The following scientific article was officially published in the journal *Artificial Intelligence in Medicine*, published by Elsevier. This article's citation is as follows:

The final publication is available via the DOI: http://dx.doi.org/10.1016/j.artmed.2012.07.002

The manuscript, in a draft version prior to being accepted by the publisher, is reproduced in the following pages.

Non Invasive Classification System of scoliosis curve types using Least-squares support vector machine.

Mathias M. Adankon[*,a,b], Jean Dansereau[a,b], Hubert Labelle[b], Farida Cheriet[a,b]

[a]*Ecole Polytechnique de Montreal,University of Montreal, 2900, boul. Edouard-Montpetit, Campus de l'Universite de Montreal, 2500, chemin de Polytechnique, Montreal (Quebec) H3T 1J4, Canada*
[b]*Sainte-Justine Hospital Research Center, 3175, Chemin de la Cote-Sainte-Catherine, Montreal (Quebec) H3T 1C5, Canada*

Abstract

Objective: To determine scoliosis curve types using non invasive acquisition, without any prior knowledge on X-ray data.

Methods: Classification of scoliosis deformities according to curve type is used to plan management of scoliosis patients. In this work, we propose a robust system that can determine the scoliosis curve type from non invasive acquisition of 3D back surface of the patients. The 3D image of back surface of the trunk is divided into patches and local geometric descriptors characterizing the surface of the back are computed from each patch and forming the features. We perform the reduction of the dimensionality by using principal component analysis and retain 53 components using overlap criterion combined with the total variance in the observed variables. In this work a multi-class classifier is built with least-squares support vector machine (LS-SVM). The original LS-SVM formulation was modified by weighting differently the positive and negative samples and a new kernel was designed in order to achieve a robust classifier. The proposed system was validated using data of 165 patients with different scoliosis curve types. A comparison of the results of a non invasive classification was done with those obtained by an expert using X-ray images.

Results: The average rate of successful classification was computed using

---

[*]Corresponding author

Email addresses: `mathias-mahouzonsou.adankon@polymtl.ca` (Mathias M. Adankon), `jean.dansereau@polymtl.ca` (Jean Dansereau),
`hubert.labelle@recherche-ste-justine.qc.ca` (Hubert Labelle),
`farida.cheriet@polymtl.ca` (Farida Cheriet)

leave-one-out cross-validation procedure. The overall accuracy of the system was 95%. Considering the correct classification rate per class, we obtained 96%, 84% and 97% respectively for thoracic major curve, double major curve and lumbar/thoracolumbar major curve.

Conclusion: This study shows that it is possible to find a relationship between the internal deformity and the back surface deformity in scoliosis with machine learning methods.The proposed system uses non invasive acquisition which is safety for the patient, no radiation. Also, a design of a specific kernel improved classification performance.

Key words: Least-squares support vector machine, 3D trunk modeling, Kernel function, Computer-Aided Diagnosis, Scoliosis.

1. Introduction

Adolescent idiopathic scoliosis (AIS) is a deformity of the spine manifested by asymmetry and deformities of the external surface of the trunk [1]. It consists of a complex curvature in the three-dimensional space: inclination in the frontal plane, rotation of vertebrae in the horizontal plane and inversion of the curves in the sagittal plane. This pathology is often visible, but it may pass as unnoticed during its development for years. There are a wide variety of deformities of the spine, however, a classification of major curve types is possible. The classification of different deformities is used to group similar curves in order to define an appropriate treatment strategy.

Currently, X-ray exam is performed in order to determine the scoliosis curve type and its severity. The scoliosis X-ray includes the entire spine image, thoracic part (upper back) and the lumbar part (lower back). Scoliosis curves are classified by their location in the spine and the magnitude of the curve expressed by the Cobb angle [2]. There are four main types of scoliosis curves or scoliosis spine shape (see Figure 1):
- The *thoracic* curve type is the most typical of the types of scoliosis curves. It affects the upper part of the spine which is often bended to the right. Since the vertebrae are rotated in these cases, this type is accompanied by the rib cage deformation.
- The *lumbar* curve type affects the lower part of the spine. This type of scoliosis also tends to the left as opposed to the right. It does not often cause noticeable deformity on the external shape of the trunk.
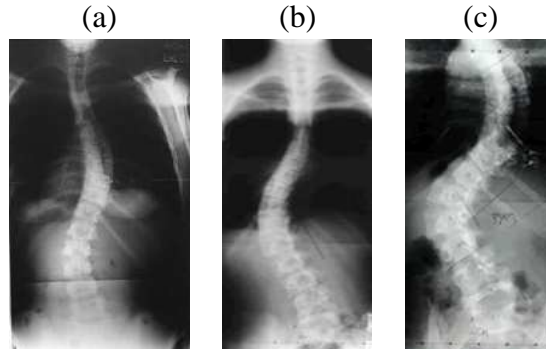
Figure 1: Scoliosis curve types (a) Thoracic, (b) Thoracolumbar and (c) Double major curve.

- The *thoracolumbar* curve type can bend left or right and effects area between the thoracic and lumbar regions of the spine. Small curves above and below accompany this long curve. This type is uncommon and can cause severe rib and trunk distortion.
- The *double major* curve type is one of the most serious scoliosis curves. It is made up of two curves with nearly equal angles and the spine is like the letter S in this case. Usually, one curve occurs on the upper spine and the other on the lower spine.

While 1 in 25 people have mild scoliosis deformities, only 1 in 200 adolescents have deformities that progress to require either bracing or surgical treatment. Since there is as yet no reliable way to predict which deformities will progress, these suspected patients are monitored with a series of X-rays acquired semi-annually during rapid adolescent growth. However, cumulative exposure to X-rays radiation significantly increases the risk for certain cancer [3].

