method to simulate the variation of the shape of the spine through time, and 3) a protocol to generate a 3D model of the spine from a volume reconstruction produced from ultrasound images.

Three main contributions were proposed toward this goal:

1) **Computer-based characterization and classification methods in AIS using 3D descriptors**. The classification system for spine deformations developed by Lenke, is a descriptive and reproducible method widely used in clinical practice. However, its main disadvantage is the use of the Cobb angle to quantify the deformation of the spine,

a measurement that does not describe the spine in the 3D space. We introduced two techniques to represent the variation of the spine in 3D. Also, we proposed to use a computer-based classification algorithm called dynamic ensemble selection to categorize spine deformations. The classification method does not depend on a specific learning algorithm or set of descriptors of the spine. It identifies the best combination of them to classify curve types. This could help clinicians to evaluate the role of each descriptor in a specific spine.

2) **Shape analysis using computer-generated models for progression of the curve of the spine through time**. Prediction of the progression of the spine deformation is one of the main concerns when treating patients with AIS. Knowing how the shape of the spine is going to evolve from the first visit of the patient, would help clinicians to improve treatment strategies. In this contribution, we proposed independent components analysis to describe the modes of variation of the spine in the 3D space, together with an approach to predict the curve progression from the first visit, every three months for a time lapse of eighteen months. The results show that our approach for curve progression is a promising technique, which can help to identify the variation of the shape of the spine through time.

3) **A preliminary study for a radiation-free 3D imaging system based on 2-D ultrasound.** Radiation is one of the clinician's critical concerns in patients with AIS. Since patients are young, there is a high risk of exposure to ionizing radiation, even with low dose systems. In this contribution, we propose the use of a freehand 3D ultrasound system to generate volume reconstructions of the spine. Ultrasound is a radiation-free technology, which could help clinicians in follow-up of patients with AIS, decreasing the need of X-rays. In this study, we were able to generate a 3D representation of the centerline of the spine, by identifying landmarks on the volume reconstruction. Our results suggest that this system can be promising for the evaluation of the shape of the spine.

0.3 Outline

This manuscript is organized as follows. In **Chapter 1,** we presented the clinical context of AIS, as well as a review of the relevant studies related to each of the contributions of this research. **Chapter 2** introduces our techniques, *leave-n-out angles* and *fan leave-n-out angles* to describe the shape of the spine in 3D space, and the dynamic ensemble selection method to categorize deformations of the spine, this work was published in the *Medical and Biological Engineering and Computing.* **Chapter 3** presents our approach to predict curve progression through time, based on 3D descriptors of the spine. This chapter was published in the *Computers in Biology and Medicine.* **Chapter 4** presents our efforts to reduce the use of X-ray imaging by introducing a freehand 3D ultrasound system to generate volume reconstructions of the spine. This work was submitted to the *Ultrasound in Medicine and Biology*. In **Chapter 5**, a summary of the main contributions of this research is presented and discusses its limitations and future work. Finally, **Appendix I** shows a complete list of works resulting from this research.

CHAPTER 1

LITERATURE REVIEW

The objective of this chapter is to present a general overview of the clinical context of Adolescent Idiopathic Scoliosis, as well as the state-of-the-art methods for the evaluation of spinal deformities in 2D and in 3D. This chapter starts describing the anatomy of the spine, followed by the clinical concepts associated with the clinical study of AIS. Then, a critical review of computer-based methods involved in the study of AIS is presented. At the end, this chapter includes a summary of the approaches proposed in the literature.

1.1 Clinical context

1.1.1 Anatomy of the spine

The spine is usually composed by articulated bones called vertebrae, which help keeping an upright or stand up posture. Being the main support of the human body, it is on charge of the movements of the head and torso, and it serves as a protection for the spinal cord. The spine can flex or rotate, but the grade of movement or function depends on the different sections that compose it: cervical, thoracic, lumbar, sacral and coccyx.

1.1.2 Sections of the spine

The spine is divided in five sections (see Figure 1.1), each of them is in charge of specific functionalities:

- **Cervical spine** (upper back): Numbered from C1-C7, is the main support of the head. It is the section with greatest range of motion, especially because the first two vertebrae are directly connected to the skull, which allow the motion of the head.
- **Thoracic spine** (middle back): Numbered from T1-T12. This section is on charge of the protection of the heart and lungs by holding the rib cage.

- **Lumbar spine** (lower back): Numbered from L1-L5. The weight of the body is supported by this region. The vertebrae that form this part of the spine are much larger in size, compared to the previous sections.
- **Sacrum**: It contains five fused vertebrae. Its principal purpose is to connect the spine to the hip bones.
- **Coccyx**: Also known as tailbone, is comprised by four fused vertebrae. It helps to keep attached the ligaments and muscles of the pelvic floor.

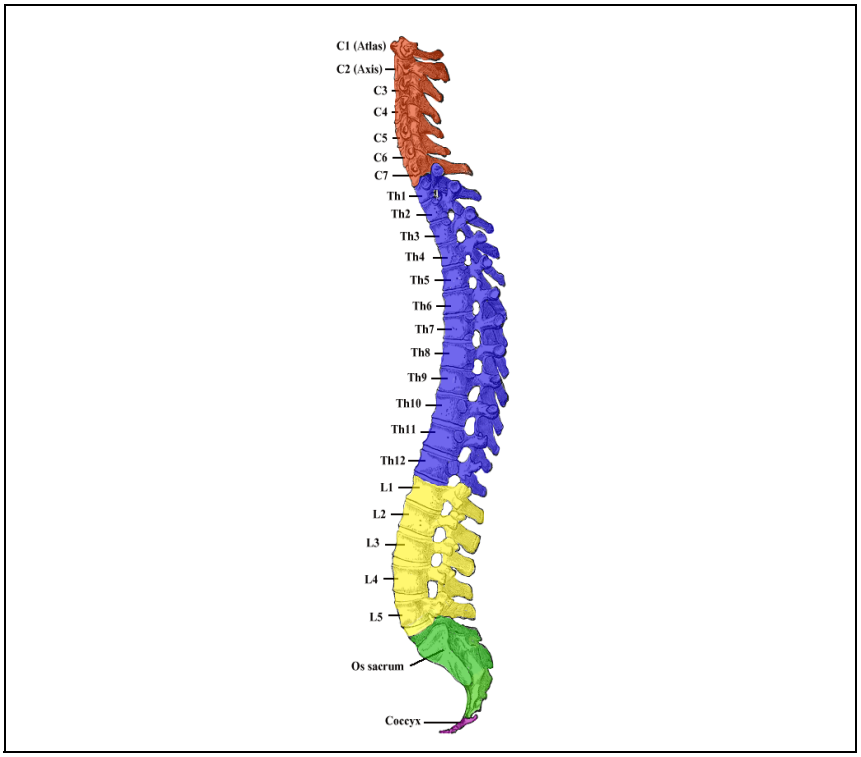


Figure 1.1 The five sections of the spine, and numbering of the vertebrae. Adapted from: Henry Gray (1918) Anatomy of the Human. Altered by User: Uwe Gille, public domain.

1.1.3 Curves of the spine

The spine naturally develops curves. When viewed from the coronal plane, it looks like a straight line. However, from the sagittal plane there are two observable curvatures in the thoracic and lumbar sections, and the aspect of the spine seems such as a soft 'S' shape (see Figure 1.1). These normal curves are known as kyphosis and lordosis, which are essential for the human body to keep the balance between the trunk and head over the pelvis. Both are considered normal to a certain extent.

Abnormal curvatures could be caused by congenital defects or triggered by degenerative diseases. These deformities occur when the natural curves of the spine are misaligned or surpass the acceptable limits (see Figure 1.2).

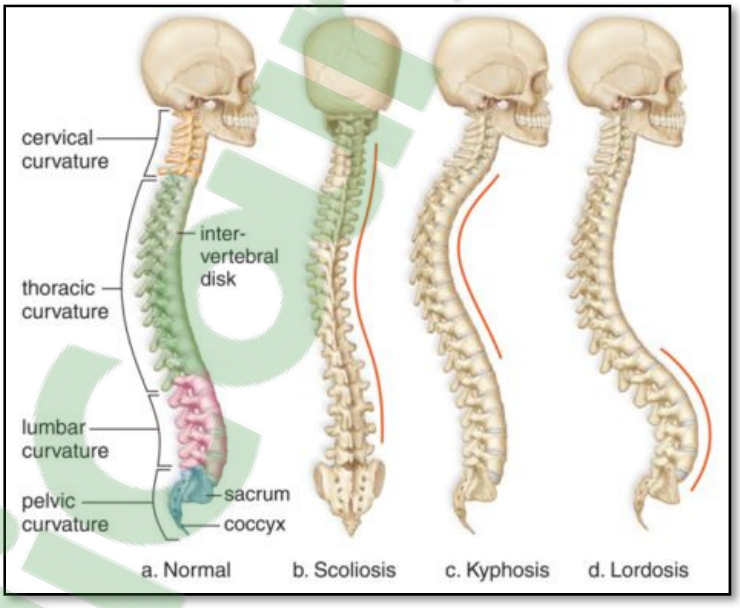


Figure 1.2 Spines and its abnormal curvatures.
Adapted from StudyForce, public domain.

1.1.4 Adolescent Idiopathic Scoliosis: characterization and classification

The gold standard method to quantify the magnitude of the curves with AIS is the Cobb angle. It receives its name from Dr. Jon R. Cobb, who in 1948, first described the curvature of the spine as a measure of the magnitude of deformities. It is measured in degrees and helps physicians to determine the severity of the deformation and to decide what treatment will be necessary for the patient.

In clinical practice, the Cobb angle is measured on the posteroanterior and lateral X-rays, the most common imaging modality to observe the spine in a standing position. The radiographs are acquired based on the global coordinate system proposed by the Scoliosis Research Society (Stokes, 1994b). The *x*-axis is the horizontal axis that runs from the rear to the front of the patient, while the *y*-axis is the horizontal axis that runs from the right to the left of the patient. The *z*-axis is the vertical axis, which goes from the bottom of the patient upward (see Figure 1.3).

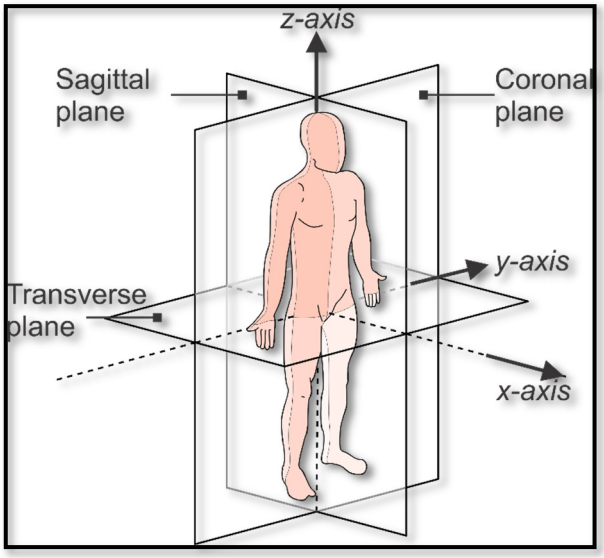


Figure 1.3 Global coordinate system. Adapted from: Wikimedia Commons used under Creative Common created by CFCF, June 2014.

The Cobb angle is calculated in the postero-anterior plane, and it is formed between a line drawn parallel to the superior endplate of the upper vertebra included in the scoliotic curve and a line drawn parallel to the inferior endplate of the lower vertebra of the same curve (see Figure 1.4). If the Cobb angle is ≥ 10°, the patient is diagnosed with scoliosis.

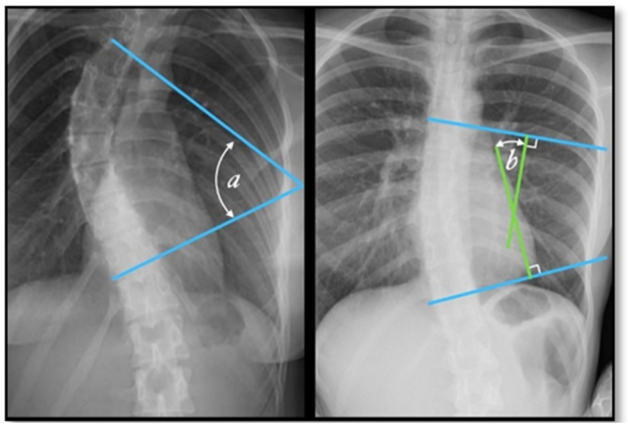


Figure 1.4 Cobb angle measurements for 2 patients using a different geometrical construction. Photo courtesy of Prof. Frank Gaillard, Radiopaedia.org, adapted under Creative Common license.

Classification methods arise as a way to ease the appreciation of common patterns. In AIS, a comprehensive classification system is relevant because it allows to identify all types of curve patterns, and hence to standardize the assessment and treatment. A classification method should have good to excellent inter- and intra- observer reliability in order to be reproducible in clinical setting. It would also provide clinicians with a common language to compare similar cases.

Concerning to AIS, King et al. proposed a classification method for severe thoracic curves (King, Moe, Bradford, & Winter, 1983). They classify the spines in 5 types, excluding thoracolumbar, lumbar, or double or triple major curves. However, poor inter- and intra-observer reliability and reproducibility of the method has been reported (Lenke et al., 1998). In 2001, a new classification system of AIS was developed by Lenke et al. Nowadays, the system has been widely accepted in clinical setups (Lenke et al., 2001). This categorization

divides the spine deformities in 6 types. It is based on the measurement of three components: type of curve (Lenke 1-6), a lumbar modifier, and sagittal thoracic modifier. These components are used to distinguish structural and nonstructural curves in the proximal thoracic, main thoracic, and thoracolumbar/lumbar sections. According to the classification of the spine deformation, treatment recommendations are also provided. The authors reported the reliability of the classification by the kappa values of the interobserver (0.74) and intraobserver (0.893).

Once the patient has been diagnosed, it is feasible to provide a treatment in order to prevent the progression of the deformation, and hence, avoid surgical operation when possible. Bracing (see Figure 1.5) is prescribed for patients between 10-15 years of age, at skeletal maturity specified by the Risser grade between 0-2, and magnitude of the main curvature between 20°-40° (Richards, Bernstein, D'Amato, & Thompson, 2005).

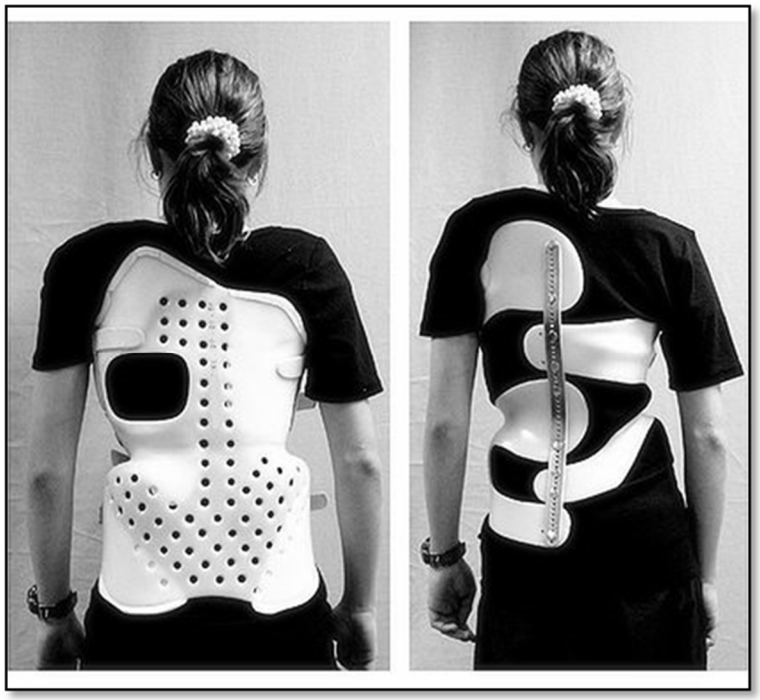


Figure 1.5 Two different braces to treat scoliosis. Chêneau brace on the left and Chêneau light on the right. Adapted from Wikimedia Commons used under Creative Common license. Created by Scolidoc (Weiss et al. Scoliosis 2007 2:2 doi:10.1186/1748-7161-2-2).

Patients at risk of curve progression during adult life, are considered for surgery (see Figure 1.6). Generally, they present curve magnitude > 50° in the thoracic section, or between 50°-60° in the thoracolumbar section. Pain, appearance and shortness of breath are symptoms used as indicators for surgery (Asher & Burton, 2006). Patients that require surgical intervention represents 0.1% of the total with AIS (Cheng et al., 2015).

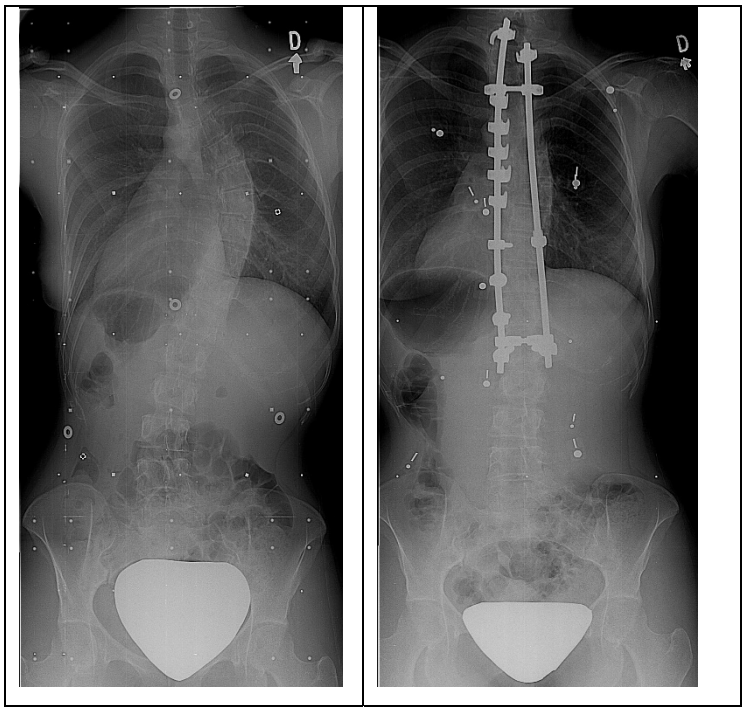


Figure 1.6 On the left, a patient before surgery. On the right, the patient after surgery. Photo courtesy of LIS3D, Sainte-Justine Hospital.

1.2 State-of-the-art on computer-based characterization and classification methods in Adolescent Idiopathic Scoliosis using 3D descriptors

A 3D classification system would help with the understanding and description of the scoliotic curves; which improve evaluation, follow-up and treatment (Labelle et al., 2011). Nowadays, technology allows to automatically collect more data, to perform measurements systematically, to generate 3D descriptors, and to create complex models to enhance the classification of the spine (Stokes, 1994a). Discovering patterns concerning the 3D space could help to introduce a

3D classification able to surpass the limitation of 2D. In addition, the Scoliosis Research Society (SRS) has accepted the need to treat AIS based on the analysis of the deformity in the 3D space (Labelle et al., 2011). Hence, the development of 3D classification method, that can be applied in everyday clinical practice, is of vital importance.

In recent studies, new classification systems based on 3D descriptors of the spine have been proposed. Negrini et al. (Negrini, Negrini, Atanasio, & Santambrogio, 2006) used an optoelectronic system (AUSCAN), generating a 3D reconstruction of the spine in real time by markers positioned on the skin of the patient. Other classification systems use 3D reconstructions of the spine obtained from standing stereographic X-rays. The reconstruction consists of 3D coordinates of particular anatomic landmarks, (Delorme et al., 2003) (see Table 1.1).

Table 1.1 Characteristics of included studies

Author	Patients	Classification	Instrument	Design
Poncet 2001	62 AIS	All Lenke	Stereo radiographies	Prospective
Negrini 2006	122 AIS 23 hyperkyphosis 4 AIS+ hyperkyphosis	All Lenke	AUSCAN	Cross-sectional
Duong 2006	409 AIS	All Lenke	Stereo radiographies	Prospective
Boisvert 2008	307 AIS	Lenke 1 and 5	Stereo radiographies	Cross-sectional
Sangole 2009	172 AIS right thoracic	Lenke 1	Stereo radiographies	Cross-sectional
Stokes 2009	110 AIS	Double curves	Stereo radiographies	Cross-sectional
Duong 2009	68 AIS	Lenke 1	Stereo radiographies	Prospective
Kadoury 2012	170 AIS right thoracic	Lenke 1	Stereo radiographies	Cross-sectional
Kadoury 2014	65 AIS 5 healthy	All Lenke	Stereo radiographies	Cross-sectional
Shen 2014	255 AIS	Lenke 1	Stereo radiographies	Cross-sectional
Thong 2015	155 AIS	Lenke 1	Stereo radiographies	Cross-sectional
Thong 2016	633 AIS	All Lenke	Stereo radiographies	Cross-sectional

The descriptors of the spine can be divided by its nature in geometric or global. Geometric are measured from the 3D reconstructions of the spine, such as apical vertebra rotation (AVR), best fit plane (BFP) , direction, geometric torsion (GT), phase, plane of maximum curvature (PMC), shift (Duong, Cheriet, et al., 2009; Kadoury, Shen, & Parent, 2014; Negrini et al., 2006; Poncet, Dansereau, & Labelle, 2001; Sangole et al., 2009; Shen, Parent, & Kadoury,

2014; Stokes, Sangole, & Aubin, 2009). In a different approach, Boisvert et al. (Boisvert, Cheriet, Pennec, Labelle, & Ayache, 2008) proposed the use of rigid transformations as geometric descriptors of the spine.

On the other hand, when the descriptors are obtained by applying a method to reduce the dimensionality of the 3D reconstructions of the spine, are called global. Some descriptors have been obtained by computing a wavelet compression technique (Duong, Cheriet, & Labelle, 2006), principal component analysis (Boisvert, Cheriet, Pennec, & Labelle, 2008), locally linear embedding (Kadoury & Labelle, 2012), and stacked auto-encoders (Thong et al., 2016; Thong, Labelle, Shen, Parent, & Kadoury, 2015) (see Table 1.2).

In order to provide a new classification, some studies used a quantitative analysis made by an expert, while others utilized clustering. Clustering analysis have been used to automatically group similar 3D curve patterns of the spine.

Table 1.2 presents the state-of-the-art in 3D classification systems as well as the type of descriptors used for the classification of spinal deformities.

Table 1.2 Methodology of included studies

Author	Classification methodology	Descriptors of the spine*	Advantages	Disadvantages
Poncet 2001	Qualitative analysis	GT	• Classification based on a 3D descriptor.	• The estimation of GT could be inaccurate due to the 3D reconstruction. • The GT only provides a measurement at vertebral level, hence, the effect on the global shape is not considered.
Negrini 2006	Qualitative visual analysis	Direction, Shift, Phase	• Quasi-3D graphic representation of the spine based on the spinal top view. • Classification based on 3D parameters.	• The AUSCAN system cannot be used in every day clinical practice. • The descriptors are not intuitive.
Duong 2006	Fuzzy clustering	Global shape descriptors based on a wavelet compression technique	• Automatic classification based on 3D curve patterns.	• The descriptors do not offer direct interpretation, which difficult their use in clinical procedures.
Boisvert 2008	Qualitative visual analysis	Principal deformation modes of articulated models	• The principal modes of variation can be interpreted. They show distinctive patterns of curves associated with Lenke 1 and 5.	• The modes of variation were computed on only two types of curves. More experiments should be performed to find out if the modes of variation can be generalized for other curve patterns.
Sangole 2009	ISOData clustering	C°, AVR, PMC, kyphosis	• Automatic classification based on PMC could be used to analyze curves in 3D. • When the PMC is used along with the daVinci view, provides a comprehensive visual representation of the deformation.	• The PMC needs to be tested on other scoliotic curves to prove its effectiveness. • The method would be more relevant if it were related to sagittal and coronal projections, common views used clinically.
Stokes 2009	K-means clustering	C°, AVL, AVR, PMC	• Automatic classification based on PMC can separate groups of 3D shapes. • PMC could be used to indicate likelihood of progression.	• The PMC is very sensitive to small changes from postural variability.
Duong 2009	K-means clustering	PMC, BFP, GT	• Automatic classification based on intuitive descriptors. • BFP was the best parameter to analyze the curves in 3D.	• The BFP is difficult to visualize from 2D radiographs.
Mezghani 2010	Self-organizing maps	C°	• Automatic classification that agrees with Lenke. • The map provides smooth transitions between Cobb angles, instead of strict cut-off used in Lenke classification.	• Limited to C°, which does not provide a 3D classification of the deformities.
Kadoury 2012	K-means clustering	Global shape descriptors extracted by applying locally linear embedding	• Automatic classification with a custom-designed similarity metric for articulated shape deformations, which allows to preserve the neighborhood relationships of similar shapes on a nonlinear manifold.	• Only tested on Lenke 1 curves. • The descriptors are not intuitive.

Table 1.2 (Continuation)

Author	Classification methodology	Descriptors of the spine*	Advantages	Disadvantages
Phan 2013	Self-organizing maps	C°	• Automatic classification that agrees with Lenke. • Improves Lenke classification accuracy and study treatment variability.	• Limited to C°, which does not provide a 3D classification of the deformities.
Kadoury 2014	Fuzzy c-means clustering	Parametric GT	• Automatic classification based on parametric GT to distinguish 3D deformations. • The parametric GT captures the estimated at the junction of the segmental curves, instead of at each vertebral level. • The descriptor can be used as 3D index to identify subgroups within Lenke classification.	• Complex to apply in clinical setups.
Shen 2014	Fuzzy c-means clustering	Parametric GT	• Automatic classification can differentiate subgroups within Lenke 1. • Allows to evaluate the spine in the thoracolumbar region, useful in surgical strategies.	• Although provides with a quantitative analysis, it is still complex to use in clinical setups.
Thong 2015	K-means++ clustering	Global shape descriptors extracted from stacked auto-encoders	• Automatic classification can differentiate subgroups using 3D descriptors. • Simplified version of the 3D reconstructions of the spine. • Could help to improve surgical strategies.	• Requires sizeable datasets to generate the low-dimensional representation of the 3D curves. • Only tested on Lenke 1 curves. • The descriptors are not intuitive or interpretable.
Thong 2016	K-means++ clustering	Global shape descriptors extracted from stacked auto-encoders	• Automatic classification can differentiate groups using 3D descriptors. • Simplified version of the 3D reconstructions of the spine. • Could help to improve surgical strategies.	• Requires sizeable datasets to generate the low-dimensional representation of the 3D curves. • There is not direct clinical interpretation of the descriptors.

Abbreviations: AVL, apical vertebra level; AVR, apical vertebrae rotation; BFP, best fit plane; C°, Cobb angle; GT, geometric torsion; PMC, rotation of the plane of maximum curvature of the main thoracic curve.

* Direction: the angle between spinal pathological and normal AP axis; Phase: the parameter describing the spatial evolution of the curve; Shift: the co-ordinates of the barycenter of the top view.

1.3 State-of-the-art on computer-generated models for progression of the curve of the spine through time

Predicting the development of the curvature of the spine through time, is one of the challenges that clinicians must face in order to improve the treatment of scoliosis. Once the patient has been clinically evaluated and diagnosed with AIS, the ideal would be to obtain a prediction of the progression of the deformation.

AIS and growth are connected. While the patient grows, the clinician needs to know how the indices of normal growth differ in patients with AIS. Also, the analysis of these indices helps to plan the best treatment, as well as to understand the progression before, during and after puberty (Dimeglio & Canavese, 2013). In clinical practice, the current indices to assess curve progression are maturity in terms of chronological, skeletal, and menarcheal age, curve location and its magnitude (Cheng et al., 2015). These indices are used to evaluate the possibility of progression. However, there is not a reliable method to predict the progression of the deformity from the first visit. The main criteria to determine the progression of the deformity is the Cobb angle increasing $\geq 6°$ between the first and the last visit (Noshchenko, 2015). Nevertheless, since the magnitude of the curve is calculated with the Cobb angle, treatment strategies, and follow-up examination are based on high variability measurements (Aubin, Labelle, & Ciolofan, 2007; Majdouline, Aubin, Robitaille, Sarwark, & Labelle, 2007).

The pubertal cycle plays a major role regarding the understanding of AIS. This is a period that last 2 years, characterized by an increase in growth rate, also called “peak height velocity”. It starts generally at bone age of 11 for girls and 13 for boys. After this phase, there is a period of 3 year of deceleration. In patients with AIS, the main curve progression occurs at the phase of peak height velocity, and it has a risk of progression associated to the initial curve angle. Patients with a curve of 5°, 10°, 20°, 30° have 10, 20, 30, and up to 100% of risk of progression respectively. In 75% of the cases with thoracic curves, in a range between 20° to 30°, are prone to progress ending up with surgery (Dimeglio & Canavese, 2013).

The assessment of the skeletal maturity is associated to the growth velocity and the cessation of the growth. Recently (Sitoula et al., 2015) showed a correlation of the Sanders Stage (SS) with the progression of the curve in AIS. This assessment of the SS for skeletal maturity is based on progressive growth and subsequent fusion of epiphyses of small long bones of the hand (SS1 to SS8), depicted from radiographs of the hand and wrist (Sanders et al., 2008).

