Taylor & Francis
Taylor & Francis Group

Computer Methods in Biomechanics and Biomedical Engineering

ISSN: 1025-5842 (Print) 1476-8259 (Online) Journal homepage: https://www.tandfonline.com/loi/gcmb20

Customized k-nearest neighbourhood analysis in the management of adolescent idiopathic scoliosis using 3D markerless asymmetry analysis

Maliheh Ghaneei, Ronald Ekyalimpa, Lindsey Westover, Eric C. Parent & Samer Adeeb

To cite this article: Maliheh Ghaneei, Ronald Ekyalimpa, Lindsey Westover, Eric C. Parent & Samer Adeeb (2019): Customized k-nearest neighbourhood analysis in the management of adolescent idiopathic scoliosis using 3D markerless asymmetry analysis, Computer Methods in Biomechanics and Biomedical Engineering, DOI: 10.1080/10255842.2019.1584795

To link to this article: https://doi.org/10.1080/10255842.2019.1584795

Published online: 08 Mar 2019.

COMPUTER METHODS IN BIOMECHANICS AND BIOMEDICAL ENGINEERING
https://doi.org/10.1080/10255842.2019.1584795

Customized k-nearest neighbourhood analysis in the management of adolescent idiopathic scoliosis using 3D markerless asymmetry analysis

Maliheh Ghaneei[a], Ronald Ekyalimpa[b], Lindsey Westover[c], Eric C. Parent[d] and Samer Adeeb[a]

[a]Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada;[b]College of Engineering, Design, Art, and Technology, Makerere University, Kampala, Uganda;[c]Department of Mechanical Engineering, University of Alberta, Edmonton, Canada;[d]Department of Physical Therapy Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Canada

ABSTRACT

Adolescent Idiopathic Scoliosis (AIS) is a 3D spinal deformity characterized by curvature and rotation of the spine. Markerless surface topography (ST) analysis has been proposed for diagnosing and monitoring AIS to reduce the X-ray radiation exposure to patients. This method captures scans of the cosmetic deformity of the torso using visible, radiation-free light. The asymmetry analysis of the torso, represented as a deviation contour map with deviation patches outlining the areas of cosmetic asymmetries, has previously been shown to predict the severity and progression of the condition in comparison with radiographs, by using classification trees. While the classification results were promising, it was reported that some mild curves were erroneously diagnosed. Furthermore, this approach is highly sensitive to threshold values selected in the decision trees. Therefore, this study aims to define a custom Neighbourhood Classifier algorithm for AIS classification to improve the accuracy, sensitivity, and specificity of predicting curve severity and curve progression in AIS. Curve severity was predicted with 80% accuracy (sensitivity = 81%; specificity = 79%) for thoracic-thoracolumbar curves and 72% (sensitivity = 93%; specificity = 53%) for lumbar curves. This represents an improvement over the previous method with curve severity accuracies of 77% and 63% for thoracic-thoracolumbar and lumbar curves, respectively. Additionally, curve progression was predicted with 93% accuracy (sensitivity = 83%; specificity = 95%) representing a substantial improvement over the previous method with an accuracy of 59%. The current method has shown the potential to further reduce radiation exposure for AIS patients by avoiding X-rays for mild and non-progressive curves identified using ST analysis.

ARTICLE HISTORY
Received 21 November 2017
Accepted 16 February 2019

KEYWORDS
Adolescent idiopathic scoliosis (AIS); surface topography; 3D markerless asymmetry analysis; classification; k-Nearest neighbour

Introduction

Spinal deformity is associated with cosmetic abnormality, difficulty in health-related quality of life, and can result in impairment such as difficulty breathing in the most severe cases (Liu et al. 2001; Tones et al. 2006). Adolescent idiopathic scoliosis (AIS) is a frequent form of such deformity (Rogala et al. 1978) which requires repetitive monitoring using X-rays (Weinstein et al. 2008). The Cobb angle is measured on radiographs to document the location and the severity of curves (Thulbourne and Gillespie 1976). This approach has two primary limitations: first, X-ray radiation has harmful effects, including an increased risk of cancer for this adolescent population (Ardran et al. 1980; Levy et al. 1996) and second, the Cobb angle is a two-dimensional measurement of a three-dimensional deformity (Duval-Beaupere et al. 1984; Thulbourne and Gillespie 1976). To reduce radiation exposure to patients, researchers are investigating the use of surface topography (ST) to assess the posture and torso shape. Of particular interest are the recent uses of the Microsoft Kinect™ system to acquire posture of subjects (Diego-Mas and Alcaide-Marzal 2014; Dutta 2012). Castro et al. (2017) utilized the Kinect™ system to assess the tendency of normal subjects to present scoliosis like asymmetries. Our team has used a system of four cameras to acquire ST data of the full torso and developed a markerless ST technique with the purpose of evaluating the patient's torso asymmetry. The proposed ST monitoring strategy suggests that patients presenting with either mild (Cobb angle $< 25°$) or non-progressive (ΔCobb angle (ΔCA) $< 5°$ increase) curves could avoid an X-ray since the typical standard of care for these patients involves only observation or no change in treatment, respectively. The ST analysis technique presented in

CONTACT Lindsey Westover lwestove@ualberta.ca

our previous work showed the potential to eliminate X-rays for some patients (Komeili et al. 2014, 2015a, 2015b).

