Classifying Vertical Facial Deformity using Supervised and Unsupervised Learning

P. Hammond, T. J. Hutton, Z. L. Nelson-Moon, N. P. Hunt, A. J. A. Madgwick
Eastman Dental Institute for Oral Health Care Sciences, University College London, UK

Summary

Objectives: To evaluate the potential for machine learning techniques to identify objective criteria for classifying vertical facial deformity.

Methods: 19 parameters were determined from 131 lateral skull radiographs. Classifications were induced from raw data with simple visualisation, C5.0 and Kohonen feature maps; and using a Point Distribution Model (PDM) of shape templates comprising points taken from digitised radiographs.

Results: The induced decision trees enable a direct comparison of clinicians' idiosyncrasies in classification. Unsupervised algorithms induce models that are potentially more objective, but their blackbox nature makes them unsuitable for clinical application. The PDM methodology gives dramatic visualisations of two modes separating horizontal and vertical facial growth. Kohonen feature maps favour one clinician and PDM the other. Clinical response suggests that while Clinician 1 places greater weight on 5 of 6 parameters, Clinician 2 relies on more parameters that capture facial shape.

Conclusions: While machine learning and statistical analyses classify subjects for vertical facial height, they have limited application in their present form. The supervised learning algorithm C5.0 is effective for generating rules for individual clinicians but its inherent bias invalidates its use for objective classification of facial form for research purposes. On the other hand, promising results from unsupervised strategies (especially the PDM) suggest a potential use for objective classification and further identification and analysis of ambiguous cases. At present, such methodologies may be unsuitable for clinical application because of the invisibility of their underlying processes. Further study is required with additional patient data and a wider group of clinicians.

1. Introduction

Individuals with very long faces and people with very short faces are examples of the two extremes of vertical facial growth. The former extreme, Long Face Syndrome or LFS (1), arises from an elongation of the face as the lower jaw rotates away from the upper jaw during growth as illustrated in Fig. 1a (2). Conversely, in Short Face Syndrome or SFS (3), the lower jaw rotates towards the upper jaw, reducing the vertical length of the face (Fig. 1c). These growth patterns not only affect the shape of the face but also result in malocclusions of varying severity where the teeth fail to meet properly. For example, in LFS only the back teeth of each jaw touch together when the mouth is closed because the lower jaw has grown away from the upper jaw.

Patients with extremes of vertical facial forms often require a combination of orthodontic realignment of the teeth and surgery on the jaws to correct their malocclusion. Different facial types require significantly different orthodontic treatment plans and respond differently to orthodontic treatment. Thus, it is important to correctly identify a subject's vertical facial form before starting treatment. Furthermore, differences in the horizontal growth of the face may confuse the diagnosis of vertical facial form.

In the UK, patients with SFS or LFS are normally referred to a specialist for treatment. Diagnosis and treatment planning are based on examination of the patient together with analysis of appropriate records, including a standardized lateral skull radiograph (cephalogram) designed specifically for use in orthodontics. Angular and linear measurements of skeletal landmarks are obtained from the cephalogram (see Fig. 2a). The clinician assesses facial form by comparing the patient's measurements with tables of relevant means and standard deviations compiled from growth studies grouped by ethnicity, age and sex (see Fig. 2b) (4). These cephalometric measurements, in conjunction with the results of the clinical assessment, act as a guide when the clinician formulates a treatment plan. In cases where the classification group is not immediately apparent, the clinician's perception of facial form, informed by clinical experience, may determine the final classification. Unfortunately, if the classification is erroneous, then orthodontic treatment can exacerbate the original problem. However, there is no internationally agreed upon objective classification of vertical facial form.

The biomechanical relationships between the skeletal components of the jaws and face, the associated musculature and the teeth may be important factors in the origins of vertical facial discrepancy. An important goal for basic science research is to identify specific developmental anomalies that cause extremes of facial growth. However, it is important that objective schema are first determined for the classification of facial form. These should allow individual subjects to be independently and reproducibly classified irrespective of the experience of the clinician. Only then can hypotheses be deduced for the biological processes involved in the origin of facial form.