Thus, during the last 30 years, many optical non invasive surface measurement systems have been developed based on a 3D reconstruction of the back or of the whole trunk. The first technique used for this purpose is Moire contour topography where a light-interference pattern projected from a fixed grid near the camera permit to draw contour lines on the back surface [4, 5]. Other techniques such as rasterstereography [6, 7], integrated shape imaging system (ISIS) [8, 9], Quantec scanner [10], were proposed in order to quantify scoliosis deformity from back surface asymmetry. However, the results obtained from the various research were mitigated because of many problems related to the limitations of the avail-

able technologies. In fact, significant correlations were found between scoliosis deformities and Moire fringe deviations, but the false-positive rate was as high as 50 percent and the Moire fringe is sensitive to small movement of the patient [4, 7]. New studies were extended to the acquisition and the analysis of the entire trunk in order to take into account the whole trunk asymmetry [11–16]. However, human torso shape has many variations and its analysis is very challenging. Thus, the main problem is how to extract the relevant features that could characterize scoliosis torso shape deformity to perform a good classification of scoliosis curve type without acquiring X-ray images.

To our knowledge, studies on non invasive classification of scoliosis type or severity using machine learning methods are limited [17–20]. In the first work [17], an artificial neural network combined with genetic algorithm is used in order to estimate the Cobb angle. The results reveal that the system evaluate the Cobb angle within 5 degrees in 65 percent on the test set. The major problem that occurs in this study is the over-fitting because the machine gives a good result on training samples but has poor estimation performance. In [18], a spinal curve was assessed based on the trunk surface image. This study has attempted to find a non-linear correlation between the trunk surface and the geometry of the underlying spine. An array of support vector regression machines is built to predict the spinal coefficients which constitute comprehensive features for spinal curve description. However, the results are moderate with 70-82% of correct evaluation. In [19], the authors proposed a system where the subjects were classified into 3 severity groups (mild, moderate, severe) using 3D back shape image combined with other indicators like sex, age, etc. And, their system achieved 69-85% accuracy in testing. In contrast, Lama et al.[20] have proposed for the first time a classification of the scoliosis curve type using only the 3D image of the trunk. In this work, the authors have considered the Lenke classification which uses also bending test; thus, the system gave a limited result, 72% of correct classification.

The previous studies have obtained moderate classification results partly because of the small number of the available data which is a critical problem in biomedical research field. In fact, the performance of many learning algorithms, as artificial neural network, are very dependent on the number of training sample. An other limit is the model selection problem, optimization of the hyperparamters, which affects significantly the performance of any designed machine.

In this paper, we propose a robust non invasive classification system of sco-

4

liosis curve types. The objective here is to develop a pattern recognition system which identify scoliosis curve type using 3D back surface image (non invasive acquisition) instead of X-ray exam. The proposed system is based on a powerful classifier, least squares support vector machine (LS-SVM) which has excellent capacity of generalization [21]. In this study, the classifier LS-SVM formulation is modified and empowered by a new kernel function [22–25] combined with an appropriate model selection strategy [26, 27] which yields a good performance. In this work, we use a special case of the weighted LS-SVM proposed early in [28, 29] in order to reduce the effects of noisy data. The 3D image of the back surface is divided into patches and local geometric descriptors are computed from each patch forming the features and classification is performed using a combination of LS-SVM classifiers. The proposed system is illustrated in Figure 2.

2. Materials and methods

2.1. Data acquisition

For many years, the acquisition of the trunk surface is part of the routine evaluation of scoliosis patients at Sainte-Justine University Hospital Center (SJUHC) in Montreal (Canada). The acquisition system is composed of four optical digitizers (Creaform, Montreal, Canada). Each optical digitizer contains one color CCD camera and a structured light projector. The acquisition process of each digitizer is as follows. Four fringe patterns, obtained by phase-shifted technique, are successively projected onto the surface. Based on the four resulting images and combined with interferometry and triangulation technique, the system computes the depth of each surface point relative to the reference plane. A fifth image, with no fringes, acquires the texture of the surface which is then mapped onto the 3D reconstruction.

For the reconstruction of the whole trunk, four scanners are placed around the patient (on the front, on the back and at $\pm 60°$ laterally in front of the patient), see Figure 3. Each digitizer acquires a portion of the trunk. During the acquisition, approximately 4 seconds, the patient stands still in the upright position with the arms slightly abducted in order not to obstruct the lateral scanners fields of view. Based on a multi-head calibration of the system that computes the rigid transformations between the digitizers, the 4 portions of the trunk are registered and merged using EM software. The resulting mesh is constituted of 40,000 to 70,000 nodes, depending on the size of the patient.
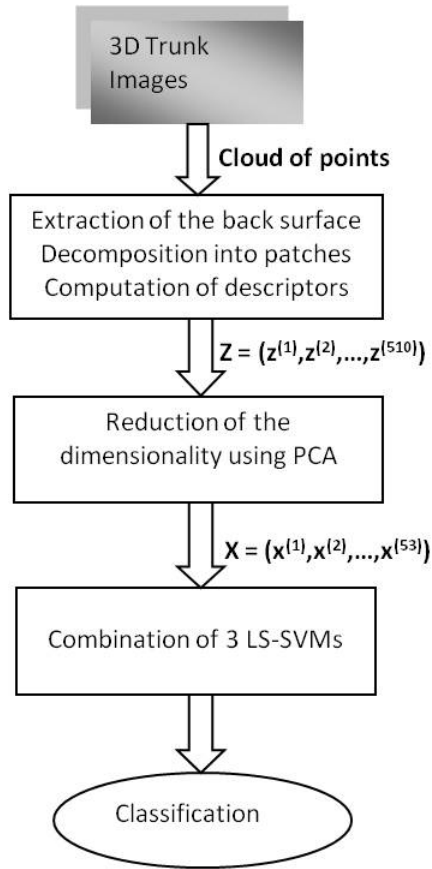
Figure 2: General overview of the classification system.

The accuracy of this system was evaluated in [30], using markers placed on a mannequin whose coordinates were previously recorded by a computer measuring machine. The results showed a reconstruction accuracy of 1.4 mm over the whole torso and of 0.56 mm over the back.