The cosmetic deformity associated with AIS involves torso asymmetry. A person with no spinal curvature is approximately symmetric across the midsagittal plane, which means that the person's torso and its reflection along this plane are almost perfectly aligned (Ho et al. 2015). However, for a person with an asymmetric torso, the sagittal plane is no longer a plane of symmetry. Our method takes advantage of the best plane of symmetry identification method introduced by Hill et al. (2014) to assess the deformity of the scoliotic spine. The best plane of symmetry is roughly aligned with the midsagittal plane; however, the actual plane is determined by minimizing the sum of distances between the patient's torso and its bilateral reflection (Komeili et al. 2014). The asymmetry is then illustrated using a deviation contour map plotted on the patient's torso. The effects of the spinal curvature are visualized in terms of dense colour areas called deviation patches containing many points whose colours represent the distance between the original and reflected torsos, depicting both areas of protrusion or depression relative to the other side (Komeili et al. 2014). The maximum and root mean square of these deviations are computed as asymmetry parameters (MaxDev and RMS, respectively). These asymmetry parameters have been compared with the Cobb angle measured in the corresponding region of the torso to create decision trees predicting curve severity on a given test day and progression between consecutive examinations (Komeili et al. 2014, 2015b; Ghaneei et al. 2018).

Decision trees were used to classify the curve severity (mild or moderate/severe) and to evaluate clinically significant progression of the curves ($\Delta CA \geq 5^\circ$) through one time interval (Komeili et al. 2015b). The results were promising for curve severity classification especially in detecting moderate/severe patients ($CA \geq 25^\circ$; sensitivity = 95%, specificity = 35%) (Ghaneei et al. 2018). However, the previous work had several limitations. Some mild curves were erroneously diagnosed indicating that the method showed high sensitivity and low specificity. Furthermore, the decision tree analysis is oversensitive to threshold values selected. In other words, a minor change in one variable leads to major changes in the subtree below or even may destabilize the tree (Rokach and Maimon 2015). Threshold values can be controlled to achieve the highest possible sensitivity to moderate/severe cases, which however may reduce the specificity by identifying mild patients as moderate/severe. The other significant drawback of the decision tree analysis is that in defining the classification trees it was assumed that as the asymmetry parameters increase, the Cobb angle also increases. However, it will be shown in this study that as the RMS increases, the MaxDev increases while the Cobb angle fluctuates within a wide range. Consequently, this assumption may result in the misclassification of the patient status.

To overcome the limitations of decision tree analysis, an appropriate classification model is sought to relate the surface topography parameters and the Cobb angle, which in general exhibit lack of correlation. The first classification rule was proposed by Fisher in 1936 in statistical classification literature, after which other classification models were proposed and applied (Raudys 2001). The k-nearest neighbour (k-NN) algorithm is one of the simplest machine learning algorithms where input parameters from a certain number (k) of the closest data points within a Training Group are mapped to an output through a systematic classification or regression process (Altman 1992). The output of the classification is a class generated from the k-neighbours (Everitt et al. 2011). The current study aims to use the k-NN method for classifying AIS patients based on their ST asymmetry parameters in terms of curve severity and curve progression to reduce the errors of such classification efforts observed when using decision trees (Ghaneei et al. 2018).

Methods

Data acquisition

ST scans and radiographs of 128 AIS patients recruited from the scoliosis clinic were used in this study. Patients included in the study had an age range of 10 to 18 years (14.4 ± 1.8 years) with a mean Cobb angle of 26.5° (from 10° to 46°). 100 patients (78%) were female and 28 (22%) were male. The sex distribution amongst the participants is representative of the prevalence of scoliosis in the larger population as the sex ratio from age 10 onward is reported at 6:1 (females:males)(Trobisch et al. 2010). Follow-up scans of 95 patients were available in a 1 year ± 3 month interval. Both baseline and follow-up data were used in the curve severity analysis, which resulted in 176 thoracic/thoraco-lumbar (T-TL) curves and 167 lumbar (L) curves, including patients with multiple curves. To monitor curve progression from baseline to follow-up, 95 ST and radiograph pairs with a total of 134 curves were used.

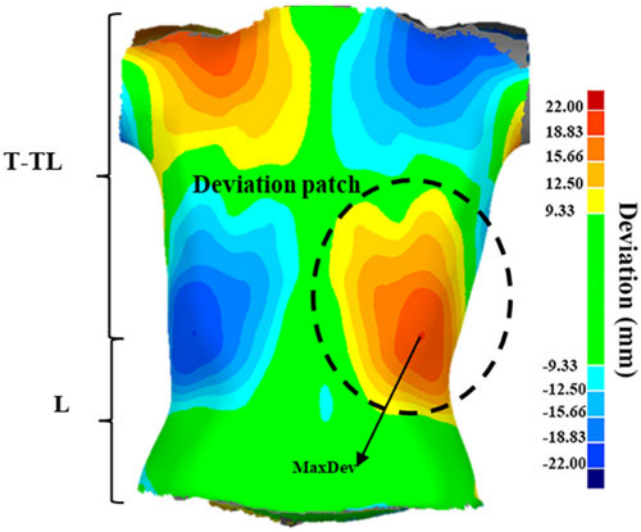

Figure 1. Deviation contour map.

The raw ST data was obtained by a standard scanning procedure. Images were acquired from the front, back, left and right of the patients with four VIVID 910 laser scanners (KONICA MINOLTA Sensing Inc.) in the clinic on the same day as their radiographs (Komeili et al. 2014). After merging the ST data to create the full 3 D model of the patient in Geomagic Control (3 D System Corporation, CA, USA), the torso was isolated by cropping the frame and unnecessary parts such as the lower extremities, head, and arms (Komeili et al. 2014). The full 3 D torso was duplicated and mirrored along the sagittal plane. Then the original torso and its reflection were aligned such that the deviations between corresponding points were minimized using the closest point approach that is built-in the Geomagic software (Hill et al. 2014). The misalignment between the two torsos was visualized in the deviation contour map plotted on the torso, which indicate areas of asymmetry (Komeili et al. 2014). The threshold between normal and abnormal deviations was 9.33 mm if the ST scan had a maximum deviation greater than 9.33 mm (Ghaneei et al. 2018), otherwise the threshold of 3 mm was used as suggested by Komeili et al. (2015a, 2015b) (see Figure 1). Blue areas indicate that the original torso is outside of the reflected torso, i.e. original torso covers the reflected torso, and red areas indicate that the original torso is inside of the reflected torso. The shades of blue and red (asymmetry patches) illustrate the areas of deviations corresponding to the level of the existing scoliosis curves. The maximum distance between the original torso and its reflection was recorded as the MaxDev and the root mean square of all the deviations within each patch was recorded as the RMS according to the following equations:

$$MaxDev = \ Max\ (Deviation_i)\ \ i = 1, 2, 3 \ldots, n \quad (1a)$$

$$RMS = \sqrt{\sum_{i=1}^{n} \frac{(Deviation_i{}^2)}{n}}\ \ i = 1, 2, 3 \ldots, n \quad (1b)$$

where, n is the number of points within the deviation patch under analysis. The ST scan of a patient's torso was divided into two parts. The lower one-third part of the torso was considered the lumbar area (L) and the upper two-thirds was considered the thoracic/thoraco-lumbar area (T-TL). The asymmetry parameters were classified as L or T-TL according to the location of the asymmetry patch from which they were extracted. To study the curve progression of those patients who had ST data at a follow-up visit, the differences of the MaxDev and RMS from baseline to follow-up were computed, i.e. ΔRMS and ΔMaxDev.

Data analysis

Preliminary data analysis

All the curves in each set were ordered according to increasing RMS and were given an identifier called curve ID. Preliminary studies on data were performed by plotting charts for asymmetry parameters and Cobb angle across the curve IDs (see Figure 5). The correlation between parameters was examined through linear trend lines plotted on the charts along with corresponding coefficients of determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - f_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \quad (2)$$

Where $\overline{y}_i = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the mean of observed data and f_i is the corresponding point at the fitted trend line.

Classification algorithms

Traditional neighbourhood classifier. The traditional neighbourhood classification method (Altman 1992) will be introduced, in general. This method consists of two phases, namely training and classification. In the training phase, all the features (independent and dependent parameters) of the existing data points are stored in a group, called the Training group. Next, in the classification phase, a new data point can be categorized based on the most frequent dependent

parameter in the Training group who has the smallest difference with the independent parameters. To calculate this difference, Euclidean Distances (Eq. 3) between the new data point and all of the Training group's members are first computed using the values for each independent parameter. Then, the data points within the Training set are ranked in ascending order based on their respective distances. Next, the first k numbers of ranked data points are collected as neighbours of the new data point to subsequently determine the probability of the output class according to the neighbourhood labels (Figure 2).

$$d_{Euclidean} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2} \quad (3)$$

Custom neighbourhood classifier. The neighbourhood classifier model developed and used in this study was unique because it had no specific parameters fitted through a formal data model training process. The training dataset was memorized in its entirety and used by the model when subsequently deployed for predictive purposes. The precision of the model lay in exposing it to a rich training set, i.e., one that had a good mix of instances – fairly balanced representation from each output class. The uniqueness in approach used in the nested neighbourhood algorithm, mainly lay in the fashion in which classification was done.

When a trained nested neighbourhood classifier model is exposed to a new case that needs to be classified, it determines the data instances that are closest to this new case, from the dataset that it memorized from the training process. These instances that are closest to the new case are referred to as the neighbours of that new case. The neighbours of the new case are determined in a two-staged process. In the first phase, the first attribute in the data is used to determine the proximity of instances memorized by the model to the new case it's exposed to. The absolute difference between values of this attribute is used as the distance measure. The data instances that are considered to be within pre-set thresholds in relation to this distance measure, were considered neighbours to this new case based on the first attribute. Then in the second and last phase, the second attribute is used in the computation of the distance measure. The absolute difference in the values of the second attribute of the new case and each of its neighbors (determined from the first phase), are computed. Data instances from the neighbors determined in the first phase, which are found to be within the thresholds

Figure 2. A flowchart summarizing the traditional Neighbourhood classification algorithm.

for the distance measure from the second attribute, are then taken as the final/true neighbors to the new case. In this study, thresholds were defined as a number of data instances closest to the new case, i.e. "k".

Finally, the class to which the new case is believed to belong is taken as the class with the highest likelihood of these final neighbours – in other words, the predominant class in the final neighbours. The name of the classifier algorithm, i.e. nested neighborhood classifier, comes from the fact that during the predictive process, neighbors are determined in a sequential fashion with each step making use of the previous neighbors as input for determining the next, with the exception of the first step which makes use of the memorized training data.

The described algorithm was used to develop the custom neighbourhood classifier for this study. The independent input features in the data, i.e. RMS and MaxDev, were utilized in a sequential order and resulted in neighbours defined and desired instance classifications obtained.

Training classifier models and classification model performance evaluation

In this study, the data points are contour patches associated with individual curves and were randomly divided into two groups, namely the Training group and the Testing group using common data splitting rules (80:20). For curve severity classification two independent parameters, RMS and MaxDev, are used. Hence, only two neighbourhood computation iterations are required. The number of neighbours in the first and second iterations are denoted as k_1 and k_2, respectively. The classification phase starts with RMS rather than MaxDev since it has illustrated stronger correlation with curve severity based on the preliminary analysis in this study as well as in the work of Ghaneei et al. (2018). The output class is the curve severity prediction (either mild or moderate/severe). The process described is summarized in the flowchart presented in Figure 3. Wolfram Mathematica (Wolfram Research, Inc., Mathematica 8.0.4.0) was used to automate this process.

