OPEN ACCESS Freely available online

PLOS ONE

Dental Informatics to Characterize Patients with Dentofacial Deformities

Seoung Bum Kim[1☯], Jung Woo Lee[2☯], Sin Young Kim[1], Deok Won Lee[2*]

1 School of Industrial Management Engineering, Korea University, Anam-Dong, Seongbuk-Gu, Seoul, Republic of Korea, **2** Department of Oral and Maxillofacial Surgery, Kyung Hee University, Sangil-Dong, Gandong-Gu, Seoul, Republic of Korea

Abstract

Relevant statistical modeling and analysis of dental data can improve diagnostic and treatment procedures. The purpose of this study is to demonstrate the use of various data mining algorithms to characterize patients with dentofacial deformities. A total of 72 patients with skeletal malocclusions who had completed orthodontic and orthognathic surgical treatments were examined. Each patient was characterized by 22 measurements related to dentofacial deformities. Clustering analysis and visualization grouped the patients into three different patterns of dentofacial deformities. A feature selection approach based on a false discovery rate was used to identify a subset of 22 measurements important in categorizing these three clusters. Finally, classification was performed to evaluate the quality of the measurements selected by the feature selection approach. The results showed that feature selection improved classification accuracy while simultaneously determining which measurements were relevant.

Citation: Kim SB1, Lee JW2, Kim SY, Lee DW (2013) Dental Informatics to Characterize Patients with Dentofacial Deformities. PLoS ONE 8(8): e67862. doi:10.1371/journal.pone.0067862

Editor: Dongxiao Zhu, Wayne State University, United States of America

Received November 4, 2012; **Accepted** May 26, 2013; **Published** August 5, 2013

Funding: This work was supported by Brain Korea 21 (Network Enterprise). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: verycutebear@khu.ac.kr

☯ These authors contributed equally to this work.

Introduction

Dental health is one of the most important factors in our lives. Although the advent of high information technology and dental devices has produced vast amounts of data, relatively little research has been conducted to retrieve meaningful information from dental data. However, this has been changing with the development of informatics that allows acquisition of relevant information to guide dental treatment increasingly becomes an important scientific discipline [1].

Among the various procedures that lend themselves to such data mining, orthodontic treatment of malocclusion patients to correct the position of teeth and improve appearance is well suited to use these techniques. Various analysis and simulators have been used to help dentists properly diagnosis and predict the outcome of intervention before actual treatment. Downs introduced Downs' analysis, the first systematized analytic diagnostic procedure for the roentgenographic assessment of craniofacial, skeletal, and dental patterns [2]. Down's analysis has been used by many orthodontists and by oral and maxillofacial surgeons. Based on the location of anatomical landmarks, various lengths and angles can be measured and compared with normal ranges [3,4]. However, the most commonly used analysis is the Steiner analysis that can provide guidelines for planning of treatment based on the prediction of changes that will occur as the result of growth and orthodontic therapy [5]. The Sassouni Cephalometric Analysis has been also beneficial to dentists in functional orthodontic treatment of TMD (temporomandibular disorders) patients [6,7]. This analysis is especially useful for determining the growth potential of these patients and in determining vertical proportions [8,9]. Wits analysis for the diagnosis of anteroposterior discrepancy was first described by [10]. McNamara's Analysis combines the anterior reference plane (a plane perpendicular to the Frankfort horizontal through the nasion) described by Burstone et al. [11,12]. McNamara's analysis is suitable to diagnosis, treatment planning, and treatment evaluation for not only conventional orthodontic patients, but also for patients with dentofacial deformities [13].

Although all of the a fore mentioned analyses, based mostly on simple skeletal analysis, can be useful in situations for which they were designed, prediction of postoperative outcomes nevertheless remains difficult. Despite the great potential of data mining algorithms for addressing a variety of problems in dental treatments, few efforts have been made to apply these techniques. Raberin et al. used a *k*-means clustering method with 278 dental casts of untreated French adults with normal occlusions to determine the main mandibular dental arch forms [14]. Similarly, Lee et al. used the same methodology with dental casts of 307 Korean subjects with normal occlusion to establish normative data on tooth size [15]. Hwang et al. employed a *k*-means clustering analysis to group 100 patients with facial asymmetry into five groups with different characteristics [16]. De Veld et al. detected oral cancer by applying a *k*-means clustering analysis and principal component analysis to the spectra obtained from autofluorescence spectroscopy [17].

The main purpose of the present study is to use data mining algorithms to characterize patients with dentofacial deformities. More precisely, we used a *k*-means clustering algorithm and

principal component analysis to detect meaningful groups based on a number of measurements related to dentofacial deformities. Further, we used the features selection algorithm to identify which of these measurements are most important in distinguishing between the different clusters. Finally, we verified the quality of the measurements identified by the feature selection algorithm.

Data

The procedures followed were in accordance with the ethical standards and approval of the Kyunghee University Institutional Review Board (KHNMC IRB 2012–089). The participants provided their written consent to participate in this study. A total of 72 patients with skeletal malocclusions who had finished the orthodontic and orthognathic surgical treatments were enrolled for data acquisition in this study. All patients had various dentofacial deformities that required single or double jaw orthognathic correction. These deformities included maxillary horizontal hypoplasia, maxillary horizontal hyperplasia, maxillary vertical hypoplasia, maxillary vertical hyperplasia, mandibular hypoplasia, mandibular hyperplasia, and facial asymmetry. A digital panoramic and cephalometric system (Eastman Kodak Co., Rochester, New York, USA) was used to obtain various landmarks and planes that characterized the size and relationships of the teeth, jaws, and cranium. Figure 1 shows landmark points and planes that generate 22 measurements related to dentofacial deformities.