The analysis of the dataset considered in this study is described in three sections. The method described in the first section was based on a straightforward visualization and provided a simple classificatory model that agreed strongly with one of the two clinicians taking part. The second section describes the use of a supervised learning
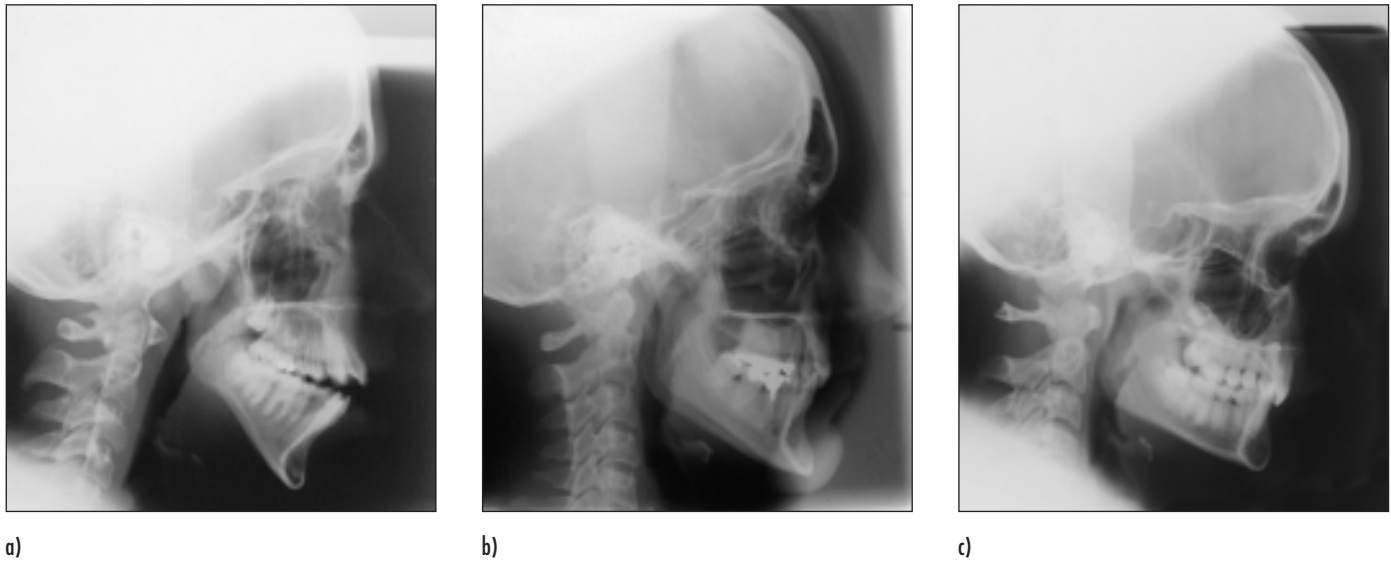
Fig. 1   Lateral radiographs of the skull illustrating a) "long face", b) "normal" and c) "short face" patients

algorithm, C5.0, with each clinician's classification, in turn, acting as a gold standard. The objective was to generate symbolic representations of the induced models for inspection and interpretation by the clinicians to clarify and corroborate the parameters upon which they place greatest emphasis. The third section focuses on two unsupervised algorithms, the first generated a Kohonen network (5) induced from the cephalometric measurements. By way of contrast, the second model was computed using a statistical technique, the Point Distribution Model (6), on point-set templates fitted to the skeletal structures and including landmarks typically used in cephalometric analysis. The aim was to determine the objectivity and utility of the clusterings induced by unsupervised techniques.

2. The Dataset and Methods for Comparing Clinicians and Induced Classifications

Data were collected for each of 131 patients attending a specialist clinic in the Eastman Dental Hospital for an on-going study into the aetiology of vertical facial discrepancy. Patients were randomly selected from medically healthy Caucasian[1] subjects undergoing orthodontic treatment with or without orthognathic surgery. Of the 131 subjects, 63.4% (n = 83) required a combination of orthodontics and surgery whereas 36.6% (n = 48) required orthodontics alone. Apart from age and sex, all measurements were taken from manual tracings of cephalograms. Two specialist clinicians independently classified the facial type of the 131 patients into one of three categories {long, normal, short} by comparing their cephalometric measurements to published means and standard deviations. Some of the available parameters were rejected by the clinicians as irrelevant or not informative for this study. The remaining sixteen
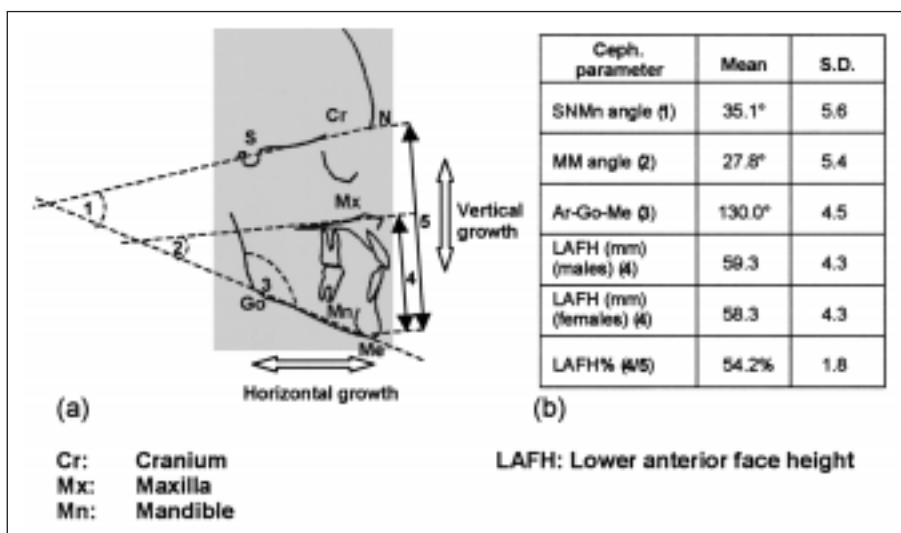

Fig. 2   Tracing of cephalogram (a) showing a selection of cephalometric measurements and (b) the published means and standard deviations (SD) for cephalometric measurements [2]. The letters N, S, Go and Me refer to skeletal landmarks identifiable in the cephalogram

parameters, including each clinician's classification of vertical facial deformity, were used for the supervised and unsupervised learning.