For curve progression, ΔRMS and ΔMaxDev are the independent parameters used in the first and second classification phases, respectively, and the output class is the progression status (either progression or non-progression).

The performance of the models developed for classifications were assessed based on sensitivity, specificity, and total accuracy. For curve severity, a positive result represents a moderate/severe curve and a negative result shows mild severity. For curve progression, a positive result indicates a progressive curve (increase in CA $\geq 5^\circ$ over the time interval) and a negative result is a non-progressive curve ($<5^\circ$). The predicted results were based on the ST analysis and the actual results were based on the radiographic measures. Table 1 presents a matrix that was used to determine the performance measures and parameters used in the classification process.

Parametric studies for the customized k-NN classification

The accuracy of the results from the customized k-NN class are dependent on the number of neighbours (k) selected for each neighbourhood (Wang et al. 2006). As such, a parametric study was performed to establish the appropriate number of neighbours that generates the best results. k_1 and k_2 were each varied sequentially from 1 to 20 and the optimal values were those that provided the best diagnostic predictive results (sensitivity, specificity, and accuracy) (Figure 4).

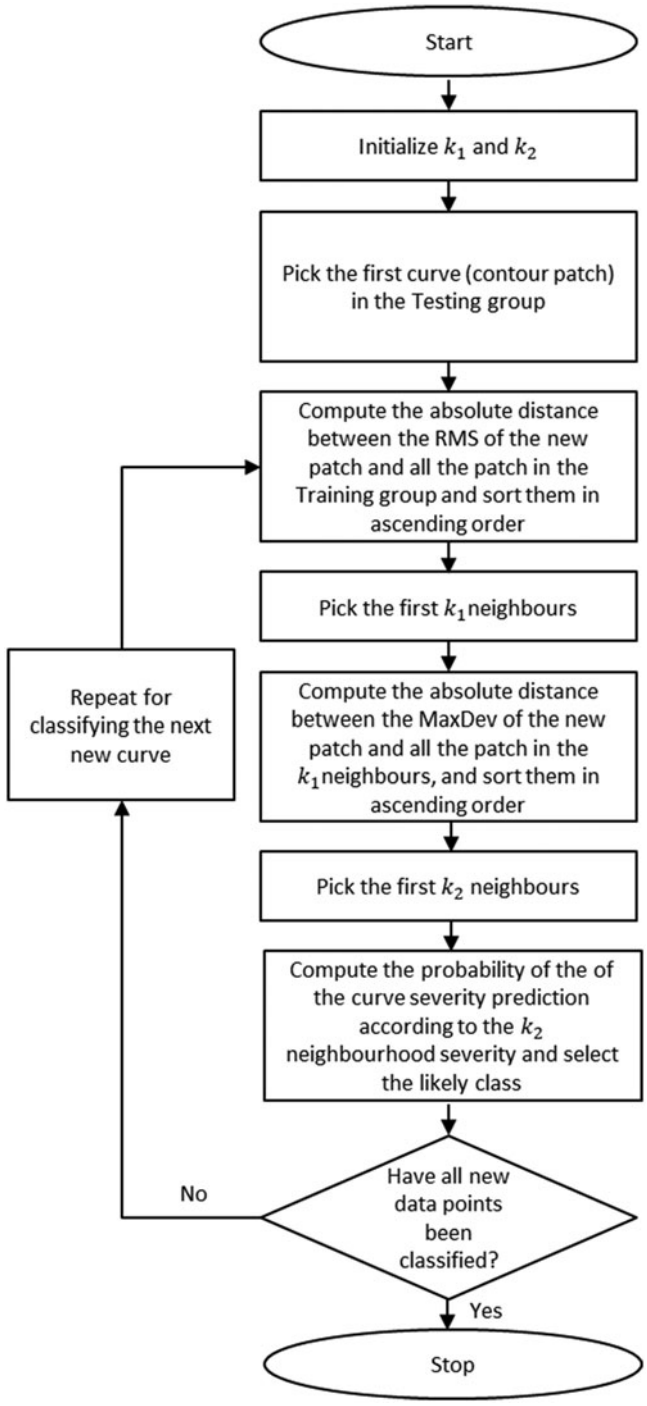

Figure 3. The customized k-NN classification algorithm for curve severity classification.

Results

Curve severity classification

The curve parameters RMS, MaxDev, and Cobb angle were plotted against the curve ID. Figure 5 shows a representative plot for the lumbar curves in the Training group and it reveals that while RMS and MaxDev are closely related to each other, the Cobb

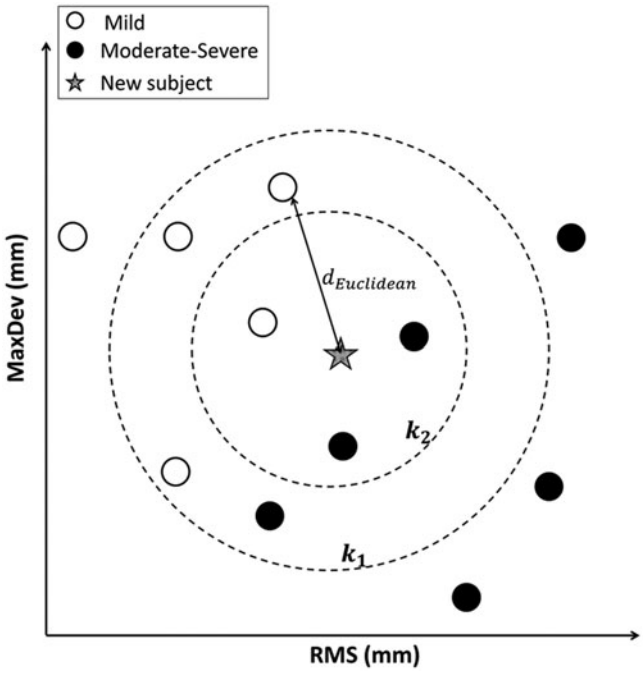

Figure 4. Simplified sketch of the customized k-NN algorithm.

angle does not follow the same trend and fluctuates within a wide range. RMS and MaxDev vary linearly with curve ID with R^2 values of 0.90 and 0.85, respectively. Conversely, the corresponding Cobb angles vary widely with curve ID and do not follow a linear trend (R^2=0.12).