Recently, Li *et al*.(Li et al., 2018) proposed a novel method to assess skeletal maturity by analyzing ossification patterns on proximal humeral epiphyseal. The novelty of this method is that the proximal humeral is present in the spine radiographs, hence there is no need of extra X-rays.

Secondary sexual characteristics are developed during puberty. Menarche status has been used as a mark of the pubertal growth spurt, and as an index to evaluate the risk of curve progression. However, it showed weak association with progression (Noshchenko, 2015; Sitoula et al., 2015).

Characterization of the spine in 3D is an important aspect to study in AIS, Labelle *et al.* (Labelle et al., 2011) have shown that similar deformities in 2D have different morphology in 3D. In recent studies, 3D descriptors to characterize the morphology of the spine have shown promising results with respect to the prediction of curve progression. These descriptors were computed from 3D reconstructions of models of the spine obtained from radiographs. In a retrospective study by Nault *et al.* (Nault et al., 2013), 5 descriptors (Cobb angles, three-dimensional wedging of vertebral body and disk, axial/sagittal/coronal rotation of the apex, upper, and lower junctional level, torsion, and slenderness) were evaluated to distinguish two groups, progressive and nonprogressive curves between the first and the last visit. From the descriptors studied, 3D wedging of apical disks, intervertebral axial rotation, spinal torsion, slenderness in the T6 vertebra, and slenderness of the whole spine, were found with statistically significant difference between the 2 groups. Later on, a prospective study by (Nault et al., 2014) was performed to evaluate 3D morphological descriptors between progressive and nonprogressive curves using the first visit of the patients. The most significant descriptors were

plane of maximal curvature, kyphosis, apical intervertebral rotation, torsion, and slenderness. However, a limitation of this work is that the values of the morphological descriptors at the first visit are small in both groups. Also, an evaluation about how accurately these descriptors can predict curve progression, and how do they change throughout puberty, must be performed.

Opposite to expert-based descriptors used in the previous studies, (Kadoury, Mandel, Roy-Beaudry, Nault, & Parent, 2017) proposed a probabilistic manifold embedding to reduce high-dimensional data to its low-dimensional representation. Based on this new representation, a spatiotemporal regression model was built to predict the evolution of the deformations. The patients were separated in two groups progressive and nonprogressive. A patient is cataloged as progressive if there is a difference of 6° in the magnitude of the curvature. This could represent a limitation, since the predicted model is based on a 2D measure, with high variability and it does not characterize the spine in 3D.

1.4 State-of-the-art on radiation-free imaging systems to reduce the use of X-rays.

Radiograph is the most common imaging method used to treat patients with AIS. They have been used to study the inside of the body to diagnose illnesses such as breast cancer, fractures, spine deformities, among others. Radiographs are acquired by applying ionizing radiation that goes through the human body, creating an image of tissues and structures inside the body on photographic plates or other detectors.

The advantage of X-rays is that it allows visualizing the spine in a standing position, hence, it is possible to analyze the full length of the spine with the effect of gravity. X-rays are the gold standard imaging technology employed to calculate the magnitude of the curve by using the Cobb angle method. However, some patients need to undergo radiographs every 4 to 6 months in follow-up visits, which results in frequent exposure to harmful radiation (Doody et al., 2000; Hoffman et al., 1989; Ronckers et al., 2010, 2008). This makes it difficult to perform close evaluations to assess progression and adequate treatments.

Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) could be used as an appealing acquisition technology compared to X-rays, however both modalities are generally not performed in standing position, which is necessary to correctly evaluate the shape of the spine. Another drawback of MRI and CT scanners is their elevated cost, which makes them inaccessible to many people.

Ultrasound (US) imaging is a possible and economical alternative to radiographs. Imaging in real time, radiation free, and low cost are its principal features. US is an inexpensive technology compared to radiographs, MRI and CT, and it does not need to have neither a special room for its installation nor security protection for its operators. In addition, the easy access to US imaging and affordable price means that most hospitals, even those with low budget, located in areas where it is difficult to transport large instruments or without room for large machines might already be equipped with at least one.

By itself, US is limited to produce only one 2D image at certain time interval. To examine complex structures like the spine this is inconvenient, since the full shape of the spine cannot be visualized. To overcome this limitation, a freehand 3d US systems have been developed. This type of system is used to generate 3D reconstructions from tracked ultrasound images. In the case of the spine, the reconstruction would represent the surface of each vertebrae.

A freehand 3d ultrasound system is non-invasive and it is composed of four devices. 1) a 2D ultrasound scanner, 2) a tracking system used to determine the position and orientation of the transducer, 3) a workstation with the software to capture, store and process the images and 4) a grabber to transmit the images from the ultrasound scanner to the workstation.

There are two common tracking systems, optical and magnetic. On the one hand, an optical tracking system (OTS) uses infrared cameras pointed to a reference and a navigation instrument with attached markers. It needs an uninterrupted line of sight to the navigation instruments. On the other hand, magnetic tracking systems (MTS) contain a magnetic field generator used to measure pulses produced by transmitters. The control unit calculates the

position of each sensor inside the magnetic field. Opposite to the OTS, it does not need a direct line of sight to the navigation instruments. However, in a clinical setup, other medical instruments could cause disturbance, which affects the accuracy and precision of the measurements.

Purnama *et al*. (Purnama et al., 2009) proposed a freehand 3D US system to generate a volume reconstruction of the spine. Only one acquisition from a healthy subject, from T4-T9 vertebrae was performed. The transverse processes (TP), superior articular processes (SAP) and laminae were the landmarks identified on the volume reconstruction. These landmarks were automatically obtained by filtering out the non-vertebral features from the reconstruction. They reported that from vertebra T3 upward, it was not possible to distinguish the landmarks. Also, the ribs caused strong reflections which made difficult to filtering them out. Additionally, two 3D measurements were determined semi-automatically, axial rotation and vertebral tilt. For this purpose, they use the center of mass of the landmarks, since exact boundaries were not easy to identify on the volume. This method was tested in only one reconstruction from a healthy individual. More experiments need to be performed to generalize the method for subjects with spine deformation.

Chen *et al*. (Chen, Lou, & Le, 2011) proposed an equivalent method to calculate the Cobb angle called center of pedicle method (CPM), based on the use of the TP and laminae from US images. They validated this method on 56 scoliotic curves from PA radiograph images. This set of images was divided into three groups based on the Cobb angle, mild, moderate and severe. Their results show an average difference between the CPM method and the Cobb angle of -0.6°, 1.7° and 2.6° respectively for each group. To validate the identification of the TPs on US images, a second experiment was performed. Two phantoms were used, a cadaver thoracic vertebra (T9) and a phantom from the T2-T12 vertebrae. For scanning purposes, the phantoms were immersed in a water-filled container. At 8mm above the phantom, a 2mm tick polypropylene sheet was placed to simulate the skin. Their result show that they were able to find the center of pedicle on the US images. However, since the images were acquired from

phantoms, the reflections were stronger, and it was easy to identify the landmarks, which could vary in acquisitions from a real patient.

Cheung *et al*. (C. W. J. Cheung, Siu-Yin Law, & Zheng, 2013) used a freehand 3D US system for acquisition of images from the spine. Four spinal phantoms, containing from L5-T1 vertebrae, were employed for acquisition of US images and X-rays. Each phantom was deformed into 4 curvatures, for a total of 16 spinal deformations. The TPs and SAPs were marked manually on the US images. Any image without these landmarks was discarded. These landmarks were projected into the three orthogonal planes. On selected vertebrae, two lines joining the TP and SP were marked. Thee lines represented the most tilted vertebrae at the top and bottom of the spine and were used to calculate the Cobb angle. This calculation was compared to the Cobb angles obtained from the X-rays. Their results showed a correlation (R^2=0.759; $p<0.005$) between both measurements. The main limitation is the manual marking in each image which is time consuming, depends on the operator and the quality of the images could vary in real subjects.

The center of the laminae (COL) method has been studied by Chen *et al*.(Chen, Lou, Zhang, Le, & Hill, 2013) to calculate the equivalent of the Cobb angle from US images. A cadaver spinal phantom, containing from L5-C1 vertebrae, was employed in the study. This was deformed to represent 30 scoliotic curves, but only from the L5-T1 vertebrae were scanned. Images with an US scanner and a laser scanner were acquired from the phantom. The COL method was used on the US images to calculate the Cobb angle. This method consists on finding the most tilted vertebrae at the top and bottom of the spine. Then, two lines were drawn joining the center of the laminae on each side of these vertebrae. The angle between these two lines was the Cobb angle (COL angle), which was compared to the Cobb angle obtained from the images from laser scanner. Their results showed an intra- and inter- observer reliability as high as the reported for Cobb angle measurements (ICC values > 0.88). An extra experiment was performed on 5 subjects who had PA X-rays. A comparison between the Cobb angle from X-rays and the COL angle was performed. As result, an average difference of 0.7° between both methods was obtained.

Koo *et al*. (Koo, Guo, Ippolito, & Bedle, 2014) proposed the posterior deformity angle to quantify scoliotic deformities based on US images. For capturing the US images, a freehand 3D US system was developed. From the tracked US images, the SPs were manually marked, and their 3D coordinates were obtained. Three cadaver spine phantoms were used to configure 30 different curvatures. PA X-rays were also acquired from these phantoms to measure the Cobb angle. To calculate the posterior deformity angle, they proposed a locally weighted polynomial regression technique to curve fit the SPs. From the fitted curve, the tangents with the most positive and negative slopes were identified at the top and bottom of the curve, and the angle between them was calculated. Their results show that their approach had a high correlation with respect to the Cobb angle (r=0.915). The limitation of this method is that there was no validation in patients with scoliosis.

Ungi *et al*. (Ungi et al., 2014) proposed a method to calculate the curvature of the spine based on TP by tracking ultrasound snapshots. One US image is taken at each side of the vertebrae finding the TP. Then, midpoints of the TP are located on the US image. The line joining these midpoints is used to calculate an angle relative to a reference line. This angle is called transverse process angle. The same angle was calculated on PA X-rays, and then compared to the one obtained from the snapshots. The method was tested on two phantoms, an adult and a pediatric spine containing 12 thoracic and 5 lumbar vertebrae. Their results show small inter-operator differences between the transverse process angle and the Cobb angle. However, the disadvantage of this method is the ability to recognize the landmarks during the acquisition. This would be challenging in patients where fat and muscles interfere the visibility of the vertebrae. Also, the change in breathing and posture could increase the difficulty of taking two images at each side of the vertebrae.

Cheung *et al*. (C. J. Cheung, Zhou, Law, & Mak, 2015) proposed a method to generate a volume projection imaging by using a freehand 3D US system. Based on this projection, curvatures of the spine were calculated using two measurement methods. In both methods, the inflection points along the projection are identified. These points are treated as the most tilted vertebrae. The first method employed the TPs of the most tilted vertebrae to calculate the angle,

while the second uses two pair of SPs from the most tilted vertebrae. These two angle measurements were compared to the Cobb angle from PA X-rays. This approach was tested on 29 subjects with different curvatures. Their results show a high correlation of $R^2 = 0.79$ ($p < 0.005$) and $R^2 = 0.78$ ($p < 0.005$) using the SPs and TPs respectively when compared to the Cobb angle. The examination of the spine is limited to the posteroanterior plane, since it is not possible to determine other landmarks that could provide information of the morphology in the 3D space of the spine.

In another study, Cheung *et al*. (C. W. J. Cheung et al., 2015) evaluated a freehand 3D US system for assessment of scoliosis. Its feasibility was validated by scanning the spine of 28 subjects. To improve the standing stability of the subject during the US sweeps, an adjustable frame support was included in the setup. This support fixed the position of the shoulders and hip. After the acquisition of the tracked US images, the TP and the SP were manually marked, and a 3D model was formed. The model was projected into a 2D plane, simulating the PA plane from X-rays. Then, the Cobb angle was calculated using this projection in an analog manner to the X-rays. This angle was compared to the one measured on X-rays. Their results show a significant correlation between both measurements ($R^2 = 0.86$; $p < 0.001$). Although their results are promising, the landmarking is time consuming and the methods to quantify the curvature of the spine tend to underestimate the deformation.

Young et al. (Young, Hill, Zheng, & Lou, 2015) validated the center of lamina (COL) as a method to approximate the Cobb angle. In this study, 20 subjects with AIS were recruited with a Cobb angle variation between 10° and 45°. X-rays and tracked ultrasound images were acquired from the L5 to C7 vertebrae. Four raters measured the Cobb angle on X-rays and its approximation from the tracked US images. Their results showed an intra-observer correlation between 0.86 to 0.96. They used previous X-rays to improve the landmarking in US, therefore, the correlation agreement between the Cobb angle and the US was high. However, without the use of previous data, the correlation was moderately reliable. The source of error was the limitation to select the end-vertebrae on the US images. They reported that for some patients, it was impossible to find their curves since the landmarks were not visible. This method has

not been in the longitudinal study, where patients with AIS could have progressed or growing up, and the previous radiographs will not correspond with the US, which would make the marking difficult.

Based on the work in (C. J. Cheung, Zhou, Law, & Mak, 2015), Zheng *et al.* (Zheng et al., 2016) evaluated the reliability and validity of a freehand 3D system called Scolioscan. This hardware uses a frontal frame and supports that can be adjusted to fix the position of the chest and hip of the subject depending on his height. The system was tested with 55 participants with AIS. Images of the spine were acquired from L5 to T1 vertebrae using the Scolioscan as well as X-rays. Their results exhibit a reliability in intraclass correlation coefficient (ICC) of 0.88 to 0.97 for angle measurement. Also, the inter-rater ICC from 0.88 to 0.93 indicates a high reliability. The support frame could have helped with the high values in reliability. Since the patients were in a fixed position, the error produced by breathing or the change in position can be reduced. However, there exists an underestimation of the angle. This can be produced for the same fixed position of the patient, as it was shown by Bellefleur et al. (Bellefleur, Dansereau, Koller, & Labelle, 2002). The natural position and the balance of the subject change when the hip or shoulder are fixed. An analysis on the platform versus barefoot should be conducted to eliminate other sources that could influence the posture, therefore the measurements.

Based on the freehand 3D US system used by Zheng *et al.* (Zheng et al., 2016), Brink *et al*. (Brink et al., 2017) tested three methods to measure the curvature of the spine from the coronal plane. These methods were 1) automatic SP angle, 2) manual SP angle and 3) manual TP angle. The angles calculated from these three methods were compared to the Cobb angle, measured from PA X-rays. In this study, 33 patients with AIS were included. As in previous studies (C. W. J. Cheung et al., 2015, 2013; Zheng et al., 2016), the angles calculated from tracked US images were 15%-37% smaller than the Cobb angle. This was because they were calculated on landmarks located in different regions of the vertebrae. However, their results showed that the three methods were reliable. The lowest linear correlation R^2=0.970 was found in the

lumbar curve by using the manual SP angle, while the highest one R^2=0.987 was found for the manual TP angle.

Table 1.3 shows the devices used in this literature review. In most of the studies, the US scanner is different. In the case of the transducer, the linear one is more frequently employed. Also, the MTS is more favored compared to the OTS.

Table 1.3 Hardware used for US image acquisition

Author	Ultrasound scanner	Transducer	Tracking system	Video capture card
Purnama 2009	Not specified	Not specified	OTS	Not specified
Chen 2011	TomoScan Focus Phased Array Ultrasound system (Olympus NDT Inc., Canada)	• Type: Linear (5L64-I1) • Frequency: 5.0 MHz Width: 38.4 mm x 10 mm	Not applicable	Not applicable
Cheung 2013	EUB-8500, Hitachi Ltd., Japan	• Type: Linear • Frequency: 5-10MHz • Width: 92mm	MTS (MiniBird, Ascension Technology Corporation, Burlington, VT, USA)	NIIMAQ PCI/PXI-1411, National Instruments Corporation, Austin, TX, USA
Chen 2013	Olympus TomoScan Focus LT Phased Array instrument (Olympus NDT Inc., Waltham, MA)	• Type: Linear transducer • Frequency: 5-MHz • Width: 64mm x 10mm	Not applicable	Not applicable
Koo 2014	Ultramark 400c; ATL Ultrasound Inc, Bothell, WA	• Type: Linear • Frequency: 6.5-10 MHz • Width: Not specified	OTS (Northern Digital Inc, Waterloo, Canada)	Frame grabber PCI-1411 Data acquisition card (PCI 6024E)
Ungi 2014	Sonix Tablet with GPS extension (Ultrasonix, Richmond, BC, Canada)	• Type: Linear • Frequency: 5 MHz • Width: Not specified	MTS (Ascension, Milton, VT, USA)	Not specified
Cheung 2015	EUB-8500 (Hitachi Ltd., Tokyo, Japan)	• Type: Linear (Hitachi L53L/10–5) • Frequency: 5–10 MHz • Width: 92mm	MTS MiniBird Model 130 (Ascension Technology Corporation, Burlington, VT, USA)	NIIMAQPCI/PXI-1411 (National Instruments Corpora- tion, Austin, TX, USA)
Cheung 2015b	EUB-8500 (Hitachi Ltd., Tokyo, Japan)	• Type: Linear (Hitachi L53L/10–5) • Frequency: 5–10 MHz • Width: 92mm	MTS MiniBird Model 130 (Ascension Technology Corporation, Burlington, VT, USA)	NIIMAQPCI/PXI-1411 (National Instruments Corpora- tion, Austin, TX, USA)
Young 2015	Sonix TABLET	• Type: Convex (C5-2/60 GPS, (Ultrasonix, BC, Canada) • Frequency: Not specified • Width: Not specified	Sonix GPS	Not specified
Zheng 2016	Custom made	• Type: Linear (custom made) • Frequency: 4–10 MHz • Width: of 100 mm	MTS (custom made)	Not specified
Brink 2017	Scolioscan, Telefield Medical Imaging Ltd, Hong Kong	• Type: Linear • Frequency: 7.5MHz • Width: 75mm	MTS (custom made)	Not specified

1.5 Summary

In this chapter, we provided the background necessary to understand the anatomy of the spine. We also described the natural curves of the spine, as well as its deformations. Then we introduced the gold standard methods used in clinical practice to characterize and classify AIS. We finished with a critical review of the state-of-the-art in three specific-objectives identified in the problem statement: 1) computer-based characterization and classification methods in AIS using 3D descriptors, 2) computer-generated models for progression of the curve of the spine through time, and 3) radiation-free imaging systems to reduce the use of X-rays. These three topics are put together to define the main blocks to form a framework. This framework can help clinicians in assessment and follow-up of patients with AIS by reducing the need of X-rays, which is the main objective of this research.

In the context of characterization and classification methods, Cobb angle and Lenke classifications are the main strategies used to diagnose and to treat AIS. However, these methods are based on 2D radiographs, which limits the description of the deformation in the 3D space. Computer-based alternatives to describe the spine in the 3D space have been proposed. However, their lack of interpretability complicates their adaptation into a clinical paradigm. In addition, new classification systems have been developed as an attempt to better categorize spine deformations based on 3D descriptors, rather than visual descriptions from radiographs. Automatic classification systems can find complex patterns to define the categories. The can also reduce the inter- and intra- observer variability associated with current classification systems. Nevertheless, its usability in clinical setups still needs to be tested.

To the best of our knowledge, the studies presented in this review did not address the problem of using easy to adopt and interpret 3D descriptors in clinical setups. Also, there is not a classification model able to dynamically find the best combination of clinical and/or computer-based descriptors to categorize individual curves.

Predicting the likely changes to occur in the spine curve through time could help clinicians in applying a more effective and patient-specific treatment. The studies based on clinical indices have been imprecise predicting the final curve deformation. 3D morphological parameters are promising; however, they must be carefully handcrafted from 3D models. Additionally, their robustness and accuracy still need to be evaluated. Moreover, existent automated models use complex descriptors which are difficult to interpret and to apply in clinical practice.

To the best of our knowledge, a model to predict the variation of the shape of the spine, from the first visit, at different intervals of time, and with 3D easy to interpret computer-based descriptors, remain as a research task.

With respect to the use of a radiation-free imaging modality, ultrasound has come as an alternative to treat patients with scoliosis. On the one hand, there have been many attempts at finding the best landmarks that can be used to calculate an angle to quantify the curvature of the spine similar to the Cobb angle obtained from PA X-rays. On the other hand, only one reconstruction of the spine based on US has been reported in the literature. This consist in tracked ultrasound images used to generate a projection of the spine on the coronal plane. These previous studies only describe the spine with 2D measurements.

To the best of our knowledge, none of the previous studies address the characterization of the spine in the 3D space. A volume reconstruction of the shape of the spine from tracked US, that facilitates the landmarking of the spine in the 3D space has not yet been proposed.

CHAPTER 2

DYNAMIC ENSEMBLE SELECTION OF LEARNER-DESCRIPTOR CLASSIFIERS TO ASSESS CURVE TYPES IN ADOLESCENT IDIOPATHIC SCOLIOSIS

Edgar García-Cano[1], Fernando Arámbula-Cosío[2], Luc Duong[1], Christian Bellefleur[3], Marjolaine Roy-Beaudry[3], Julie Joncas[3], Stefan Parent[3], Hubert Labelle[3]

[1] Software and IT Engineering Department,
École de technologie supérieure

[2] Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas,
Universidad Nacional Autónoma de México

[3]Research Center, Sainte-Justine Hospital

This article was published in *Medical & Biological Engineering & Computing*, Springer, in June 2018. DOI: 10.1007/s11517-018-1853-9

2.1 Abstract

While classification is important for assessing adolescent idiopathic scoliosis (AIS), it however suffers from low interobserver and intraobserver reliability. Classification using ensemble methods may contribute to improving reliability using the proper 2D and 3D images of spine curvature features. In this study, we present two new techniques to describe the spine, namely, *leave-n-out* and *fan leave-n-out.* Using these techniques, three descriptors are computed from a stereoradiographic 3D reconstruction to describe the relationship between a vertebra and its neighbors. A dynamic ensemble selection method is introduced for automatic spine classification. The performance of the method is evaluated on a dataset containing 962 3D spine models categorized according to three curve types. With a log loss of 0.5623, the dynamic ensemble selection outperforms voting and stacking ensemble learning techniques. This method can improve intraobserver and interobserver reliability, identify the best combination of descriptors for characterizing spine per curve type, and provide assistance to clinicians in the form of information to classify borderline curvature types.

Keywords: Spine Classification, Descriptors of the Spine, Adolescent Idiopathic Scoliosis, Dynamic Ensemble Selection, Machine Learning

2.2 Introduction

Adolescent Idiopathic Scoliosis (AIS) is a medical condition involving a 3D spinal deformity. It causes the shape of the spine on the posteroanterior plane to take an "S" or a "C" form, instead of a straight line. AIS affects between 1 and 3% of the population, and 1 out of 1000 patients will require a surgical treatment. When treating scoliosis, it is crucial to find out the characteristics that best describe each specific deformity in order to provide patients with an optimal treatment, as well as to monitor their progress. To date, the evaluation of the spine has relied mainly on observations of posteroanterior and sagittal radiographs, which constitute the most common imaging modality used in clinical practice to observe the spine in a standing position. The Cobb angle is the standard measurement of scoliosis severity in radiographs, and is a 2D index that measures the curvature of the spine (Stokes, 1994a). The analysis of these radiographs to assess a diagnosis is not deterministic and may vary from expert to expert.

Two classification models have been used in an effort to identify curve patterns. In 1983, the King classification (King et al., 1983) model was proposed, and described five thoracic curve classes. Its low intra- and interobserver reliability and limitation to only thoracic curves were its principal disadvantages. In 2001, the Lenke classification of AIS (Lenke et al., 2001) emerged as a new paradigm to characterize the deformation of the spines, and it became the most clinically accepted and widely used model. This paradigm classifies scoliosis into six types (Lenke 1-6). It relies on 2D radiographs of the posteroanterior and sagittal planes and Cobb angle measurements, which are used to discern structural and nonstructural curves in the proximal thoracic, main thoracic, and thoracolumbar/lumbar regions. It also provides treatment recommendations according to the type. Although the Cobb angle and Lenke classifications are the prevailing primary strategies used to define and treat scoliosis, these methods do not allow a full understanding of the deformity.

Computer-based classification models have been used to study AIS. Phan et al. (Phan, Mezghani, Wai, De Guise, & Labelle, 2013) proposed the use of self-organizing maps based on eight Cobb angle measurements to classify AIS and highlight treatment patterns. Clustering techniques, such as ISO Data, K-means, Fuzzy k-means, K-means++, have been used to propose new categorizations (Duong et al., 2006; Duong, Mac-Thiong, Cheriet, & Labelle, 2009; Kadoury & Labelle, 2012; Kadoury et al., 2014; Sangole et al., 2009; Shen et al., 2014; Thong et al., 2015). These methods use different descriptors of the spine, such as parametric 3D curve representation (Duong et al., 2006), 3D reconstructions (Kadoury & Labelle, 2012; Sangole et al., 2009; Thong et al., 2016, 2015), plane of maximum curvature (Duong, Mac-Thiong, et al., 2009; Sangole et al., 2009), best fit plane (Duong, Mac-Thiong, et al., 2009), torsion estimator (Duong, Mac-Thiong, et al., 2009; Kadoury et al., 2014; Shen et al., 2014), Cobb angle (Sangole et al., 2009), and axial rotation (Sangole et al., 2009). Some of these works use techniques, such as wavelet-based, non-linear manifold and stacked auto-encoders (Duong et al., 2006; Kadoury & Labelle, 2012; Thong et al., 2016, 2015), to reduce high dimensionality as a pre-step before clustering. The importance of these classification systems resides in their attempt to better categorize the severity and progression of AIS to allow better treatments. However, although these descriptors could complement the Lenke classification, there is no consensus among experts on which of them to use in everyday clinical practice. The descriptors used in the studies referenced are based on the choices made by researchers on how to tackle a specific clinical problem (Donzelli et al., 2015).

In this paper, we present the first approach for an automated classification of spinal deformities based on a dynamic ensemble selection of descriptors to characterize the spine. A set of eight descriptors are employed, from which we propose the *leave-one-out* and the *fan leave-one-out* angle measurement techniques used to calculate three different descriptors, one based on the existing Cobb angle measurement, and the others based on two new proposed variations of angle calculations among the vertebrae. Considering that each descriptor characterizes different aspects of the spine, the objective of this work is to contribute through new descriptors to characterize the spine, as well as through a computer-based model to assist clinicians in the classification of spine deformities. Three classes are defined for this task: 1) spines with a main

thoracic (MT) major curve and a non-structural thoracolumbar/lumbar (TL/L) curve; 2) spines with an MT major curve, and a structural TL/L curve; and 3) spines with a TL/L major curve and a non-structural proximal thoracic (PT) curve. The dynamic ensemble selection is carried out using Random Forest (RF) as base classifier.

2.3 Methods

2.3.1 Dataset

A dataset consisting of 962 3D spine models provided for this work by the *Centre hospitalier universitaire Sainte-Justine* (Sainte-Justine University Hospital Center) in Montreal, Canada, was used. This dataset contained 3D spine models reconstructed from stereoradiographic X-rays, as described in (Delorme et al., 2003). Each 3D model in the dataset consisted of at least 17 vertebrae (T1 to L5). For each vertebra, six points (superior and inferior endplates; left and right superior pedicles, and left and right inferior pedicles) were identified by a trained technician. The dataset was divided into three main classes, as described in Table 2.1**,** and their distribution is shown in Table 2.2. Two criteria were used to define a curve as structural or non-structural: 1) the mayor curve is always defined as structural, and 2) a minor curve that bends less than 25° on side bending radiographs is always non-structural.

Table 2.1 Description of the classes in the dataset

	Proximal thoracic	**Main thoracic**	**Thoracolumbar/Lumbar**
Class 1	Structural/non-structural	Major curve	Non-structural
Class 2	Structural/non-structural	Major curve	Structural
Class 3	Non-structural	Structural/non-structural	Major curve

Table 2.2 Distribution of the dataset

	Class 1	**Class 2**	**Class 3**
Samples	329 (34%)	327 (34%)	306 (32%)

2.3.2 Descriptors of the spine

A descriptor is a measurement that characterizes the spine. Clinical measurements were not available in the dataset, and therefore, a set of eight descriptors were automatically estimated for our experiments (see Table 2.3).

Table 2.3 Descriptors generated from a dataset of 3D spine models

Name	Description	Number of features
Descriptor 1	Cartesian coordinates in yz (posteroanterior plane)[a]	204
Descriptor 2	Cartesian coordinates in xz (sagittal plane)[a]	204
Descriptor 3	Cartesian coordinates in xy (transverse plane)[a]	204
Descriptor 4	Axial rotation of each vertebra[b]	17
Descriptor 5	The first and second derivatives of the centroid of each vertebra[c]	34
Descriptor 6	*Leave-n-out* angles estimated with n=0 to 10 (see section 2.3.2.1)	108
Descriptor 7	*Leave-n-out* angles estimated with n=0 to 17 (see section 2.3.2.1)	136
Descriptor 8	*Fan leave-one-out* angles estimated with n=0 to 17 (see section 2.3.2.2)	64

[a] Coordinates extracted directly from 3D models of spines.

[b] Calculated by applying the method proposed by Stokes et al. (Stokes, Bigalow, & Moreland, 1986).

[c] Method used by Duong et al. (Duong, Cheriet, & Labelle, 2010) to deduce curves in King's classification.