These 22 measurements can be summarized as follows:

1. **SN to FH**: An angle between the sella-nasion (SN) line and the Frankfort horizontal (FH). The SN is a line connecting the sella to the nasion. The FH is a horizontal line connecting the cephalometric porion and orbital landmarks.
2. **SN to PP**: An angle between the SN line and palatal plane (PP). PP is a line joining the posterior nasal spine and anterior nasal spline.
3. **SN to mandibular**: An angle between the SN line and mandibular plane (MP). The MP is a line/plane connecting the gonion and menton, representing the inferior border of the mandible in the sagittal plane. The mandibular plane may also be drawn as a tangent to the interior border of the mandible.
4. **FH to occlusal**: An angle between the FH and occlusal planes (OP). The OP is a line on the cephalometric radiograph representing an imaginary plane at the level of the dental occlusion.
5. **FH to mandibular**: An angle between the FH and mandibular planes (MP). The FH is a horizontal line connecting the cephalometric porion and orbital landmarks. MP is a line/plane connecting the gonion and menton, representing the inferior border of the mandible in the sagittal plane. The mandibular plane may also be drawn as a tangent to the interior border of the mandible.
6. **SNA**: An angle made up of three points: sella, nasion and point A. Point A (or ss, subspinale) is the point at the deepest midline concavity on the maxilla between the anterior nasal spine and prosthion.
7. **FH to NA**: An angle between the FH plane and the NA line.
8. **Convexity**: A distance from point A to the N-Pog line. The N-Pog line, also called the facial plane, is a line connecting the nasion and the pogonion.
9. **SNB**: An angle composed of three points: sella, nasion, and point B. Point B is the point at the deepest midline concavity on the mandibular symphysis between the infradentale and the pogonion (unilateral).
10. **SNPog**: An angle composed of three points: sella, nasion, and pogonion. The pogonion is the point of tangency of a perpendicular from the mandibular plane to the most prominent convexity of the mandibular symphysis.
11. **FH to NB**: An angle between the FH plane and the NB line. The NB is a line connecting the nasion and point B.
12. **Facial angle**: An inferior inside angle between the FH plane and the N-Pog line.
13. **Y axis**: An acute angle between the FH plane and the S-gnathion line.

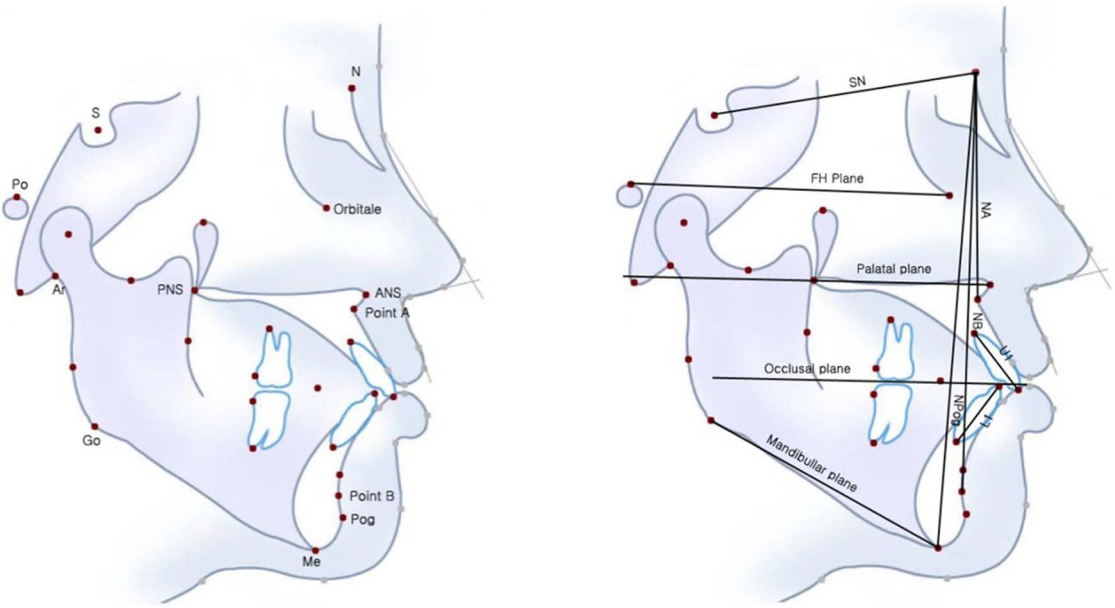

Figure 1. Landmark points (left figure) and planes (right figure) of lateral cephalometric radiograph.
doi:10.1371/journal.pone.0067862.g001

Table 1. Results of the Rand index and adjusted Rand index methods to determine *k*.

k	2	3	4	5
Rand index	0.92	0.92	0.74	0.73
Adjusted Rand index	0.80	0.82	0.34	0.26

doi:10.1371/journal.pone.0067862.t001

14. **Gonial angle**: An angle between the mandibular plane and the posterior border of the mandible body.
15. **ANB difference**: The ANB angle ($3\pm2^{\circ}$) measures the relative position of the maxilla to the mandibleThe ANB angle can be measured or calculated by the formula: ANB = SNA – SNB.
16. **Palatal to mandibular**: An angle between the palatal and mandibular planes.
17. **FH to U1**: The angle between the FH plane and the U1; the U1 is a line connecting the incisal edge and the root apex of the most prominent maxillary incisor.
18. **FH to L1**: The angle between the FH plane and L1; L1 is a line connecting the incisal edge and the root apex of the most prominent lower incisor.
19. **Interincisal angle**: The angle between U1 and L1.
20. **Mandibular to L1**: The angle between the mandibular plane and L1.
21. **NP to U1**: Distance from upper the incisal edge to the N-Pog line.
22. **NP to L1**: Distance from the lower incisal edge to the N-Pog line.