There were a total of 176 T-TL curves and 20% of them (35 curves) were randomly selected to establish the Testing group while the rest (141 curves) were considered as the Training group. The ultimate goal of the proposed method is the proper detection of moderate/severe patients, which is equivalent to achieving the highest sensitivity. With this goal in mind, the optimum values for neighbours in the first and second iterations were found to be 3 for both (k_1 =k_2=3). Thus, it is only needed to consider the first three ranked data points from the Training group and the severity of the patient under study is decided based on the most frequent class labels (mild or moderate/severe) among such ranked data.

Of the 35 T-TL curves in the Testing group, 16 were moderate/severe and 19 were mild based on radiographic measures. Table 2 illustrates the curve severity prediction for the curves in the Testing group. As it can be seen the overall sensitivity is 81% which indicates that 13/16 moderate/severe curves were accurately classified on the basis of the ST parameters. Moreover, 15/19 mild curves were classified correctly with a specificity of 79%. This suggests that 79% of the mild curves could have avoided an X-ray examination based on the algorithm introduced in this study. Consequently, the overall accuracy in predicting the severity of T-TL curves based on this method is 80%.

Table 1. Calculation of the accuracy, sensitivity, and specificity.

		Radiograph	
		+	−
Surface topography	+	True positive (TP)	False positive (FP)
	−	False negative (FN)	True negative (TN)
Accuracy $(TP+TN)/(TP+FN+FP+TN)$		Sensitivity $TP/(TP+FN)$	Specificity $TN/(FP+TN)$

With the same procedure, 20% of L curves were stored in the Testing group (32 curves) leaving 135 curves in Training group. The optimum values for the number of neighbours in the first and second iterations respectively were found to be k_1=18 and k_2=5.

Of the 32 L curves in the Testing group, 15 were moderate/severe and 17 were mild based on radiographic measures. 14/15 moderate/severe curves were correctly identified with a sensitivity of 93% (Table 2). Furthermore, 9/17 of the mild curves were diagnosed correctly for a specificity of 53% suggesting that more than half of the mild curves could have avoided an X-ray examination. The overall accuracy of the proposed method for L curves was 72%.

Curve progression

The curve progression was studied for those patients who have follow-up scans (134 curves) with ΔRMS and ΔMaxDev being the independent variables. Twenty-seven curves were included in the Testing group and 107 curves were included in the Training group. The optimum value for k_1 was found to be 17, i.e. 17 curves with the lowest absolute difference of ΔRMS with respect to the data point under study were used. The optimum value for k_2 was 1, indicating that the closest neighbour in terms of ΔMaxDev predicts whether the current curve experienced progression or not. 5/6 progressive curves were identified correctly (Sensitivity = 83%) and 20/21 non-progressive curves were identified correctly (Specificity = 95%) (Table 3.) The high sensitivity confirms that the proposed method can accurately distinguish the progressive curves (low number missed) giving clinicians confidence in the method's ability to identify patients who could skip radiographs without missing treatment opportunities.

The capability to classify the curve severity of AIS curves though using the customized k-NN algorithm was compared with previous work (Ghaneei et al.

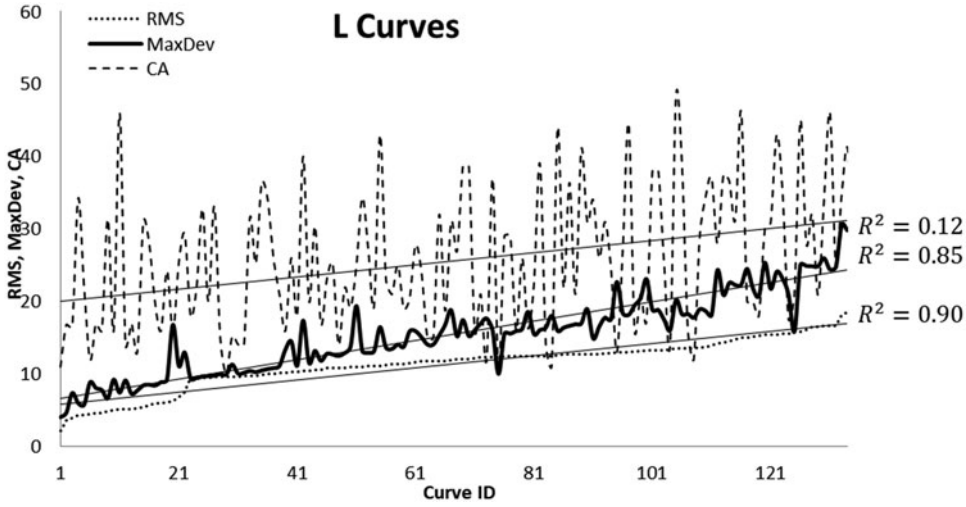

Figure 5. Comparing the associations between RMS, MaxDev, and Cobb angle with curve ID ranked by RMS magnitude for the L curve Training group.

Table 2. Severity classification results for the testing group.

		Radiograph				Radiograph	
T-TL		(+)	(–)	L		(+)	(–)
ST	(+)	13	4	ST	(+)	14	8
	(–)	3	15		(–)	1	9
Accuracy		Sensitivity	Specificity	Accuracy		Sensitivity	Specificity
80%		81%	79%	72%		93%	53%

Table 3. Curve progression classification results for the testing group.