The first five descriptors were estimated straightforwardly. Descriptors 1 to 3 correspond to the normalized values of the coordinates in each plane (6 points for each of the 17 vertebrae). Descriptor 4 corresponds to the axial rotation of each vertebra, and Descriptor 5 was estimated using the first and second derivatives of the centroids of the vertebrae. To estimate descriptors 6 to 8, we are proposing two new techniques named *leave-n-out* and *fan leave-n-out* angles.

2.3.2.1 Leave-n-out angles

The general idea behind this technique is to take advantage of the spine being a sequence of vertebrae to automatically estimate the angle that one vertebra has with respect to its neighbors. For example, considering the sequence of vertebrae from T1 to T4, the *leave-**0**-out angles* of

this sequence would be computed as the angles between the centroids of T1 and T2, T2 and T3, and T3 and T4 (see Figure 2.1a); the *leave-**1**-out angles* would be the angles between the centroids of T1 and T3, and T2 and T4 (skipping **one** vertebra in between; see Figure 2.1b); while the *leave-2-out angles* would be the angles between the centroids of T1 and T4 (skipping two vertebrae in between; see Figure 2.1c), and so forth. The assumption is that the inter-vertebrae angles of spines with similar curve types will have similar measurements.

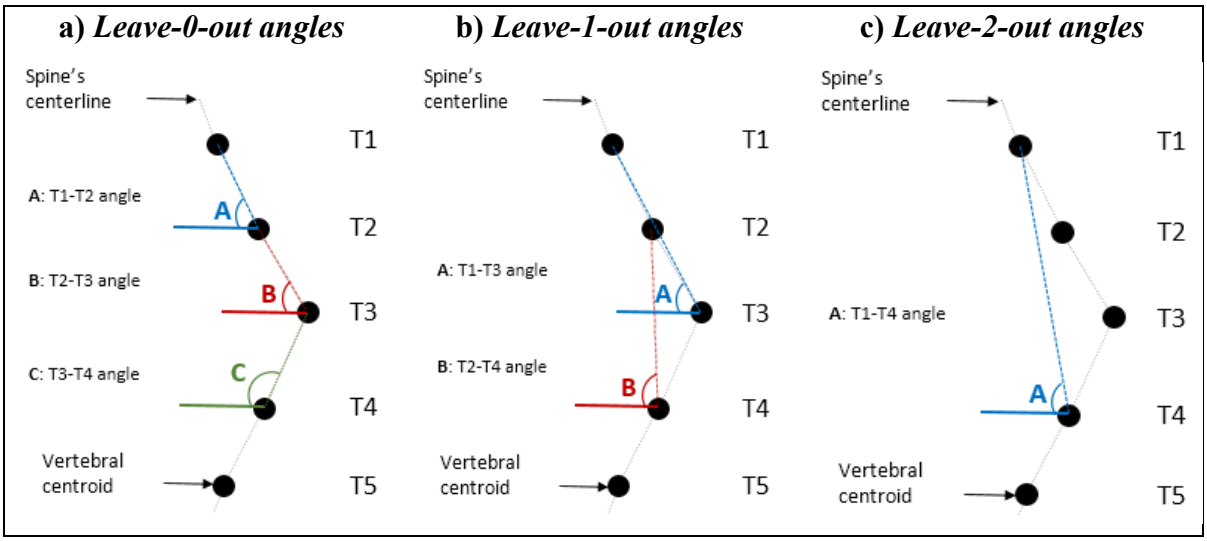


Figure 2.1 Leave-n-out angles calculation with respect to the horizontal axis, with n=0 to n=2

Based on this idea, two descriptors are proposed. Descriptor 6 estimates the *leave-n-out* angles using the Cobb angle method. For descriptor 7, the angles are estimated from the line connecting two vertebral centroids with respect to the horizontal axis (see Figure 2.1).

2.3.2.2 Fan leave-n-out angles

Based on the notion of the *leave-n-out* angles, we propose an estimation of the inter-vertebrae angle relationship among three vertebrae centroids, e.g., T1, T2 and T3, by estimating the angle formed between two lines, T1-T2 and T2-T3, forming a fan-like shape (see Figure 2.2). Considering the sequence of vertebrae from T1 to T5, the *fan leave-0-out angles* of this sequence would be computed as the angles formed by the line segments T1-T2 and T2-T3, T2-

T3 and T3-T4, and T3-T4 and T4-T5 (see Figure 2.2a). In the T1 to T6 sequence, the *fan leave-1-out angles* would be computed as the angles formed by the line segments T1-T3 and T3-T5; and T2-T4 and T4-T6 (leaving **one** vertebra in between for each line segment; see Figure 2.2b).

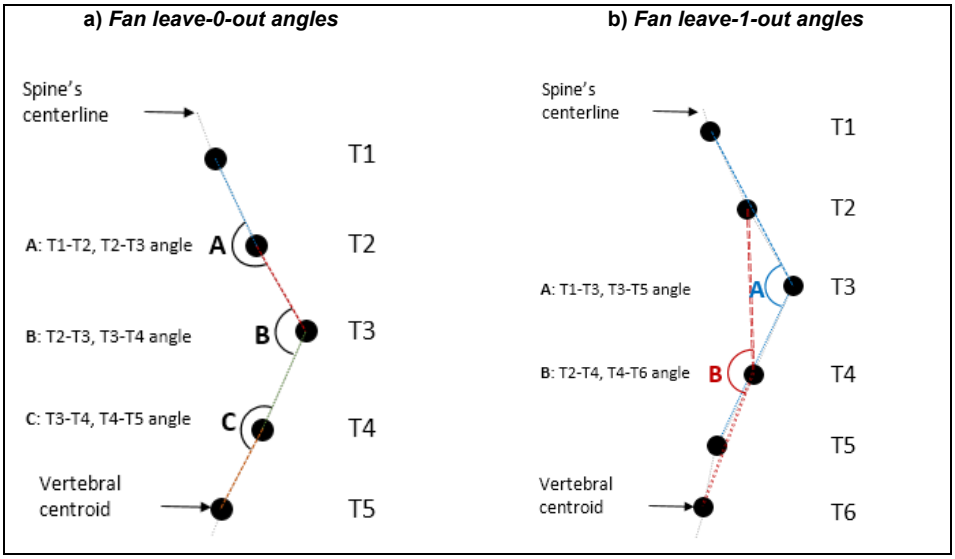


Figure 2.2 Fan Leave-n-out angles calculation for n=0, and n=1

2.3.3 Ensemble learning

Given the great diversity of models used by machine learning algorithms to learn from data, the results obtained from applying these algorithms on the same dataset may differ significantly from one case to the other. In some cases, one algorithm can perform better than others for certain types of patterns or classes, but not sufficiently for others. Hence, it is to be expected that in some cases, a single classifier will not be able to cover the whole variability in a dataset. A multiple classifier system consists of a set of different classifiers that combine their individual decisions into a more accurate and robust consensual prediction (Woźniak, Graña, & Corchado, 2014). An example of this type of system is RF, which combines different decision tree classifiers to perform the final prediction. In this combination model, it is assumed that the classifiers are independent and that the errors made by the individual classifiers are uncorrelated (Dietterich, 2000). Britto et al. (Britto, Sabourin, & Oliveira, 2014) decompose a multiple classifier system into three phases: generation, selection and integration.

In the *generation* phase, a *pool* of base classifiers is produced. A pool combining base classifiers which use a single learning algorithm, with each base classifier trained with different subsets of the training data, is known as a homogenous pool. On the other hand, where the base classifiers use different learning algorithms, with all of the former trained with the same data, we have a heterogeneous pool.

In the *selection* phase, the best base classifiers are selected from the pool. There are two different types of selection, static and dynamic. In static selection, the best classifiers are chosen during the training phase, based on how competent they are in discriminating. In contrast, with dynamic selection, the classifiers are selected during the test phase. A complete review of the dynamic selection of classifiers is available in (Britto et al., 2014).

The *integration* phase consists in combining the selected classifiers to categorize each test sample, under a determined strategy (Kittler, Hatef, Duin, & Matas, 1998). Woźniak et al. (Woźniak et al., 2014) describe this phase in three strategies: 1) *class label fusion*, where a majority voting scheme is applied; 2) *support function fusion*, which provides a decision based on the estimated likelihood for each class computed by all individual classifiers; and 3) *trainable fuser*, which involves a learning process to reach a consensus.

2.3.3.1 Dynamic ensemble selection of learner-descriptor classifiers

For our work, we selected the dynamic ensemble selection (DES) method to combine descriptors that are routinely used in clinical settings with other computational methods (Donzelli et al., 2015), in order to find the most suitable descriptors for characterizing a spine. This method consists in dynamically finding the best learner-descriptor classifiers (LDCs) that can be brought together as an ensemble to assess the curvature type of a spine. Here, a *descriptor* is a measurement that characterizes the spine, and a *learner* refers to the algorithm used to find patterns in the data. Since the idea behind DES is to select the most accurate classifiers for each pattern, a decision was made to train one learning algorithm per descriptor,

with the aim of determining the best LDC or their combination to define a particular type of deformity.

The method proposed in this work is an adaptation of the *K-nearest-oracle*, with the *eliminate* fusion scheme of the best classifiers (KNORA-E) proposed by Ko et al. (Ko, Sabourin, & Britto, 2008). In KNORA-E, the dataset is divided into three parts: training (*Tr*), validation (*Va*) and test (*Te*) sets. It has a pool of base classifiers (trained with a *Tr* set) that evaluates a test sample *t* (from the *Te* set) as follows: if the entire pool reaches a unanimous decision, the test sample *t* is labeled accordingly. Otherwise, the k-nearest neighbors (*knn*) to the test sample are found in the *Va* set using the Euclidean distance. The *knn* are classified by each algorithm in the pool. The classifiers that correctly categorize all *knn* are brought together as an ensemble to perform the final classification of the test sample.

Our proposal includes three modifications to the KNORA-E, while still maintaining its essence. First, the base algorithm for the LDCs is RF, and each LDC is trained with a different descriptor (see Table 2.3). Second, during the training step, a ranking of the best LDCs per class is performed. For this strategy, a 3-fold cross-validation (CV) is computed, and the confusion matrix is obtained in each fold. From the confusion matrix, the accuracy for every class is calculated. At the end of the CV, the 3-fold mean accuracy per class is obtained for each LDC. Then, one list per class is created, and the LDCs are ranked from highest to lowest, in accordance with their mean accuracy in *Tr*. The lists will be used to select the best classifiers for the dynamic ensemble in the test phase. Third, only the nearest neighbor is used as a reference from the *Va* set.

A description of the steps to follow to perform the dynamic ensemble selection of learner-descriptor classifiers is presented in Algorithm 2.1.

Algorithm 2.1 Dynamic ensemble selection of learner-descriptor classifiers (LDCs)

Input: Set of *d* descriptors calculated from the original set of spines.
Output: *Te* set labeled by class according to the dynamic ensemble of LDC.
Method:

1. For each descriptor, the samples are divided into three sets: *Tr*, *Va* and *Te*.
2. An LDC is trained with each *Tr* set of every descriptor, generating one pool.
3. Compute a list of ranked LDCs per class using 3-fold CV.
4. For each test sample *t* in the *Te* set:
    a. Obtain the decision of each LDC in the pool.
    b. If all LDCs coincide, the test sample *t* is labeled accordingly.
    c. Else
        i. Look in the *Va* set for the nearest neighbor of *t*.
        ii. Using the LDCs ranking list for the *nn* class, select the *m* highest ranked LDCs (*m* is specified by the user).
        iii. Classify the *nn* using the selected LDCs.
        iv. If at least one LDC correctly classified the *nn:*
            1. Generate the ensemble with the LDCs that correctly classified the *nn.*
            2. Classify *t* according to the maximum sum of the predicted probabilities of the ensemble.
        v. Else
            Use the highest ranked LDC for the *nn* class to classify *t*.

Three scenarios are possible when classifying a test sample (step 4, Algorithm 2.1). In the first scenario, all the LDCs agree on the class type; hence, the classification is unanimously decided (step 4.b in Algorithm 2.1). In the second scenario, not all the LDCs agree, and therefore, the *nn* must be found in the *Va* set. The best *m* highest ranked LDCs capable of categorizing the *nn* form an ensemble that predict the class of the *t* (step 4.c.iv Algorithm 2.1). Table 2.4 shows an example of a classification of a test sample with *nn* and the best 3 highest ranked LDCs per class. Table 2.4a displays the ranking of the LDCs per class. For this example, only the best three LDCs per class are used during the testing phase. Table 2.4b exhibits the LDCs used to classify the *nn,* which belongs to class 1*,* and their predictions. *LDC_1* is within the first three places in the ranked list for class 1, hence it performs a prediction. In the case of *LDC_2*, it is not in the first three places in the ranked list for class 1, and therefore, it is discarded. At the end, the LDCs that correctly categorize the *nn* (*LDC_1*, *LDC_3*) are used to form the ensemble. In the third scenario, there is neither unanimity nor an ensemble capable of making a prediction. This leads us to use the highest ranked LDC for the *nn* class to classify *t*.

Table 2.4 Example of the analysis of one test sample using the 3 highest ranked LDCs per class. a) displays the rank of 6 LDCs per class, computed during the training phase. b) shows the classifiers that were employed to perform a prediction on the nn, which belongs to class 1 *(LDC_1, LDC_2 and LDC_3)*. Then, the DES ensembles LDC_1 and LDC_3 to perform the prediction of the test sample, which were the ones that predicted correctly the *nn*. The crossed LDCs were discarded for the prediction.

a. Rank of LDC per class

Class 1	Class 2	Class 3
LDC_1	LDC_4	LDC_5
LDC_5	LDC_5	LDC_1
LDC_3	LDC_6	LDC_4
~~LDC_4~~	~~LDC_1~~	~~LDC_6~~
~~LDC_2~~	~~LDC_3~~	~~LDC_3~~
~~LDC_6~~	~~LDC_2~~	~~LDC_2~~

b. Ensemble

	True label (nn)	Prediction	Ensemble?
LDC_1	1	1	Yes
~~LDC_2~~	-	-	-
LDC_3	1	1	Yes
~~LDC_4~~	-	-	-
LDC_5	1	3	No
~~LDC_6~~	-	-	-

2.4 Experimental setup

2.4.1 Base learning algorithm

Random Forests is an ensemble of classifier algorithms built on decision trees. It was first introduced by Breiman (Breiman, 2001), and was selected as the base learning algorithm for this work. It is an ensemble classification method, which means that it is composed of multiple decision trees that are encapsulated in a single classifier. Every decision tree is a learner that votes on a category for a sample. The category with the most votes is then chosen as the final classification for that sample. RF has been widely used, and its performance has been proven in various automatic classification problems (Fernández-Delgado, Cernadas, Barro, & Amorim, 2014).

2.4.2 Feature selection

Eight descriptors were generated from the original dataset. They were standardized to the (−3:3) range for experimentation. We took advantage of RF to compute feature selection and reduce dimensionality based on the importance of the features. Algorithm 2.2 describes the steps for extracting the best features for one descriptor, and this procedure is repeated for each of them.

Algorithm 2.2 Selection of the best features

Input: 1 descriptor.
Output: 1 descriptor with only best features selected.
Method:
1. Generate 10 folds.
    1.1. Train one learner with 9 folds performing inner cross-validation of the hyper-parameters based on the internal out-of-bag error.
    1.2. Select the best features based on their importance scores.
        1.2.1. Compute the scores.
        1.2.2. Generate 15 intervals to select the features between min and max values of the sorted importance scores.
        1.2.3. Train and evaluate a new model with the best features selected for each interval.
        1.2.4. Select the best features with the highest accuracy scores.
2. Return the reduced version of the descriptor with the features that were selected during the 10 folds.

2.4.3 Classification

The DES was used to perform the classification. Its performance was compared to two other ensemble learning techniques. The first technique was based on stacking. It combined multiple classification learners via a meta-classifier. In the technique, one base learner is trained on one of the eight descriptors, (eight base learners in total). Then, the meta-classifier is trained with the outputs of the base learners as features and is used for the final prediction. The second technique was based on voting ensembles. In this as well, eight base learners are trained with one of the eight descriptors, and the final prediction is made based on the argmax of the sums of the predicted probabilities. Two metrics were employed to evaluate all three ensemble

learning techniques, accuracy and log loss. Algorithm 2.3 shows the steps involved in the classification task of each learning technique.

Algorithm 2.3 Evaluation process for an ensemble learning technique

Input: 8 descriptors.

Output: The accuracy and log loss scores of a 10-fold cross-validation for the ensemble.

Method:

1. Generate 10 folds.
   1.1. Train the ensemble with 9 folds performing inner cross-validation of the hyper-parameters based on the internal out-of-bag error.
   1.2. Test the ensemble learning technique with the fold that was left out.
2. Return the accuracy and log loss scores of each of the 10 folds.

2.5 Results

2.5.1 Feature selection

The features were selected before the classification, based on the procedure described in Algorithm 2.2. The set of features with the highest accuracy were the ones chosen for experimentation. Table 2.5 shows the final dimension of each descriptor.

Table 2.5 Feature selection based on importance scores

Name	Original number of features	Reduced number of features
Descriptor 1	204	204
Descriptor 2	204	51
Descriptor 3	204	204
Descriptor 4	17	17
Descriptor 5	34	17
Descriptor 6	108	89
Descriptor 7	136	86
Descriptor 8	64	25

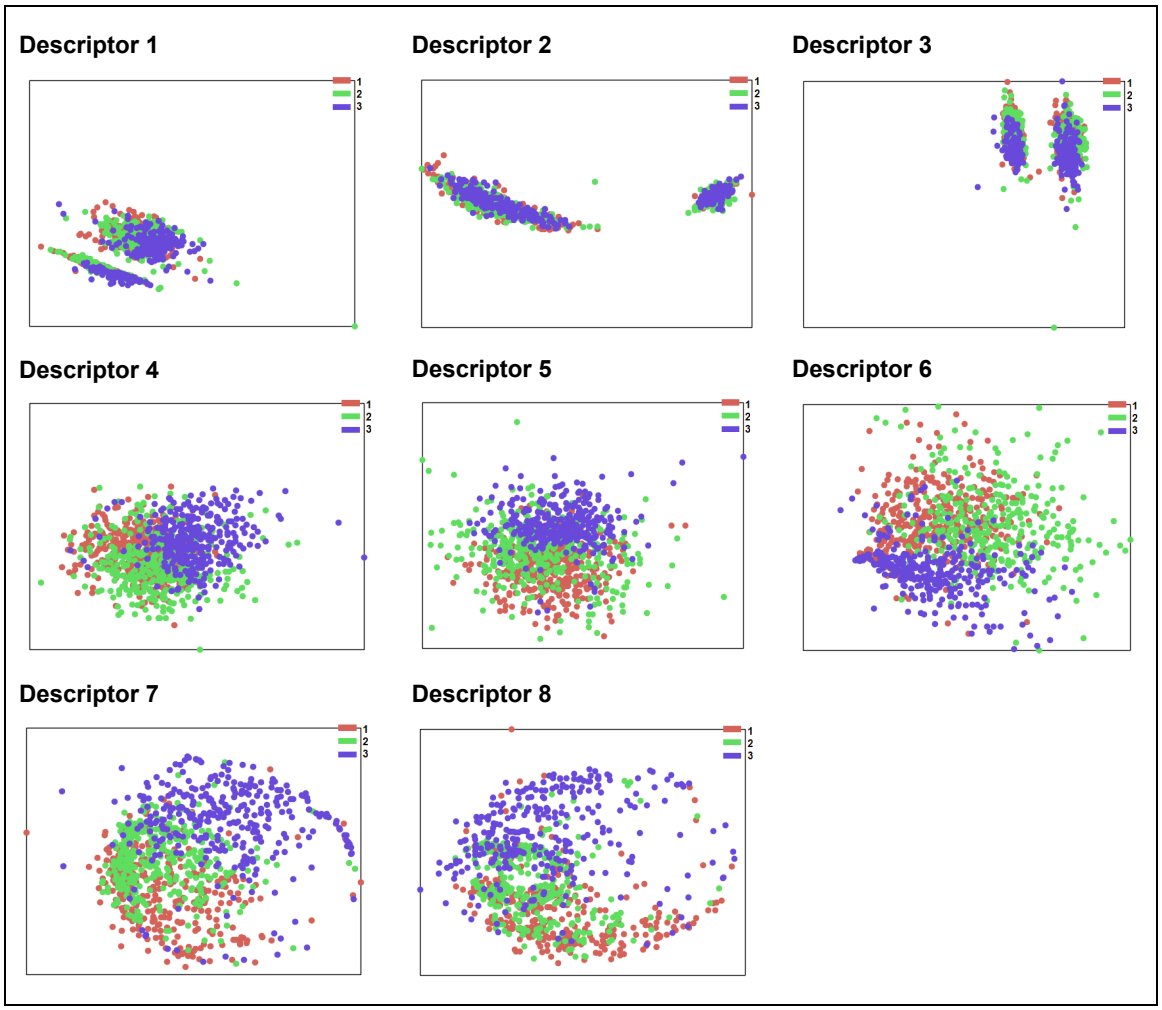


Figure 2.3 Visualization of the best features of each of the eight descriptors employing multidimensional scaling

2.5.2 Classification

The performance of the algorithms was tested using a 10-fold cross-validation (see Algorithm 2). The DES employed the best four LDCs per class obtained in *Tr* to evaluate the *nn* in the test phase. Table 2.6 shows the most effective LDCs. The results of the classification are shown in Tables Table 2.7 and Table 2.8.

Table 2.6 Best descriptors employed in the test phase of the DES (steps 4.c.i to iv in Algorithm 2.1)

	LDC_1	LDC_2	LDC_3	LDC_4	LDC_5	LDC_6	LDC_7	LDC_8
Class 1	0	0	1	0	9	10	10	10
Class 2	1	0	2	1	10	8	9	9
Class 3	0	0	1	0	10	9	10	10

Table 2.7 Accuracy of the classification

Folds	Voting	Stacking	DES
1	0.7423	0.7423	0.7732
2	0.7113	0.6598	0.6907
3	0.8041	0.8041	0.8247
4	0.8351	0.8041	0.8041
5	0.7423	0.7113	0.7423
6	0.8041	0.8144	0.8144
7	0.6979	0.7188	0.7396
8	0.7368	0.7368	0.7789
9	0.8105	0.7789	0.7789
10	0.8298	0.8298	0.8191
Avg.	0.7714	0.7600	**0.7766**

Table 2.8 Log loss of the classification

Folds	Voting	Stacking	DES
1	0.5966	1.6358	0.5808
2	0.6792	1.7748	0.6575
3	0.5273	0.8891	0.4852
4	0.5438	0.4779	0.4886
5	0.6116	1.2755	0.6249
6	0.5651	1.5285	0.5405
7	0.6520	1.6699	0.6188
8	0.6016	0.9350	0.5255
9	0.5656	0.8599	0.5588
10	0.5532	1.1493	0.5429
Avg.	0.5896	1.2196	**0.5623**

Table 2.9 Descriptive statistics of accuracy and log loss

	Mean	Std. Deviation	Minimum	Maximum
Accuracy				
Voting	0.7714	0.0506	0.6979	0.8351
Stacking	0.7600	0.0548	0.6598	0.8298
DES	0.7766	0.0425	0.6907	0.8247
Log loss				
Voting	0.5896	0.0484	0.5273	0.6792
Stacking	1.2196	0.4291	0.4779	1.7748
DES	0.5623	0.0578	0.4852	0.6575

2.6 Discussion

We presented two new techniques, leave-one-out and the fan leave-one-out, to calculate three descriptors for use in characterizing the spine. These descriptors calculate the relation of one vertebra and its neighbors in three ways: applying the Cobb angle method, using the line that connects two vertebral centroids with respect to the horizontal axis, and the angle formed between two-line segments that connect three vertebral centroids forming a fan-like shape. In addition, we employed the dynamic ensemble selection approach to perform an automatic classification of spines with scoliosis. The classification aims not only to distinguish among three classes, but also to find the LDC or sets of them that best describe each spine. Our objective is to provide clinicians with new descriptors and computational assistance for an impartial and consistent characterization of spine deformations.

We performed a feature selection process to reduce dimensionality. Table 2.5 shows that Descriptors 2, and 5 to 8 reduced their size significantly, while Descriptors 1, 3 and 4 maintained their size. This reduction was based on the importance of their features.

Figure 2.3 shows that Descriptors 4 to 8 seem to provide the best partition of the classes. To perform the classification, the best four LDCs per class, ranked during the training, were used in the test phase to generate the dynamic ensemble. Since the DES depends on the *nn* to perform the prediction, there is no general ensemble that works for all types of curvatures. In Table 2.6**,** LDCs 5 to 8 are notably the most relevant. This is almost in agreement with the visualization in Figure 2.3. While LDCs 1 to 4 were practically disregarded during the dynamic ensemble, our proposed LDCs, 6, 7, and 8, were always in the list of the best ranked LDCs per class and were consistently used to generate the final ensemble.

We compared the performance of the DES classification against two ensemble learning techniques, stacking and voting. Tables Table 2.7 and Table 2.8 show the performance of the classifiers with two metrics, accuracy and log loss, respectively. Table 2.9 shows the descriptive statistics for the two metrics, accuracy and log loss. At first glance, DES slightly

outperforms voting and stacking in both metrics. In accuracy, DES (**0.7766)** has a small advantage when compared to voting (**0.7714**) and stacking (**0.7600**). Regarding log loss, the advantage of DES (**0.5623**) is higher when compared to stacking (**1.2196**), and more modest in comparison to voting (**0.5896**).

To evaluate the significance of these differences, the non-parametric Friedman's test was applied. A null-hypothesis was formulated, stating that the DES outperforms the voting and stacking ensemble methods. The results of the test are shown in Table 2.10. The following parameters were used: level of significance, $\alpha = 0.05$; corresponding critical value, $p_\alpha = 5.99$; and 2 degrees of freedom.

Table 2.10 Results of Friedman's test, considering 2 degrees of freedom, significance level α = 0.05, and critical value pα = 5.99

Average ranks					
Metric	**Voting**	**Stacking**	**DES**	**F_t**	**Null-hypothesis**
Accuracy	2.05	1.60	2.35	0.205	Rejected
Log loss	2.00	2.80	1.20	0.002	Accepted

When comparing the three ensemble methods using the accuracy metric, there is no significant difference between the three approaches. However, when using the log loss metric, the difference is significant (F_t=0.002).

To provide a one-to-one comparison of the log loss metric among the classifiers, a post hoc analysis with the Wilcoxon signed-rank test was conducted with a Bonferroni correction applied on a pairwise combination. The significance level was set to $p < 0.017$. The performance between all three approaches was significantly different, as seen in Table 2.11.

Table 2.11 Results of Wilcoxon Sign-Rank Test, on the results using the log loss metric, with a Bonferroni correction and significance level $p < 0.017$

	Stacking vs. Voting	**DES vs. Voting**	**DES vs. Stacking**
Z	-2.701	-2.497	-2.701
Significance	0.007	0.013	0.007

The log loss metric evaluates the performance of a classifier by penalizing misclassifications. Based on the uncertainty of the prediction, it quantifies how much the prediction differs from the correct class. Log loss tends to zero as predictions are closer to being correct. In contrast, the accuracy metric only estimates the percentage of correct classifications. In our view, the log loss metric provides a more thorough comparison between classification techniques.

The DES outperformed the other algorithms despite being at a small disadvantage: it sacrifices 20% of the training data to create the *Va* set where the *nn* is searched.

2.7 Conclusions

Our main contribution in this work is to present the leave-n-out and the fan leave-n-out angle measurement techniques, which automatically determine the position of each vertebra with respect to its neighbors. Three different descriptors were obtained to characterize the curvature of the spine using these techniques. The first descriptor was based on the Cobb angle, a clinical measurement used in the classification of AIS. The second and third descriptors were two new proposed variations of angle calculations carried out among vertebrae. In our experiments, these descriptors were the most relevant on the DES, being the most frequently picked for performing the classification.

When performing automatic classifications of spine deformities, each type of deformation may be characterized by a specific descriptor or combination of them. In this paper, we proposed the use of a classifier based on a dynamic ensemble selection. For our implementation, we also

proposed an adaptation of the KNORA-E to incorporate the ranked lists of the best LDCs that classify the nearest neighbor of each sample.

Our proposed approach does not depend on a specific learning algorithm or descriptor, but rather, it dynamically selects the best combination of LDCs for each individual case. To this end, the dynamic ensemble selects the specific LDCs that have proven to best characterize similar curvatures, instead of using a learner that has been generically trained with many different cases that may or may not be related to the one at hand. This versatility offers the opportunity to combine well-known clinical-based descriptors with computer-generated ones to provide a more intuitive and consistent insight into a specific curvature type, allowing the user to analyze the role that each LDC plays during the classification. To the best of our knowledge, this is the first attempt at dynamically combining independent descriptors of the spine to perform a classification.

Despite sacrificing 20% of the training data to create the validation set to find the nearest neighbor for each sample, in our experiments, the DES method modestly outperformed the stacking and voting ensemble techniques in terms of the log loss metric. However, the potential of a DES approach for characterizing spine deformities extends beyond just classification applications. Since the combination of certain learning algorithms with particular descriptors allows for a better discernment of specific curvature types, these combinations could be analyzed to determine how they complement each other in order to improve the study of a particular type of deformation. This could also help clinicians to extend their current gold standard methods with 3D descriptors. In addition, regarding the best discerning descriptors, a 3D sub-classification of AIS could be performed by using a clustering technique.