2.2. Features extraction

The 3D trunk image is divided into h equal parts using horizontal planes as shown in Figure 4. Using cylindrical coordinates (ρ, ϕ, z), each transversal slice is divided in n patches by varying the azimuth ϕ (see Figure 5). Thus, the whole
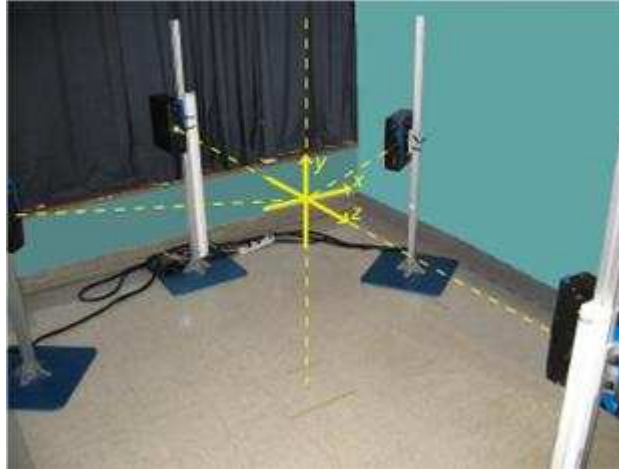
Figure 3: Configuration of Creaform digitizers at Sainte-Justine hospital research center.

3D image is split up into $n*h$ patches and the geometric descriptors are calculated from each patch. First, we approximate each patch (piece surface) by a plane and the normal vector (n_x, n_Y, n_z) of this plane is kept as a descriptor. Then, from each 3D trunk image, we compute $3nh$ descriptors.
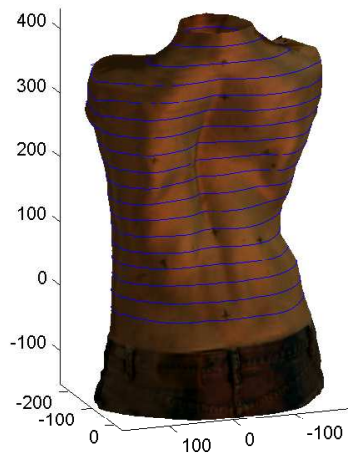

Figure 4: Decomposition of trunk into slices.

In this study, we consider the thoracic and lumbar region (17 vertebrae) and set $h = 17$; and we evaluate only the back surface by fixing ϕ between 0 and π. The number n of patches per slide is chosen by evaluating the overlap rate between classes (classification error rate evaluated with 1-Nearest Neighbor technique [31]). Figure 6 shows the variation of the overlap rate with respect to the number of the patches per slide. Following this procedure, we set $n = 10$ corresponding to the minimum of the overlap rate.

We divided the trunk in two regions: thoracic region and lumbar region. The back surface has been divided into 17 equal parts using horizontal planes. The first 10 slices from the top represent the thoracic region and the remaining 7 slices represent the lumbar region [1]. Thus, we obtained respectively 300 ($3 \times 10 \times 10$) and 210 ($3 \times 10 \times 7$) descriptors for thoracic and lumbar regions. In sum we have 510 descriptors for each 3D image.
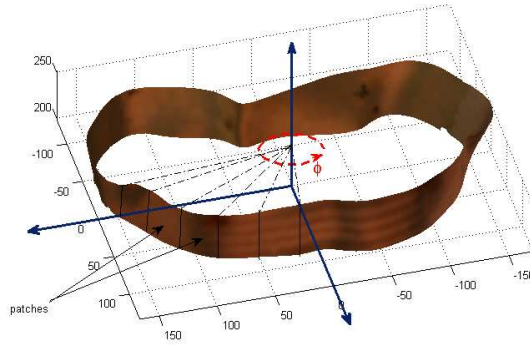
Figure 5: Decomposition of transversal slice into patches.

In general, building a classifier based on statistical methods with a large number of features is not recommended because working in high-dimensional space involves the curse of dimensionality problem. Thus, for this work, we apply principal component analysis (PCA), which is an orthogonal linear transformation, in order to reduce the dimensionality of the data. The number of components is selected based on an overlap criterion (classification error rate evaluated with 1-

---

[1]Considering T and L respectively the size of the thoracic and lumbar region; the relationship between them is approximatively T/L=1.4 according to Jean Cruveilhier, Anatomie descriptive, Volume 1, 1837, page 49.

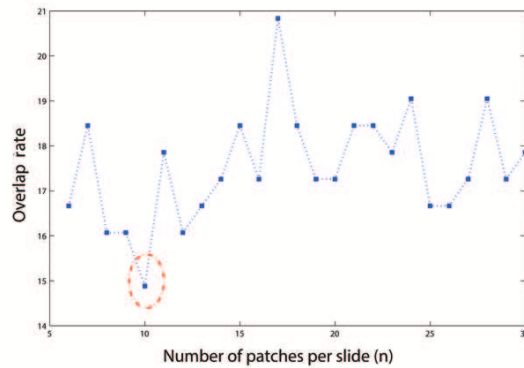Figure 6: Variation of overlap rate w. r.t. the number of the patches per slide.

Nearest Neighbor technique [31]) and the rate of the total variance in the observed variables retained by the selected principal components. In fact, we select the number of components which provides more than 80% of the total variance and yield a better overlap rate. In Figure 7, the rate of total variance in the observed variables with respect to the number of principal components in each region is represented and in Figure 8, we plot the variation of the overlap rate according to the number of components $n1$ (for thoracic region) and $n2$ (for lumbar region). After investigation, we note that the overlap rate is small and stable for $n1 = 23$ and $n2 = 30$. Thus, we retain the first 23 components for the thoracic region and 30 components for the lumbar region. Therefore, each trunk is represented by 53 features.
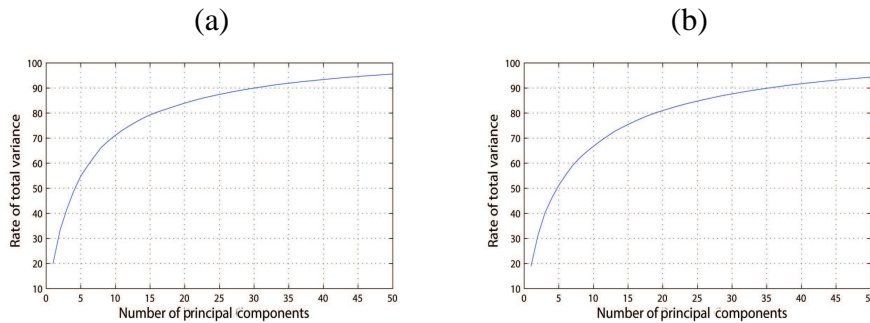
(a)   (b)

Figure 7: Rate of total variance in the observed variables w.r.t. the number of principal components (a)Thoracic region , (b)Lumbar region.