		Radiograph	
		(+)	(–)
ST	(+)	5	1
	(–)	1	20
Accuracy		Sensitivity	Specificity
93%		83%	95%

2018) in markerless ST (Figure 6). For the case of T-TL curves, the specificity in the curve severity prediction showed a substantial increase compared to the previous work (Ghaneei et al. 2018) from 58% to 79%. However, our approach failed to distinguish moderate/severe curves as well as the traditional classification tree with a decrease in sensitivity from 100% to 81%. The overall accuracy slightly improved with an increase from 77% to 80%.

Considering the curve severity of L curves, we showed an improvement in sensitivity from 87% to 93% along with an increase in specificity from 41% to 53% compared to the previous work (Ghaneei et al. 2018).

Considering curve progression, the current method showed a substantial improvement over the previous work (Ghaneei et al. 2018) (Figure 7). Progressive curves were identified with 83% accuracy, exhibiting an increase in sensitivity from 67% with the other method. Furthermore, the overall accuracy of the predicting curve progression significantly increased from 59% to 93%.

Discussion and conclusions

The motivation of the present study lies in the potential of classifying the severity and progression of curves in AIS patients based on ST thereby potentially reducing the use of radiographs. More precisely, the k-NN analysis was conducted to further minimize the risk of missing moderate/severe or progressive curves and reduce the number of mild and non-progressive curves misclassified while maximizing the sensitivity compared to the classification tree analysis previously reported in the literature (Ghaneei et al. 2018). Given the errors in radiographic measurements, it is difficult to define a threshold of acceptable values for specificity, sensitivity, and accuracy. The method presented here has been designed to maximize sensitivity, which is a conservative approach meant to minimize the number of moderate/severe or progressive curves that would be missed.

The arbitrary lower and upper boundaries that could have been used to define the size of a training dataset are influenced by different factors which also vary with context. Consequently, there are no hard and fast rules/criteria for defining these boundaries as universally accepted sharp thresholds using crisp values or definitive formulae. On the lower side, the feasibility of access to abundant data affects things while the upper side is controlled by the avoidance of an overly fit/trained model. Some guidelines exist within the machine learning domain that facilitate the

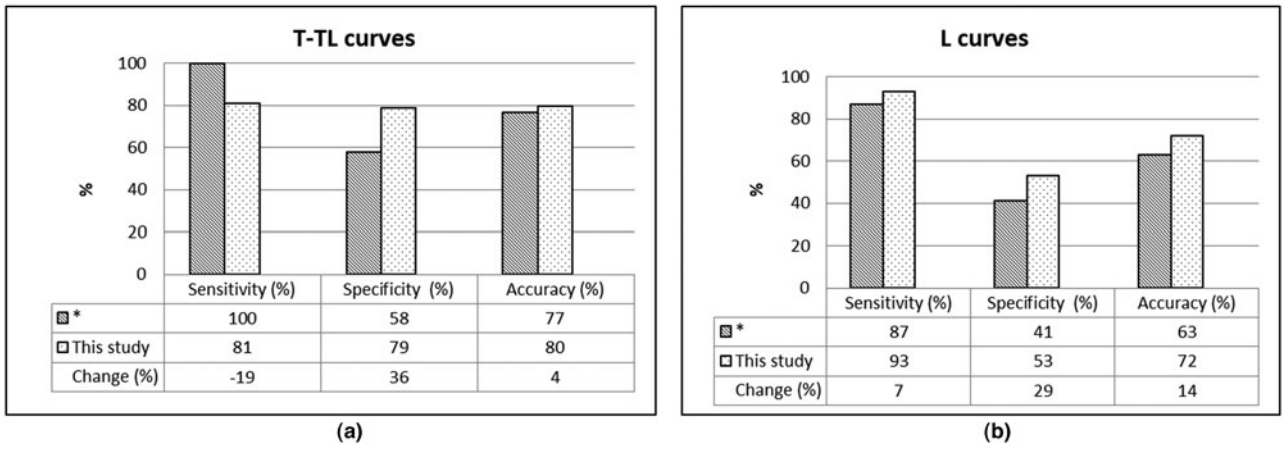

Figure 6. Comparing the performance of customized k-NN with a previous study using classification tree analysis (Ghaneei et al.2018) in prediction the severity for (a) T-TL and (b) L curves in the Testing group.

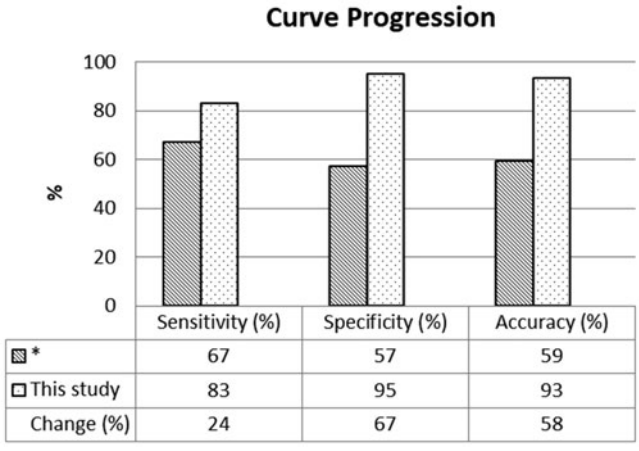

Figure 7. Comparing the performance of the customized k-NN with classification tree analysis reported in a previous study (Ghaneei et al.2018) in predicting the progression of scoliosis curves.

estimation of the size of training datasets. The three which are popularly utilized are presented next.