There are two disadvantages to consider when using DES. First, it requires a large quantity of data since it sacrifices a percentage of the data for the Va set. Second, DES is a time-consuming algorithm because it has to find the nn, evaluate and classify it, and then perform the ensemble to classify each test sample. Hence, it is not recommended for use in real-time applications.

For this work, no clinical measurements were available to use as descriptors. As part of our future work, we will be experimenting with clinical descriptors in a bid to improve the performance of our DES. Additionally, we consider that a bigger dataset could improve the performance of the dynamic ensemble, which sacrifices a significant portion of data to generate the validation subset, thus reducing the amount of data available for training.

CHAPTER 3

PREDICTION OF SPINAL CURVE PROGRESSION IN ADOLESCENT IDIOPATHIC SCOLIOSIS USING RANDOM FOREST REGRESSION

Edgar García-Cano[1], Fernando Arámbula-Cosío[2], Luc Duong[1], Christian Bellefleur[3], Marjolaine Roy-Beaudry[3], Julie Joncas[3], Stefan Parent[3], Hubert Labelle[3]

[1] Software and IT Engineering Department,
École de technologie supérieure

[2] Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas,
Universidad Nacional Autónoma de México

[3]Research Center, Sainte-Justine Hospital

This article was published in *Computers in Biology and Medicine*, Elsevier, in October 2018.
DOI: 10.1016/j.compbiomed.2018.09.029

3.1 Abstract

Background: The progression of the spinal curve represents one of the major concerns in the assessment of Adolescent Idiopathic Scoliosis (AIS). The prediction of the shape of the spine from the first visit could guide the management of AIS and provide the right treatment to prevent curve progression.

Method: In this work, we propose a novel approach based on a statistical generative model to predict the shape variation of the spinal curve from the first visit. A spinal curve progression approach is learned using 3D spine models generated from retrospective biplanar X-rays. The prediction is performed every three months from the first visit, for a time lapse of one year and a half. An Independent Component Analysis (ICA) was computed to obtain Independent Components (ICs), which are used to describe the main directions of shape variations. A dataset of 3D shapes of 150 patients with AIS was employed to extract the ICs, which were used to train our approach.

Results: The approach generated an estimation of the shape of the spine through time. The estimated shape differs from the real curvature by 1.83, 5.18, and 4.79 degrees of Cobb angles in the proximal thoracic, main thoracic, and thoraco-lumbar lumbar sections, respectively.

Conclusions: The results obtained from our approach indicate that predictions based on ICs are very promising. ICA offers the means to identify the variation in the 3D space of the evolution of the shape of the spine. Another advantage of using ICs is that they can be visualized for interpretation.

Keywords: Prediction of spinal curve progression, Adolescent Idiopathic Scoliosis, Independent Component Analysis, Machine Learning, Random Forest

3.2 Introduction

Adolescent idiopathic scoliosis (AIS) is a complex 3D deformation of the spine which looks like an “S” or “C” shape from the posteroanterior plane. It is called idiopathic because its cause is unknown. It is the most common type of scoliosis, with a high prevalence in adolescents between 10 and 18 years of age. AIS affects between 1 and 4% of adolescents, mainly females (Cheng et al., 2015). In a meta-analysis, Cheng *et al.* (Fong et al., 2010) showed that the global prevalence of AIS with the main curvature ≥ 10° was 1.34%. Currently, the evaluation of the spine relies mainly on the observation of conventional posteroanterior and sagittal X-rays, which constitute the most common imaging modalities for observing the spine in a standing position in clinical practice.

The Cobb angle represents the gold standard method for measuring the curvature of the spine. Its measurement is based on the most tilted vertebrae, at the top (upper vertebra) and at the bottom (lower vertebra) of the curve. The angle is formed by the line parallel to the superior endplate of the upper vertebra and the inferior endplate of the lower vertebra. It should however be noted that the Cobb angle has certain limitations. First and foremost, it is a measurement of a 3D spinal deformity from 2D radiographs. This is noteworthy because two spines with

radically different 3D morphologies could yield similar Cobb angle estimations (Labelle et al., 2011). Furthermore, it is known that Cobb angle measurements could vary by up to 10 degrees (Majdouline et al., 2007). This is relevant since two spines with similar curves may render different recommendations for treatment (Labelle et al., 2011).

Predictions of the progression of a spinal curve should provide valuable insights into how the deformation is going to evolve and should greatly assist in guiding treatment strategies. Maturity (chronological, skeletal, and menarcheal age), curve magnitude, and curve location (Cheng et al., 2015) have traditionally been the main clinical indices used to assess spinal curve progression, with treatment decision based mainly on the curve magnitude: because the Cobb angle is normally used to assess the curve magnitude, this therefore means that the treatment depends on high-variability measurements.

Other clinical indices, such as different body length dimensions (sitting height, subischial leg length, and foot length or shoe size), secondary sexual characteristics, skeletal age in different areas, the Risser index, status of the triradiate cartilage, and electromyography ratios of the paraspinal muscle activity, have also been considered as predictors of curve progression (Busscher, Wapstra, & Veldhuizen, 2010; J Cheung et al., 2004; John Cheung, Veldhuizen, Halberts, Sluiter, & Horn, 2006; Little, Song, Katz, & Herring, 2000; Sanders, 2007; Sanders et al., 2006, 2007). Additionally, the relationship between a rapid growth of the patient and the evolution of the spinal deformity has been widely studied (Busscher et al., 2010; J Cheung et al., 2004; John Cheung et al., 2006; Little et al., 2000; Lonstein & Carlson, 1984; Ran et al., 2014; Tan, Moe, Vaithinathan, & Wong, 2009). Noshchenko *et al.* (Noshchenko, 2015) carried out a systematic review of 25 studies presenting clinical parameters that are statistically significantly associated with the progression of AIS. However, the parameters presented a limited or little evidence as predictors of the final deformation.

Studying the analysis of the spine in 3D is of vital importance, since it can lead to a more relevant and reliable 3D classification method for assessing and treating AIS (Labelle et al., 2011). In this respect, computerized clinical indices (Stokes, 1994a) and geometric descriptors

(Duong et al., 2006; Kadoury et al., 2017, 2014; Shen et al., 2014; Thong et al., 2016, 2015) have been proposed to capture the complexity of the spinal deformity. However, characterizing the spine in 3D space with meaningful descriptors is still challenging. This characterization must be capable of retaining the most significant information, not only in order to achieve the highest classification performance, but also to be clinically relevant.

In statistical shape analysis, methods such as Active Shape Models or Active Appearance Models have been used to study the main directions of shape variations (Cootes, Hill, Taylor, & Haslam, 1994; Cootes, Taylor, Cooper, & Graham, 1995; Cootes & Taylor, 1999) with the objective of mapping high-dimensional feature vectors onto lower-dimensional representations, while maintaining most of the variability of the original dataset. Usually, these models use Principal Component Analysis (PCA) to derive the low-dimensional representation of the data. The eigenvectors with the highest variance are used as modes of shape variations. The main disadvantage of PCA is the assumption of a Gaussian distribution of data, which could lead to incorrect descriptions.

Using support vector machines, Assi *et al.* (Assi, Labelle, & Cheriet, 2014) analyzed several dimensional reduction techniques, which were used before surgery to predict the postoperative appearance of a patient's trunk. Recently, a supervised model based on discriminant manifolds was proposed to study the 3D morphology of the curve progression (Kadoury et al., 2017). The samples in the dataset of the latter were labeled as progressive and non-progressive, based on the Cobb angle. However, since there are many forces acting simultaneously in the curve progression, the prediction could fail if only patterns related to the Cobb angle are considered, which may not necessarily characterize the progression in a 3D space in sufficient detail.

Independent Component Analysis (ICA) is another technique that has been used in shape analysis to obtain the modes of shape variations (Rogez, 2005; Ruto, Lee, & Buxton, 2006; Üzümcü, Frangi, Reiber, & Lelieveldt, 2003; Zhao et al., 2014). Unlike PCA, ICA generates independent non-Gaussian components. It also takes into account higher-order moments of

data distribution, instead of variance maximization, as in PCA. Hence, ICA could obtain more representative modes of variation from the dataset.

In this work, we propose an approach for predicting the progression of spinal curvatures using ICA to capture the modes of shape variation of 3D models of the spine from a cohort of patients with AIS. We compared the performance of shape variation modes obtained with ICA against a low-dimensional representation of 3D models of the spine, generated from Stacked Denoising Autoencoders (SDAE).

3.3 Methods

3.3.1 3D spine models

For this study, we selected 150 unique patients from a database of 3D spine models collected at the *Centre hospitalier universitaire Sainte-Justine*, Montreal, Canada. The inclusion criteria for our research were: (1) all patients must have at least three visits; (2) these visits must be pre-surgery (if surgery was performed); (3) the Cobb angle > 10°; (4) all patients should have a Risser index of 0 or 1; (5) the patients ought to have posteroanterior and lateral radiographs at each visit.

The gold standard measurement to quantify the curvature of the spine is the Cobb angle, which was performed from radiographs. At the first and following visits, lateral and posteroanterior spine radiographs were acquired. Usually, the follow-up ended when the patient reached skeletal maturity (Risser 4) or underwent surgery. In severe cases, patients undergo radiographs every 4 to 6 months during follow-up, which results in higher exposure to ionizing radiations. This is the main impediment performing close evaluations to assess progression.

In our approach, for each patient, 3D spine models were reconstructed from each visit using stereoradiographic 3D reconstructions from conventional X-rays. All 3D models comprised vertebral levels ranging from C7 to S1. A trained technician identified 6 landmarks in each

vertebra (superior and inferior endplates, left and right superior pedicles, and left and right inferior pedicles).

We aimed to simulate the shape of the spine every three months. However, two aspects had to be considered with respect of the patients' visits: 1) not all the patients had the same number of visits, and 2) the time between one visit and the next was not always the same for all patients. In order to overcome these limitations, we linearly interpolated transitional 3D spine models that represented intermediate visits for each patient as needed. We estimated an interpolated value among visits for every landmark on each vertebra. This interpolation was based on the speed at which each landmark changes over the span between visits. Finally, a set of 7 3D spine models per patient was formed, with each model representing a visit separated by a span of three months.

When generating the interpolated models, we always favored preserving models obtained from actual visits. Three considerations were taken into account for the interpolations: 1) the 3D spine model at the first visit was always preserved; 2) among patients with only three visits, we only considered those who had the second and third visits within a ±30-day range from 9 and 18 months after the first visit, respectively, and 3) if a patient had more than three visits, we picked the two visits closest to 9 and 18 months from the first one. As for the remaining visits, we included in the model those that were closest (within a ±30-day span) to 3, 6, 12 and 15 months from the first visit. If no visit was made within that span, we then proceeded to interpolate one visit based on the nearest actual ones.

Figure 3.1 shows three examples of patients for whom interpolated 3D spine models were generated. Case 1 represents the worst-case scenario, where there are only 3 actual models, separated by 9 months: models i_3 and i_6 were interpolated based on models r_0 and r_9, while models i_{12} and i_{15} were interpolated using models r_9 and r_{18}. Case 2 had 4 actual models (r_0, r_6, $r_{9\text{-}12}$, and r_{18}): model i_3 was interpolated using models r_0 and r_6. At 6 months, the actual model r_6 was preserved since it was within the ±30 days range of difference to the current interval. Model i_9 was interpolated using models r_6 and $r_{9\text{-}12}$. Models i_{12} and i_{15} were interpolated with

the closest models at $r_{9\text{-}12}$ and r_{18}. Model r_{18} was preserved. In case 3, there were 5 actual models (r_0, $r_{3\text{-}6}$, r_9, r_{15} and r_{18+}). Model i_3 was interpolated by models r_0 and $r_{3\text{-}6}$. Model i_6 was interpolated using $r_{3\text{-}6}$ and r_9. Model i_{12} was interpolated using models r_9 and r_{15}. Finally, model i_{18} was interpolated using models r_{15} and r_{18+}.

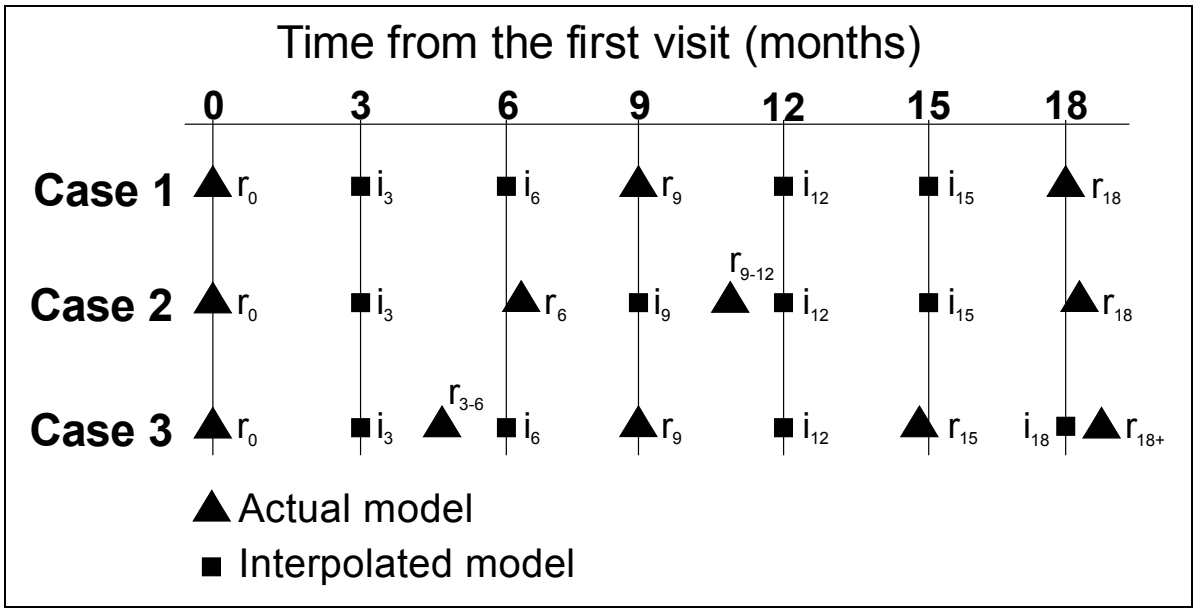


Figure 3.1 Three cases of interpolated 3D models of the spine. The black triangles correspond to models of actual visits. If they were within 30 days of the cut-off time for an interval, they were preserved in the dataset as the models for that interval. Black squares represent interpolated models generated based on the nearest actual models.

The 3D models were represented in the Scoliosis Research Society reference frame, where *x* is the horizontal axis that runs from the rear to the front of the patient, *y* is the horizontal axis that runs from right to left, and *z* is the vertical axis, which goes from the bottom of the patient upward (Stokes, 1994a). To align the models of the spines, we followed a process used in previous works (Duong et al., 2006; Thong et al., 2016, 2015). All 3D spine models were normalized according to the patient height, calculated along the axis defined by C7-S1. Also, a rigid transformation was computed to consider the centroid of the superior endplate of S1 as the origin of the reference frame. Each 3D spine model consisted of a vector of 306 values, corresponding to the 3D coordinates in x, y and z of the 6 landmarks on each vertebra.

3.3.2 Descriptors of the spine

Two sets of descriptors were computed for this work. We used Independent Component Analysis (ICA) and Stacked Denoising Autoencoders (SDAE) to obtain a simplified version of the 3D models of the spines. The aim here was to capture the main variation of the shape of the spine in 3D space.

3.3.2.1 Independent Component Analysis (ICA)

Independent Component Analysis (Comon, 1994; Hyvärinen & Oja, 2000) is a statistical model in which the centered shapes x are described in terms of a linear combination of statistically independent components, also called latent variables s, and an unknown constant mixing matrix A ($\boldsymbol{x} = \boldsymbol{As}$). In addition, the independent latent variables must have a non-Gaussian distribution. By estimating the mixing matrix A, it is possible to obtain the Independent Components (ICs) s by computing W as the inverse of A ($\boldsymbol{s} = \boldsymbol{Wx}$). The ICs are used as the modes of shape variations. As in the case of PCA, the modes of shape variations can be interpreted by the linear model that combines the mean shape ($\dot{\boldsymbol{x}}$) and each IC (s) ($\boldsymbol{x} \approx \dot{\boldsymbol{x}} + \boldsymbol{sb}$), where **b** is the weighted coefficient vector. By modifying **b**, we can observe variations with respect to the mean shape. A method to calculate **b** is proposed in (Üzümcü et al., 2003).

We used the *MetICA* (Liu et al., 2016) algorithm to compute the ICs from the 3D models of spines. Unlike other algorithms, it statistically evaluates the reliability of the ICs. In *MetICA,* PCA is computed on the centered data, and the denoised matrix X_d is obtained while preserving 95% of relevant information. *FastICA* is performed multiple times on X_d to estimate the demixing matrix W. Then, for each run, the sources are estimated to form the combined source matrix S. Spearman's correlation coefficient is used to describe the similarity between the components from different runs, and the correlation matrix is transformed into a distance matrix D. Hierarchical clustering analysis is computed on D. By cutting the dendrogram, a number of clusters are obtained. The centrotypes of each cluster are considered as convergence points of FastICA. The reliability of the centrotypes is evaluated based on bootstrapping

validation. A score is provided to measure how similar the centrotypes are to the estimated sources of the bootstrapped data. The centrotypes can be sorted based on the bootstrap score, and the higher the score, the more statistically significant it is.

3.3.2.2 Stacked Denoising Autoencoders (SDAE)

An autoencoder is an unsupervised machine learning algorithm based on feedforward neural networks. It is fundamentally characterized by its input vector *x* being ideally equal to its output vector *x'* (Bengio, 2009). The process involves two parts. The first part is *encoding*, in which the input is compressed into a low-dimensional representation. The second part is *decoding*, where the output is reconstructed from that low-dimensional representation. In the case of denoising autoencoders, the input is corrupted, and the objective is to recover the clean input while extracting useful features that capture the structure of the data (Vincent, Larochelle, Bengio, & Manzagol, 2008). A typical stacked denoising autoencoder contains several layers of nodes. In the encoder, each subsequent layer has fewer nodes than the last one , while in the decoder, the number of nodes increases symmetrically to the encoder (Vincent, Larochelle, Lajoie, Bengio, & Manzagol, 2010). For dimensionality reduction, the layer that is shared between the *encoder* and *decoder* (code layer) contains the low-dimensional representation of the input, also called codes. An implementation of SDAE for this work was performed based on the Keras library, version 2.1.3 (Chollet, 2015) for Python.

3.3.3 Spinal curve shape prediction

In this study, an approach using chained predictors to estimate the shape of the spine from the first visit is proposed. We consider two schemes (***a*** and ***b***) to perform the shape prediction of the spine (see Figure 3.2). Each scheme has six chained layers, meaning that the output of one layer is the input of the next. The difference between schemes is that ***a*** only considers the data of the immediate previous visit (short memory), while ***b*** considers the data of all previous visits (long memory).

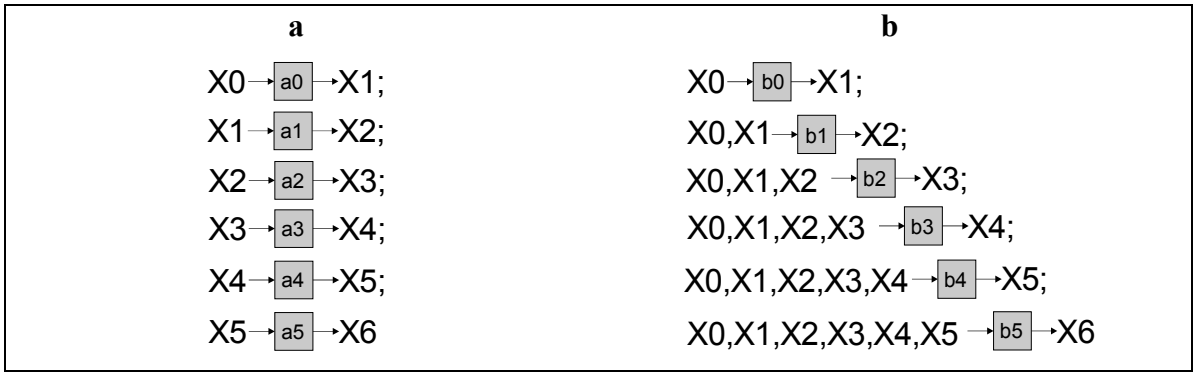


Figure 3.2 Schemes for shape prediction. Scheme *a* uses only the immediate output of the past visit as input for the next layer. Scheme *b* takes all previous outputs as input for the next layer.

Each layer consists of two tasks and represents the changes from one visit to the next within a three-month time lapse. The first task receives as input the ICs obtained from ICA or the codes obtained from SDAE. Its output are the estimated ICs or codes of the next visit. The second task takes the same input, but the output is the estimated 3D shape of the spine.

Random Forests (RF) was applied to model the changes between visits in both tasks. Breiman (Breiman, 2001) introduced RF as an ensemble of decision tree predictors that apply bagging and random selection of features at each split of each tree. Predictions are made by combining decisions from a set of decision trees, which are constructed independently using a different subsample of the data. RF is able to capture non-linear relationships between the features and the target, and can be applied to classification, regression and feature selection tasks. Regressions models usually predict only one target. If predictions of more targets are desired, it is possible to concatenate multiple regression models. However, the relationships between the models are not considered. Conversely, RF has the characteristic of performing multi-output regressions natively. It captures the dependencies between the different target variables, as opposed to other models that build a predictor for each target (Borchani, Varando, Bielza, & Larrañaga, 2015). These characteristics make RF suitable to be applied as a predictor in the layers of the schemes. The RF regression implementation of the Python library scikit-learn version 0.19.1 (Pedregosa et al., 2011) was used in this work.

3.4 Results

3.4.1 Descriptors of the spine

An independent component analysis was performed on 1050 3D models of the spine (150 patients x 7 3D spine models each) to describe the main variations of the shape of the spine. A set of 9 ICs was obtained from the dataset. These ICs captured 95% of the variability of the shape of the spines.

Table 3.1 presents the modes of variation of the shape and the positions of the spines in the posteroanterior (PAP), sagittal (SP) and apical planes (AP) with respect to the mean shape. The shapes are projected onto each IC, and a histogram is obtained from the projections. The variance of the ICs was estimated by using the width of the histogram *w***.** The value of *w* is varied between ±w/2 to visualize the modes of variation with respect to the mean shape, as proposed in (Üzümcü et al., 2003).

For the stacked denoising autoencoders, we used the architecture shown in Figure 3.3. The hyper-parameters were optimized by grid search cross-validation. The code layer applies linear activation, while the other layers use a rectified linear unit $f(x) = log(1 + exp\ x)$. A random Gaussian noise of 0.4 was added to the 3D models of the spines used as input.

Table 3.1 Variation of independent components, obtained from the 3D models of the spines with respect to the mean shape.

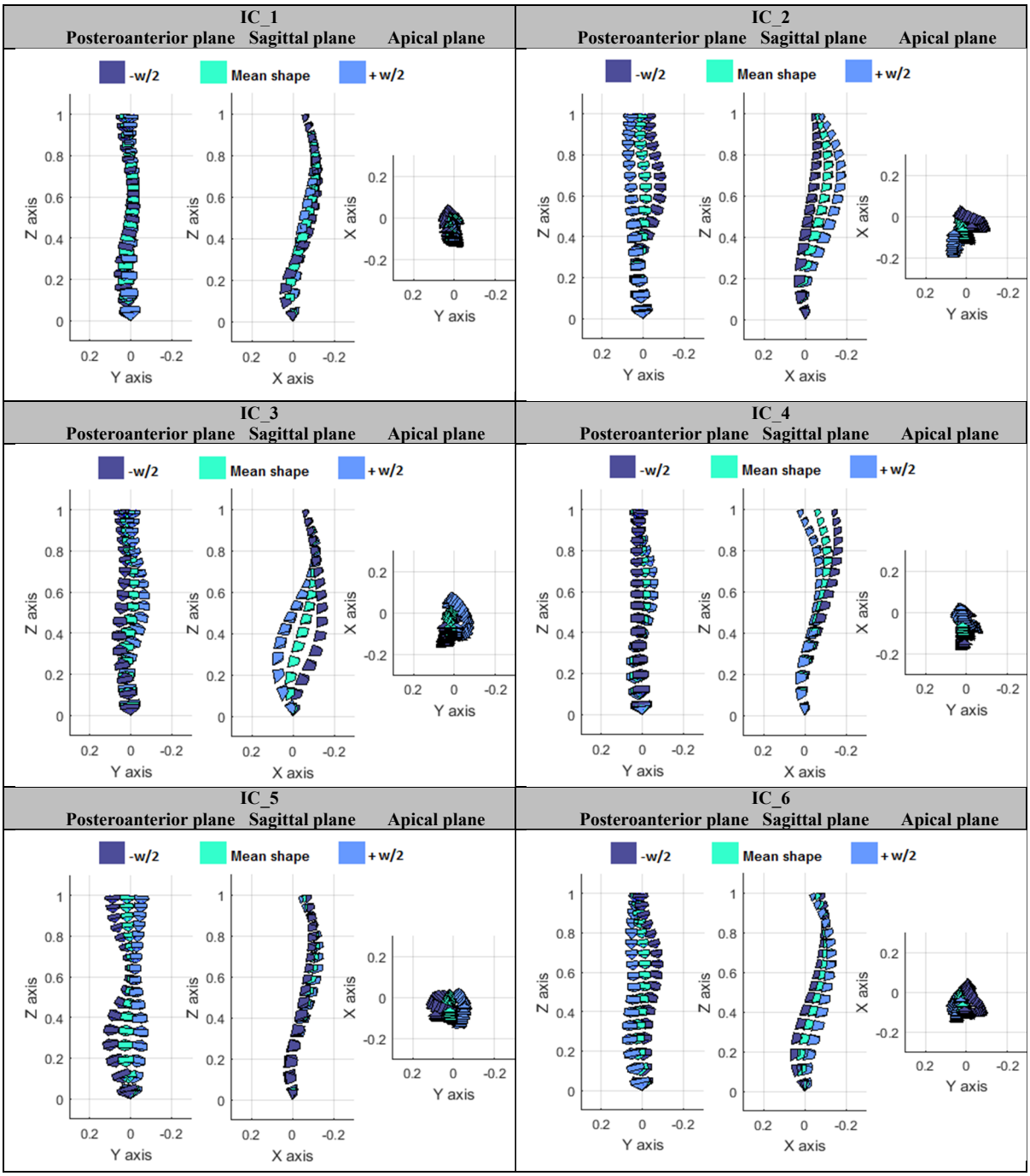
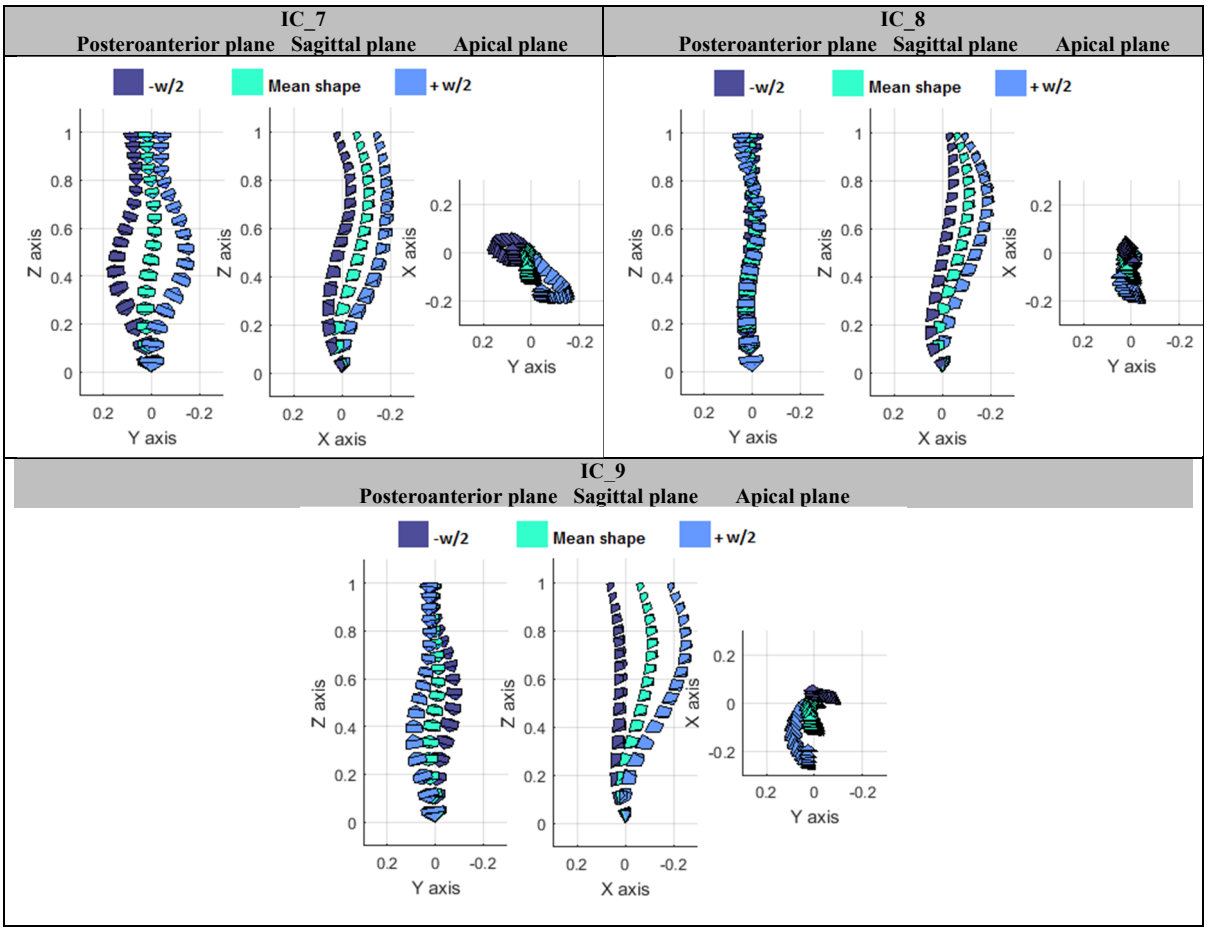


Table 3.1 (Continuation)



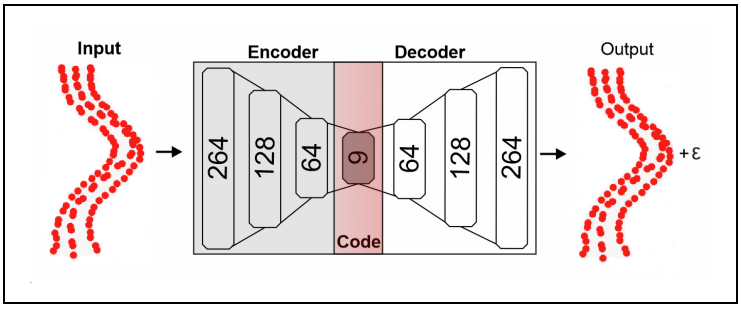


Figure 3.3 Architecture of the stacked denoising autoencoders. The layer in the center (9) is the coded representation of the 3D models of the spine.