Table 1 indicates a raw classificatory agreement of 84 out of 131 for the two specialist clinicians or, equivalently, a proportional agreement of 0.64. Raw agreement is too simplistic so the kappa statistic (κ), or chance-corrected proportional agreement, is used instead (7). The kappa statistic takes into account the chance occurrence of agreement and also where in the frequency table the agreement occurs (see Appendix a). However, κ treats all disagreements equally, which is not appropriate here since the categories {long, normal, short} can be ordered, and disagreements can be weighted according to their "distance" from the diagonal of the frequency table. For Table 1, the weighted kappa (κ_w) is 0.58, earning an agreement categorisation of "moderately good" on Altman's suggested interpretive scale (7). Throughout the remainder of the paper, the weighted kappa statistic will be used as a convenient means for summarising a comparison of classification schemes. However, as Altman recommends, the frequency table is always provided in addition to support a detailed comparison.

Upon completing the analysis of the models, the clinicians were asked to list the parameters used in their cephalometric diagnosis. They were then asked to comment upon the results presented below and for their view of the suitability of these models for clinical diagnosis.

Table 1 Frequency table and weighted kappa statistics of 131 patients by two specialist clinicians (see Appendix for a full definition of the kappa statistics and an example calculation for the data in Table 1)

Clinician 1 κ_w = 0.58	Clinician 2 long	normal	short	Interpretation of κw values
long	37	1	0	<0.20 = poor agreement
normal	35	24	7	0.21–0.40 = fair agreement
				0.41–0.60 = moderate agreement
short	0	4	23	0.61–0.80 = good agreement
				0.81–1.00 = very good agreement

Table 2 Linear combination (SNMn+17*MM/28) compared with both clinicians

SNMn/ combination	Clinician 1 κ_w = 0.96 long	normal	short	Clinician 2 κ_w = 0.57 long	normal	short
long	37	0	0	31	2	0
normal	1	63	0	33	42	5
short	0	3	27	2	5	23

Table 3 Parameter sets nominated by the two clinicians as being used in their respective classification schemes. The starred parameters were typically used in ambiguous or borderline cases

Clinician 1	Clinician 2
MM	MM
SNMn	SNMn
ArGoMe	ArGoMe
PFH/AFH%	PUDH
LAFH%	LAFH%
	*LAFH (mm)
	*OB

a simple linear combination $SNMn + 17*MM/28$, with the ranges "<44", "44 to 60.27", "> = 60.27" chosen to minimise misclassification and to provide putative

short, normal and *long* classification respectively. For Clinician 2, the plot of *SNMn* against *MM* (Fig. 3) separated the three groups, but less obviously than for Clinician 1. The agreement between each clinician and the classification by the linear combination of *SNMn* and *MM* are shown in Table 2. Thus, there was very good agreement, for this dataset, between a simple combination of two parameters and Clinician 1. For Clinician 2, the agreement was less good. No single variable or linear combination of two variables could be found to separate the groups for Clinician 2 as well as for Clinician 1.

Both clinicians were asked independently to identify the variables used in their cephalometric analysis (see Table 3).
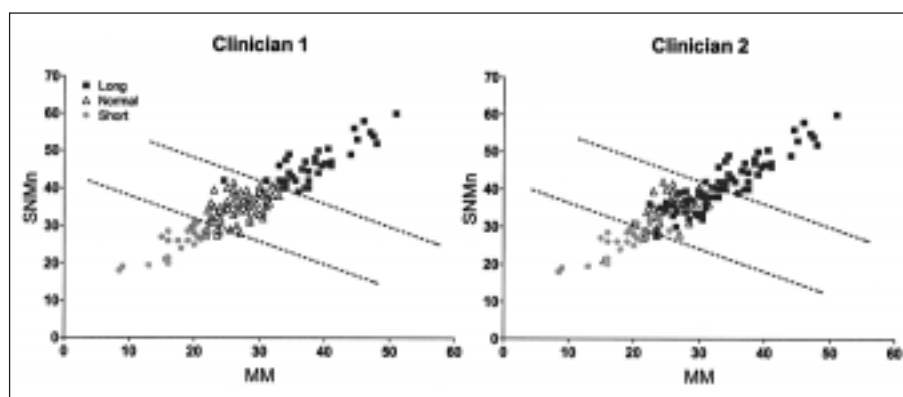
3. Use of Data Visualisation to Identify a Simple Classificatory Model