Methods

K-means clustering algorithm

We performed a clustering analysis to group 72 patients with facial deformities into several groups according to specific characteristics. Clustering analysis partitions the data by minimizing within-group variation and maximizing between-group variation [18]. These variations can be measured by various distance metrics between observations in a dataset.

In the present study we used a *k*-means clustering algorithm mainly because it is the most well-known clustering method and has been used in various applications including previous dental studies [14,15,16,17]. Our procedure requires a brief summary of the *k*-means clustering algorithm. Given k seed points, each observation is assigned to one of the k seed points near the observation. This creates k clusters. Next, the seed points are replaced with the mean of the currently assigned clusters. This procedure is repeated with updated seed points until the assignments do not change. The results of the *k*-means clustering algorithm depend upon three parameters: distance metrics, the number of clusters (k), and the location of seed points.

Numerous distance metrics are available. These include the Euclidian, Manhattan, Mahalanobis, and correlation distance

Table 2. Basic statistics of features (measurements) in each cluster (n = the number of patients).

Feature (Measurement)	Cluster 1(n=17)	Cluster 2(n=30)	Cluster 3(n=25)
	Mean±SD		
1 S-N to FH	9.3±3.0	8.8±3.3	8.8±2.9
2 S-N to palatal	8.1±2.9	8.6±3.3	9.6±3.1
3 S-N to mandibular	34.7±4.4	39.7±4.9	38.7±3.7
4 FH to occlusal	6.7±4.2	9.7±3.6	11.8±3.0
5 FH to mandibular	25.4±5.0	30.9±4.6	30.0±4.1
6 SNA	83.1±2.3	79.4±3.2	80.4±2.5
7 FH to NA	92.4±2.4	88.2±2.5	89.1±1.9
8 Convexity	−0.3±8.0	−4.7±6.5	8.8±4.0
9 SNB	83.2±2.5	81.3±3.2	76.0±2.6
10 SN Pog	83.2±2.7	81.7±3.2	76.4±2.6
11 FH to NB	92.5±3.1	90.0±2.7	84.8±2.0
12 Facial angle	92.5±3.4	90.5±2.9	85.1±1.9
13 Y axis	59.3±3.4	61.5±2.9	65.4±2.5
14 Gonial angle	126.5±8.1	131.5±6.4	124.2±7.1
15 ANB difference	−0.1±3.3	−1.8±2.7	4.4±1.7
16 Palatal to mandibular	26.6±5.2	31.1±4.7	29.2±4.1
17 FH to U1	122.8±9.1	116.6±7.2	113.7±9.8
18 FH to L1	59.2±6.4	63.8±6.8	52.1±4.8
19 Interincisal angle	116.4±8.7	127.2±8.7	118.4±11.6
20 Mandibular to L1	95.4±5.6	85.2±5.8	97.9±4.9
21 NP to U1 (mm)	9.1±4.3	4.8±2.5	12.1±3.1
22 NP to L1 (mm)	9.3±3.8	5.9±3.4	7.3±2.5

doi:10.1371/journal.pone.0067862.t002

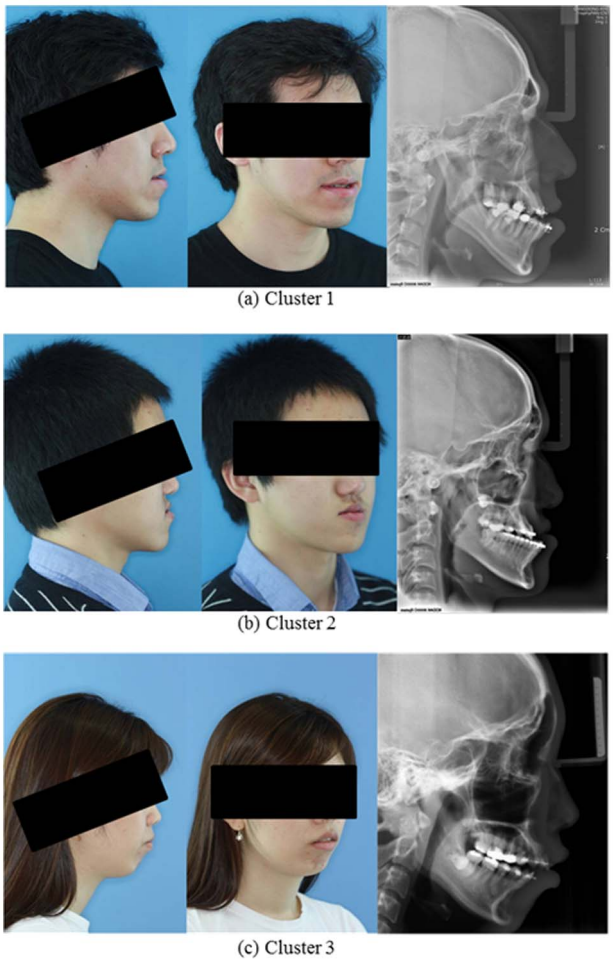

Figure 2. The photos and X-ray images representing three clusters identified by a k-means clustering algorithm.
doi:10.1371/journal.pone.0067862.g002