9
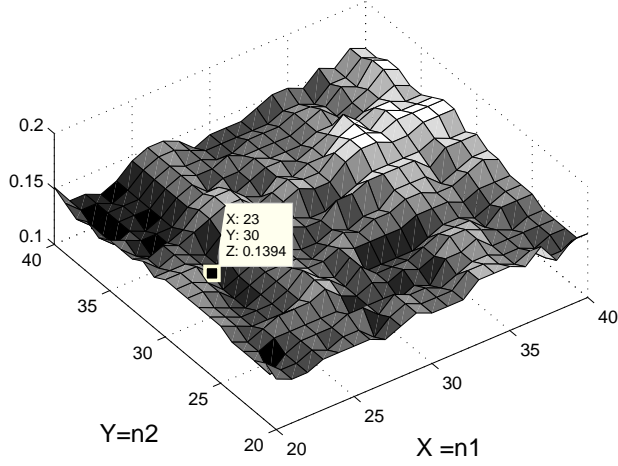
Figure 8: Variation of overlap rate w. r.t. the number of the principal component $n1$ (for thoracic region) and $n2$ (for lumbar region).

2.3. Optimization of the classification system

As classifier, we used LS-SVM which is based on the margin maximization principle[32]. Considering a binary classification problem involving a dataset $\{(x_1, y_1), \ldots, (x_\ell, y_\ell)\}$ with $x_i \in \mathcal{R}^d$ and $y_i \in \{-1, 1\}$; nonlinear LS-SVM classifiers use the kernel trick to produce nonlinear boundaries. The decision function given by an LS-SVM is :

$$f(x) = sign[w'\phi(x) + b] \tag{1}$$

where w and b are the classifier parameters, w' represents transpose of w, ϕ is the explicit projection function corresponding to the kernel $k(x_i, x_j) = \phi(x_i).\phi(x_j)$.

The LS-SVM classifier parameters are found by resolving the following optimization problem which expresses the maximization of the margin $2/\|w\|$ and the minimization of the training error:

$$\min_{w,b,\xi} \frac{1}{2}w'w + \frac{1}{2}\gamma \sum_{i=1}^{\ell} \xi_i^2 \tag{2}$$

$$\text{subject to} : \xi_i = y_i - [w'\phi(x_i) + b] \quad \forall i = 1, ..., \ell \tag{3}$$

10

LS-SVM is a variant of the standard support vector machine (SVM) where the original SVM formulation is modified at two points. First, the inequality constraints with the slack variable ξ_i are replaced by equality constraints. Second, a squared loss function is considered in the objective function. These two essential modifications simplify the problem, which becomes a linear system.

The Lagrangian of problem (2) is expressed by :

$$\mathcal{L} = \frac{1}{2}w'w + \frac{1}{2}\gamma \sum_{i=1}^{\ell} \xi_i^2 - \sum_{i=1}^{\ell} \alpha_i\{y_i - [w'\phi(x) + b] - \xi_i\}$$

where α_i are Lagrange multipliers, which can be positive or negative because of equality constraints.

The system arising from the Karush-Kuhn-Tucker conditions is linear, and the system of linear equations is expressed in the following matrix form:

$$\begin{pmatrix} K + \gamma^{-1}I & \vec{1}' \\ \vec{1} & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ b \end{pmatrix} = \begin{pmatrix} Y \\ 0 \end{pmatrix} \tag{4}$$

where $K_{ij} = \phi(x_i).\phi(x_j)$; $Y = (y_1, ..., y_\ell)'$; $\alpha = (\alpha_1, ..., \alpha_\ell)'$; and $\vec{1} = (1, ..., 1)$.

The decision function of the LS-SVM becomes :

$$f(x) = sign[\sum_{i=1}^{\ell} \alpha_i k(x_i, x) + b] \tag{5}$$

2.3.1. Regularization of LS-SVM with γ_1 and γ_2

In this study, the impact of the hyperparameters γ which controls the tradeoff between training error minimization and margin maximization, is modified in order to balance the training error. We propose to use γ_1 and γ_2 for balancing the influence of the number of positive samples with respect to the number of negative samples contained in the training set. Thus, γ_1 and γ_2 penalize respectively the training error made by the positive samples and the negative samples. Therefore, we reformulate the LS-SVM as follows:

$$\min_{w,b,\xi} \frac{1}{2}w'w + \frac{1}{2}\gamma_1 \sum_{i/y_i=1} \xi_i^2 + \frac{1}{2}\gamma_2 \sum_{i/y_i=-1} \xi_i^2 \tag{6}$$

$$\text{subject to} : \xi_i = y_i - [w'\phi(x_i) + b] \quad \forall i = 1, ..., \ell \tag{7}$$

11

The formulation (6) is the special case of the weighted LS-SVM [29, 33]. In the previous work, the values of γ_1 and γ_2 are set using the number of positive and negative samples; but in this study, these values are optimized during the model selection process.

In dual form, we obtain also a system of linear equations expressed in the following matrix form:

$$\begin{pmatrix} K + \Gamma & \vec{1}' \\ \vec{1} & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ b \end{pmatrix} = \begin{pmatrix} Y \\ 0 \end{pmatrix} \tag{8}$$

where $K_{ij} = \phi(x_i).\phi(x_j)$; $Y = (y_1, ..., y_\ell)'$; $\alpha = (\alpha_1, ..., \alpha_\ell)'$; $\vec{1} = (1, ..., 1)$; and Γ is diagonal matrix containing ℓ^+ times γ_1 and ℓ^- times γ_2 with ℓ^+ and ℓ^- respectively the number of positive samples and negative samples.