A visual inspection of a matrix of scatter graphs (not shown) of parameters in the dataset identified two of these parameters (the angles *SNMn* and *MM,* see explanation in Fig. 2) as providing an excellent basis for agreement with Clinician 1. A plot of *SNMn* against *MM* (Fig. 3) overlaid by Clinician 1's classification suggested

Fig. 3 Graph of MM against SNMn overlaid by each clinician's classification and by boundary separators (dashed lines) for the values 44.00 and 60.27 of SNMn + 17*MM/28

They both listed *MM* and *SNMn* and suggested that they were significant factors in the diagnosis of facial form. This may explain the agreement identified between the clinicians and their linear combination as described above. However, the clinicians commented that other parameters were given weight in ambiguous or borderline cases. Thus, the lack of perfect agreement suggests that other parameters may fine-tune the diagnostic process.

4. Supervised Learning

Although it is the overall clinical assessment of an individual patient that determines corrective treatment, cephalometric analysis contributes significantly to the classification of vertical facial deformity, especially for research purposes. However, both clinical and cephalometric assessments are prone to subjectivity (reflecting training and experience) and lead to differences in categorisation between clinicians. Because of this, the aim was to determine a more objective assessment of vertical facial form from the cephalogram. The induction of symbolic models of clinical characterisations of facial form may identify similarities and differences between how cephalometric measurements appear to be used and how clinicians suggest they are used. They may also help to articulate differences between clinicians.

4.1 C5.0 for Inducing Decision Trees

The C5.0 algorithm (8) was used within the CLEMENTINE machine learning environment to induce decision trees with each clinician's classification, in turn, acting as a gold standard. In both cases, the entire dataset was used to induce a decision tree for preliminary discussion with the clinician. The algorithm was executed with various penalty weightings for incorrect classifications of normal as short or long (and vice versa) and short as long (and vice versa). A penalty weighting of 2 produced the best performing trees. A 10-fold cross validation, with the same penalty-weighting scheme, was also performed to induce 10 additional
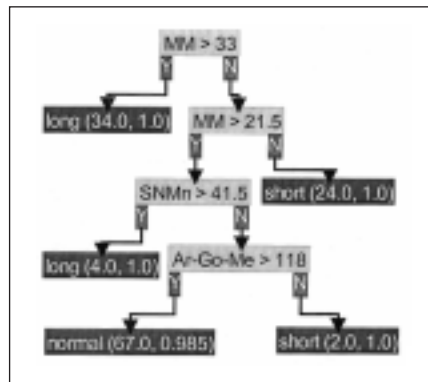

Fig. 4 Decision-tree induced by C5.0 from the entire dataset with clinician 1 as gold standard. The pair of numbers in brackets before each arrow gives the number of patients satisfying the conditions in that branch and, of those, the proportion of agreement between the clinician's and the induced classifications

Table 4 Frequency table for the decision tree (Fig. 4) induced with C5.0 using the entire dataset and Clinician 1 as gold standard

C5.0 (entire data)	Clinician 1		
	long	normal	short
long	38	0	0
normal	0	66	1
short	0	0	26

decision trees for comparison and for further interpretation. Mean weighted kappa values were computed for both the training and testing performance of the cross validation trees.

With Clinician 1 acting as gold standard, a decision tree was induced from the entire dataset. The tree and its frequency table are shown in Fig. 4 and Table 4 respectively. Given the simple model derived from visualization alone, one would expect to induce a simple decision tree based largely on the two parameters *MM* and *SNMn*. Furthermore, the critical values for *MM* and *SNMn* (33 and 41.5 respectively) are close to the means ± 1 SD for these parameters (see Fig. 2). These are the values that are in general clinical use. Of the 10 cross validation trees induced, seven were isomorphic apart from only minor variation to the tree induced from the entire dataset. The mean weighted kappa value for the cross-validation for testing was 0.89. Clinician 1 accepted the common structure to 8 of the 11 trees induced despite the apparent redundancy of 2 of the 5 parameters listed in Table 3.