metrics. In the present study we relied on the widely used Euclidean distance metric. Several methods have been developed to determine the appropriate number of clusters. These include silhouette plot, gap statistics, Rand index, and adjusted Rand index methods [19,20,21,22]. However, no consensus exists about which of them best satisfies all conditions. We used the Rand index and adjusted Rand index methods to determine the number of clusters. With an appropriate number k, the clustering algorithm that reproduces consistent clustering results would be considered the better one. The Rand index and adjusted Rand index measure the stability (i.e., consistency) of cluster results [23]. To calculate cluster stability with the Rand index and adjusted Rand index, we divided the data into three datasets. With two datasets, we conducted k-means clustering and got two sets of seed points. If k is optimal, these two sets of seed points must be similar. This means two sets of seed points with the same data should produce similar results. At this point, we have two different sets of seed points. We then split the remaining third dataset into k with these seed points. Finally, we used the Rand index and the adjusted Rand index to calculate cluster stability. Note that the results of both the Rand index and the adjusted Rand index lie between 0 and 1. When a cluster algorithm reproduces the same clustering results, both the Rand index and the adjusted Rand index will converge to 1 because they consider the probability of chance as the determinant

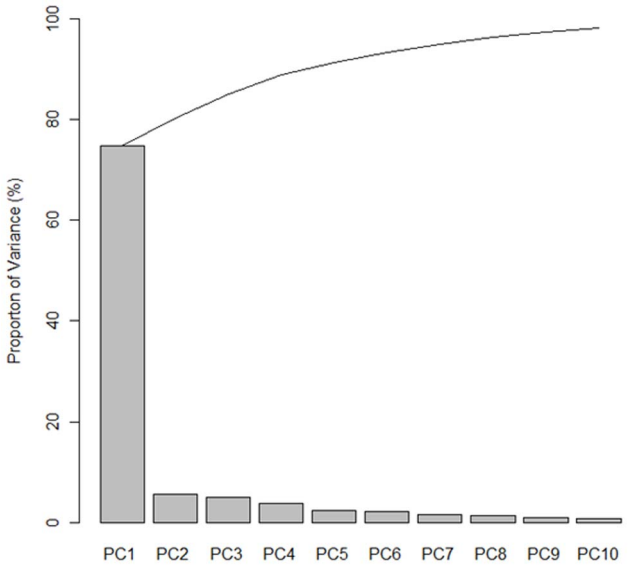

Figure 3. A scree plot to determine the number of PCs.
doi:10.1371/journal.pone.0067862.g003

of which cluster results are consistent [22]. As for determining the location of seed points, we used a random selection approach available in R software (www.r-project.org). In this study we used the "kmeans", "randIndex", and "adjustedRandIndex" functions in R software to implement the k-means clustering, Rand index, and adjusted Rand index algorithms, respectively.

Principal component analysis

Principal component analysis (PCA) is one of the mostly widely used multivariate statistical methods for dimensionality reduction and visualization of high dimensional data [24]. PCA reduces the dimensionality of a dataset by linear combination of the original features, called principal components (PCs). Extracted PCs are uncorrelated with each other, and typically the first few PCs are sufficient to represent most of the variability in the high-dimensional original data [25,26]. Thus, the PCA plot of observations using these first few PC axes facilitates the visualization of high-dimensional datasets. These PCs can be represented by a linear of combination of the original features ($\mathbf{X_1}$, $\mathbf{X_2}$, …, $\mathbf{X_p}$)

$$\begin{aligned}\mathbf{PC_1} &= a_{11}\mathbf{X_1} + a_{12}\mathbf{X_2} + \cdots a_{1p}\mathbf{X_p}\\ \mathbf{PC}_2 &= a_{21}\mathbf{X_1} + a_{22}\mathbf{X_2} + \cdots a_{2p}\mathbf{X_p}\\ &\vdots\\ \mathbf{PC}_p &= a_{n1}\mathbf{X_1} + a_{n2}\mathbf{X_2} + \cdots a_{np}\mathbf{X_p}\end{aligned} \quad (1)$$

The coefficients of each PC, called loading value, can be calculated by eigenvector decomposition of the covariance (or correlation) matrix of the original data. For example, the loading values of the first PC (a_{11}, a_{12}, …, a_{1p}) are the components of the eigenvector that corresponds to the largest eigenvalue of the covariance (or correlation) matrix. Determination of the appropriate number of PCs to retain can be subjective. Typically, a scree plot that exhibits the proportion of variance caused by each PC can be used. In a scree plot, the number of PCs to retain can be identified at an elbow point at which the proportion of variation

begins to stabilize [26]. We used the "princomp" function in R software (www.r-project.org) to generate the PCA results.