2.3.2. Design of a specific kernel

The idea behind kernels is to map training data nonlinearly into a higher-dimensional feature space via a mapping function Φ and to construct a separating hyperplane which maximizes the margin. The construction of the linear decision surface in this feature space only requires the evaluation of dot products $\phi(x_i).\phi(x_j) = k(x_i, x_j)$, where the application $k : \mathcal{R}^d \times \mathcal{R}^d \to \mathcal{R}$ is called the kernel function [22].

When we consider LS-SVM like other kernel classifiers, the choice of the kernel corresponds to choosing a function space for learning. The kernel determines the functional form of all possible solutions. Thus, the choice of the kernel is very important to construct a good machine. So, in order to obtain a good performance of the LS-SVM classifier, we need first to design or choose a type of kernel and then to optimize the hyperparameters for improving the capacity of generalization. In our context, we design a new kernel derivative from the Gaussian kernel where the features for the thoracic region and those for the lumbar region are weighted differently.

Considering two examples $x = (x_{(1)}, ..., x_{(T)}, x_{(T+1)}, ..., x_{(T+L)})$ and $y = (y_{(1)}, ..., y_{(T)}, y_{(T+1)}, ..., y_{(T+L)})$ where each element is formed by T features obtained from the thoracic region and L features from the lumbar region, we define our kernel function by:

$$k(x, y) = exp\left[-\beta_T \sum_{i=1}^{T} (x_{(i)} - y_{(i)})^2 - \beta_L \sum_{j=T+1}^{T+L} (x_{(j)} - y_{(j)})^2 \right] \tag{9}$$

12

where $\beta_T > 0$ and $\beta_L > 0$ represent respectively the width of the kernel for the thoracic and lumbar regions.

The previous kernel can be viewed as the special case of automatic relevance determination (ARD) method. In fact, ARD adresses the features selection problem by using regularisation function which prunes away redundant or superfluous features. Therefore, the previous kernel can be considered as the extended version of the Gaussian for ARD and the parameters can be inferred in the third level as presented in [21].

2.3.3. Model selection

Model selection for the LS-SVM consists of selecting the hyperparameters that yield the best performance of the machine. The LS-SVM classifier has two types of hyperparameters: the regularization parameters which control the trade-off between training error minimization and margin maximization, and the kernel parameters that define the given kernel function.

In this study we have for each LS-SVM two regularization parameters γ_1 and γ_2 and two kernel parameters β_T and β_L. Since selecting the kernel corresponds to choosing a function space for learning, we set the couple (β_T, β_L) of kernel parameters equal for all LS-SVM classifiers. In fact, we need to have the same function space for learning in order to combine the output of the three classifiers without bias.

Thus, we need to find eight best values for our hyperparameters, six for the regularization parameters and two for the kernel parameters. We perform model selection for setting the hyperparameters through generalization performance. The details of the model selection procedure are described in [26].

2.3.4. Classification algorithm

LS-SVM like classical SVM was designed first for two-class problem where the labels of the first class samples are assigned to $+1$ and the labels of the second class are assigned to -1. Concerning multi-class problem, the classification task is divided in several two-class problems with adequate combination at the end. Following this idea, many techniques are proposed in the literature with their advantages and their inconvenients. For our multi-class problem, we used

13

one-against-all technique with rejection when all the LS-SVM outputs are negative. Thus, three LS-SVMs are trained, each separating one class from the others. During the test phase, the label is found by the following equation:

$$y = \underset{i}{\operatorname{argmax}}[w_i'\phi(x) + b_i] \tag{10}$$

where w_i and b_i are the corresponding parameters of the LS-SVM trained for separating the samples of the class i from the others.

2.4. Validation

Our classification system is validated on the dataset of 3D torso images of 165 patients with adolescent idiopathic scoliosis with an age between 10-18 years and Cobb angle measurement of the main curve is greater than $35°$. The 3D trunk image and the radiography acquisitions have been done at the same visit for each patient. Based on the common clinical classification, the scoliosis curve type (thoracic major curves, thoracolumbar major curves, lumbar major curves and double major curves) was determined from the X-ray images by an expert. The number of lumbar major curves being too small, we mixed them with the thoracolumbar major curves. The repartition of samples in these different classes is shown in Table 1.

Table 1: Distribution of the patients among the classes

Classes	Curve type	# patients
Class 1	Thoracic major	102
Class 2	Double major	26
Class 3	Lumbar/thoracolumbar major	37
Total		165

Considering the size of the dataset (N=165) and the small number of samples in each class (only 26 samples for class 2), it is not interesting to reduce the training set in order to form the separate validation set. Therefore, we choose "leave-one-out" cross-validation (LOOCV) to estimate the performance of our classification system. LOOCV is used because it is almost unbiased and its error should be relatively informative about the generalization error of the classifier, see [34, 35].

The cross-validation procedure is a good technique for evaluating classifier generalization performance, the idea being to test the generalization capacity of the classifier through unseen data. In k-fold cross-validation, we divide the available training data into k subsets. We train the machine k times, each time leaving out one of the subsets from training, and use only the omitted subset to compute the given error criterion. The average error rate obtained over the k operations, gives an estimation of the classifier generalization capacity. The variance of the result is reduced when k is increased. If k equals the size of the training set, the maximum value, this is called "leave-one-out" cross-validation, or LOO cross-validation.

Indeed, we used double cross-validation where the LOOCV methodology is repeated at two stages: first stage for performing model selection (selecting the hyperparameters $(\beta_T, \beta_L, \gamma_1, \gamma_2)$) and second stage for evaluating the global system performance. See Algorithm 1 for the details.