With Clinician 2 acting as gold standard, the same scheme of supervised learning was completed. The decision tree induced from the entire dataset and its corresponding frequency table are shown in Fig. 5 and Table 5 respectively. Clinician 2 accepted this decision tree as a reasonable reflection of the classification scheme. However,
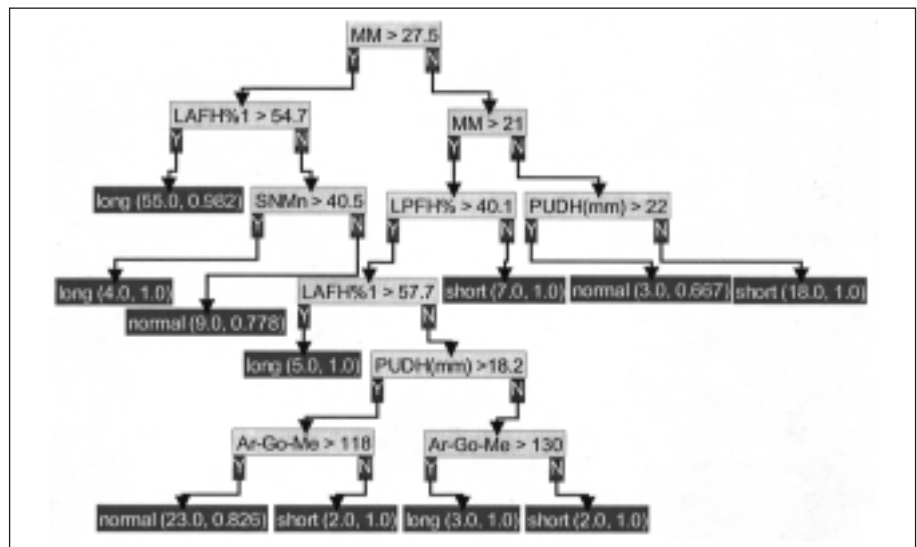

Fig. 5 Decision-tree induced by C5.0 from the entire dataset with clinician 2 as gold standard. The pair of numbers in brackets before each arrow gives the number of patients satisfying the conditions in that branch and, of those, the proportion of agreement between the clinician's and the induced classifications

one parameter *(LPFH%)* appearing in the decision tree is certainly not one of those listed in Table 3 by Clinician 2. The mean weighted kappa value for the cross-validation for testing was 0.78. Common structures in the corresponding decision trees were infrequent. Thus, while the cross validation produced good weighted kappa statistics, there was inconclusive support for the validity of the decision tree induced from the entire dataset.

In terms of potential clinical use, both clinicians preferred an equivalent, rule-based representation of these decision trees. Because of space limitations, only the more compact decision-tree format is given. The subdivided graph of Fig. 3 would be more easily applied in a clinical environment. However, this may reflect the fact that two parameters happen to capture Clinician 1's reasoning so well.

5. Unsupervised Learning

5.1 Kohonen Networks

Kohonen networks of various shapes and sizes were induced [see (5) for a detailed description of the underlying algorithm]. Typically, in the 2D plot of clusters provided by the Clementine environment, there were always large clusters leaving smaller ones with only a few members, and often of mixed class. This makes their interpretation difficult and suggested that the net topology be kept quite small. Indeed, 3×3 topologies performed best.

By overlaying the two-dimensional output from the Kohonen modelling with the classification scheme of each clinician, agreement between clinician and clustering can be visualised. Fig. 6 suggests that the particular Kohonen net considered (and this was the case for all those generated) is in closer agreement with Clinician 1 than with Clinician 2. Furthermore, the discrepancy between the classification of *long* and *normal* by both clinicians is again highlighted [see cluster (2,0) in Fig. 6]. The overlaying of a clinician's classification partitions each cluster into subgroups of *small*, *normal* and *long* cases. The standard entropy calculation for a cluster partition

Table 5 Frequency table for the decision tree (Fig. 5) induced with C5.0 using the entire dataset and Clinician 2 as gold standard

C5.0 (entire data)	Clinician 2		
	long	normal	short
long	66	1	0
normal	6	28	1
short	0	0	29

$(-\sum p_c \log_2 p_c$, where p_c is the proportion of class c in a cluster and the summation is over $c \in \{small, normal, long\}$) can be added to give an overall measure of ambiguity of the match between the clusters and the clinician's classification. For a cluster comprising a single class, the entropy is 0. For the two sets of clusters in Fig. 6, the sum of the entropies of the constituent clusters is 4.85 and 7.05 for Clinician 1 and Clinician 2 respectively. Thus the visually observed bias towards Clinician 1 is endorsed by this informal use of the entropy measure.

5.2 Point Distribution Model (PDM)

The previous unsupervised learning method used the cephalometric angles and ratios to induce a clustering of the 131 examples. The approach described here used

instead the raw tracing of each cephalogram. A shape template of 148 landmark points was manually placed on each image, as in Fig. 7a. The template included points spaced along important structures, such as the mandible, as well as the standard cephalometric landmarks. Thus the data here implicitly contained the angles and distances presented as parameters to the C5.0 and Kohonen algorithms. The connectivity between template points is included solely for display purposes.