3.4.2 Shape prediction

The prediction of the spine was performed according to two schemes. Scheme ***a*** considered only the descriptors of the previous visit, while scheme ***b*** used all the descriptors of the preceding visits (see Figure 3.2). The performance of each scheme was evaluated using a 10-fold cross-validation procedure. The dataset was divided into 10 sets, with 9 out of the 10 used to train the schemes, while the remaining one was used to test them. The predicted shape was evaluated against its original representation, and a prediction error calculated using the distance between points. The root-mean-squared error (RMSE) metric was used to evaluate the shape prediction.

The internal out-of-bag error was applied to validate the hyper-parameters of RF: n_estimators [500, 1000] (number of trees in the forest), max_features [0.1, 0.2, 0.3, 0.4, 0.5] (variables randomly chosen as candidates at each split), max_depth [5, 10, 15, 20, 25, 30, None] (maximum depth of the tree), and min_samples_leaf, [1, 2, 4, 6, 8, 10] (minimum number of samples required to be at a leaf node). Table 3.2 shows the RMSE results of evaluating the two schemes using the sets of descriptors from ICA and SDAE. Table 3.3 displays the prediction of the spines of two patients after 18 months from their first visit.

Table 3.2 Average scores by layer of the prediction models using the descriptors obtained from ICA and SDAE after 10-fold cross-validation. Four root-mean-squared errors (RMSE) were calculated. *3D* indicates the error in the three-dimensional space. *PAP*, *SP* and *AP* show the RMSE in the posteroanterior, sagittal and apical planes, respectively. Each row indicates a layer in the scheme.

ICA

	RMSE (mm), scheme *a*					**RMSE (mm), scheme *b***			
Layer	**3D**	**PAP**	**SP**	**AP**		**3D**	**PAP**	**SP**	**AP**
1	7.93	7.71	6.31	9.04		7.92	7.69	6.31	9.03
2	9.77	9.65	7.57	11.16		8.38	8.02	6.76	9.55
3	10.35	10.23	8.11	11.74		8.41	8.26	6.72	9.53
4	10.91	10.72	8.63	12.37		9.09	8.93	7.22	10.27
5	11.19	11.05	8.76	12.68		9.40	9.22	7.50	10.57
6	12.01	11.75	9.62	13.57		9.51	9.26	7.66	10.74
Avg.	**10.36**	**10.18**	**8.17**	**11.76**		**8.78**	**8.56**	**7.03**	**9.95**

SDAE

	RMSE (mm), scheme *a*					**RMSE (mm), scheme *b***			
Layer	**3D**	**PAP**	**SP**	**AP**		**3D**	**PAP**	**SP**	**AP**
1	7.47	7.08	6.63	8.25		7.46	7.06	6.62	8.24
2	9.53	9.26	7.94	10.66		7.87	7.39	6.91	8.75
3	10.13	9.94	8.23	11.35		7.70	7.38	6.65	8.57
4	11.31	11.05	9.22	12.70		8.33	8.13	7.03	9.28
5	11.80	11.65	9.49	13.19		8.46	8.31	7.10	9.39
6	12.57	12.34	10.16	14.06		8.57	8.34	7.27	9.55
Avg.	**10.47**	**10.22**	**8.61**	**11.70**		**8.06**	**7.77**	**6.93**	**8.96**

Table 3.3 Prediction of the shape of the spines (centerline) of two patients in the posteroanterior (PAP), sagittal (SP) and apical (AP) planes. The first and second rows belong to the schemes using descriptors from ICA, while the third and fourth rows correspond to the descriptors from SDAE. Shapes in rows 1 and 3 were obtained with scheme *a*, and shapes in rows 2 and 4 were obtained with scheme *b*. The grey shape represents the original shape, while the black shape depicts the predicted one.

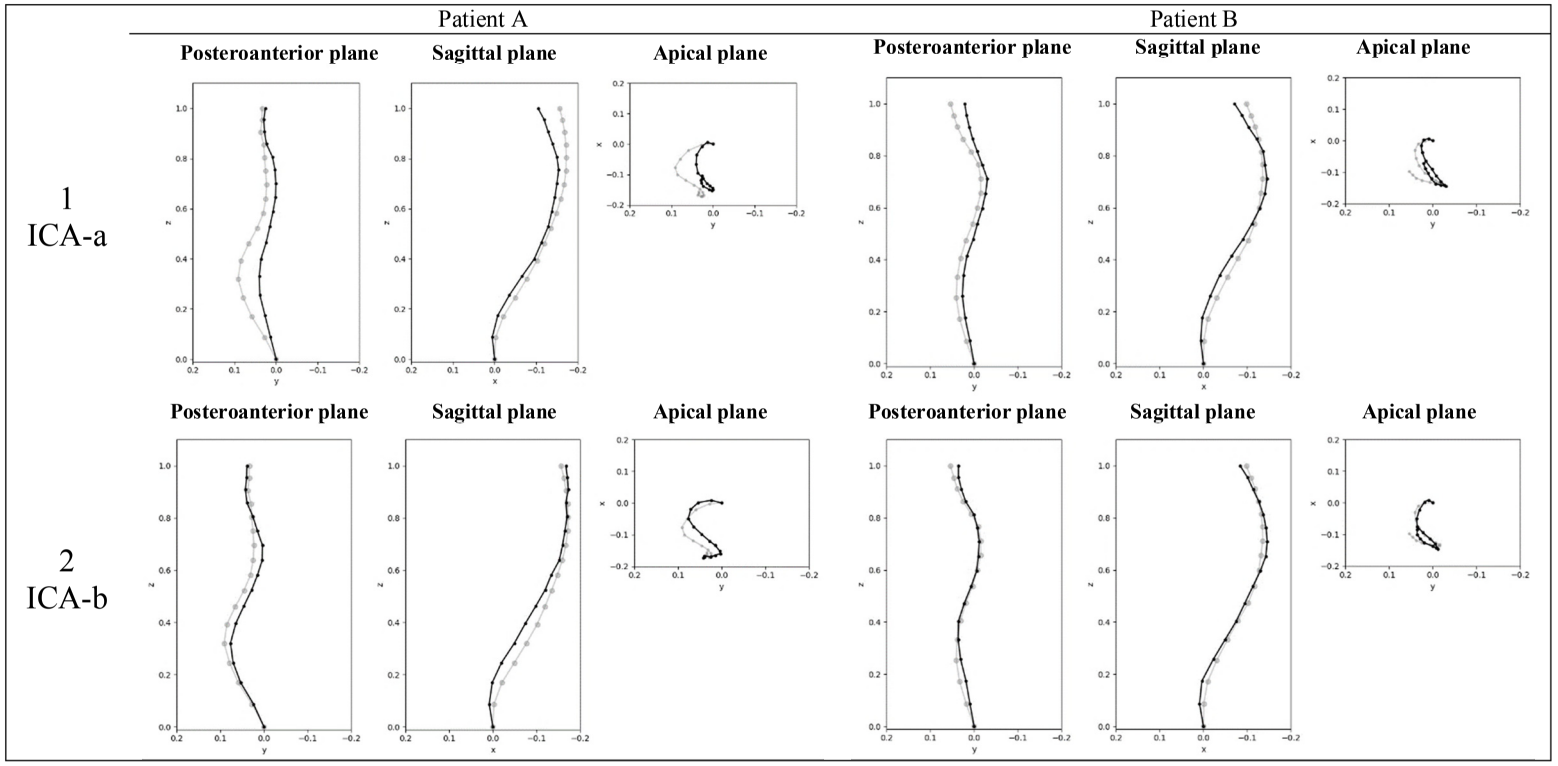
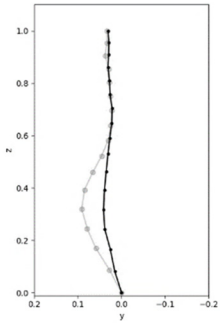
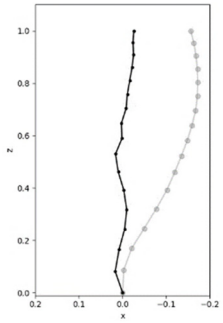
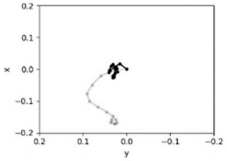
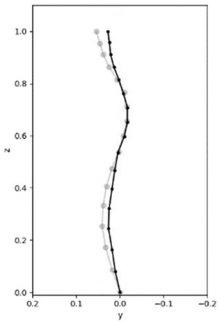
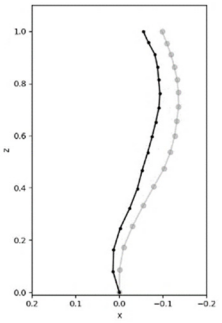
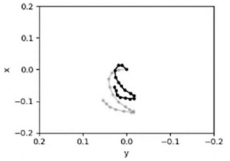
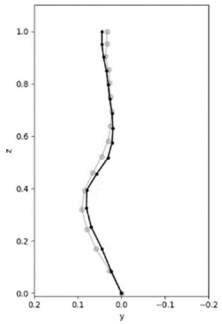
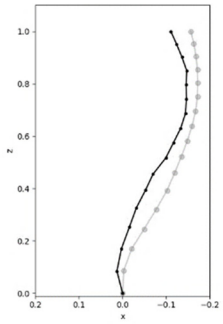
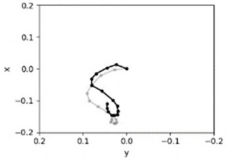
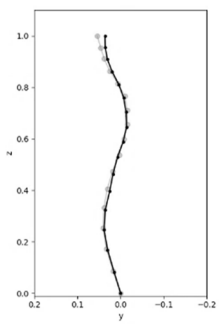
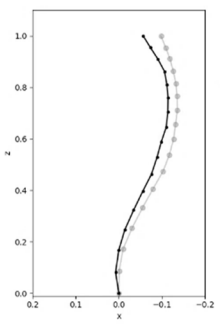
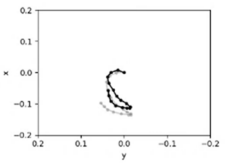


Table 3.3 (Continuation)

	Patient A			Patient B		
	Posteroanterior plane	**Sagittal plane**	**Apical plane**	**Posteroanterior plane**	**Sagittal plane**	**Apical plane**
3 SDAE-a						
	Posteroanterior plane	**Sagittal plane**	**Apical plane**	**Posteroanterior plane**	**Sagittal plane**	**Apical plane**
4 SDAE-b						

We took the scheme ***b***, which was the one with the lowest RMSE with both sets of descriptors and compared the magnitude of the main curvature between the predicted spine and the original one. Table 3.4 shows the averages of the differences in Cobb angles after the 10-fold cross-validation.

Table 3.4 Averages and standard deviations of the differences in Cobb angles, in the proximal thoracic (PT), main thoracic (MT) and thoraco-lumbar lumbar (TL/L) sections, between the predicted and the original shapes of the spine after the 10-fold cross-validation. Each row indicates a layer in the scheme.

	ICA scheme b			**SDAE scheme b**		
Layer	**PT (degrees)**	**MT (degrees)**	**TL/L (degrees)**	**PT (degrees)**	**MT (degrees)**	**TL/L (degrees)**
1	-1.65 ± 1.96	-4.71 ± 2.02	-4.61 ± 2.25	-2.34 ± 1.69	-5.68 ± 2.56	-5.75 ± 1.83
2	-1.69 ± 1.40	-4.94 ± 2.81	-4.40 ± 2.33	-2.22 ± 1.59	-5.46 ± 3.52	-5.35 ± 2.62
3	-1.76 ± 1.46	-5.25 ± 2.58	-4.57 ± 2.67	-2.48 ± 1.72	-6.11 ± 2.96	-5.48 ± 2.68
4	-1.75 ± 1.81	-5.45 ± 3.09	-4.91 ± 2.60	-2.36 ± 1.68	-5.87 ± 2.53	-5.14 ± 3.15
5	-1.98 ± 2.00	-5.43 ± 2.63	-5.16 ± 2.64	-2.82 ± 1.78	-6.42 ± 3.04	-6.05 ± 3.15
6	-2.15 ± 2.37	-5.27 ± 3.22	-5.11 ± 2.95	-2.86 ± 1.57	-7.06 ± 3.83	-6.42 ± 3.58
Avg.	**-1.83 ± 1.83**	**-5.18 ± 2.72**	**-4.79 ± 2.57**	**-2.51 ± 1.67**	**-6.10 ± 3.07**	**-5.70 ± 2.84**

3.5 Discussion

In AIS, the deformation prognosis varies from patient to patient. Adolescents are in a period of growth, which means that their tissues and skeleton are immature. Furthermore, the way the shape of the spine changes through time is different from patient to patient as well. For optimal treatment, there is a need to identify which patients are at higher risk of curve progression at the early stages of the disease.

In this study, we modeled the geometric progression of the spinal curvature based on prior observations of retrospective visits. Knowing such progression patterns could assist clinicians in monitoring, following up and adequately treating patients according to their specific needs. Extracting the relevant geometric components to describe changes in the spine is very important for quantification of curvature progression. Labelle et al. (Labelle et al., 2011)

showed that the morphology of two similar deformities in 2D are different when they are characterized in 3D. Recently, in a retrospective study, expert-based 3D morphological descriptors of the spine were proposed to differentiate between two groups of progressive and non-progressive curves (Nault et al., 2013). Then, in a prospective evaluation, these 3D morphological descriptors were analyzed in order to find a significant difference between the two groups at each patient's first visit (Nault et al., 2014). One limitation of the local descriptors presented in (Nault et al., 2013, 2014) is that they can vary depending on the patient's posture during the acquisition of radiographs, which could lead to inaccuracies if used as predictors.

On the other hand, computer-based methods such as wavelet compression techniques (Duong et al., 2006), manifold characterization (Kadoury & Labelle, 2012; Kadoury et al., 2017), and staked auto-encoders (Thong et al., 2016, 2015) have been proposed to characterize the 3D models of the shape of the spine by reducing the high-dimension set of features to a new low-dimension representation. Most of these cited works offer a new 3D classification system of AIS based on clustering techniques. From these approaches, only (Kadoury et al., 2017) has been proposed for predicting the evolution of AIS. Descriptors are extracted from 3D models of the spine and divided into progressive and non-progressive curves based on the magnitude of the Cobb angle. Then, a spatiotemporal regression model is computed to predict the progression of the spine deformation. The robustness of this method could be affected by the well-known high variability of the Cobb angle and its limitation in characterizing the spine in 3D space. In addition, in spite of its high association with progression, this measure has a limited prognostic capacity (Noshchenko, 2015).

Computer-based descriptors offer the advantage of capturing the complex nature of the 3D models in low-dimensional space. However, their interpretation is difficult for application in everyday clinical practice. In this work, we found ICA to be a promising technique for obtaining the principal modes of variation of the 3D models of the spine. It reduces the complexity of the 3D models, in addition to providing an interpretation of the principal modes of variation in 3D space, which represents an advantage over the techniques mentioned earlier.

ICA is a generative model, which identifies the ICs resulting from a process of mixture of components. Building on this concept, we considered the dataset of shapes of spines as a mixture of components from which we obtained the ICs, which we used to describe the main shape variation.

ICA nevertheless has some disadvantages. It is a stochastic method, which means that with it, the same algorithm will not always produce the same output when repeated. Additionally, unlike PCA, ICs are not sorted by their significance. MetICA is an implementation of ICA that addresses these points by applying a heuristic method, hierarchical clustering and bootstrapping validation.

We used MetICA to reduce the input space from 306 coordinates of 3D reconstructions to 9 ICs. We used these representations to perform a prediction of the shape of the spine from the first visit, without using any other clinical descriptor derived from the 3D models of the spine or 2D projections.

Our 9 ICs describe 95% of the 3D variability of the shapes in the posteroanterior, sagittal and apical planes, as can be seen in Table 3.1. For all the ICs, the apical plane captures the direction of the main curvature in the posteroanterior plane. IC_1 mainly captures minor changes in thoracic curves in the PAP, and lordosis in the SP. IC_2 captures changes in thoracic curves to the right in the PAP, and kyphosis in the SP. IC_3 shows changes in thoracic curves to the right in the PAP, and lordosis in the SP. IC_4 displays changes in thoracic curves in the PAP, and alignment and kyphosis in the SP. IC_5 represents double curvatures in the PAP, where the lumbar curve is significant, as well as kyphosis in the SP. IC_6 shows double curvatures in the PAP, where the thoracic curve is the significant, and also shows kyphosis in the SP. IC_7 displays major changes in thoracic curves in the PAP, and lordosis and alignment in the SP. IC_8 mainly captures kyphosis in the SP. Finally, IC_9 presents thoracolumbar/lumbar curves in the PAP, and alignment and kyphosis in the SP.

Previously, Thong et al. (Thong et al., 2016, 2015) evaluated the relevance of using autoencoders (SDAE) for dimensionality reduction in the classification of 3D spinal deformities. Based on the encoded versions of the 3D spines, they performed cluster analysis and found subgroups within Lenke types. We decided to compare the performance of our proposed ICs approach to SDAE. To this end, we configured the code layer in the SDAE architecture (see Figure 3.3) to obtain 9 codes in order to match the number of ICs. This meant that the same number of components was used as input for shape prediction in both approaches. The code layer captures the most relevant information into a compressed representation. However, due to the nature of neural networks, there is no direct interpretability of the codes. Unlike the codes, ICs can provide an intuition of how the shape changes by visualizing its variation modes.

We presented two schemes to compute the shape prediction. Scheme ***a*** was a short-memory strategy that used the immediate output of the past visit as input for the next layer. As a long-memory strategy, scheme ***b*** took all previous outputs as input for the following layers. The results of comparing both schemes using the two sets of descriptors are shown in Table 3.2. After a 10-fold cross-validation, it can be seen, through the different layers of the approaches, that the error spreads over time. Comparing both schemes, we see that ***b*** has the smaller RMSE with both sets of descriptors. Hence, incorporating all the information from previous layers helps improve the prediction of the following layer. This can be better observed in Table 3.3 (rows 2 and 4).

Since the shape of the spine develops in 3D space, we evaluated the predictions of the shapes in the posteroanterior, sagittal and apical planes. Table 3.2 shows that RMSE scores from SDAE present a modest advantage over those from ICA.

When comparing two particular cases of patients (Table 3.2) using scheme ***b***, the descriptors from SDAE provided a better prediction on the posteroanterior plane, while the descriptors from ICA showed more favorable approximations on the sagittal and apical planes.

After performing the 10-fold cross-validation, we observed that scheme ***b*** generated a conservative prediction. Independently of the descriptors, the magnitude of the Cobb angle in the predicted shape was smaller than the original in all the sections of the spine (Table 3.4). Using ICs produces a model with lesser curve magnitude differences with respect to the originals on average when compared to SDAE.

Our approach could be used clinically to monitor the progression of current patients and to evaluate new patients. By simulating the shape of the spine, we could identify which curve pattern might be a candidate for progression. Hence, we could help clinicians plan a treatment based on our estimations. Also, by analyzing the treatment of the curve pattern that progresses up to surgery, we could guide clinicians to identify the proper treatment or surgical management based on the progression of the curve pattern rather than on the curve pattern alone.

From the 150 patients included in this study, 66 patients were found to have progressive (P) curves, which means that there was a Cobb angle difference of 6° or more between the first and the last visit. The other 84 patients had non-progressive (NP) curves. We used a 6° difference cut-off to determine the progression based on the confidence level of the measurement error in radiographs (Nault et al., 2014).

In addition to shape prediction, we compared the ICs at the first visit between P and NP patients. There were 53 patients with main thoracic curves and 13 with main thoracolumbar curves in the P group, while in the NP group, there were 43 patients with main thoracic and 41 with thoracolumbar main curves. We found a higher prevalence of progression in main thoracic curves (55%) as compared to thoracolumbar curves (24%). This finding was in agreement with previous works (Dimeglio & Canavese, 2013; Nault et al., 2013). This dataset was not evenly distributed, and had a relatively small size, and as such, these percentages may not be representative of the general population. A specific study focused on these types of curves could help illustrate a more accurate prevalence for each type of curve.

Table 3.5 shows the significance (p-value) of each IC in order to differentiate P and NP after performing t-tests. Significant ICs (p-value < 0.05) are presented with an asterisk. We found a statistically significant positive correlation of NP with IC_5 (double curvatures in the PAP with major lumbar curve and kyphosis in the SP), and IC_9 (thoracolumbar/lumbar curves in the PAP, and alignment and kyphosis in the SP) values at the first visit.

Table 3.5 Significance of correlation of ICs at first visit with progression

Independent component	p-value
IC_1	0.104
IC_2	0.102
IC_3	0.547
IC_4	0.815
IC_5	0.011*
IC_6	0.217
IC_7	0.808
IC_8	0.194
IC_9	0.019*

We consider that in a prospective cohort study, the significance of the ICs could be confirmed, not only to evaluate the ICs at the first visit, but also to know how the changes of the ICs through time is related to the magnitude of the curve. Also, since every IC describes a mode of variation of the spine in 3D, an unsupervised analysis could be performed to automatically group the components at each visit, and to evaluate if there is a significance level of the groups that can provide new insights between the ICs and progression. This knowledge could help improve patient follow-up and treatment.

Growth is an essential factor in studying patients with AIS. The main curve progression occurs at the peak height velocity phase during puberty. Mainly, thoracic curves are prone to progression among 20° to 30° of pre-pubescent children (Dimeglio & Canavese, 2013). A curve with a magnitude of more than 30° at peak height velocity has a high probability of progression (Little et al., 2000). However, the limitations of using the peak height velocity

include the fact that they are a function of several continual height measurements of the patient, and they vary according to gender.

Bone age is one of the main parameters used to monitor growth. The Risser index is a common method used by clinicians to grade bone age. However, it has been found to be inaccurate for this task (Dimeglio & Canavese, 2013; Sitoula et al., 2015), since patients at peak height velocity could be cataloged as Risser 0, along with patients that are not yet in this phase. Recently, Sitoula et al. (Sitoula et al., 2015) found a correlation between Sanders' skeletal maturity and Cobb angle in determining curve progression.

Most patients with AIS are female, and as a result, menarche has been used to assess progression. However, it is not a reliable indicator since it occurs at a median of 7 months after the peak height velocity (Noshchenko, 2015; Sitoula et al., 2015).

Different genes have also been associated with curve progression. Nevertheless, the way the studies are designed, along with their replicability, represent the main limitations for their use as prognostic descriptors (Noshchenko, 2015).

Given the limitations inherent in using demographic factors to accurately evaluate growth, in this study, we proposed a predictive approach to generate the shape of the spine through time, considering only geometric descriptors. We decided to generate transitional models between real visits in order to visualize changes in the shape of the spine through time.

Evaluating the progression of AIS depends on assessments of radiographs captured between 4 and 6 months prior to and until skeletal maturity. Hence, the samples in our study were not acquired according to a fixed schedule, which meant that we had to interpolate 3D spine models linearly. This implies that a possible dependency among the models that were interpolated, which could lead to a reduction in the amount of prediction errors. To minimize this effect, we always interpolated each visit based on actual ones, and we favored actual visits as inputs to our approach.

We evaluated the errors using interpolated spine models versus actual visits, and performed the evaluation at different intervals on 15 patients with at least 5 visits each. The patients chosen were the ones with the most visits. This helped remove intermediate actual visits at certain intervals. Then, interpolated models were generated, replacing the actual visits that had been removed. We evaluated the error between the actual and the interpolated visits (average ± Std. Dev.). We obtained an RMSE of the 3D shape of 7.26 ± 3.17 (mm). We also calculated the error in terms of the Cobb angle (degrees) at 1.31 ± 2.50, -0.44 ± 4.05, and -1.28 ± 3.56 of the proximal thoracic, main thoracic and thoracolumbar-lumbar sections, respectively. Although the change in the shape of the spine is not linear, for patients with less than 5 visits, the approximation of the interpolated models between actual visits did not produce significant errors.

This study provides a basis for further investigations into the significance of the ICs and curve progression in AIS. We consider that the inclusion of peak height velocity and skeletal maturity to our approach could be very useful in improving the generation of interpolated transitional models, as well as the prediction of spine curve progression.

3.6 Conclusion

The ability to predict the evolution of spine curves among patients could help clinicians detect patients who may have progressive curves. This could help them devise patient-specific treatments, which could in turn lead to better outcomes. Currently, the gold standard for evaluating AIS patients is the Cobb angle, which presents high variability of measurements and does not capture the 3D morphology of the spine. Computer-generated descriptors offer the advantage of using standardized data, which eliminates the variability of manual measurements and improves reproducibility. In this paper, we propose an approach for predicting the progression of spinal curves. Our predictions show the possible development of the shape of the spine right from the first visit, and for every three months thereafter, up to a period of 18 months.

We propose the use of the Independent Component Analysis to capture the variation modes in a dataset of 3D spine models of patients with AIS, and compare it with an approach based on autoencoders. Although both approaches have the potential to simulate the development of the spine in 3D space, one advantage of using ICA over SDAE is that in the former, the descriptors can be visualized for interpretation (see Table 3.1). This information could provide clinicians with a better insight into how the shape of the spine is expected to evolve through time.

Our proposed work makes predictions based only on 3D models obtained from radiographs taken at the first visits by patients. We compare two schemes to generate the predictions, one short-memory and one long-memory. The long-memory scheme provides the 3D reconstructions closest to the real evolution of patients' spine curves. This means that inputting the information on subsequent visits would potentially significantly improve the predicted models.

In future work, we aim to further this study by exploring how the variation modes are related to the progression of the curvature in AIS. We will also include local vertebra information to the model in order to improve the prediction of the vertebral bodies. Additionally, we will incorporate radiographic and other growth indicators. Finally, to improve the robustness of the approach, a larger dataset would contribute more diverse curve patterns.

CHAPTER 4

A FREEHAND ULTRASOUND FRAMEWORK FOR SPINE ASSESSMENT IN 3D: A PRELIMINARY STUDY

Edgar García-Cano[1], Fernando Arámbula Cosío[2], Fabian Torres Robles[3], Zian Fanti[2], Christian Bellefleur[4], Julie Joncas[4], Hubert Labelle[4], Luc Duong[1]

[1] Software and IT Engineering Department,
École de technologie supérieure

[2] Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas,
Universidad Nacional Autónoma de México

[3]Centro Virtual de Computación,
Universidad Nacional Autónoma de México

[4]Research Center, Sainte-Justine Hospital

This article was submitted to *Ultrasound in Medicine and Biology*, in November 2018.

4.1 Abstract

X-ray imaging is the current gold standard technology for the assessment of spinal deformities. The purpose of this study is to evaluate a freehand 3D ultrasound system for volumetric reconstruction of the spine. A setup consisting of an ultrasound scanner with a linear transducer, an electromagnetic measuring system and a workstation was used. We conducted 64 acquisitions of US images of 8 adults in natural standing position, and we tested three setups: 1) Subjects are constrained to be close to a wall, 2) Subjects are unconstrained, and 3) Subjects are constrained to performing fast and slow acquisitions. The spinous processes were manually selected from the volume reconstruction from tracked ultrasound images to generate a 3D point-based model depicting the centerline of the spine. We defined three measurements to quantify the variation of the landmarks of the 3D point-based models: the distance between spinous processes, and the angles of two adjoint spinous processes with respect to the horizontal in the posteroanterior (PA) and sagittal (Sa) planes. Based on these measurements, we calculated the root-mean-square error (RMSE) between the models of each subject. On

average, we obtained an RMSE of 6.50 mm, 5.66 degrees in PA, and 6.94 degrees in Sa for the thoracic section, and an RMSE of 4.54 mm, 6.74 degrees in PA, and 7.37 degrees in Sa for the lumbar section. The landmarks in the lumbar section were more difficult to identify since this section contains more muscles. The results suggested that a freehand 3D ultrasound system can be suitable for representing the spine. Volumetric reconstructions can be computed and landmarking can be performed to model the surface of the spine in the 3D space. These reconstructions are promising to generate computer-based descriptors to analyze the shape of the spine in the 3D space.