There are data modellers that experiment with the error generated by the model during the training and predictive phases. This is referred to as a learning curve approach and involves varying the size of the training dataset and plotting the model's predictive error against the training error. The convergence of these two errors gives insights into the appropriate size training data to make use of. Also, there exist rules of thumb that have been developed by those within the machine learning domain. These provide heuristics on the adequacy of a certain dataset size in training a model for a particular problem (Jain and Chandrasekaran 1982; Hastie et al. 2009). The majority of these are based on the dimensionality of the problem. An example of such a rule that is in popular use, is one that recommends that a dataset used for training models be no less than 10 times the number of input dimensions in the problem. This one is referred to as the "10x" rule. Similar heuristics also exist which are based on the number of output classes.

It is worth-noting that none of these fore-mentioned guidelines can be used on a "one size fits all" basis. As such, data modellers need to approach the problem of determining the appropriate size of training dataset with the following mind: (1) the nature of the data - dimensionality, (2) the quantity of data, (3) the type of model etc. In this study, a heuristic approach that makes use of the dimensionality of the input features was used in determining a training dataset size. A "10x" rule applied to the 2 input features yielded a minimum threshold requirement of 20 data instances for training. The number of instances used for training the neighbourhood classifier in this study were significantly higher than this threshold value.

For L curve severity classification, by using the new classification method introduced in this study, sensitivity, which represents the probability of identifying moderate/severe curves, was improved from 87% to 93% compared to when using classification tree analysis. It should be noted that the only misclassification of a moderate/severe curve was specific to the only patient with a double curve, in which the upper curve in the T-TL area was detected correctly as severe. In a clinical setting, the severe T-TL curve would lead the patient to undergo further X-ray investigation, thus the missed L curve would not have impacted the clinical care of the patient. In a double curve spine, having a large curvature in the T-TL section reduces the cosmetic deformity in the lumbar area, i.e. the asymmetry parameters corresponding to

the L curve on the deviation contour map are smaller than they could be if there was no T-TL curve. Further investigation is required to study the interconnected effects of T-TL and L curves on a larger sample set of patients with double curves.

Another advantage of the proposed technique is the increase in the specificity of curve severity classification for lumbar curves from 41% to 53%. This increase in specificity would result in a decrease in the number of patients with mild curves being exposed to X-ray radiation. The corresponding increase in sensitivity from 87% to 93% shows that this would occur while missing even fewer moderate/severe curves requiring the clinician's attention. These results substantiate the fact that the new method is better able to diagnose L curves by improving the overall accuracy of curve severity prediction from 63% to 72%. Hence, L curve severity can be assessed with higher confidence.

For T-TL curve severity, the sensitivity obtained in this study was lower than the former classification method (81% compared to 100%). The three moderate/severe curves that were inaccurately identified as mild in the k-NN classification analysis belong to patients with high body mass indices (BMI). These patients had BMI $>$ 25 which is the standard threshold of overweight. Hence, this misclassification likely resulted from a reduction in the torso asymmetry due to the fatty tissue masking the underlying deformity.

To investigate the effect of BMI $>$ 25, all the patients meeting this criterion were eliminated from the dataset and the analysis with the same features was performed again. The sensitivity and accuracy for patients with BMI $\leq$ 25 reached to 100% and 83%, respectively. In particular, all the moderate/severe curves were successfully identified by the new analysis. The modified dataset did not reveal any changes on the number of mild misclassified curves (4 curves); however, the specificity reduced to 67% due to the decrease in the number of patients in the modified dataset. The results, in particular the significant improvements of sensitivity and accuracy after excluding the subjects with BMI $>$ 25, suggests that excess body fat can considerably influence the asymmetry parameters.

In a clinical setting, if radiographs were not ordered for patients identified with a mild curve, the classification trees presented in Ghaneei et al. (2018) could reduce the number of X-rays by 24% (9/37) while the customized k-NN presented here could reduce the number of X-rays by 34% (12/37). In this study, the k-NN classification correctly identified 55% (12/22) of the mild patients.

Furthermore, this study proposed a new method for monitoring AIS curves over time, which remarkably improved the accuracy of identifying curve progression. Only one curve was misclassified and considered as non-progressive after the one-year follow-up interval. For this curve, however, the increase in Cobb angle was exactly equal to 5 degrees which is on the margin between progressive and non-progressive and is within the Cobb angle measurement error. In addition, one curve was misdiagnosed as progressive when the increase in Cobb angle was actually less than 5 degrees. The increase in sensitivity from 67% to 83% compared to the previous study by Ghaneei et al. (2018) means fewer progressive curves would be missed by using the customized k-NN classification approach. The identification of non-progression curves (specificity) also increased from 57% to 95%, demonstrating a clear advantage of the current method in terms of protecting patients against unnecessary radiographs.

In a clinical setting, if radiographs were not ordered for patients identified as non-progressive, the customized k-NN could reduce the number of X-rays by 74% (14/19) while the classification trees presented in Ghaneei et al. (2018) could reduce the number of X-rays by 42% (8/19). It is important to note that 100% (14/14) of the non-progressive patients in the Testing group were correctly identified by k-NN while only 57% (8/14) were identified correctly using the classification trees (Ghaneei et al. 2018). These are clinically important results and indicate that our markerless surface topography technique, combined with the customized k-NN classification method is far superior to previous methods in terms of the ability to reduce the X-ray exposure for AIS patients.

One of the limitations of this study is that some curves came from the same patients which may lead to data overfitting. The results presented here can be further validated on a larger sample of patients. In addition, further work is required to investigate the low specificity reported in the L curves which could be related to the curve type of these patients. If lumbar curves were more often secondary than thoracic curves, this region may require a different classification strategy. Initial curve type classification, for example using a surface topography asymmetry classification system (Komeili et al. 2014), may be important in future attempts at further improving the method presented here.