The Procrustes algorithm (9) was used to align the templates, giving a mean. Principal Components Analysis (PCA) was then applied to derive the major modes of deformation (see Appendix b). Together, the mean and the modes formed the PDM. The first three modes accounted for 64% of the total variation in shape seen across the examples (Fig. 7b). The first mode had captured the horizontal variation and accounted for 37%. The second mode showed change in vertical form and accounted for 19% of the total variation. The third mode of 9% showed variation in the position of the molars and incisors. It must be emphasized that these deformation modes were computed directly from the set of templates and that no other information was given.
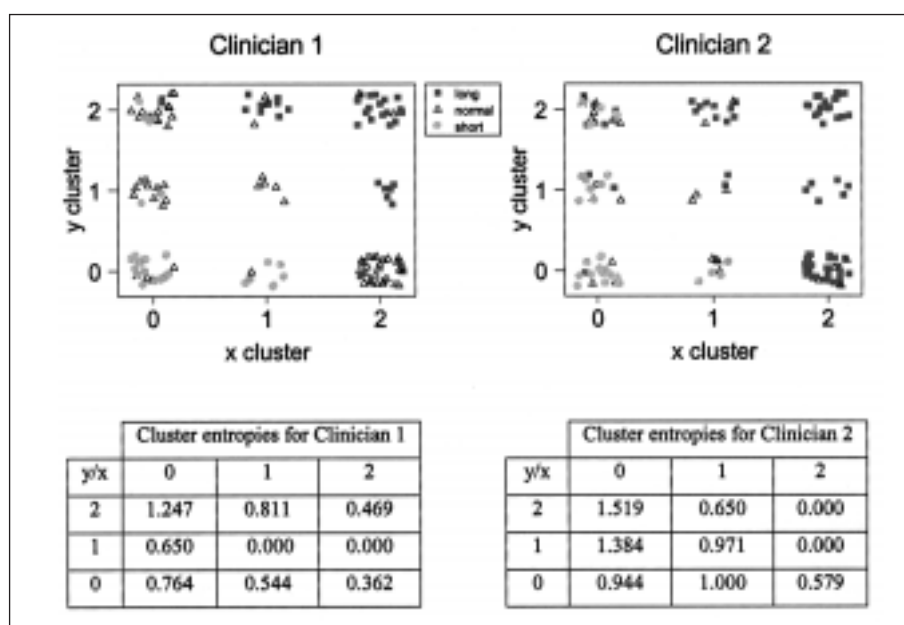

Fig. 6 2D clustering for an induced 3×3 Kohonen net overlaid with each clinician's classification (*short, normal, long*). The entropies for each individual cluster are tabulated below the plots

PDM Mode 2 $\kappa_w = 0.62$	Clinician 1 long	normal	short
long	24	8	0
normal	14	54	9
short	0	4	18

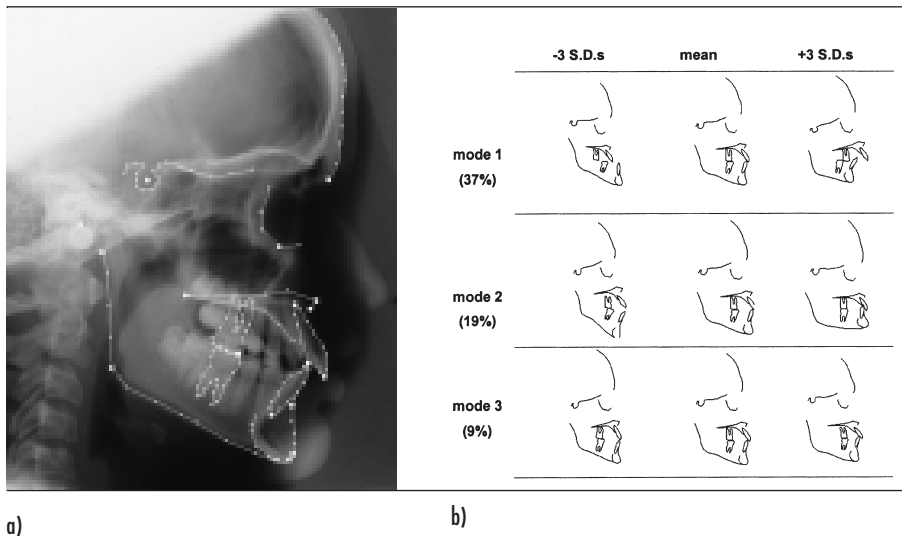| PDM Mode 2 $\kappa_w = 0.70$	Clinician 2 long	normal	short
long	62	11	0
normal	7	15	6
short	3	3	24

Fig. 7 Cephalogram with overlaid template a) and examples of the three major modes of variation showing the means and their extremes b). The figures in parentheses represent the percentages of the total variation accounted for by each of the modes