Algorithm 1 Double LOO cross-validation

---

for $i = 1 \rightarrow \ell$ **do**
    $A = \{(x_1, y_1), \ldots, (x_\ell, y_\ell)\} \setminus \{(x_i, y_i)\}$ //First LOO cross-validation
    for each $(\beta_T, \beta_L, \gamma_1, \gamma_2)$ combination **do**
        for $j = 1 \rightarrow \ell - 1$ **do**
            $B = A \setminus \{(x_j, y_j)\}$ //Second LOO cross-validation
            Train LS-SVM with B using different value of $(\beta_T, \beta_L, \gamma_1, \gamma_2)$
            Test the classifier on x_j
        end for
    end for
    Select the optimal value of $(\beta_T, \beta_L, \gamma_1, \gamma_2)$ which gives the minimum average error on $x_{j(j=1:\ell-1)}$
    Train the classifier with A using the optimal value of $(\beta_T, \beta_L, \gamma_1, \gamma_2)$
    Test the classifier on x_i
end for
Compute the overall performance on $x_{i(i=1:\ell-1)}$

We validate the proposed system on the previous dataset to assess its performance. In fact, our system is built by improving many parts of the classification process. First, the data representation is improved by using PCA technique where the features dimension is reduced. Second, we design a new kernel which makes

possible to weight differently the different features according to their origin (thoracic region or lumbar region). Third, we introduce the balanced regularization parameters for the positive and negative samples in training formulation of the classifier. Thus, we design experiments where the data representation or the classifier is modified in order to confirm the influence of each strategy on the proposed method:

1) To quantify the effect of the dimensionality reduction, experiments are performed using row data and Wrapper method which is an iterative dimensionality reduction technique comparative with PCA.

2) To show the influence of the new kernel, a standard Gaussian kernel and linear kernel are tested.

3) To point out the impact of the regularization parameters γ_1/γ_2, a classical LS-SVM is tested with equal value for the parameters γ_1 and γ_2.

Finally, we make comparison with other classifiers. The proposed classifier is compared with the standard LS-SVM, with SVM and with artificial neural network.

3. Results and discussion

3.1. Performance of our system

We perform the LOOCV procedure through the 165 samples with our classification system which gives 95% of correct classification. The optimal hyperparameters value of each classifier are set through the double-cross-validation procedure. The optimal value of the hyperparameters (mean value estimated through LOOCV) are shown in Table 2.

Table 2: Hyperparameters value

Kernel parameter	Regularization parameter for class 1	Regularization parameter for class 2	Regularization parameter for class 3
$\beta_T = 0.0595$	$\gamma_1 = 1$	$\gamma_1 = 0.05$	$\gamma_1 = 1$
$\beta_L = 0.0190$	$\gamma_2 = 1$	$\gamma_2 = 0.08$	$\gamma_2 = 2$

The interesting result of good classification 95%, shows that it is possible to identify the scoliosis curve type using only 3D back surface image of the patients.

16

Also, the result confirms that the selected features extract the relevant information from the 3D image for a scoliosis curve type classification task.

In Table 3, we report the correct classification rate per class. Considering these results, it is remarkable that the prediction of double major curve is the most difficult task. Further analysis of the results shows that almost the misclassification of double major curve samples are predicted as thoracic major curve. Thus, the major confusion of the classifier is produced between the double major curve and thoracic major curve. In fact, we know that it is also very difficult for an expert to separate these two groups using radiographs of the spine when the main curvature of the double major curve is thoracic. Thus, the clinicians in addition to the difference of the Cobb angle, develop the classification based on the curve reduction between X-ray acquired in standing position and those acquired in lateral flexion position [36]. On the other hand, the moderate predictive capacity of our classification system for class 2 comes also from the small number of the patients who have double major curve in the cohort.

Table 3: Classification rate per class

Curve type	Classification rate
Thoracic major	96.1%
Double major	**84.2%**
Lumbar/thoracolumbar major	97.2%

3.2. Impact of the dimensionality reduction

Dimensionality reduction is an interesting preprocessing in machine learning. However, sometimes a sub-optimal result can be observed. Thus, we performed a comparison with row data without any dimensionality reduction. Also, we tested the wrapper method which is an iterative technique used for feature subset selection [37]. The performances of each system with LOOCV procedure are reported in Table 4. The results outpoint that the dimensionality reduction is very useful; because without any dimensionality reduction, the recognition rates for classes 2 and 3 decrease with respect to the systems using dimensionality reduction method. In addition, we noted that PCA outperforms the Wrapper method by giving an interesting results for each class. Therefore, we can conclude that PCA is more

17

suitable than Wrapper method for our problem.

Table 4: Data representation comparison

Curve type	Row data	Wrapper method	PCA
Thoracic major	96.1%	97.0%	96.1%
Double major	**61.9%**	**68.8%**	**84.2%**
Lumbar/thoracolumbar major	94.4%	96.9%	97.2%

3.3. Impact of the new kernel

In order to quantify the impact of the new kernel with respect to the standard Gaussian kernel, we evaluate our system using respectively the two types of kernels. Table 5 summarizes the results obtained using the different Gaussian kernel widths. The Gaussian kernel width is varied between $min(\beta_T, \beta_L)$ and $max(\beta_T, \beta_L)$ where β_T and β_L are optimal values found for the new kernel. We note that the recognition rate within the class 2, double major curve, was improved from 73.7% to 84.2%. Therefore, the overall recognition rate increased also. Thus, the results demonstrate that the new kernel is more suitable than the Gaussian kernel for this problem and permit to ameliorate the classification rate of the double major curve.

Moreover, the proposed kernel is compared with the standard linear kernel and modified linear kernel. This last kernel is built by weighting differently the features computed from each trunk region as in the case of proposed kernel. The results reported in Table 6 show that the linear kernel is not appropriate for our problem. This confirms that the classification problem is not linear. Therefore, linear classification gives a sub-optimal performance.