A multiple hypothesis testing procedure controlling the false discovery rate

We employed a multiple hypothesis testing procedure that controls the false discovery rate (FDR) to identify the subset of features important to distinguishing the different clusters from each other. The FDR procedure has been used to identify the significant features in high-dimensional data such as microarray, mass spectra, nuclear magnetic resonance spectra, and pairwise amino acids [27,28,29,30]. First we begin with the definition of FDR, followed by the FDR procedure for feature selection. An FDR, a useful measure of the error rate in a multiple hypothesis test, is defined as the expected proportion of false positives among the all hypotheses rejected [31].

To apply FDR for feature selection, we first construct a hypothesis for each feature. More precisely, a null hypothesis, stating that the average value of the feature is equal between k different clusters, is established for each feature, and these hypotheses are tested simultaneously. In our study, we can construct the following multiple hypotheses for 22 features:

$$\begin{aligned} &H_{01} \;:\; \mu_{11} = \mu_{12} = \cdots = \mu_{1k} \quad \text{vs} \quad H_{A1} : \mu_{1i} \neq \mu_{1j} \\ &\text{for some } i \text{ and } j \\ &H_{02} \;:\; \mu_{21} = \mu_{22} = \cdots = \mu_{2k} \quad \text{vs} \quad H_{A2} : \mu_{2i} \neq \mu_{2j} \\ &\text{for some } i \text{ and } j \\ &\qquad\qquad \vdots \\ &H_{022} : \mu_{221} = \mu_{222} = \cdots = \mu_{22k} \quad \text{vs} \quad H_{A22} : \mu_{22i} \neq \mu_{22j} \\ &\text{for some } i \text{ and } j, \end{aligned} \tag{2}$$

where k is the number of clusters. Assuming that the data follow a normal distribution, we can employ an F-test for each feature by using the following test statistic:

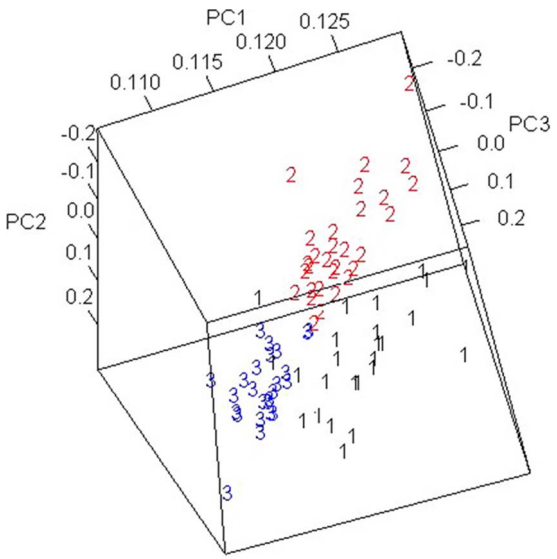

Figure 4. Three-dimensional PCA score plot of 72 patients with facial deformities.
doi:10.1371/journal.pone.0067862.g004

Table 3. Selected features by FDR procedures for $\alpha = 0.01$ and 0.05.

$\alpha = 0.01$		$\alpha = 0.05$	
Feature	***p*-value**	**Feature**	***p*-value**
FH to occlusal	0.000	FH to occlusal	0.000
Convexity	0.000	Convexity	0.000
SNB	0.000	SNB	0.000
SN Pog	0.000	SN Pog	0.000
FH to NB	0.000	FH to NB	0.000
Facial angle	0.000	Facial angle	0.000
Y axis	0.000	Y axis	0.000
ANB difference	0.000	ANB difference	0.000
FH to L1	0.000	FH to L1	0.000
		FH to NA	0.001
		FH to U1	0.002

doi:10.1371/journal.pone.0067862.t003

$$F_p = \frac{\dfrac{\sum_{i=1}^{k} n_i(\bar{x}_{i.} - \bar{x}_{..})^2}{k-1}}{\dfrac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{..})^2 - \sum_{i=1}^{k} n_i(\bar{x}_{i.} - \bar{x}_{..})^2}{N-k}}, \tag{3}$$

for $p = 1, 2, \ldots, 22$. $\bar{x}_{i.}$ and n_i are, respectively, the sample mean and the sample size of the ith cluster of the pth feature. x_{ij} is the value for the ith cluster and the jth observation. $\bar{x}_{..}$ is an overall mean of the observations. Based on statistical theory, F_p follows an F distribution with degrees of freedom k-1 and N-k. Combining this with the observed F_p yields the p-value for each feature. Once we obtained a collection of p-values for a total of 22 features, we can use the FDR procedure that can be summarized as follows [31]:

Consider a series of p-values and ordered p-values, denoted, respectively, as p_i and $p_{(i)}$, for $i = 1, 2, \ldots, 22$.

- Choose an FDR level α with a range between 0 and 1.
- Find $\tau = \max\left[i : p_{(i)} \leq \frac{i}{m} \cdot \frac{\alpha}{\pi_0}\right]$ where m is the total number of features (here $m = 22$), π_0 denotes the proportion of a true null hypothesis. In general, $\pi_0 = 1$ is the most conservative choice [32]. As a consequence, we used $\pi_0 = 1$.
- Let the p-value threshold be $p_{(\tau)}$. Declare the feature significant if and only if $p_{(i)} \leq p_{(\tau)}$.

In this study we used the R software (www.r-project.org) to implement the FDR procedure.