For this study, it was the second mode that was of primary interest. Each cephalogram could be classified by correlating it with this deformation mode, yielding a score that typically ranged between –3 and +3 standard deviations from the mean. If the classification of a clinician was plotted for each example against this value, it could be seen that there was some agreement (see Fig. 8). Indeed, for negative values of the second mode parameter there was more frequent concurrence with Clinician 2's "long" classification than with that of Clinician 1. The PDM methodology gave a dramatic visualisation of two modes separating horizontal and vertical facial growth. It is interesting that it is the facial shape in mode 1 at -3 SD (Fig. 7b) that causes the greatest disagreement between Clinician 1 and Clinician 2. Clinician 1 would classify this facial shape as having "normal" vertical form with an horizontal problem that generates a large chin and secondarily increases the length of the face. Clinician 2 would include this particular facial shape in the "long" category. A classification based on values of mode 2 could be constructed using each clinician's classification scheme but required specific cut-off values to separate the three classes. For each clinician, a pair of cut-off values of mode 2 were chosen to minimise the number of inconsistent classifications. The contingency tables and kappa statistics for the two comparisons are shown in tables 6 and 7. The slight bias towards one clinician, this time Clinician 2 (as seen in the histograms in Fig. 8), was confirmed by the weighted kappa values.
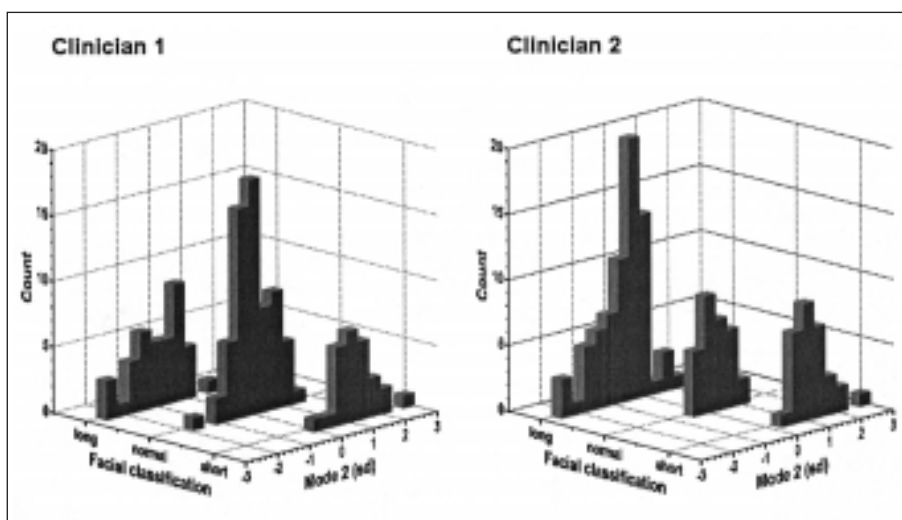
6. Discussion

This study has considered how machine learning techniques can contribute to the long-term goal of identifying an objective process for the classification of vertical facial discrepancy that is suitable for both research and clinical use. Although supervised learning algorithms, such as C5.0, are necessarily subject to a gold standard against which models are induced, the decision trees generated do capture a

Fig. 8 Frequency histograms showing the relationship between mode 2 derived from the PDM model and the facial classifications determined by the two clinicians. The scale for Mode 2 is expressed in standard deviations (sd) from the mean of Mode 2

symbolic model of each clinician's classification scheme. The induced trees, or rules (as were preferred by both clinicians), enable a direct comparison of idiosyncrasies in classification. Although unsupervised algorithms induce models that are potentially more objective, their blackbox nature does not inspire confidence in their clinical application. Moreover, unless the induced clusterings are sharply differentiating, it is still necessary to make some subjective interpretation of the output.

The PDM methodology gives a dramatic visualisation of two modes separating horizontal and vertical facial growth. Indeed, the PDM approach has highlighted the fact that horizontal growth may be a confounding factor when classifying vertical facial discrepancy. The orthodontic significance of this is beyond the remit of this paper, but it does reveal the power of these techniques as a research tool. Why Kohonen appears to favour one clinician and PDM the other is still a matter of speculation. Importantly, the response of the clinicians suggests that while Clinician 1 places greater weight on 5 of 6 parameters, Clinician 2 may rely on a greater number of parameters that in some way "capture" facial shape during the diagnostic process. Such processes are mirrored in the induced models where the data entry into the two models differed. The Kohonen model worked from a rather limited number of cephalometric values, while the PDM model was able to deal with many more reference points in close proximity to one another. This use of fewer variables in the Kohonen model may have led to some bias in the interpretation of lower facial growth. However, the limited numbers of clinical opinions tested in this way preclude any firm interpretation of this data.

7. Conclusions and Further Work

While machine learning and statistical analyses can classify subjects for vertical facial height, they have limited application in their present form. The supervised learning algorithm C5.0 has proved effective for generating rules for individual clinicians but its inherent bias invalidates its use for objective classification of vertical facial form for research purposes. On the other hand, promising results from the unsupervised strategies (especially the PDM) suggest a potential use for objective classification of form and further identification and analysis of ambiguous cases. At present, such methodologies may be unsuitable for clinical applications because of the invisibility of their underlying processes. Further study is required with additional patient data and a wider group of clinicians.