Keywords: Freehand 3D ultrasound, tracked sonography, tracked ultrasound, ultrasound images, spine reconstruction, spinous process, adolescent idiopathic scoliosis

4.2 Introduction

Adolescent idiopathic scoliosis (AIS) is a common deformation of the spine that affects 1 to 4% of adolescent population, with a greater prevalence among females (Cheng et al., 2015). Patients are diagnosed with AIS when the Cobb angle, the angle between the two most rotated vertebrae is greater than 10 degrees. The Cobb angle and X-ray imaging are currently the gold standard to assess spinal deformities. X-ray imaging allows visualizing the full spine in standing position, including the pelvis, and C7 vertebrae.

The treatment of AIS depends on the severity of the curvature and progression. Generally, the curve magnitude increases over time. However, how much the magnitude increases depends on each individual patient. Patients with Cobb angles of 20 degrees or less usually remain under observation, while in whom the angle lies between 20 and 40 degrees are eligible for bracing treatment. Patients are candidates for corrective surgery when they have a thoracic Cobb angle > 50 degrees or between 40 to 45 degrees in the thoracolumbar section. These patients represent 0.1% of the total population with AIS (Cheng et al., 2015).

There are three main limitations with the current gold standard when it comes to assessing AIS: 1) It has been reported that the Cobb angle measurement could have a variation of up to 10 degrees (Majdouline et al., 2007); 2) 2D radiographs present an oversimplification of the entire 3D shape of the spine, and 3) Patients with a high risk of curve progression are usually closely monitored, with follow-ups every 4 to 6 months. This results in frequent exposure to potentially harmful ionizing radiations, and, consequently, an increased risk for breast or lung cancer (Doody et al., 2000; Ronckers et al., 2010, 2008). Therefore, a radiation-free imaging method for assessing and following up patients would be very beneficial. Magnetic resonance imaging could represent a very good alternative to radiography, but does not allow imaging in standing position, hence modify the normal posture.

Ultrasound (US) is one of the most inexpensive and widely used radiation-free diagnostic image technologies in medicine. It provides images from within the body by applying high-frequency sound waves on the skin. The waves are reflected to the transducer by the organs as echoes. Then, the received signals are processed and displayed as images on the screen. The elapsed time from the emission of the wave to its reception from the body is used to create the images. These images can be used to diagnose and treat several medical conditions. Prenatal health is its most common application; however, it has been exploited to evaluate more complex information concerning, for example, cancer, flow of blood, bones and tissues.

Since US in B-mode only produces one 2D image at a given time, it is not suitable for analyzing the volume of structures. This would be the main limitation with examining spine deformations using US imaging. However, freehand 3D US systems have been developed and applied to augment the capabilities of US. This is a non-invasive and low-cost technique that makes it possible to generate a 3D view of the anatomy. It combines a tracker sensor that captures the position and orientation of the probe, while the 2D images are being acquired in real time. The result is a 3D volume that represents the topography of the anatomy.

Recently, freehand 3D US systems have been proposed as an alternative to characterize curvatures of the spines (C. J. Cheung, Zhou, Law, & Mak, 2015; C. W. J. Cheung et al., 2015,

2013; Ibrahim, Usman, Mohktar, & Ahmad, 2016; Zheng et al., 2016). Approaches to identify landmarks on US images, such as the spinous process, transverse process, superior articular process and the center of laminae have been reported (Chen et al., 2011; C. W. J. Cheung et al., 2013; Koo et al., 2014; Vo, Lou, Le, & Huynh, 2015). All these studies used these landmarks as a means to find an equivalent angle to characterize curvatures like the Cobb angle. However, a measure conducted in the 2D space is not suitable for performing a full description of a 3D deformity. The Scoliosis Research Society has identified the analysis of the spine in 3D space as a step forward to improve the assessment, follow-up and treatment of AIS (Labelle et al., 2011).

In recent studies, computer-based descriptors (Duong et al., 2006, 2010; Duong, Mac-Thiong, et al., 2009; García-Cano et al., 2018a, 2018b) have been proposed to characterize the 3D nature of the spine. These descriptors are based on 3D spine models reconstructed from stereoradiographic X-rays. Likewise, automatic methods have been proposed to predict the curve progression (García-Cano et al., 2018b). By extracting the visible landmarks (spinous process, transverse process, superior articular process and laminae) from freehand 3D US reconstructions, it could be possible to produce computer-based descriptors to model the spine in 3D, similar to the aforementioned approaches, complementing the current 2D measurements. Also, a prediction curve deformation model could be applied to personalize each patient's treatment.

In this paper, we investigated a freehand 3D US system using hardware available in clinics, as well as, a free, open-source software for data acquisition, pre-processing, calibration to reconstruct the shape of the spine of healthy subjects. The paper is organized as follows: In Section 2, we present the hardware of our freehand 3D US system, the acquisition protocol, as well as our methodology for identifying landmarks and the metrics for evaluating the posture of the subject. Section 3 shows the results of our experiments in three setups. Section 4 discusses our findings, challenges and limitations. Finally, Section 5 concludes the work and identifies areas for prospective future work.

4.3 Materials and Methods

4.3.1 Freehand 3D ultrasound system

The freehand 3D US system consisted of an US scanner Toshiba Xario with a linear transducer with a width of 38 mm (Toshiba PLT-704AT/5-11MHz). A USB video capture card (Dazzle DVD Recorder HD, Pinnacle) was used to save the digital images produced by the US scanner on a computer. An electromagnetic measuring system (EMS) Aurora V2 (NDI Ontario, Canada) was used to record the 3D position and orientation of the US transducer in real time. This information was synchronized with the US image acquisition. According to the information provided by the manufacturer, the tracking system generates a magnetic field in the shape of a cube measuring 50 cm per side, with a root mean square error (RMSE) of 0.70 mm for position accuracy, and an RMSE of 0.20 degrees for orientation accuracy.

The Open-Source Toolkit for Ultrasound-Guided Intervention Systems (PLUS) (Lasso et al., 2014) was used to perform the temporal and spatial calibrations between the transducer and the tracking sensor attached to it. This framework provides a convenient user interface to perform each of the steps involved in the calibration, as well as the functionalities for the acquisition of the US images, and the volume reconstruction. In this study, we used the PlusApp-2.6-Win64 version. The software was tested on an Intel Core i7 3.6 GHz workstation with 16 GB of RAM.

4.3.2 Study subjects

A total of 8 healthy adults (5 women and 3 men; mean age, 30±5.13 years) were recruited for this study. The study was evaluated and approved by our institution's research ethics committee. All participants were informed of the acquisition protocol by the first author, and they signed a written consent form before being enrolled in this study. The inclusion criteria used in this study were: 1) subjects should have no spinal deformation, 2) subjects should have no metallic implants, and 3) subjects should not be overweight.

4.3.3 Acquisition protocol

Acquisitions were performed in a controlled environment two days per week for two weeks; each time, they were done twice on the same day, once in the morning and once in the afternoon, for a total of 8 acquisitions per subject. The subjects were asked to use a gown to cover the front of their body, while leaving the back exposed. Prior to the first acquisition, we measured each subject's height, weight and waist circumference. Furthermore, two pictures of the trunk were taken, the first picture from the posteroanterior plane, and the second from the sagittal plane. We tried as much as possible to avoid taking pictures of any identifiable features such as the face, hair, birthmarks, tattoos, scars, and other recognizable markings.

In this study, we performed three experiments. The first was done during the first week. We acquired the tracked US images by positioning the subject in front of a wall, at a distance of 15 cm from it (*constrained setup*). A vertical line was drawn on the wall, and was used as the reference to center the subject in the setup (see Figure 4.1). In all the experiments, we adjusted the height of the magnetic field generator with respect to the height of the subject, positioning the former close to the subject's region of interest. Four acquisitions were performed for each subject using this setup. Since the transducer of the US was not wide enough to capture the whole vertebral body, we performed three sweeps in each acquisition. The first one was to the left of the centerline (L-sweep) of the spine (tip of the spinous process), the second one directly on the centerline of the spine (C-sweep), and the last one, to the right of the centerline of the spine (R-sweep). This can be seen in Figure 4.2.

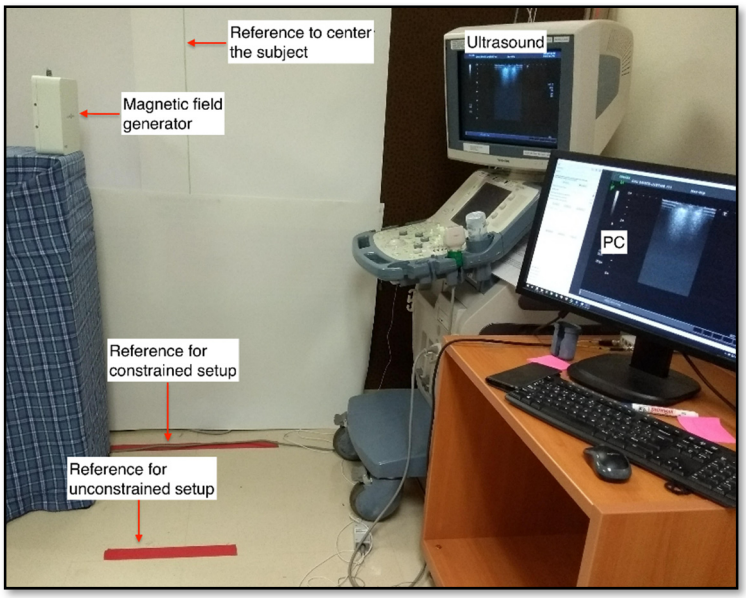


Figure 4.1 Image acquisition setup: the electro-magnetic measurement system, US scanner, and workstation.

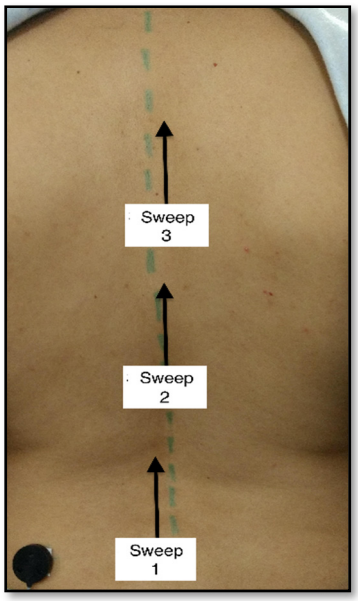


Figure 4.2 One acquisition has 3 sweeps. The white boxes indicate the positions of the probe in each sweep.

In the second week, we performed the second and third experiments. For the second experiment, we separated the subjects from the wall. This implied that the patients were not limited by the closeness to the wall or aligned relative to any reference (*unconstrained setup*). We only put a red line on the floor as reference for the subjects, so that they knew where to stand up (see Figure 4.1). Two acquisitions per subject were performed the same day, one in the morning and one in the afternoon.

For the third experiment, we carried out the acquisitions using the same arrangement as for the constrained setup. However, in this session the acquisitions were completed in two modes, *fast* and *slow*. The fast mode lasted approximately 20 seconds per scanning, while the slow mode took approximately 140 seconds. Only one acquisition per mode per patient was obtained. Before scanning the subjects, an identification and marking of the vertebrae was performed by the operator through palpation of the spinous process (see Figure 4.3). First, the subject was sitting on a chair with the head bended forward, exposing the C6 and C7 vertebrae, which were identified and marked with a water-based marker. These vertebrae have the most prominent

spinous process in the cervical section of the spine. Then, the subject curved the spine and, by palpation of the iliac crest bones, the intervertebral space between vertebrae L4 and L5 was marked (Figure 4.4). By counting downward from the C7 vertebra, the tips of the thoracic (T1 to T12) and lumbar (L1 to L4) vertebrae were identified and marked. Later, to validate the initial markings, a counting upward from the vertebra L4 to C7 was performed. This procedure was validated by two physicians.

Figure 4.3 Identification of vertebrae by the operator

Figure 4.4 Localization of the intervertebral space between the L4 and L5 vertebrae

Figure 4.5 Fix reference sensor on the subject, 3 inches to the left from the centerline

Before performing the acquisitions, temporal and spatial calibrations were carried out. All the acquisitions were performed by the same operator. The acquisition of the US data was made in B-mode, with the subject in a natural standing position, barefoot and without any support or platform that could alter the standing stability of the subject. Also, it is important that patients continue breathing naturally to maintain the normal shape of the spine. However, breathing and other involuntary movements of subjects could change their position during the acquisition. Hence, a reference tool that is part of the tracking system was attached to the subject. The reference tool is used to capture any unintentional shifting, and its location is used to correct the position and orientation of the tracked data. This reference tool was fixed

three inches to the left of the intervertebral space between the vertebrae L4 and L5 on each subject (see Figure 4.5).

Once all the vertebrae were identified with a water-based marker, the operator applied US gel on the region of interest to ensure image quality. He then requested the subject to stand still behind one of the red lines (depending on the experiment) on the floor, to be centered according to the vertical line on the wall (only for the constrained setup), to breathe shallowly, and to keep the sight forward with the arms relaxed. Prior to the acquisition, the subject was requested to remove any metallic object. For long-haired subjects, we asked them to arrange their hair in an updo to have an unobstructed view of the spine.

The calibration of the probe's frequency was set at 6.6 MHz, and the depth was fixed at 6 cm. The gain and the dynamic range were adjusted depending on the subject to enhance the quality of the images of the vertebrae. During each scanning, the operator moved the probe upward, starting at the fourth lumbar (L4). The position of the transducer was always adjusted to ensure that the spinous process was visible in the images. At the end of each acquisition, all subjects were questioned whether they experienced any inconvenience or discomfort during the procedure.

4.3.4 3D reconstruction of the spine

The data of each sweep was saved in one raw image file containing the raw images with the transformations required to perform the volume reconstruction. As part of the pre-processing, all the data that belonged to the US configuration was removed from each of the images acquired, and only the region of interest was saved (see Figure 4.6).

The collected sweeps were used to generate a freehand 3D reconstruction. The reconstruction consists of arranging every US image into a 3D volume. Then, the value of each voxel is determined by the weighted average of all the coinciding pixels, or simply by the last coinciding pixel. A method based on the interpolation of nearby voxels was included to

compute hole-filling. This process was computed using the image utilities of the PLUS software (Lasso et al., 2014).

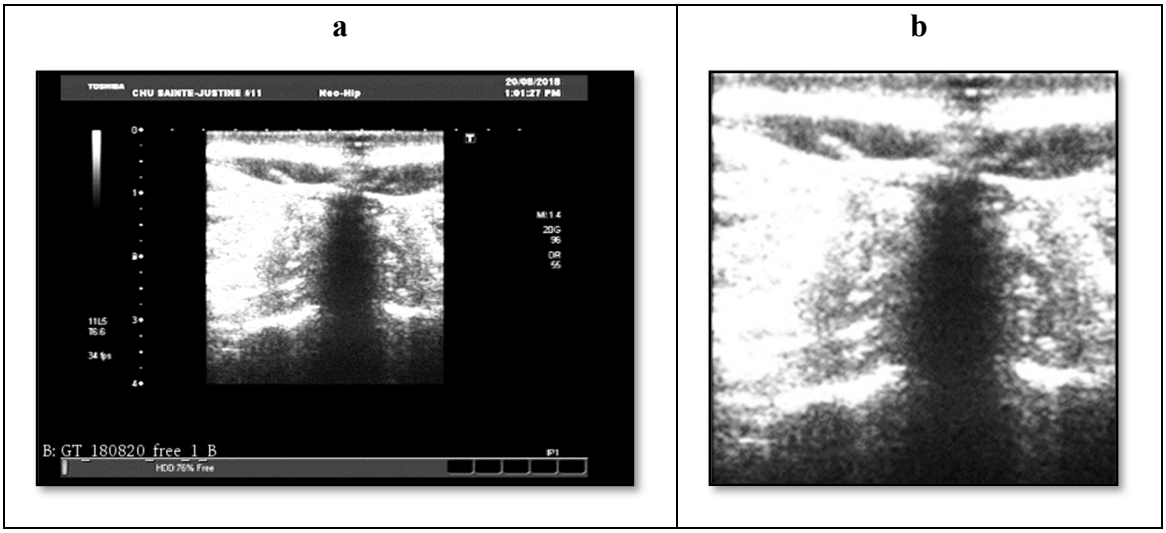


Figure 4.6 a) original raw image with dark margins and configuration from the US. b) cropped image showing the region of interest.

4.3.5 Anatomical landmark identification

Once the reconstruction was generated, the volume was displayed in 3D Slicer (Fedorov et al., 2012) using the volume rendering module. Then, guided by a physician, the operator manually identified the spinous processes on the volume reconstructions as anatomical landmarks. When the US waves go into the body, most of them are absorbed, and the rest are reflected to the transducer, which are used to generate the images. In the case of vertebrae, these reflect most of the sound waves, producing a bright section on the image. Also, since the waves cannot penetrate the osseous matter, an acoustic shadow is presented behind each vertebral body (Abu-Zidan, Hefny, & Corr, 2011).

From the sagittal view, we divided the reconstruction in two by identifying the inflexion point that divides the lumbar and thoracic sections of the spine. The vertebra L4 was the starting point at the bottom of the reconstruction. Also, since we marked and took a note of how many vertebrae were acquired of all subjects, we knew how many vertebrae we should identify in

the reconstruction. To recognize the spinous process, we looked for the acoustic shadows on the reconstruction. Using the three sweeps, we used the sweep 1 or 3 to locate the vertebrae in the sagittal plane (see Figure 4.7a). By modifying the volume rendering of sweeps 1 or 3, we were able to see in more detail the structure of the surface of the vertebrae. Figure 4.7b and Figure 4.7c show the depth of the spinous process in the sagittal view by modifying the display values in 3D Slicer. Finally, using sweep 2, we aligned the landmarks to the centerline of the spine (see Figure 4.7d).

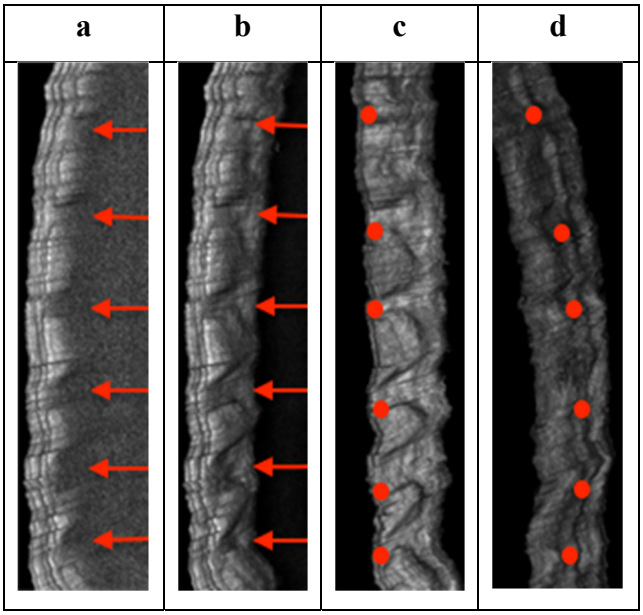


Figure 4.7 Identification of spinous processes

4.3.6 Posture quantification

Based on the 3D point-based model of the centerline of the spine, we divided the spine into thoracic and lumbar sections. To quantify the posture of the subject in each section, we used three measurements, the distance between spinous processes, and the angles of two adjoint spinous processes with respect to the horizontal in the posteroanterior and sagittal planes (see Figure 4.8). The root-mean-square error was calculated to evaluate the anatomical landmarks from each acquisition.

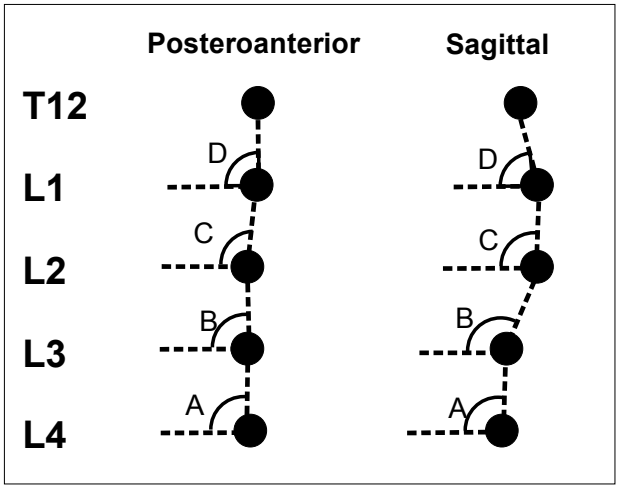


Figure 4.8 Calculation of the angle formed by two adjoint vertebrae (black dots) with respect to the horizontal axis in the lumbar section, from two planes

4.4 Results

Table 4.1 shows the anthropometric characteristics of the sample. Two indices were calculated based on these measurements. Body Mass Index (BMI), estimated by dividing the weight in kilograms by the square of the height in meters, and Waist-to-height Ratio (WtHR), calculated by dividing the waist circumference by the height.

Table 4.1 Anthropometric characteristics of the subjects involved in this study

Subject	Genre	Age	Waist (cm)	Height	Weight	BMI	WtHR (%)
1	F	33	80	168.5	61.6	21.70	47.48
2	F	32	74.5	172.5	59.9	20.13	43.19
3	M	37	70	178.5	60	18.83	39.22
4	M	32	89	182.5	74.8	22.46	48.77
5	F	28	77	158	55.5	22.23	48.73
6	F	33	76.5	159	51.6	20.41	48.11
7	M	25	73	179	69.6	21.72	40.78
8	F	21	71	166.5	57.3	20.67	42.64
Average:		**30±5.3**	**76.4±6.1**	**171.7±8.2**	**61.3±7.5**	**21.02±1.2**	**44.87±3.9**

Each acquisition was composed of three sweeps. In total, there were eight acquisitions per subject. The average time for the first acquisition was 20 minutes. For the following acquisitions, the average time was of 12 minutes, since no anthropometric characteristics or pictures were taken. Preparing the subject for each acquisition took around 5 minutes. This time included: providing initial instructions, changing of upper clothes for a gown, marking of the vertebrae, positioning of the subject and application of US gel. Table 4.2 shows the averages of the acquisition time and frames per sweep in three different setups. As well as the average of the disk space used to save the raw and pre-reconstruction data. Also, it displays the average time to generate a reconstruction and the disk space to save it.

Table 4.2 Statistics per one sweep in different setups. As part of the acquisition, time (seconds), number of frames and disk space (megabytes) used are presented. Also, disk space (megabytes) after selection the region of interest is displayed, together the reconstruction time and disk space for each computed reconstruction.

Setup	Average time per acquisition (sec.)	Average frames	Raw images (disk space)	Crop images (disk space)	Average reconstruction time (sec.)	Average reconstruction (disk space)
Exp 1*	40	1460	86	72	27	2
Exp 2†	40	1370	80	71	27	2
Exp 3*	20	691	40	36	30	1.8
Exp 3*	140	3800	226	200	50	2.1

*Constrained setup
†Unconstrained setup

4.4.1 Volume reconstructions

Figure 4.9 shows the L-sweep volume reconstruction of three subjects along with their corresponding picture from the sagittal plane. It can be observed that the reconstruction captures the shape of the spine of the subject.

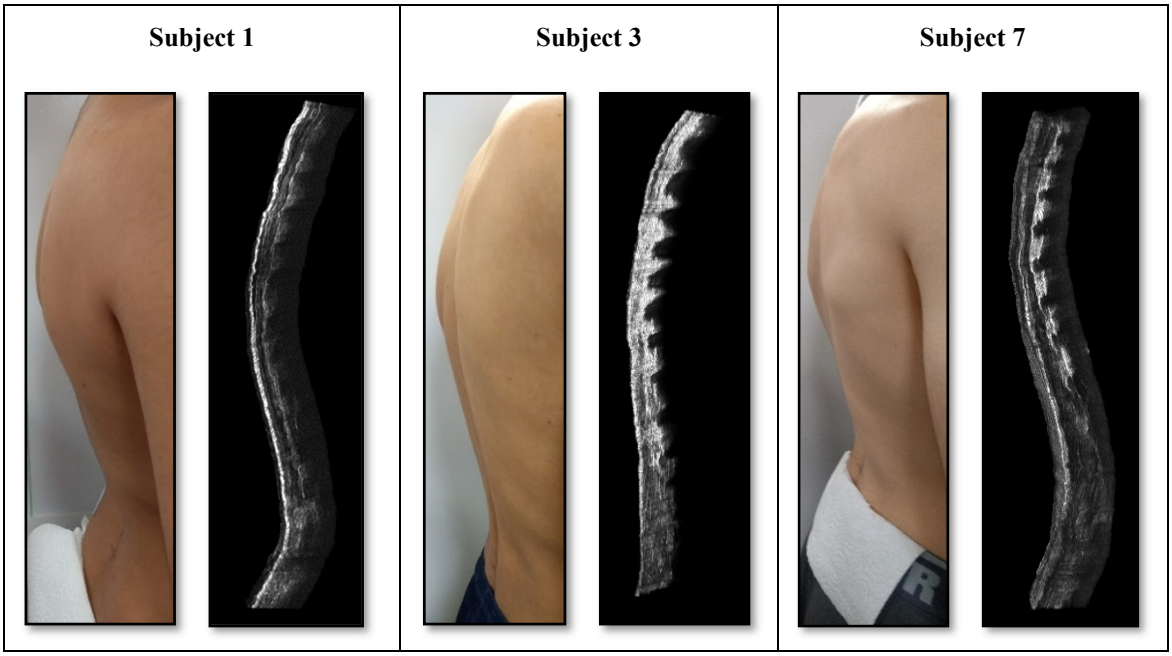


Figure 4.9 Volume reconstruction of the spine of three subjects

In the experiment 1 (constrained setup), we quantified the differences in the location of the spinous process marked on the freehand 3D reconstructions across different acquisitions. Three measurements were evaluated, distance between spinous processes, and the angle of two adjoint spinous processes in the posteroanterior and sagittal planes. We quantified the variation of the first three 3D-point-based models of the spine with respect to the fourth. The results of this evaluation are shown in Table 4.3.

In experiment 2 (unconstrained setup), we performed the evaluation of the two 3D-point-based models similarly as for experiment 1, the results are shown in Table 4.4.

In experiment 3, we evaluated the two 3D-point-based models of the spine obtained from the freehand 3D reconstructions, where the US images were acquired in fast and slow mode. The fast and slow models were contrasted against the fourth model from experiment 1, Table 4.5 and Table 4.6 respectively.