K-nearest Neighbors

A k-nearest Neighbors (KNN) algorithm is one of the most widely used algorithms for both classification and regression problems [33]. KNN does not require a trained model. Given a query point, the k closest points are determined. A variety of distance measures can be applied to calculate how close each point is to the query point. Then the k-nearest points are examined to find which of the most categories belong to the k-nearest points [33]. In the present study we used a KNN algorithm to computationally evaluate the features selected by an FDR

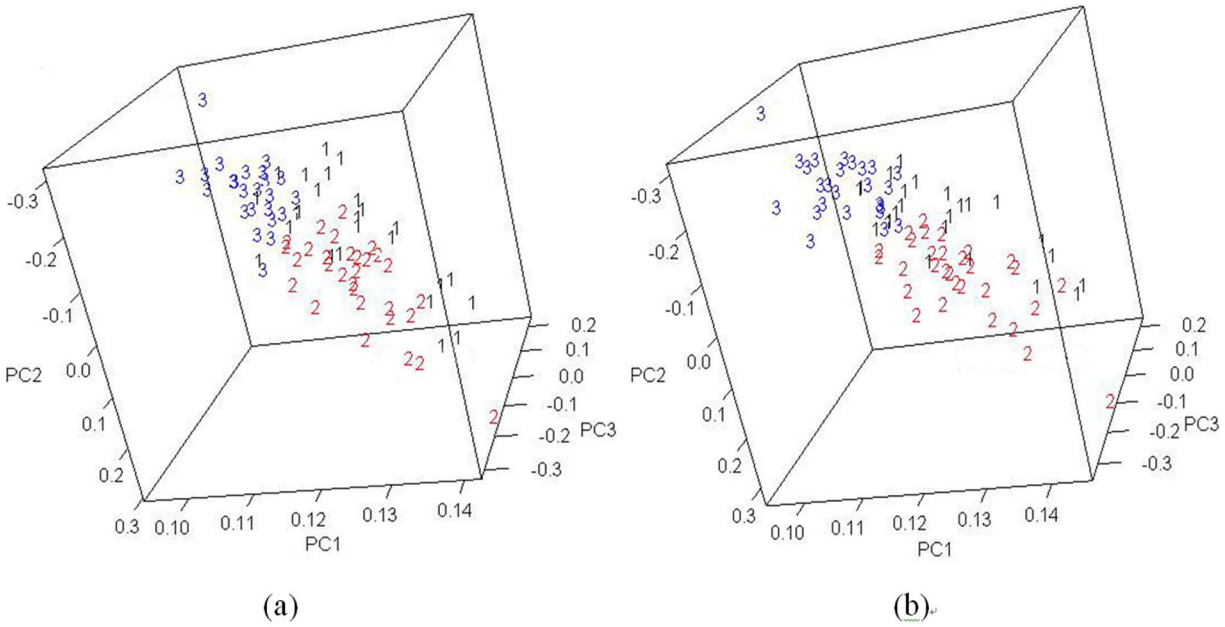

Figure 5. PCA plots using the features selected by (a) FDR level = 0.01 and (b) FDR level = 0.05.
doi:10.1371/journal.pone.0067862.g005

procedure. We used the "knn" function in R software (www.r-project.org) to implement a KNN algorithm.

Results

Clustering of patients with facial deformities

The *k*-means clustering algorithm using Euclidean distance was conducted on 72 patients with facial deformities. In order to determine the appropriate number *k*, we used the Rand index and adjusted Rand index approaches described in Section 3.1. Table 1 shows the resulting Rand index and adjusted Rand index for different *k* ($k = 2, 3, 4, 5$), indicating that both methods yielded large index values when $k = 2$ or 3. We thought that using $k = 2$ was too small to capture the important grouping of the data. Thus, we chose $k = 3$ for this study.

The *k*-means clustering method partitioned 72 patients into three clusters in which the first, second, and third clusters contain 17, 30, and 25 patients, respectively. Table 2 shows the descriptive statistics of the 22 measurements for each cluster.

Figure 2 shows the photos and X-ray images representing three clusters identified by a *k*-means clustering algorithm. The patients in the first cluster tend to have larger values of "SNB," "SN Pog," "FH to NB," and "Facial angle," but have smaller values of "FH to occlusal" and "Y axis." In particular, "ANB difference" value is almost zero. This characteristic can be categorized into the skeletal Class III type caused by excessive antero-posterior and less vertical growth of mandible. Therefore, the patients in the first cluster require surgical treatment such as orthognathic surgery of mandible. Patients in the second cluster have smaller values of "Convexity" and "ANB difference," but have higher values of "FH to L1" than other clusters. This is the main characteristic of the skeletal Class III type caused by the combination of maxillary deficiency and mandibular overgrowth. Consequently, these patients require bi-jaw surgery for maxillary advancement and mandibular setback. In the third cluster, the patients have larger values of "FH to occlusal," "Convexity," "Y axis," and "ANB difference" than appear in other clusters. This is the main characteristic of the skeletal Class II caused by the mandibular undergrowth. Thus, the patients in the third cluster require surgical treatment for mandible advancement and genioplasty.

Table 4. Misclassification rate of KNN ($k = 2, 4, 8, 16$) for the datasets used with different numbers of features.