3.4. Impact of regularization parameters γ_1/γ_2

In this section, we run two sets of experiments in order to quantify the impact of the regularization parameters γ_1 and γ_2 on the prediction rate. First, we tested our proposed system without the regularization parameters γ_1 and γ_2 (This means using the same value for the regularization parameters $\gamma_1 = \gamma_2$ like in standard LS-SVM) and we compute the different classification rates. Second, we used the regularization parameters with the different value for γ_1 and γ_2 for each classifier. Table 7 shows the summary of the results where the classification rate per class

18

Table 5: Prediction rate according to the kernel choice

Curve type:	Thoracic major	Double major	Lumbar/thoracolumbar major	Overall
Gaussian kernel(0.0190):	96.0%	**70.6%**	99.9%	**94.2%**
Gaussian kernel(0.0235):	96.0%	73.7%	97.2%	93.6%
Gaussian kernel(0.0280):	96.0%	73.7%	97.2%	93.6%
Gaussian kernel(0.0325):	96.0%	73.7%	97.2%	93.6%
Gaussian kernel(0.0370):	96.0%	73.7%	97.2%	93.6%
Gaussian kernel(0.0415):	96.0%	73.7%	97.2%	93.6%
Gaussian kernel(0.0460):	97.0%	72.2%	97.2%	94.1%
Gaussian kernel(0.0505):	97.0%	**73.7%**	97.2%	**94.2%**
Gaussian kernel(0.0550):	97.0%	68.2%	97.2%	93.0%
Gaussian kernel(0.0595):	97.0%	68.2%	97.2%	93.0%
Proposed kernel: ($\beta_T = 0.0595; \beta_L = 0.0190$)	96.1%	**84.2%**	97.2%	**95.0%**

Table 6: Comparison with linear kernel

Curve type	Linear kernel	Linear modified kernel	Proposed kernel
Thoracic major	96.0%	96.1%	96.1%
Double major	**52.9%**	**60.0%**	**84.2%**
Lumbar/thoracolumbar major	94.4%	94.4%	97.2%

and the overall rate are reported. We note that the classification rate of the class 1, Thoracic major curve, was increased by one percent; Therefore, the global classification rate was improved too.

3.5. *Comparison with other classifiers*

The main goal of this study is to build a strong system of scoliosis type prediction based on 3D trunk image. Thus, the comparison with the classical classifiers are performed. We tested the multilayer perceptron (MLP) which is an artificial

Table 7: Impact of regularization parameters γ_1/γ_2

Curve type	Proposed system with $\gamma_1 = \gamma_2$	Proposed system with $\gamma_1 \neq \gamma_2$
Thoracic major	**95.1%**	**96.1%**
Double major	84.2%	84.2%
Lumbar/thoracolumbar major	97.2%	97.2%
Overall	94.2%	95.0%

neural network model designed for non linear problems. In this work, MLP with one hidden layer is used and the number of neurons in the hidden layer is optimized through LOOCV process. Also, the classical margin kernel classifiers SVM and LS-SVM are tested. The hypeparameters used for these two last classifiers are reported in Table 8.

The comparative results with MLP, LS-SVM and SVM are reported in Table 9. The results show that the MLP gives a sub-optimal solution with respect to the kernel classifiers. Our proposed system outperforms clearly this classifier. Considering the good classification on double major curve, the proposed system outperforms all the other classifiers. Therefore, the results demonstrate that our system yield a good performance comparative with the other classifiers as MLP, SVM and the standard LS-SVM. This confirms the robustness and the strong stringency of our system.

Table 8: Hyperparameters value of SVM and LS-SVM

Classifier	Kernel parameter	Regularization parameter for class 1	Regularization parameter for class 2	Regularization parameter for class 3
SVM	0.0370	1	10	1
LS-SVM	0.0505	1	0.05	1

Table 9: Comparison with other classifiers: classification rate per class

Curve type	MLP	SVM	LS-SVM	Proposed system
Thoracic major	94.1%	97.0%	97.0%	96.1%
Double major	**72.0%**	**80.0%**	**73.7%**	**84.2%**
Lumbar/thoracolumbar major	91.4%	97.3%	97.2%	97.2%

4. Conclusion

In this work, we proposed to build a computer-aided diagnosis (CAD) system which is able to classify the scoliosis curve type using 3D back surface images which are obtained from non invasive acquisitions. Our system is validated on 165 patients and we obtained promising results. This shows that it is possible to find a relationship between the internal deformity and the torso shape deformity in scoliosis with machine learning methods. Also, the results indicate that relevant features based on geometric approach combined with artificial intelligence techniques can be extracted from 3D torso image in order to build a CAD system with a high capacity of classification.

In addition, our system is very fast during the testing phase because only 53 principal components are selected and used by the classifier. The small number of features makes it possible to reduce the complexity storage of the classifier and to avoid the curse of dimensionality problem.

However, we note that the classification capacity of our system for the double major curve could be improved. This part of the proposed system could be ameliorated by including the side-bending acquisitions with additional set of descriptors. We expect that further work in this direction will yield better results and will improve the performance of the proposed system.

References

[1] B.V. Reamy and J.B. Slakey. Adolescent idiopathic scoliosis: review and current concepts. *American family physician*, 64(1):111, 2001.

[2] J.R. Cobb. Outline for the study of scoliosis. *Instr Course Lect*, 5:261–275, 1948.

[3] M. M. Doody, J. E. Lonstein, M. Stovall, D. G. Hacker, N. Luckyanov, and C. E. Land. Breast cancer mortality after diagnostic radiography - findings from the us scoliosis cohort study. *Spine*, 25(16):2052–2063, Aug 2000.

[4] MS. Moreland, MH. Pope, I. Wilder, DG. Stokes, and JW. Frymoyer. Moiré fringe topography of the human body. *Med Instrum.*, 15(2):129–132, 1989.

[5] N. Suzuki, G. W. D. Armstrong, and J. Armstrong. Application of moiré topography to spinal deformity. *Moreland, M. S., Pope, M. H., and Armstrong, G. W. D. (Eds): Moiré fringe topography and spinal deformity*, pages 225–240, 1981.

[6] E. Hierholzer and W. Frobin. Rasterstereography measurement and curvature analysis of the body surface of patients with spinal deformities. *Moreland, M. S., Pope, M. H., and Armstrong, G. W. D. (Eds): Moiré fringe topography and spinal deformity*, pages 267–276, 1981.