Table 4.3 Differences of the first three 3D point-based models of the spine with respect to the fourth in experiment 1

	Distance (mm)				**Posteroanterior angle (degrees)**				**Sagittal angle (degrees)**			
	Thoracic		Lumbar		Thoracic		Lumbar		Thoracic		Lumbar	
Subject	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD
1	5.82	5.44	6.84	6.41	6.94	6.09	7.59	5.96	7.92	5.64	10.23	8.29
2	5.32	5.19	4.44	4.05	6.66	6.48	2.39	1.73	10.42	9.98	11.79	10.51
3	10.48	9.85	5.19	4.31	5.12	4.01	4.40	4.29	5.52	4.97	4.07	2.80
4	7.57	7.38	3.67	2.14	3.50	3.32	5.70	5.45	5.64	4.84	4.65	4.50
5	5.52	5.08	5.23	4.05	5.81	5.73	10.41	3.91	7.28	7.16	6.26	4.88
6	5.29	4.81	4.69	3.27	6.97	6.51	8.16	6.57	5.50	5.16	11.12	5.78
7	6.62	5.57	3.48	3.18	3.78	3.35	9.40	8.86	5.30	4.56	5.77	4.77
8	5.40	5.23	2.78	2.73	6.51	6.16	5.87	5.17	7.97	7.70	5.08	4.80
Average	**6.50**	**6.07**	**4.54**	**3.77**	**5.66**	**5.21**	**6.74**	**5.24**	**6.94**	**6.25**	**7.37**	**5.79**

Table 4.4 Differences between two models obtained from an unconstrained setup

	Distance (mm)				**Posteroanterior angle (degrees)**				**Sagittal angle (degrees)**			
	Thoracic		Lumbar		Thoracic		Lumbar		Thoracic		Lumbar	
Subject	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD
1	5.03	5.01	2.00	2.98	2.83	2.83	4.06	1.50	7.03	7.00	8.17	7.00
2	6.04	5.76	9.65	1.87	5.23	5.16	3.22	2.26	10.56	10.50	12.83	8.15
3	4.45	4.43	5.09	3.49	6.03	5.68	3.61	1.56	4.57	3.92	3.62	4.55
4	6.16	6.07	7.33	1.98	3.67	2.84	6.92	1.01	8.45	8.04	9.04	5.24
5	6.67	6.57	6.73	2.58	9.56	9.46	6.47	6.06	8.79	8.61	6.53	7.39
6	7.97	7.86	5.42	2.86	6.66	5.33	9.42	4.73	6.54	6.50	9.92	8.46
7	3.20	3.14	7.68	2.40	9.85	9.56	5.03	3.23	3.91	3.91	7.70	7.82
8	3.52	3.48	2.97	2.21	8.64	8.64	5.12	3.57	8.85	8.80	7.05	5.66
Average	**5.38**	**5.29**	**5.86**	**2.54**	**6.56**	**6.19**	**5.48**	**2.99**	**7.34**	**7.16**	**8.11**	**6.78**

Table 4.5 Differences between the fast and the fourth model of experiment 1

	Distance (mm)				**Posteroanterior angle (degrees)**				**Sagittal angle (degrees)**			
	Thoracic		Lumbar		Thoracic		Lumbar		Thoracic		Lumbar	
Subject	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD
1	6.08	5.96	2.74	2.48	6.58	6.20	3.14	1.98	4.84	4.84	3.38	3.38
2	7.43	6.90	6.19	5.62	5.48	5.36	2.23	1.98	11.87	10.92	14.08	10.40
3	5.33	5.13	5.29	3.07	4.31	4.24	3.98	3.90	4.42	4.42	5.24	5.05
4	9.07	7.76	4.31	3.93	2.19	2.17	3.98	3.92	5.78	5.53	7.90	7.67
5	4.29	4.28	2.01	1.87	4.62	3.42	5.15	4.45	8.34	8.21	2.89	2.88
6	3.96	3.02	4.55	2.73	6.58	6.58	7.38	7.19	4.89	4.87	11.23	1.86
7	7.87	7.46	3.87	3.70	2.58	2.39	8.55	8.22	5.44	5.26	6.25	4.71
8	5.52	5.44	6.37	6.22	5.48	5.36	7.42	7.36	7.76	7.66	6.79	6.23
Average	**6.19**	**5.74**	**4.41**	**3.70**	**4.73**	**4.47**	**5.23**	**4.87**	**6.67**	**6.46**	**7.22**	**5.27**

Table 4.6 Differences between the slow and the fourth model of experiment 1

	Distance (mm)				**Posteroanterior angle (degrees)**				**Sagittal angle (degrees)**			
	Thoracic		Lumbar		Thoracic		Lumbar		Thoracic		Lumbar	
Subject	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD
1	4.19	4.19	4.30	3.90	6.55	6.53	7.92	7.81	6.21	6.21	3.49	3.49
2	4.91	4.91	6.91	6.25	7.23	7.22	2.87	2.48	9.76	7.72	8.72	7.35
3	4.12	4.10	6.67	5.90	2.28	2.26	5.76	5.75	4.25	4.24	6.19	4.43
4	8.43	6.03	8.07	6.68	4.10	3.11	3.73	2.88	6.67	6.28	8.44	8.01
5	3.84	3.81	2.72	2.60	5.33	4.73	5.80	3.59	8.00	7.19	10.74	3.60
6	3.95	3.07	5.23	1.51	6.21	5.66	7.23	7.10	8.20	7.94	10.99	6.72
7	6.50	5.93	3.82	3.64	3.90	3.89	9.65	8.34	4.53	4.31	9.42	8.20
8	4.52	4.51	4.43	3.99	7.19	7.18	5.68	5.45	6.30	6.30	10.58	10.25
Average	**5.06**	**4.57**	**5.27**	**4.31**	**5.35**	**5.07**	**6.08**	**5.42**	**6.74**	**6.27**	**8.57**	**6.50**

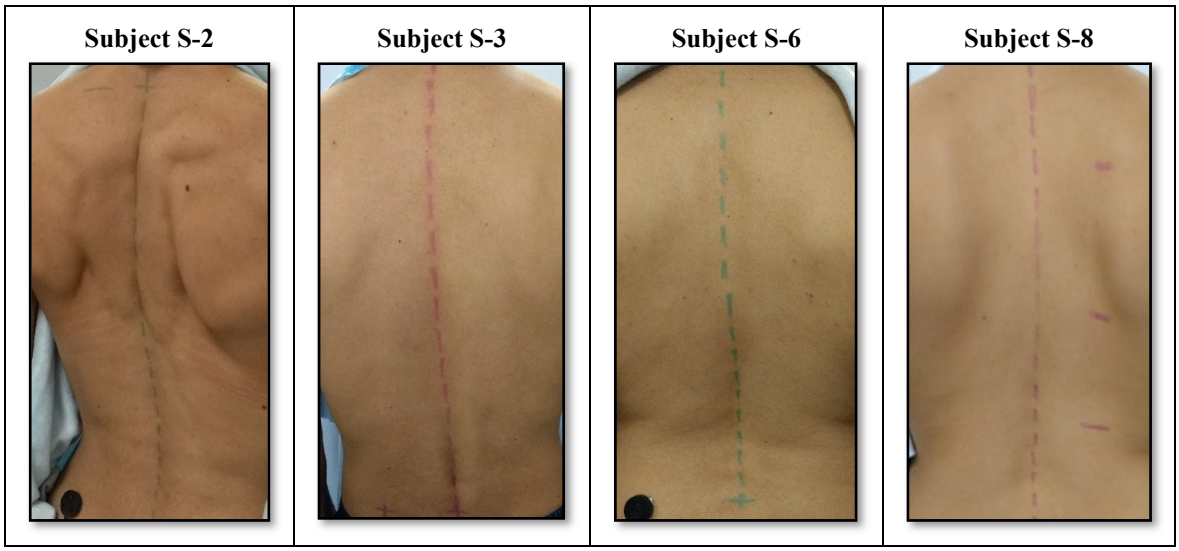


Figure 4.10 Subjects with different body composition. Subject S-2 has the leanest mass. Subject S-3 is not muscular, but slim build. Subjects 6 and 8 have healthy body compositions.

4.5 Discussion

In this study, we modeled the shape of the spine by using a freehand 3D US system. The objective is to examine the challenges of these alternative systems to characterize the spine in the 3D space. These systems are composed by three main devices, a 2D US scanner, a tracking system and a computer. There is no agreement regarding the characteristics of the hardware to setup these systems to scan the spine. However, in most of the studies, linear transducers (Chen et al., 2011, 2013; C. J. Cheung et al., 2015; C. W. J. Cheung et al., 2015, 2013; Koo et al., 2014; Ungi et al., 2014; Young et al., 2015) were more frequently used in comparison to convex ones (Zheng et al., 2016). Also, magnetic trackers (Chen et al., 2011, 2013; C. J. Cheung et al., 2015; C. W. J. Cheung et al., 2015, 2013; Ungi et al., 2014; Young et al., 2015) were more recurrent than optical trackers (Koo et al., 2014; Purnama et al., 2009). In most of these studies, a custom made software was implemented for the acquisitions and processing of the data; the exception was the work of Ungi *et al.* (Ungi et al., 2014), who used the PLUS library (Lasso et al., 2014).

The patient is required to be in standing position when modeling the shape of spine. This position is the gold standard in clinical practice to evaluate its morphology in the 3D space. In comparison to X-rays, the acquisition of US images from the spine takes more time. While sweeping the spine, an involuntary motion is produced, either by the operator when pushing the transducer on the spine, or by the subject's natural breathing movements and while trying to maintain a stable position. Zheng et al. (Zheng et al., 2016) used a support frame to fix the shoulders and the hips during their acquisitions. However, Bellefleur et al. (Bellefleur et al., 2002), showed that fixing the hips or shoulders produce a change in the natural position of the subject. As in (Ungi et al., 2014; Young et al., 2015), we preferred to adopt a natural position in our experiments.

We used a freehand 3D US system comprised of an US scanner, an EMS, a grabber connected to simple workstation. We employed the PLUS library, which provides the software and phantom models to ensure the reproducibility of the setup. One of the advantages of this framework is that it allows carrying out the acquisition protocol with different hardware without modifying the software.

In our cohort of 8 subjects, all of them were young healthy adults. Most of the participants fitted in the category of healthy weight by either of two indices, BMI or WtHR. According to the BMI, a healthy weight category is between 18.5-24.9. The categories for WtHR change depending on the gender. For a healthy weight in females, the ratio is between 42-48%, while for males it is between 43-50%. Subjects with values lower than these ranges are considered as underweight.

From the cohort, two of the subjects, S-2 and S-7 were the most muscular. Subject S-3 was not muscular, but slim. These three subjects were the ones with more lean mass. In particular, subjects S-3 and S-7 were categorized as underweight according to their WtHR. The rest of the subjects were in healthy weight under both BMI and WtHR (see Table 4.1).

A total of 64 acquisitions were performed. The acquisitions were simple to repeat each time following the protocol. To the best of his ability, the operator tried to scan the subject at the same frame rate during each sweep. However, as seen in Table 4.2, for experiments 1 and 2, there was an unavoidable speed change during the sweeps, hence, a variability in the number of frames acquired. The operator preformed three sweeps per acquisition, which were used to generate the reconstruction of the spine (see Figure 4.9). In general, the reconstruction can be made in less than a minute, and the disk space use to store them is similar to the one used for radiograph images (see Table 4.2).

For most of the cases, the operator only applied US gel one time, covering the region of interest before the acquisition. However, due to the stiffness and thickness of the muscles in subjects S-2 and S-7, the operator had to put more US gel and apply more pressure on the transducer to assure that the vertebrae were visible in the US images, mainly in the thoracolumbar/lumbar region (see Figure 4.11).

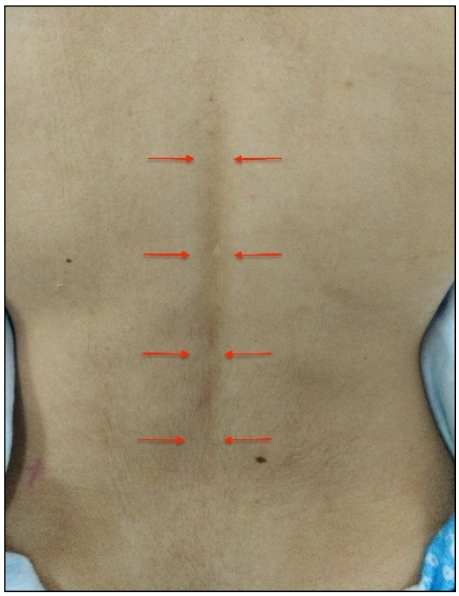


Figure 4.11 Problematic region on thoracolumbar/lumbar section of the spine due to the stiffness of the muscles in subjects S-2 and S-7.

Electromagnetic measuring systems are important tools for computer-assisted interventions. These systems have been used to determine the position and orientation of some sensors or medical instruments relative to the anatomy of a patient. In our case, the EMS allowed us to know the position and orientation of the transducer. According to the manufacturer, the transducer can be tracked in a magnetic field in the shape of a cube of 50 cm per side. During the acquisitions, we were able to capture the images at a length of 45 cm approximately. This was the main limitation to capture the full shape of the spine of the subjects. Also, the height of the subject played an important role in the number of vertebrae that we acquired (see Table 4.7).

Table 4.7 Number of vertebrae acquired by subject

Subject	**Height**	**Vertebrae**
1	168.5	16
2	172.5	15
3	178.5	14
4	182.5	14
5	159	16
6	158	15
7	179	16
8	166.5	14

An advantage of the EMS is that there is an uninterrupted line of sight between the subject and the navigation tools, so it is possible to use the EMS in applications such as tracking endoscopes or minimally-invasive procedures (Schicho et al., 2005). Nevertheless, in clinical setups, the disturbance caused by metal medical instruments, must be considered in order to know how they will affect the accuracy and precision of the measurements. In our setup, we detected that the EMS's system control unit caused noisy images when it was in a range of less than 100 cm from the US scanner.

The limitation of the length of the field of view could be overcome by using an optical tracker, as long as the transducer is visible at all the time. Recently, Prevost et al. (Prevost et al., 2018) proposed a method based on convolutional neural network aimed to perform a freehand 3D US reconstructions by incorporating an internal measurement unit instead of and EMS or an optical tracking system. This method could be an option to replace either of the trackers, but its efficiency on spine reconstructions requires further research.

Before computing the volume reconstructions, the US images were cropped and only the region of interest was left (see Figure 4.6). Then, the volume reconstruction of each of the three sweeps for every subject was generated. When the US waves go through the skin, some of them are reflected to the transducer as an echo as soon as they come into contact with tissues or osseous matter. These echoes then are processed to form images. Every interface in the body reflects the echoes in different amount depending on the density of the tissue and the speed of the sound wave. This is called *acoustic impedance*. Fat, muscle and bone have acoustic impedances of 0.138 $g \cdot cm^{-1} \cdot s^{-1}$, 0.170 $g \cdot cm^{-1} \cdot s^{-1}$ and 0.78 $g \cdot cm^{-1} \cdot s^{-1}$ respectively (Wagner, 2013). Since the impedances of fat and muscle are lower than bones, the former will produce weaker echoes. On the US images, weaker reflections appear as grey pixels, while stronger reflection appear brighter. In the cases of solid structures like the vertebrae, they do not absorb the US waves, which produces an acoustic shadow (Wagner, 2013). Figure 4.9 shows the sagittal profile of the reconstructions along with the correspondent image of the subject in the same profile. Subject S-3, the participant with the lowest BMI and WtHR, had the brightest US images and the acoustic shadows generated by the vertebrae are more evident. Also, the vertebrae of subject S-7, one of the two with more muscle mass, are less visible compared to subject S-3, but clearer in contrast to subject S-1.

Each of the three sweeps per acquisition can be reconstructed individually, or they can be put together to generate a unique volume. The disadvantage of a unique volume is that identifying the structure of the vertebrae is more time-consuming and less evident in comparison to the three individual reconstructions. For this reason, we decided to use the three separate reconstructions to perform the vertebral-level identification.

Landmark identification is not a trivial work. The thoracic section contains the ribs, which produce similar reflections to those of the vertebrae. The lumbar section contains more muscles in comparison to the thoracic section. Therefore, the muscles tend to occlude the vertebrae. In contrast to radiographs, in US images only the surface of certain regions of the vertebrae are visible. The spinous process(Brink et al., 2017; C. J. Cheung et al., 2015; Koo et al., 2014; Zheng et al., 2016), transverse process (Brink et al., 2017; Chen et al., 2011; C. J. Cheung et al., 2015; C. W. J. Cheung et al., 2015, 2013; Koo et al., 2014; Purnama et al., 2009; Ungi et al., 2014), superior articular process (C. W. J. Cheung et al., 2015, 2013; Purnama et al., 2009) and laminae (Chen et al., 2011, 2013; Purnama et al., 2009; Young et al., 2015) are the common landmarks used in US images.

In this study, we were able to mark the spinous processes to generate a 3D point-based model of the spine. This was performed manually by the operator using the volume rendering module of 3D Slicer. We evaluated these models in the thoracic and lumbar sections by calculating the distance between each spinous process, as well as the angles of two adjoint spinous processes with respect to the horizontal axis in the posteroanterior and sagittal planes. The angle in the posteroanterior plane indicates the displacement of the spinous process on the *y* axis, while the angle in the sagittal plane indicates the depth of the spinous process.

For experiment 1, we used the fourth model as reference to compare the first three. We used this model because, by the time of the fourth acquisition, the operator had gained experience and was able to adjust more efficiently the parameters of the US scanner and the EMS.

Table 4.3 and Table 4.4 contrast the measurements between the experiment 1 (constrained setup) and experiment 2 (unconstrained setup) respectively. The errors were similar between both setups. However, in the reconstructions we noticed that motion of the subjects was more evident in the unconstrained setup. Figure 4.12 shows two reconstructions of the C-sweep from subject S-4. The operator detected that this individual was more prone to move during the sweeps. For subject S_2, the identification of the spinous process in the lumbar section was more difficult, which is reflected as higher errors in Tables 4.3 and 4.4. This was due to the

stiffness and thickness of the muscles, which did not permit a good contact with the transducer. In both setups, it can be seen that the highest errors were produced at the lumbar section.

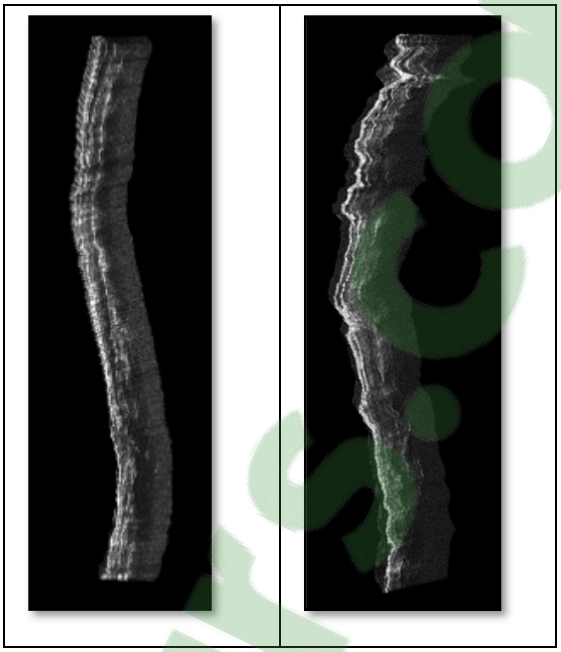


Figure 4.12 Motion in reconstructions from constrained (left) and unconstrained (right) setups

During the acquisition, the subjects reported that they felt more comfortable in the constricted setup, since they had a reference while looking forward and could maintain the position. Also, the operator manifested that the motion of the subject during the acquisitions with the unconstrained setup was more noticeable. Thus, we used the constrained setup for experiment 3. For the acquisitions in this experiment, the operator put extra marks on the back of the subject (see subject 8 in Figure 4.10). These marks helped the operator to control the motion of the transducer, and to cover regions of the spine equitably in both the fast and slow sweeps.

We compared the models obtained from the fast and slow acquisitions with the fourth model from experiment 1. Table 4.5 and Table 4.6 show the results of these comparisons. Although the results are similar in terms of the error in the three measurements, for the subjects, it was tedious to keep the relaxed standing position for 140 seconds per sweep. Hence, the subjects

tended to move more. Also, it was more tiring for the operator to move the transducer upward in a slow fashion.

The greater or lesser number of frames per sweep influenced the resolution of the volume reconstruction. Since the operator cannot keep the same pace during the sweeps, the number of slices vary in different regions of the spine. When the spacing between slices is large, the resolution of the volume is low. On the other hand, when the slices are close, a higher resolution volume reconstruction can be generated. Since the errors are similar in either of the three experiments, we considered that 40 seconds per acquisition allows a good compromise between time, subject's and operator's comfort, and number of frames per sweep (around 1400).

Recently, a volume projection imaging was proposed by Cheung et al. (C. J. Cheung et al., 2015), to generate a coronal representation of the spine from tracked US images. This method has been implemented in the freehand 3D system proposed by Zheng et al. (Zheng et al., 2016). They use this projection to calculate an angle equivalent to the Cobb angle. However, the angles based on this projection underestimate the spinal deformity compared to the Cobb angle. This implies a double disadvantage for the method, since it is well known that Cobb angle has already a high-variability of measurements (Majdouline et al., 2007). Since AIS is a 3D deformation, a 2D measurements cannot describe spine in enough detail.

For our approach, we used a volume reconstruction and we marked the spinous processes to generate a 3D representation of the spine. Although the landmark identification is a time-consuming process, we were able to identify the spinous processes in the 3D space. We also found out that analyzing in more detail the reconstructions' geography in some subjects, the laminae could be identified (see Figure 4.13). The correct location of the center of the laminae will be investigated in future work.

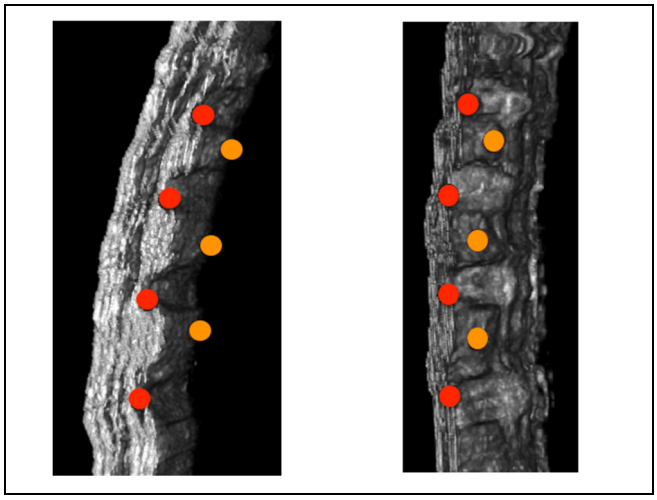


Figure 4.13 On the left, a sagittal view of four thoracic vertebrae. On the right a frontal view of the same vertebrae. Red dots indicate the spinous processes, and orange dots indicate the laminae

In future work, we will use a wider transducer to perform only one sweep per subject, which will reduce the acquisition time. Also, the transducer should have a higher penetration capability, to evaluate if landmarks can be extracted from freehand 3D reconstructions on overweight subjects. Other tracker systems will be evaluated to capture the full shape of the spine.

The manual placement of the markers on the volume reconstruction is a time-consuming process, which depends on the interpretation of the operator, mainly in the lumbar section where the vertebrae are less visible. Practice may improve US image acquisition by the operator, as well as reduce the variability of the landmarking.

We aim to improve the quality of the reconstructions by performing an automatic segmentation of the vertebrae similar to the one done in (Berton, Cheriet, Miron, & Laporte, 2016). The correct segmentation of the vertebrae could help to reduce the manual work and subjectivity when placing the landmarks.

4.6 Conclusions

In this study, we made a proof of concept of using a freehand 3D US system for 3D modeling of the spine. For this, we acquired US images from the back of eight healthy subjects. This is a radiation-free alternative to X-rays for describing the spine shape. From each subject, we obtained 3D models that represent the centerline of the vertebral column, based on the spinous process of each vertebrae.

The errors in the landmarks in either of the three setups tested were similar. This can be caused because the subjects do not maintain the exact same position throughout the different acquisitions, the operator moved the transducer at different speeds, or simply because there is a margin of error when placing the landmarks on the reconstructions.

Examining the results of our experiments, it is our belief that our proposed method could be used to help in the assessment and monitoring of AIS. This assessment could be performed by analyzing the landmarks detected on the 3D reconstruction of the spine, which is generated by tracking US images from the back of the subject and provide more information about the morphology of the spine than 2D measurements. For such an assessment, we consider that the constrained setup, would favor the evaluation of patients holding a stable natural position. The acquisition could be performed in 40 seconds. This time provides a good trade-off between the number of frames per acquisition, and the comfort of patients and operators.

Additional studies with more participants are required to support the identification of the landmarks and to ensure the reproducibility of the protocol. Ideally, participants should present different types of spine curvatures and body composition. Further studies with a cohort of AIS patients would help to validate the application of this method in the follow-up of this disease. It is also necessary to investigate the potential application of computer-based descriptors to characterize the deformation of the spine based on models obtained from tracked US images. This could help to improve the comprehension of the 3D nature of spinal deformations, as well

as to assist in the adequation of treatments and follow up of patients by using a radiation-free technology.

After our experimentations, we consider that, ideally, a freehand 3D US framework should able to:

- Guarantee its usability in individuals with different body compositions.
- Capture in a single sweep the full length of the spine, and the width of each vertebra.
- Generate a fast volume reconstruction with enough quality to easily identify the landmarks of the spine. This could be accomplished by performing preprocessing steps to improve the identification of the vertebrae on the US images.
- Automatize the landmarking process to decrease variability in the generation of 3D models from US images.

Since this framework uses a radiation-free technology, patients could be examined more frequently and decrease the use of X-rays for follow-ups, which could help clinicians to adapt more effectively patients' treatments.

CHAPTER 5

DISCUSSION AND CONCLUSION

Considerable research has been made to understand the complexity of adolescent idiopathic scoliosis. The gold standard methods for assessing spinal deformities are limited since they are based on the analysis of radiographs, which implies that the observations are restricted to 2D measurements, limiting the characterization of the full morphology of the spine. Moreover, the frequent exposition to radiation during the follow-up period on young patients could lead to other health problems, such as an increased risk of cancer.

Patients with AIS must undergo constant monitoring, a process that could last a few years depending on how early the deformation was detected, and on the magnitude of the curve. If the curve increases with time, the patients will need more regular examinations. In severe cases, the patients will need surgical correction. The uncertainty in curve progression brings with it concerns, both, in patients and their families, and in the clinicians. On the patient and family side, knowing a possible outcome of the treatment could help to cope with the mental stress related to the use of bracings. On the medical side, if clinicians could know beforehand how the shape of the spine of each patient will vary through time, they would be able to provide a patient-specific treatment. This could improve outcomes and reduce treatment times, the number of medical visits, and therefore, the exposure to radiation.

New computer-based studies have been proposed to address the limitations in key areas such as early detection and 3D classification of the spinal deformation, prediction of the curve progression, and reduction of radiation during follow-ups. However, the adaptation of these studies in a clinical setup is still challenging since they are difficult to interpret and to apply in everyday practice.

In this research we delineated as main objective the design of a framework to characterize and model the variation of the 3D shape of the spine through time. In Chapter 1, we presented a literature review, which highlighted the state-of-the-art and its limitations with regard to the main objective. From here, we defined three specific objectives: 1) 3D characterization and classification of spinal curves, 2) prediction of the shape of the spine, and 3) radiation-free imaging for modeling the shape of spine in 3D.

In Chapter 2, we established our first contribution, which comprises two parts, characterization of the spine and a classification method to categorize curvatures. We detected a disconnection between the gold standard method to quantify spinal deformations and the computer-based descriptors used to characterize the spine. Therefore, instead of proposing another automated descriptor, we introduced two new measurement techniques. These techniques use angles to characterize the spine in 3D. The use of angles for this task is well known in clinical practice.

The two angle-measurement proposed techniques were called *leave-n-out* and *fan leave-n-out*. These techniques consist of determining the angle of one vertebra with respect to the adjacent vertebrae in any of the 3 planes, assuming that similar spinal curvatures would share similar angles. The *leave-n-out* measurement has the advantage that it could be estimated using the same technique employed for estimating the Cobb angle, which is a well-known technique among clinicians and facilitates the understanding and adopting of this method in clinical practice.

The Lenke classification uses 2D measurements to categorize spinal deformations. To include 3D descriptors in this method could help to improve the categorization of the spines. Compared to new 3D classification methods proposed in literature, which are mainly based on cluster techniques, we aimed to design a computer-based method that can be used to extend the current Lenke classification.

We presented a novel method to classify spinal curvatures. This method was called dynamic ensemble selection of learner-descriptor classifiers. This consisted into first, training a set of

predictors with different descriptors. And second, categorizing test samples. This categorization was performed by automatically selecting the most suitable learner-descriptor classifiers. Therefore, the method does not depend on a specific descriptor or predictor, but rather, on the best combination of them that describes each test sample. This flexibility allows the user to combine clinical or computer-based descriptors to analyze how each of them is involved in the classification of a specific curvature, or in a set of similar curvatures.

The clinical relevance of this contribution lies on providing clinicians with an approach that can extend their current gold standard methods. On the one side, the angles are easy to interpret and help describing the spine in the 3D space. On the other hand, the dynamic selection method could aid to disentangle complex or borderline deformations by providing the more suitable descriptors to define them. This could assist clinicians to improve the characterization of these deformities by identifying patient-specific descriptors, and hence, provide tailored treatments.

The classification method has the limitation of needing a large quantity of data. It requires to exclude part of the training data to create a validation set. Also, the prediction of a complex test sample can be time-consuming. This is because the algorithm must look for the *k*-nearest-neighbors, classify them, and then generate the final ensemble to classify a test sample. This presents a limitation for real-time prediction applications.

The main reason for performing follow-ups of patients is the concern of clinicians for the risk of an increase in the magnitude of the deformation. In Chapter 3, we presented our second contribution, which consisted in longitudinally simulate the changes in the shape of the spine. When the Cobb angle increases 6 degrees in the major curvature between the first and the current visits, the curve is considered as progressive. The current state-of-the-art does not provide methods to visualize how this progression of the curve is modifying the 3D shape of the spine in-between visits. Therefore, our proposed approach projects the shape of the spine every three months from the first visit, for up to 18 months. We aimed to model the 3D shape of the spine every three months to provide information about how the spine changes in the short term.

Our approach was trained with modes of variation obtained from a dataset containing 3D reconstruction models of patients with AIS. The modes of variation were computed applying Independent Component Analysis. We decided to apply ICA because it is effective to identify the sources that produce the variations. We compared the modes of variation with a low-dimensional representation of the same 3D models of the spine, produced from Stacked Denoising Autoencoders. Our simulations with both representations generated similar results. However, the advantage of using the modes of variation over the low-dimensionality representation is that the first ones can be visualized for interpretation and comparison.

We compared two schemes for the simulation. The first, called *short-memory*, was trained with data of the immediate previous visit. The second, named *long-memory*, was trained taking into account all previous visits. Our experiments showed that the *long-memory* scheme foretells 3D models of the spine closest to the real progression of patients in comparison to the *short-memory* one.

Our clinical contribution of simulating the shape of the spine from the first visit is to help clinicians to detect curves that might progress through time. Identifying patients at risk of progression could aid clinicians to plan better treatments based on our predicted models. In addition, we can easily adapt our approach to simulate the shape of the spine in shorter or larger periods of time if necessary.

The main limitation of our second contribution is that we had to generate interpolated models when a patient does not have an actual model at a specific time span. The ideal situation would be to have a very short gap in between actual visits.

In our first contribution, we characterized the spine and then we classified deformations. In the second contribution, we simulated the shape of the spine through time. In both contributions we used a dataset containing 3D model reconstructions obtained from radiographs. Therefore, we were motivated to find a radiation-free imaging modality for generating similar 3D model reconstructions. The acquisition of the images should be performed in a standing position.