	KNN($k=2$)	KNN($k=4$)	KNN($k=8$)	KNN($k=16$)
All Features	0.22(0.10)	**0.19**(0.10)	0.20(0.11)	0.25(0.12)
FDR($\alpha=0.05$)	0.23(0.10)	0.19(0.10)	**0.17**(0.10)	0.20(0.11)
FDR($\alpha=0.01$)	0.23(0.10)	0.19(0.10)	**0.18**(0.10)	0.21(0.10)

Average standard errors from 1,000 experiments are shown inside the parentheses; boldface values indicate in each dataset the KNN models with minimum error rates.
doi:10.1371/journal.pone.0067862.t004

Visualization of clustering results

PCA can be used as a test of the validity of the groupings obtained by the *k*-means clustering analysis based on $k = 3$. The scree plot shows that the first three PC accounted for 85% of the variability of the original data (Figure 3). Thus, we used three PCs.

Figure 4 shows a three-dimensional PCA score plot of PC1, PC2, and PC3. It clearly demonstrates that the separation of the 72 patients with facial deformities into three groups hinged on three PCs. This grouping result is consistent with the *k*-means clustering analysis.

Identification of important features

The FDR procedure was performed to test for each feature with significant differences between the clusters at FDR levels(α) = 0.01 and 0.05. The cutoffs ($p_{(\tau)}$) when $\alpha = 0.01$ and 0.05 are 0 and 0.002, respectively.

Table 3 shows the results of feature selection using the FDR approach at $\alpha = 0.01$ and at $\alpha = 0.05$. Different choices of FDR levels lead to selection of different numbers of features. A higher

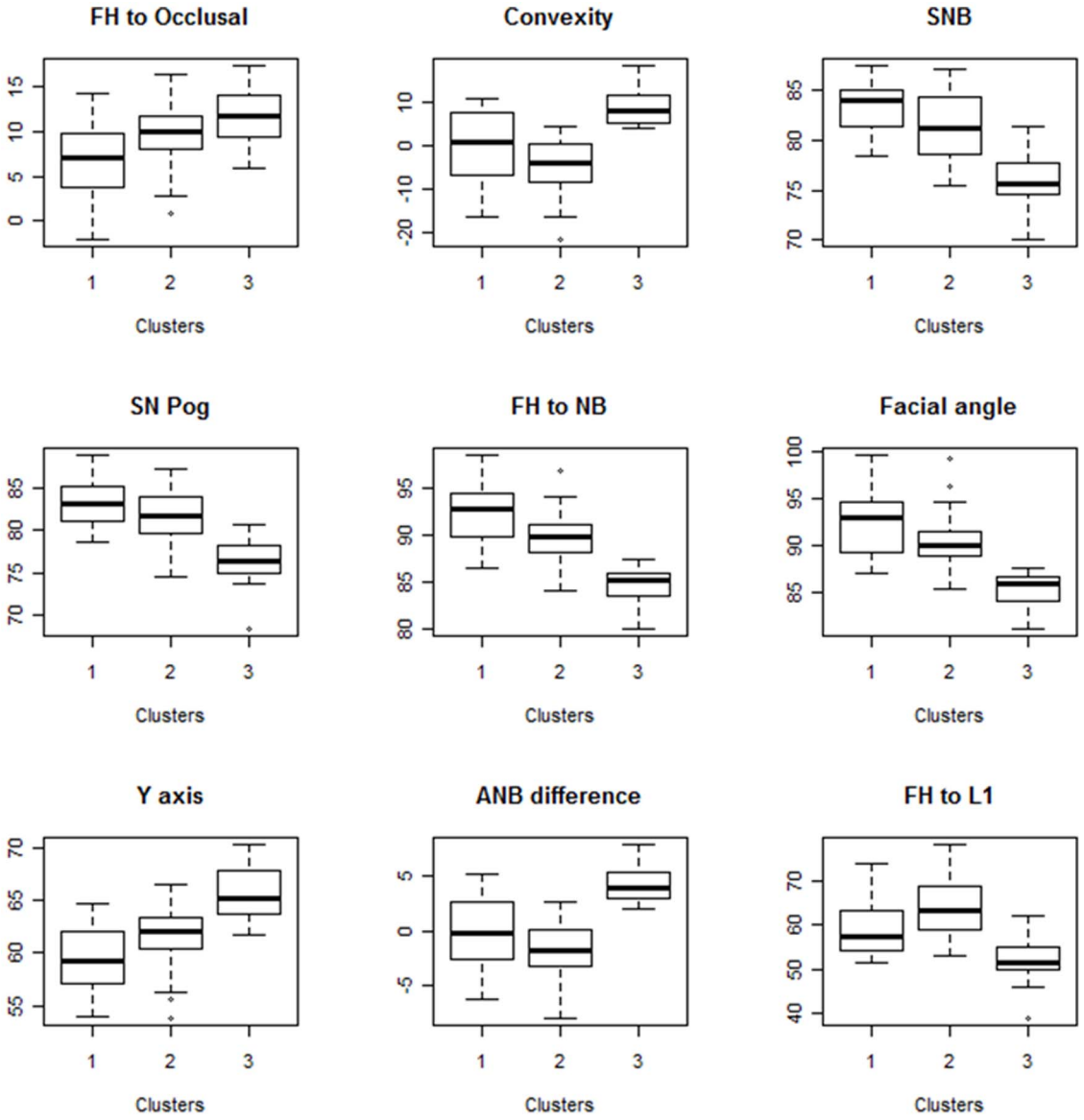

Figure 6. Box plots of different clusters using nine features selected by the FDR-based feature selection approach using $\alpha = 0.01$.
doi:10.1371/journal.pone.0067862.g006

FDR level increases the number of features selected, which results in more false positives at the same time it increases the capability to identify which features are significant. Conversely, a lower FDR level decreases the occurrence of false positives but diminishes the power to identify significant features. Here, the power is defined as the ability to correctly identify the significant features. Interpretation of our results for a case in which $\alpha = 0.05$ shows that on average less than one ($0.05 = 11 \cdot 0.05$) feature is falsely identified as significant (which is termed "false discovery") out of the 11 features selected by the FDR procedure.