[7] I. A. Stokes and M. S. Moreland. Concordance of back surface asymmetry and spine shape in idiopathic scoliosis. *Spine (Phila Pa 1976)*, 14(1):73–78, Jan 1989.

[8] Alan R. Turner-Smith, J.Derek Harris, Gregory R. Houghton, and Rosalind J. Jefferson. A method for analysis of back shape in scoliosis. *Journal of Biomechanics*, 21(6):497 – 509, 1988.

[9] I. Weisz, R. J. Jefferson, A. R. Turner-Smith, M. A. Hougthon, and M. A. Harris. Isis scanning: a useful assessment technique in the management of scoliosis. *Spine (Phila Pa 1976)*, 13(4):405–408, 1988.

[10] X.C. Liu, J.G. Thometz, R.M. Lyon, and J. Klein. Functional classification of patients with idiopathic scoliosis assessed by the quantec system: a discriminant functional analysis to determine patient curve magnitude. *Spine*, 26(11):1274–1278, 2001.

[11] EG. Dawson, MA Kropf, G. Purcell, JM. Kabo, LE. Kanim, and C. Burt. Optoelectronic evaluation of trunk deformity in scoliosis. *Spine (Phila Pa 1976)*, 18(3):326–331, Mar 1993.

[12] P. Poncet, S. Delorme, JL. Ronsky, J. Dansereau, G. Clynch, J. Harder, RD. Dewar, H. Labelle, PH. Gu, and RF. Zernicke. Reconstruction of laser-scanned 3d torso topography and stereoradiographical spine and rib-cage geometry in scoliosis. *Computer Methods in Biomechanics and Biomedical Engineering*, 4(1):59–75, 2000.

[13] J L Jaremko, P Poncet, J Ronsky, J Harder, J Dansereau, H Labelle, and R F Zernicke. Estimation of spinal deformity in scoliosis from torso surface cross sections. *Spine (Phila Pa 1976)*, 26(14):1583–1591, 2001.

[14] Jacob L. Jaremko, Philippe Poncet, Janet Ronsky, James Harder, Jean Dansereau, Hubert Labelle, and Ronald F. Zernicke. Indices of torso asymmetry related to spinal deformity in scoliosis. *Clinical Biomechanics*, 17(8):559 – 568, 2002.

[15] P.O. Ajemba, N.G. Durdle, D.L. Hill, and V.J. Raso. Classifying torso deformity in scoliosis using orthogonal maps of the torso. *Medical and Biological Engineering and Computing*, 45(6):575–584, june 2007.

[16] P.O. Ajemba, N.G. Durdle, and V.J. Raso. Characterizing torso shape deformity in scoliosis using structured splines models. *Biomedical Engineering, IEEE Transactions on*, 56(6):1652 –1662, june 2009.

[17] J. L. Jaremko. *Estimation of Scoliosis Severity from the Torso Surface by Neural Networks*. Ph.D. Thesis, Dept. Biomed. Eng., Univ. Calgary, 2003.

[18] Charles Bergeron, Farida Cheriet, Janet Ronsky, Ronald Zernicke, and Hubert Labelle. Prediction of anterior scoliotic spinal curve from trunk surface using support vector regression. *Eng. Appl. of AI*, 18(8):973–983, 2005.

[19] L. Ramirez, N.G. Durdle, V.J. Raso, and D.L. Hill. A support vector machines classifier to assess the severity of idiopathic scoliosis from surface

topography. *Information Technology in Biomedicine, IEEE Transactions on*, 10(1):84 –91, 2006.

[20] L. Seoud, M.M. Adankon, H. Labelle, J. Dansereau, and F. Cheriet. Prediction of scoliosis curve type based on the analysis of trunk surface topography. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pages 408 –411, apr. 2010.

[21] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific, Singapore, 2002.

[22] Bernhard E. Boser, Isabelle Guyon, and Vladimir Vapnik. A training algorithm for optimal margin classifiers. In *Computational Learing Theory*, pages 144–152, 1992.

[23] B. Scholkopf and A.J. Smola. *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.

[24] Nello Cristianini and John Shawe-Taylor. *An introduction to support Vector Machines and other kernel-based learning methods*. Cambridge University Press, Cambridge, 2000.

[25] John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, Cambridge, 2004.

[26] Mathias M. Adankon and Mohamed Cheriet. Model selection for the ls-svm. application to handwriting recognition. *Pattern Recognition*, 42(12):3264–3270, 2009.

[27] Mathias M. Adankon and Mohamed Cheriet. Support vector machine. In *Encyclopedia of Biometrics, Li, Stan Z. (Ed.)*, pages 1303–1308. Springer, US, 2009.

[28] J.A.K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle. Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing*, 48(1-4):85–105, 2002.

[29] Gavin Cawley. Leave-one-out cross-validation based model selection criteria for weighted ls-svms. In *proceedings IJCNN 2006, Vancouver, Canada*, pages 1661 – 1668, July 2006.

[30] V. Pazos, F. Cheriet, L. Song, H. Labelle, and J. Dansereau. Accuracy assessment of human trunk surface 3d reconstructions from an optical digitising system. *Medical and Biological Engineering and Computing*, 43(1):11–15, 2005.

[31] T. Cover and P. Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967.

[32] J. A. K. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3):293–300, 1999.

[33] Anneleen Daemen, Marco Signoretto, Olivier Gevaert, Johan A. K. Suykens, and Bart De Moor. Improved microarray-based decision support with graph encoded interactome data. *PLoS ONE*, 5(4):1–16, 04 2010.

[34] A. Luntz and V. Brailovsky. *On estimation of characters obtained in statistical procedure of recognition*. Technicheskaya Kibernetica, 1969.

[35] V. Vapnik and O. Chapelle. Bounds on error expectation for support vector machines. *Neural Computation*, 12(9):2013–2036, 2000.

[36] D.D. Aronsson, I.A.F. Stokes, P.J. Ronchetti, and B.S. Richards. Surgical correction of vertebral axial rotation in adolescent idiopathic scoliosis: prediction by lateral bending films. *Journal of Spinal Disorders & Techniques*, 9(3):214–219, 1996.

[37] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003.