Ultrasound technology fulfilled all these requirements. In Chapter 4, we present our third contribution, a protocol to model the spine from tracked US images.

US images were acquired using a freehand 3D US system. This system was comprised of a US scan, an electromagnetic measurement system, and a workstation. In our experimentation, we were able to generate a 3D representation of the centerline of the spine. This representation was derived by marking the spinous processes on a 3D reconstruction of the spine, which was generated from tracked ultrasound images. We tested different setups varying the acquisition time. We also tested a setup in which patients were standing up with a visual reference to help them maintain a stable position, and another one without the point of reference.

After our experimentation, we considered that an acquisition of the full shape of the spine can be achieved in 40 seconds. This is a tolerable time for the patient to maintain a stable natural position, and sufficient to capture enough frames for generating the volume reconstruction. We also found that patients felt more comfortable and maintained a more stable standing position when provided with a visual reference.

We evaluated patients with different body compositions. For subjects with more lean mass, the images were difficult to acquire in the lumbar section since they had thicker muscles, and the transducer did not make good contact with the spine. However, the vertebrae were easier to identify on the images. On the contrary, the image acquisition process was easy in subjects with an average body composition, but the spinous processes were not as clear on the images as with leaner patients. Nevertheless, with both types of body compositions we were able to generate 3D models.

The medical significance of our third contribution lies on supplying clinicians with a protocol that could be integrated in clinical setups for the assessment and monitoring of AIS. This protocol provides guidance on a setup for US image acquisitions. Since this protocol uses a radiation-free technology, the radioactive exposure during monitoring could be reduced.

Our protocol has two limitations with respect to the hardware devices. The magnetic field was insufficient to scan the full shape of the spine, and the transducer was not wide enough to capture the full length of the vertebrae.

In contrast to scanning systems, non-contact 3D optical systems have been used to reconstruct the body surface to reduce exposure to radiation. These systems could help to assess the external asymmetry of the patients, as well as their progression. However, body composition will play a role in the evaluation of the asymmetry in extremely obese individuals. The fast acquisition of these system is one of its advantages, compared with tracked ultrasound systems. Moreover, lighting conditions and involuntary movement could interfere with the accuracy of the 3D reconstructions. Tracked ultrasound and non-contact 3D optical could complement each other to improve the monitoring of the shape of the spine using radiation-free technologies.

Radiation-free acquisitions could be intercalated during regular monitoring. Our framework is a step forward on the inclusion of radiation-free technologies to reduce potential adverse effects caused by the constant use of X-rays in the immature tissue of young patients. In addition, our proposed methods for characterizing the spine curvature and its progression could provide clinicians not only with automated classification of curves, but also with interpretable information that may lead to an insight to help them design patient-specific treatments.

In our framework, we designed methods to characterize and model the variation of the spine. However, further studies with patients with AIS still need to be performed to evaluate the appropriateness of the protocol for follow-ups of this condition. At least two things should be evaluated in further studies, 1) its usability for assessing spine deformities, and 2) its reproducibility from the first visit, up to the end of the follow-up.

Also, the inclusion of an automated segmentation of the vertebra on the US images could help to ease the visualization of more landmarks on the surface of the reconstruction of the spine. Automated landmarking could be performed to reduce the variability of manual marking, and the speed to generate the shape model of the spine.

Skeletal maturity is an important index for predicting progression in AIS. A volume reconstruction of the humeral head can help in the identification of ossification patterns to assess skeletal maturity. An extension of our current protocol could include the acquisition of tracked US images of the humeral head for this purpose.

Our first two contributions were made by using 3D reconstructions of stereographic radiographs. As a future work, it would be highly valuable to evaluate the effectiveness of our proposed methods with 3D models of the shape of the spine from tracked US images.
Another open venue of research is to perform clustering analysis through time using the modes of variation to classify progressive and non-progressive spines. We hypothesize that spines with low progression will remain in the same cluster, while spines with higher progression will move to other clusters containing spines with higher curvatures. The advantage of this method would be to use progression curve patterns instead of a strict 6 degrees cut-off from 2D measurements.

APPENDIX I

Publications during Ph.D.

Journals

1. **Garcia-Cano, E.**, Arámbula Cosío, F., Torres Robles, F., Fanti, Z., Bellefleur, C., Joncas, J., Labelle, H., Duong, L. "A freehand ultrasound framework for spine assessment in 3D: a preliminary study". In *Ultrasound in Medicine and Biology* (under revision)
2. **Garcia-Cano, E**., Arámbula Cosío, F., Duong, L., Bellefleur, C., Roy-Beaudry, M., Joncas, J., … Labelle, H. (2018). "Prediction of spinal curve progression in Adolescent Idiopathic Scoliosis using Random Forest regression". In *Computers in Biology and Medicine*, *103*(April), 34–43. http://doi.org/10.1016/j.compbiomed.2018.09.029
3. **Garcia-Cano, E**., Arámbula Cosío, F., Duong, L., Bellefleur, C., Roy-Beaudry, M., Joncas, J., … Labelle, H. (2018). "Dynamic ensemble selection of learner-descriptor classifiers to assess curve types in adolescent idiopathic scoliosis". In *Medical and Biological Engineering and Computing*, 1–11. http://doi.org/10.1007/s11517-018-1853-9

Conferences

1. **Garcia-Cano, E**., Arámbula Cosío, F., Duong, L. "Shape Analysis of the Spine from Ultrasound Images ". In *2nd International Forum of Mexican Talent, Innovation Match*, Mexico City. 2017.
2. **Garcia-Cano, E**., Arambula-Cosío, F., Duong, L., Bellefleur, C., Roy-Beaudry, M., Joncas, J., … Labelle, H. (2016). Dynamic ensemble selection of learner-descriptor classifiers to assess curve types in adolescent idiopathic scoliosis: a preliminary study. In *46e Réunion Annuelle de la Société de la Scoliose du Québec*. Estérel, QC.

3. **Garcia-Cano, E**., Arambula-Cosío, F., Duong, L. (2015). “3D classification of Adolescent Idiopathic Scoliosis using an ensemble-of-classifiers approach”. In *45e Réunion Annuelle de la Société de la Scoliose du Québec*. Estérel, QC.

Other publications

1. **Garcia-Cano, E**., Arámbula Cosío, F., Duong, L., Bellefleur, C., Roy-Beaudry, M., Joncas, J., … Labelle, H. (2018). “Dynamic Classification of Abnormal Spinal Curves”. In *Substance ÉTS*.

BIBLIOGRAPHY

Abu-Zidan, F. M., Hefny, A. F., & Corr, P. (2011). Clinical ultrasound physics. *Journal of Emergencies, Trauma and Shock*, *4*(4), 501–503.

Asher, M. A., & Burton, D. C. (2006). Adolescent idiopathic scoliosis: Natural history and long term treatment effects. *Scoliosis*, *1*(1), 1–10.

Assi, K. C., Labelle, H., & Cheriet, F. (2014). Statistical model based 3D shape prediction of postoperative trunks for non-invasive scoliosis surgery planning. *Computers in Biology and Medicine*, *48*(1), 85–93.

Aubin, C.-E., Labelle, H., & Ciolofan, O. C. (2007). Variability of spinal instrumentation configurations in adolescent idiopathic scoliosis. *European Spine Journal*, *16*(1), 57–64.

Bellefleur, C., Dansereau, J., Koller, A., & Labelle, H. (2002). Evaluation of the efficiency of patient stabilization devices for 3D X-ray reconstruction of the spine and rib cage. *Studies in Health Technology and Informatics*, *88*(Figure 1), 127–131.

Bengio, Y. (2009). Learning Deep Architectures for AI. *Foundations and Trends® in Machine Learning*, *2*(1), 1–127.

Berton, F., Cheriet, F., Miron, M. C., & Laporte, C. (2016). Segmentation of the spinous process and its acoustic shadow in vertebral ultrasound images. *Computers in Biology and Medicine*, *72*, 201–211.

Boisvert, J., Cheriet, F., Pennec, X., & Labelle, H. (2008). Principal Deformations Modes of Articulated Models for the Analysis of 3D Spine Deformities. *Electronic Letters on Computer Vision and Image Analysis*, *7*(4), 13–31.

Boisvert, J., Cheriet, F., Pennec, X., Labelle, H., & Ayache, N. (2008). Geometric variability of the scoliotic spine using statistics on articulated shape models. *IEEE Transactions on Medical Imaging*, *27*(4), 557–68.

Borchani, H., Varando, G., Bielza, C., & Larrañaga, P. (2015). A survey on multi-output regression. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *5*(5), 216–233.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Brink, R. C., Wijdicks, S. P. J., Tromp, I. N., Schlösser, T. P. C., Kruyt, M. C., Beek, F. J. A., & Castelein, R. M. (2017). A reliability and validity study for different coronal angles using ultrasound imaging in adolescent idiopathic scoliosis. *Spine Journal*.

Britto, A. S., Sabourin, R., & Oliveira, L. E. S. (2014). Dynamic selection of classifiers - A comprehensive review. *Pattern Recognition*, *47*(11), 3665–3680.

Busscher, I., Wapstra, F. H., & Veldhuizen, A. G. (2010). Predicting growth and curve progression in the individual patient with adolescent idiopathic scoliosis: design of a prospective longitudinal cohort study. *BMC Musculoskeletal Disorders*, *11*(1), 9.

Chen, W., Lou, E. H. M., & Le, L. H. (2011). Using ultrasound imaging to identify landmarks

in vertebra models to assess spinal deformity. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, (August), 8495–8498.

Chen, W., Lou, E. H. M., Zhang, P. Q., Le, L. H., & Hill, D. (2013). Reliability of assessing the coronal curvature of children with scoliosis by using ultrasound images. *Journal of Children's Orthopaedics*, *7*(6), 521–529.

Cheng, J. C., Castelein, R. M., Chu, W. C., Danielsson, A. J., Dobbs, M. B., Grivas, T. B., Burwell, R. G. (2015). Adolescent idiopathic scoliosis. *Nature Reviews Disease Primers*, *1*, 15030.

Cheung, C. J., Zhou, G., Law, S., & Mak, T. (2015). Ultrasound Volume Projection Imaging for Assessment of Scoliosis. *IEEE Transactions on Biomedical Engineering*, *34*(8), 1760–1768.

Cheung, C. W. J., Siu-Yin Law, & Zheng, Y.-P. (2013). Development of 3-D ultrasound system for assessment of adolescent idiopathic scoliosis (AIS): And system validation. *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6474–6477.

Cheung, C. W. J., Zhou, G. Q., Law, S. Y., Lai, K. L., Jiang, W. W., & Zheng, Y. P. (2015). Freehand three-dimensional ultrasound system for assessment of scoliosis. *Journal of Orthopaedic Translation*, *3*(3), 123–133.

Cheung, J., Veldhuizen, A. G., Halberts, J. P. K., Sluiter, W. J., & Horn, J. R. Van. (2006). Geometric and electromyographic assessments in the evaluation of curve progression in idiopathic scoliosis. *Spine (Phila Pa 1976)*, *31*(3), 322–329.

Cheung, J., Veldhuizen, A. G., Halbertsma, J. P., Maurits, N. M., Sluiter, W. J., Cool, J. C., & Van Horn, J. R. (2004). The relation between electromyography and growth velocity of the spine in the evaluation of curve progression in idiopathic scoliosis. *Spine (Phila Pa 1976)*, *29*(9), 1011–1016.

Chollet, F. and others. (2015). Keras. Github.

Comon, P. (1994). Independent component analysis—a new concept? *Signal Processing*, *36*(3), 287–314.

Cootes, T. F., Hill, A., Taylor, C. J., & Haslam, J. (1994). Use of active shape models for locating structures in medical images. *Image and Vision Computing*, *12*(6), 355–365.

Cootes, T. F., & Taylor, C. J. (1999). A mixture model for representing shape variation. *Image and Vision Computing*, *17*(October 1997), 567–573.

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models-their training and application. *Computer Vision and Image …*, *61*.

Delorme, S., Petit, Y., De Guise, J. A., Labelle, H., Aubin, C.-E., & Dansereau, J. (2003). Assessment of the 3-d reconstruction and high-resolution geometrical modeling of the human skeletal trunk from 2-D radiographic images. *IEEE Transactions on Biomedical Engineering*, *50*(8), 989–98.

Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In *Multiple Classifier Systems. MCS 2000. Lecture Notes in Computer Science* (Vol. 1857, pp. 1–15). Springer, Berlin, Heidelberg.

Dimeglio, A., & Canavese, F. (2013). Progression or not progression? How to deal with adolescent idiopathic scoliosis during puberty. *Journal of Children's Orthopaedics*, *7*(1), 43–49.

Donzelli, S., Poma, S., Balzarini, L., Borboni, A., Respizzi, S., Villafane, J. H., Negrini, S. (2015). State of the art of current 3-D scoliosis classifications : a systematic review from a clinical perspective. *Journal of NeuroEngineering and Rehabilitation*, *12*(1), 1–11.

Doody, M. M., Lonstein, J. E., Stovall, M., Hacker, D. G., Luckyanov, N., & Land, C. E. (2000). Breast cancer mortality after diagnostic radiography: findings from the U.S. Scoliosis Cohort Study. *Spine*, *25*(16), 2052–2063.

Duong, L., Cheriet, F., & Labelle, H. (2006). Three-dimensional classification of spinal deformities using fuzzy clustering. *Spine*, *31*(8), 923–30.

Duong, L., Cheriet, F., & Labelle, H. (2010). Automatic Classification of Spinal Deformities Using Support Vector Machines. *IEEE Transactions on Biomedical Engineering*, *57*(5), 1143–1151.

Duong, L., Cheriet, F., Labelle, H., Cheung, K. M. C., Abel, M. F., Newton, P. O., Stokes, I. A. F. (2009). Interobserver and intraobserver variability in the identification of the lenke classification lumbar modifier in adolescent idiopathic scoliosis. *Journal of Spinal Disorders and Techniques*, *22*(6), 448–455.

Duong, L., Mac-Thiong, J.-M., Cheriet, F., & Labelle, H. (2009). Three-dimensional subclassification of Lenke type 1 scoliotic curves. *Journal of Spinal Disorders & Techniques*, *22*(2), 135.

Fedorov, A., Beichel, R., Kalphaty-Cramer, J., Finet, J., Fillion-Robbin, J.-C., Pujol, S., Kikinis, R. (2012). 3D slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging*, *30*(9), 1323–1341.

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research*, *15*(1), 3133–3181.

Fong, D. Y., Lee, C. F., Cheung, K. M., Cheng, J. C., Ng, B. K., Lam, T. P., Luk, K. D. (2010). A meta-analysis of the clinical effectiveness of school scoliosis screening. *Spine*, *35*(10), 1061–1071.

García-Cano, E., Arámbula Cosío, F., Duong, L., Bellefleur, C., Roy-Beaudry, M., Joncas, J., Labelle, H. (2018a). Dynamic ensemble selection of learner-descriptor classifiers to assess curve types in adolescent idiopathic scoliosis. *Medical and Biological Engineering and Computing*, 1–11.

García-Cano, E., Arámbula Cosío, F., Duong, L., Bellefleur, C., Roy-Beaudry, M., Joncas, J., Labelle, H. (2018b). Prediction of spinal curve progression in Adolescent Idiopathic Scoliosis using Random Forest regression. *Computers in Biology and Medicine*,

103(April), 34–43.

Hoffman, D. A., Lonstein, J. E., Morin, M. M., Visscher, B., & Harris III, S. H. (1989). Breast cancer in women with scoliosis exposed to multiple diagnostic x rays. *National Cancer Institute*, *81*(17), 1307–1312.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*(4–5), 411–430.

Ibrahim, F., Usman, J., Mohktar, M. S., & Ahmad, M. Y. (2016). Quantitative Evaluation of Spinal Coronal Curvature for Scoliosis Using a Fast 3-D Ultrasound Projection Imaging Method. *IFMBE Proceedings*, *56*, 313–317.

Kadoury, S., & Labelle, H. (2012). Classification of three-dimensional thoracic deformities in adolescent idiopathic scoliosis from a multivariate analysis. *European Spine Journal : Official Publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society*, *21*(1), 40–9.

Kadoury, S., Mandel, W., Roy-Beaudry, M., Nault, M.-L., & Parent, S. (2017). 3D Morphology Prediction of Progressive Spinal Deformities from Probabilistic Modeling of Discriminant Manifolds. *IEEE Transactions on Medical Imaging*, *0062*(c), 1–1.

Kadoury, S., Shen, J., & Parent, S. (2014). Global geometric torsion estimation in adolescent idiopathic scoliosis. *Medical & Biological Engineering & Computing*, *52*(4), 309–19.

King, H. A., Moe, J. H., Bradford, D. S., & Winter, R. B. (1983). The selection of fusion levels in thoracic idiopathic scoliosis. *The Journal of Bone & Joint Surgery*, *65*(9), 1302–1313.

Kittler, J., Hatef, M., Duin, R. P. W., & Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(3), 226–239.

Ko, A. H. R., Sabourin, R., & Britto, A. S. (2008). From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognition*, *41*(5), 1735–1748.

Koo, T. K., Guo, J. Y., Ippolito, C., & Bedle, J. C. (2014). Assessment of scoliotic deformity using spinous processes: Comparison of different analysis methods of an ultrasonographic system. *Journal of Manipulative and Physiological Therapeutics*, *37*(9), 667–677.

Labelle, H., Aubin, C.-E., Jackson, R., Lenke, L., Newton, P. O., & Parent, S. (2011). Seeing the spine in 3D: how will it change what we do? *Journal of Pediatric Orthopedics*, *31*(1 Suppl), 37–45.

Lasso, A., Heffter, T., Rankin, A., Pinter, C., Ungi, T., & Fichtinger, G. (2014). PLUS: Open-source toolkit for ultrasound-guided intervention systems. *IEEE Transactions on Biomedical Engineering*, *61*(10), 2527–2537.

Lenke, L. G., Betz, R. R., Bridwell, K. H., Clements, D. H., Harms, J., Lowe, T. G., & Shufflebarger, H. L. (1998). Intraobserver and Interobserver Reliability of the Classification of Thoracic Adolescent Idiopathic Scoliosis. *The Journal of Bone & Joint Surgery*, *80*(8), 1097–1106.

Lenke, L. G., Betz, R. R., Harms, J., Bridwell, K. H., Clements, D. H., Lowe, T. G., & Blanke, K. M. (2001). Adolescent Idiopathic Scoliosis. A New Classification to Determine Extent

of Spinal Arthrodesis. *The Journal of Bone & Joint Surgery*, *83–A*(8), 1169–1181.

Levy, A. R., Goldberg, M. S., Mayo, N. E., Hanley, J. A., & Poitras, B. (1996). Reducing the lifetime risk of cancer from spinal radiographs among people with adolescent idiopathic scoliosis. *Spine*, *21*, 1540–1547; discussion 1548.

Li, D. T., Cui, J. J., Devries, S., Nicholson, A. D., Li, E., Petit, L., Smith, B. G. (2018). Humeral Head Ossification Predicts Peak Height Velocity Timing and Percentage of Growth Remaining in Children. *Journal of Pediatric Orthopaedics*, *38*(9), 546–550.

Little, D., Song, K., Katz, D., & Herring, J. (2000). Relationship of Peak Height Velocity to Other Maturity Indicators in Idiopathic Scoliosis in Girls*. *The Journal of Bone and Joint Surgery. American Volume*, *82*(5), 685–93.

Liu, Y. Z., Smirnov, K., Lucio, M., Gougeon, R. D., Alexandre, H., & Schmitt-Kopplin, P. (2016). MetICA: independent component analysis for high-resolution mass-spectrometry based non-targeted metabolomics. *Bmc Bioinformatics*, *17*, 14.

Lonstein, J. E., & Carlson, J. M. (1984). The prediction of curve progression in untreated idiopathic scoliosis during growth. *The Journal of Bone and Joint Surgery. American Volume*, *66*(7), 1061–1071.

Majdouline, Y., Aubin, C.-E., Robitaille, M., Sarwark, J. F., & Labelle, H. (2007). Scoliosis Correction in Adolescent Idiopathic Scoliosis. *Journal Of Pediatric Orthopedics*, *27*(7), 775–781.

Nault, M.-L., Mac-Thiong, J.-M., Roy-Beaudry, M., De Guise, J. A., Labelle, H., & Parent, S. (2013). Three-dimensional spine parameters can differentiate between progressive and nonprogressive patients with AIS at the initial visit: a retrospective analysis. *Journal of Pediatric Orthopedics*, *33*(6), 618–23.

Nault, M.-L., Mac-Thiong, J.-M., Roy-Beaudry, M., Turgeon, I., De Guise, J. A., Labelle, H., & Parent, S. (2014). Three-Dimensional Spinal Morphology Can Differentiate Between Progressive and Nonprogressive Patients With Adolescent Idiopathic Scoliosis at the Initial Presentation. *Spine*, *39*(10), E601–E606.

Negrini, S., Negrini, A., Atanasio, S., & Santambrogio, G. C. (2006). Three-dimensional easy morphological (3-DEMO) classification of scoliosis, part I. *Scoliosis*, 1:20.

Noshchenko, A. (2015). Predictors of spine deformity progression in adolescent idiopathic scoliosis: A systematic review with meta-analysis. *World Journal of Orthopedics*, *6*(7), 537.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 1825–2830.

Phan, P., Mezghani, N., Wai, E. K., De Guise, J. A., & Labelle, H. (2013). Artificial neural networks assessing adolescent idiopathic scoliosis: Comparison with Lenke classification. *Spine Journal*, *13*(11), 1527–1533.

Poncet, P., Dansereau, J., & Labelle, H. (2001). Geometric torsion in idiopathic scoliosis: three-dimensional analysis and proposal for a new classification. *Spine*, *26*, 2235–2243.

Prevost, R., Salehi, M., Jagoda, S., Kumar, N., Sprung, J., Ladikos, A., Wein, W. (2018). 3D freehand ultrasound without external tracking using deep learning. *Medical Image Analysis*, *48*, 187–202.

Purnama, I. K. E., Wilkinson, M. H. F., Veldhuizen, A. G., Van Ooijen, P. M. A., Sardjono, T. A., Lubbers, J., & Verkerke, G. J. (2009). Following scoliosis progression in the spine using ultrasound imaging. *IFMBE Proceedings*, *25*(2), 600–602.

Ran, B., Chen, X., Zhang, G., Shen, F., Chen, J., Wu, J., Li, M. (2014). Comparison of the sagittal profiles among thoracic idiopathic scoliosis patients with different Cobb angles and growth potentials. *Journal of Orthopaedic Surgery and Research*, *9*, 19.

Richards, B. S., Bernstein, R. M., D'Amato, C. R., & Thompson, G. H. (2005). Standardization of criteria for adolescent idiopathic scoliosis brace studies: SRS Committee on Bracing and Nonoperative Management. *Spine*, *30*(18), 2068–2075.

Rogez, G. (2005). Human figure segmentation using independent component analysis. ... *and Image Analysis*, 1–8.

Ronckers, C. M., Doody, M. M., Lonstein, J. E., Stovall, M., & Land, C. E. (2008). Multiple diagnostic X-rays for spine deformities and risk of breast cancer. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *17*, 605–613.

Ronckers, C. M., Land, C. E., Miller, J. S., Stovall, M., Lonstein, J. E., & Doody, M. M. (2010). Cancer mortality among women frequently exposed to radiographic examinations for spinal disorders. *Radiation Research*, *174*(1), 83–90.

Ruto, A., Lee, M., & Buxton, B. (2006). Comparing principal and independent modes of variation in 3D human torso shape using PCA and ICA. *ICA Research Network*, 3–6.

ers, J. O. (2007). Maturity indicators in spinal deformity. *The Journal of Bone and Joint Surgery. American Volume*, *89 Suppl 1*, 14–20.

Sanders, J. O., Browne, R. H., Cooney, T. E., Finegold, D. N., McConnell, S. J., & Margraf, S. a. (2006). Correlates of the peak height velocity in girls with idiopathic scoliosis. *Spine*, *31*(20), 2289–95.

Sanders, J. O., Browne, R. H., McConnell, S. J., Margraf, S. A., Cooney, T. E., & Finegold, D. N. (2007). Maturity Assessment and Curve Progression in Girls with Idiopathic Scoliosis. *J Bone Joint Surg Am*, *89*(1), 64–73.

Sanders, J. O., Khoury, J. G., Kishan, S., Browne, R. H., Mooney, J. F., Arnold, K. D., … Finegold, D. N. (2008). Predicting scoliosis progression from skeletal maturity: A simplified classification during adolescence. *Journal of Bone and Joint Surgery - Series A*, *90*(3), 540–553.

Sangole, A. P., Aubin, C.-E., Labelle, H., Stokes, I. A. F., Lenke, L. G., Jackson, R., & Newton, P. O. (2009). Three-dimensional classification of thoracic scoliotic curves. *SPINE*, *34*(1), 91–9.

Schicho, K., Figl, M. L., Donat, M., Birkfellner, W., Seemann, R., Wagner, A., … Ewers, R. (2005). Stability of miniature electromagnetic tracking systems. *Physics in Medicine and*

Biology, *50*(9), 2089–98.

Shen, J., Parent, S., & Kadoury, S. (2014). Classification of Spinal Deformities using a Parametric Torsion Estimator. In *Computational Methods and Clinical Applications for Spine Imaging. Lecture Notes in Computational Vision and Biomechanics* (Vol. 17, pp. 67–78).

Sitoula, P., Verma, K., Holmes, L., Gabos, P. G., Sanders, J. O., Yorgova, P., Shah, S. A. (2015). Prediction of curve progression in idiopathic scoliosis: Validation of the sanders skeletal maturity staging system. *Spine*, *40*(13), 1006–1013.

Stokes, I. A. F. (1994a). Three-dimensional terminology of spinal deformity. A report presented to the Scoliosis Research Society by the Scoliosis Research Society Working Group on 3-D terminology of spinal deformity. *SPINE*.

Stokes, I. A. F. (1994b). Three-dimensional terminology of spinal deformity. *SPINE*, *19*(2), 236–48.

Stokes, I. A. F., Bigalow, L. C., & Moreland, M. S. (1986). Measurement of axial rotation of vertebrae in scoliosis. *Spine (Phila Pa 1976)*, *11*(3), 213–218.

Stokes, I. A. F., Sangole, A. P., & Aubin, C.-E. (2009). Classification of scoliosis deformity three-dimensional spinal shape by cluster analysis. *Spine (Phila Pa 1976)*, *34*(6), 584–590.

Tan, K.-J., Moe, M. M., Vaithinathan, R., & Wong, H.-K. (2009). Curve progression in idiopathic scoliosis: follow-up study to skeletal maturity. *Spine*, *34*(7), 697–700.

Thong, W., Labelle, H., Shen, J., Parent, S., & Kadoury, S. (2015). Stacked Auto-encoders for Classification of 3D Spine Models in Adolescent Idiopathic Scoliosis. In *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging. Lecture Notes in Computational Vision and Biomechanics* (Vol. 20, pp. 13–26). Springer, Cham.

Thong, W., Parent, S., Wu, J., Aubin, C.-E., Labelle, H., & Kadoury, S. (2016). Three-dimensional morphology study of surgical adolescent idiopathic scoliosis patient from encoded geometric models. *European Spine Journal*, *25*(10), 3104–3113.

Ungi, T., King, F., Kempston, M., Keri, Z., Lasso, A., Mousavi, P., Fichtinger, G. (2014). Spinal curvature measurement by tracked ultrasound snapshots. *Ultrasound in Medicine & Biology*, *40*(2), 447–54.

Üzümcü, M., Frangi, A. F., Reiber, J. H. C., & Lelieveldt, B. P. F. (2003). Independent component analysis in statistical shape models. *Proceedings of SPIE*, *5032*(0), 375–383.

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning - ICML '08*, (January), 1096–1103.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *Journal of Machine Learning Research*, *11*(3), 3371–3408.

Vo, Q. N., Lou, E., Le, L. H., & Huynh, L. Q. (2015). Investigation of the Optimal Freehand

Three-Dimensional Ultrasound Configuration to Image Scoliosis: An In-vitro Study, *46*, 163–166.

Wagner, D. R. (2013). Ultrasound as a tool to assess body fat. *Journal of Obesity*, *2013*.

Woźniak, M., Graña, M., & Corchado, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion*, *16*(1), 3–17.

Young, M., Hill, D. L., Zheng, R., & Lou, E. (2015). Reliability and accuracy of ultrasound measurements with and without the aid of previous radiographs in adolescent idiopathic scoliosis (AIS). *European Spine Journal*, *24*(7), 1427–1433.

Zhao, Q., Okada, K., Rosenbaum, K., Kehoe, L., Zand, D. J., Sze, R., … Linguraru, M. G. (2014). Digital facial dysmorphology for genetic screening: Hierarchical constrained local model using ICA. *Medical Image Analysis*, *18*(5), 699–710.

Zheng, Y.-P., Lee, T. T.-Y., Lai, K. K.-L., Yip, B. H.-K., Zhou, G.-Q., Jiang, W.-W., Lam, T.-P. (2016). A reliability and validity study for Scolioscan: a radiation-free scoliosis assessment system using 3D ultrasound imaging. *Scoliosis and Spinal Disorders*, *11*(1), 13.