The devised methodology in the present work provided a substantially improved accuracy compared to literature. Further improvements may be achieved by

increasing the Training group size. Future work will focus on further investigation of double curve spinal deformities as well as patients with BMI greater than 25. Also, in future studies, BMI can be applied as an independent variable along with RMS and MaxDev in the analysis. We are confident that our research will serve as a base for future studies on AIS analysis based on ST monitoring and we have shown that this method has the potential to significantly reduce the number of X-rays required during clinical follow-up of patients with AIS.

Acknowledgements

The authors gratefully acknowledge the financial support from the Scoliosis Research Society (SRS), the Women and Children's Health Research Institute (WCHRI), the Stollery Children's Hospital Foundation, and Sick Kids Foundation.

References

Ardran GM, Coates R, Dickson RA, Dixon-Brown A, Harding FM. 1980. Assessment of scoliosis in children: low dose radiographic technique. Br J Radiol. 53(626): 146–147.

Altman NS. 1992. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat. 46(3): 175–185.

Castro APG, Pacheco JD, Lourenco C, Queiros S, Moreira AHJ, Rodrigues NF, Vilaca JL. 2017. Evaluation of spinal posture using Microsoft Kinect™: a preliminary case-study with 98 volunteers. Porto Biomed J. 2(1): 18–22.

Diego-Mas JA, Alcaide-Marzal J. 2014. Using Kinect™ sensor in observational methods for assessing postures at work. Appl Ergon. 45(4):976–985.

Dutta T. 2012. Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. Appl Ergon. 43(4):645–649.

Duval-Beaupere G, Lespargot A, Grossiord A. 1984. Scoliosis and trunk muscles. J Pediatr Orthop. 4(2): 195–200.

Everitt BS, Landau S, Leese M, Stahl D. 2011. Miscellaneous Clustering Methods. In: Shewhart WA, Wilks SS, editors. Cluster analysis. 5th ed. Chichester: John Wiley & Sons Ltd; p. 215–255.

Ghaneei M, Komeili A, Li Y, Parent E, Adeeb S. 2018. 3D Markerless Asymmetry Analysis in the Management of adolescent idiopathic scoliosis. BMC Musculoskelet Disord. 19(1):385

Hastie T, Tibshirani R, Friedman J. 2009. The elements of statistical learning: data mining, inference, and prediction. 2nd Ed. New York: Springer. ISBN: 978-0-387-84857-0

Hill S, Franco-Sepulveda E, Komeili A, Trovato A, Parent E, Hill D, Lou E, Adeeb S. 2014. Assessing asymmetry using reflection and rotoinversion in biomedical engineering applications. Proc Inst Mech Eng H. 228(5): 523–529.

Ho C, Parent EC, Watkins E, Moreau MJ, Hedden D, El-Rich M, Adeeb S. 2015. Asymmetry assessment using surface topography in healthy adolescents. Symmetry. 7(3):1436–1454.

Jain AK, Chandrasekaran B. 1982. 39 Dimensionality and sample size considerations in pattern recognition practice. Handbook of Statistics. 2:835–855.

Komeili A, Westover LM, Parent EC, Moreau M, El-Rich M, Adeeb S. 2014. Surface topography asymmetry maps categorizing external deformity in scoliosis. Spine J. 14(6):973–983.e2.

Komeili A, Westover L, Parent EC, El-Rich M, Adeeb S. 2015. Correlation between a Novel surface topography asymmetry analysis and radiographic data in Scoliosis. Spine Deform. 3(4):303–311.

Komeili A, Westover L, Parent EC, El-Rich M, Adeeb S. 2015. Monitoring for idiopathic scoliosis curve progression using surface topography asymmetry analysis of the torso in adolescents. Spine J. 15(4):743–751.

Levy AR, Goldberg MS, Mayo NE, Hanley JA, Poitras B. 1996. Reducing the lifetime risk of cancer from spinal radiographs among people with adolescent idiopathic scoliosis. Spine (Phila Pa 1976). 21(13):1540–1547. discussion 1548.

Liu XC, Thometz JG, Lyon RM, Klein J. 2001. Functional classification of patients with idiopathic scoliosis assessed by the Quantec system: a discriminant functional analysis to determine patient curve magnitude. Spine (Phila Pa 1976). 26(11):1274.

Raudys S. 2001. Statistical and neural classifiers: an integrated approach to design. London: Springer-Verlag.

Rogala EJ, Drummond DS, Gurr J. 1978. Scoliosis: incidence and natural history. A prospective epidemiological study. J Bone Joint Surg Am. 60(2): 173–176.

Rokach L, Maimon O. 2015. Data Mining with Decision Trees: Theory and Applications. 2nd ed. Toh Tuck Link, Singapore: World Scientific Publishing Co. Pte. Ltd.

Thulbourne T, Gillespie R. 1976. The rib hump in idiopathic scoliosis. Measurement, analysis and response to treatment. J Bone Joint Surg Br. 58(1):64–71.

Tones M, Moss N, Polly Jr. DW. 2006. A review of quality of life and psychosocial issues in scoliosis. Spine (Phila Pa 1976). 31(26):3027–3038.

Trobisch P, Suess O, Schwab F. 2010. Idiopathic scoliosis. Dtsch Arztebl Int. 107(49):875–884.

Wang J, Neskovic P, Cooper LN. 2006. Neighborhood size selection in the k-nearest-neighbor rule using statistical confidence. Pattern Recognit. 39(3):417–423.

Weinstein SL, Dolan LA, Cheng JCY, Danielsson A, Morcuende JA. 2008. Adolescent idiopathic scoliosis. Lancet. 371(9623):1527–1537.