Appendices

(a) Chance corrected proportional kappa statistic

Agreement between observations is quantified as the chance-corrected proportional or *kappa statistic* (κ). In a square frequency table $[f_{ij}]$ of n observations and g categories, $\kappa = (p_0 - p_e)/(1 - p_e)$ where p_0, the observed proportional agreement, is $\sum f_{ii}/n$ and p_e, the expected index of agreement by chance, is $\sum r_i c_i/n^2$ (r_i and c_i are the row and column totals for the ith category). Using the weighting scheme $w_{ij} = 1 - |i - j|/(g-1)$, the *weighted kappa statistic*, κ_w, is defined as $(p_0(w) - p_e(w))/(1 - p_e(w))$ where $p_0(w) = \sum \sum w_{ij}f_{ij}/n$ and $p_e(w) = \sum \sum w_{ij}r_ic_j/n^2$.

Example calculation of κ and κ_w (weighted kappa) for data in Table 1

$$p_0 = (37+24+23)/131$$
$$p_e = (38*72 + 66*29 + 27*30)/131^2$$
$$\kappa = 0.47$$

$$p_0(w) = (1*37+0.5*1+0*0+0.5*35+1*24+0.5*7+0*0+0.5*4+1*23)/131$$
$$p_e(w) = [(1*38+0.5*66+0*27)*72+(0.5*38+1*66+0.5*27)*29+(0*38+0.5*66+1*27)*30]/131^2$$
$$\kappa_w = 0.58$$

(b) Procrustes Alignment and Principal Components Analysis

The Procrustes algorithm aligns the set of templates together in a least-squares sense:
1. align (rotate, scale and translate) all the templates with the first template.
2. compute the mean of the set of templates.
3. align the mean template with the first template.
4. align all the templates to this mean.
5. repeat from step 2 until no further change is observed.

After alignment, the difference between each shape and the mean is used to analyse the variation in shape that is seen across the training set. Each template has 148 points, the x and y co-ordinates of these are concatenated into a vector of length 296: $x_1\ y_2\ x_2\ y_2 \ldots x_{148}\ y_{148}$. The shape vector for the mean shape is subtracted from each template in the training set, leaving a set of residual vectors. The covariance matrix of this set of residuals is then computed by summing the outer product of these vectors with themselves: $C = \sum(x_i-x_{mean})(x_i-x_{mean})^T$. Karhunen-Loeve decomposition is then performed on the covariance matrix (which is real and symmetric), yielding the eigenvectors and eigenvalues. The eigenvectors are sorted by the magnitude of their corresponding eigenvalues. The size of an eigenvalue as a proportion of the sum of all the eigenvalues determines the percentage of variation associated with the eigenvector ($p_k = \lambda k*100/\sum \lambda_i$).

The eigenvectors can be added to the shape vector for the mean template in varying proportions to create new shapes. Similarly, any given shape can be correlated against the various eigenvectors to yield a parameterisation in terms of the weighting required to synthesise that particular shape.

(c) Kohonen feature maps (following description in [5])

The Kohonen self-organising algorithm transforms input patterns of arbitrary dimension into a two dimensional map preserving topological relationships. Each training pattern is presented to a neural network in succession, giving rise to a localized region of high activity in the feature map compared to lower levels of activity

elsewhere. Once trained, the presentation of a test pattern should stimulate a localized group of network nodes giving a clustering of similar patterns.

(d) C5.0 decision tree generation

Quinlan's C5.0 algorithm constructs decision trees top-down by recursively choosing the next attribute to test as that best classifying the training examples handed down to the current node in the tree. The evaluation of each attribute is performed using a statistical measure known as *information gain*.

References

1. Schendel SA, Eisenfeld J, Bell WH, Epker BN, Mishelevich DJ. The Long Face Syndrome: vertical maxillary excess. Am J Orthod 1976; (70): 398-408.
2. Bjork A. Prediction of mandibular growth rotation. Am J Orthod 1969; 55: 585-99.
3. Opdebeeck H, Bell WH. The short face syndrome. Am J Orthod 1978; 73: 499-511.
4. Bhatia SN, Leighton BC. A manual of facial growth. A computer analysis of longitudinal cephalometric growth data. Oxford: Oxford University Press; 1993.
5. Tarassenko L. A guide to neural computing applications. London: Arnold 1998.
6. Cootes JF, Taylor CJ, Cooper DH, Graham J. Active shape models – their training and application. Comput Vis Image Underst 1995; 61 (1): 38-59.
7. Altman D. Practical statistics for medical research. London: Chapman & Hall; 1991.
8. Quinlan JR. C4.5: Programs for machine learning. San Mateo, CA: Morgan Kaufman 1993.
9. Goodall C. Procrustes methods in the statistical analysis of shape. J R Stat Soc 1991; B 53 (2): 285-339.

Correspondence to:
Professor Peter Hammond
Eastman Dental Institute for Oral Health Care Sciences
University College London
256 Gray's Inn Road, London WC1X 8LD
United Kingdom
E-mail: p.hammond@eastman.ucl.ac.uk