Validation of the features selected

To demonstrate the validity of the feature selection results, we generated a PCA score plot using only the features selected by the FDR approach.

Figure 5 demonstrates that the PCA score plots produced by using the features selected by the FDR approach yielded results almost as good as the visualization capability created by using all features. This indicates that the FDR-based feature selection approach reduced the number of features required without degrading clustering performance.

The classification model is another approach to evaluating feature selection. In the present study we employed a KNN algorithm. We used Euclidean distance to determine the neighborhoods and tested different values of k (2, 4, 8, 16). To ensure classification accuracy, we used 80% of the dataset for training the KNN model and 20% for testing. We conducted this test 1,000 times and computed an average of 1,000 testing error rates to arrive at the final testing error rate. The datasets with different numbers of features were used for the KNN algorithm. First, we used the full dataset containing all the features. In our second and third tests we used the datasets containing the 11 and 9 features identified by the FDR approaches using $\alpha = 0.05$ and $\alpha = 0.01$, respectively. Table 4 shows the misclassification rates from KNN ($k = 2, 4, 8, 16$) with different numbers of features.

This table shows that misclassification error rates are comparable for all three datasets, indicating that the subsets of features identified by the FDR-based feature selection approach achieve as good misclassification rates as methods that use all features. In conclusion, the FDR-based feature selection approach reduced the dimensionality of the original data without deteriorating classification accuracy.

To further explore the feature selection results (visually), Figure 6 shows the box plots of different clusters using nine features selected by the FDR-based feature selection approach using $\alpha = 0.01$. We can see that at least two clusters can be distinguished by each of nine features.

Conclusions

This paper aimed to use data mining to characterize orthodontic data. We employed a *k*-means clustering algorithm to group 72 patients with facial deformities into several groups according to their characteristics. A statistical point of view suggests that these facial deformities fit into three clusters. To investigate each cluster's characteristics, we used FDR to select the measurements important to this categorization. To interpret the validity of the results of this identification of the selected features, we used visualization and classification. PCA shows that the selected measurements yield good visualization ability by using all measurements. KNN results suggest that use of FDR reduced the dimensions involved without loss of information. These results imply that the selected features are potentially useful for understanding the pattern of facial deformities.

We believe the selected features will be a great help in diagnosis. We hope that the present study increases awareness within the dental community of efficient methodologies to improve predictive diagnosis of dental treatment.

Acknowledgments

The authors thank the editor and the referees for constructive comments and suggestions that greatly improved the quality of this paper.

Author Contributions

Conceived and designed the experiments: SB DW. Performed the experiments: SB SY JW. Analyzed the data: SB SY. Contributed reagents/materials/analysis tools: SB SY DW. Wrote the paper: SB SY JW DW.

References

1. Schleyer T and Spallek H (2001) Dental informatics, A cornerstone of dental practice. Journal of the American Dental Associations 132: 605–613.
2. Downs WB (1948) Variations in facial relationships; their significance in treatment and prognosis. American Journal of Orthodontics 34(10): 812–840.
3. Higurashi N, Kikuchi M, Miyazaki S, Itasaka Y(2001) Comparison of Ricketts analysis and Downs-Northwestern analysis for the evaluation of obstructive sleep apnea cephalograms. Psychiatry and Clinical Neurosciences 55(3): 259–260.
4. Cotton WN, Takano WS, Wong WM (1951) The Downs analysis applied to three other ethnic groups. The Angle orthodontist 21(4): 213–220.
5. Al-Jasser NM (2005) Cephalometric evaluation for Saudi population using the Downs and Steiner analysis. Journal of Contemporary Dental Practice. 6(2): 52–63.
6. Sassouni V (1969) A classification of skeletal facial types. American Journal of Orthodontics 55(2): 109–123.
7. Nanda SK, Sassouni V (1965) Planes of reference in roentgenographic cephalometry. The Angle orthodontist 35(4): 311–319.
8. Gerber JW, Magill T (2006) NFO diagnostics: a modified Sassouni Cephalometric Analysis. Funct Orthod 23(2): 32–34, 36–37.
9. Reissmann E (1989) Sassouni analysis-validation of normal values. Z Stomatol, 86(6): 305–322.
10. Jacobson A (1975) The "Wits" appraisal of jaw disharmony. American Journal of Orthodontics 67: 125–138.
11. Burstone CJ, James RB, Legan H, Murphy GA, Norton LA(1978) Cephalometrics for orthognathic surgery. J Oral Surg 36(4): 269–277.
12. McNamara CM (1989) A retrospective cephalometric study of the effects of the Harvold appliance in the treatment of 20 patients with a Class II division 1 malocclusion. J Ir Dent Assoc 35(1): 36–38.
13. Wu J, Hagg U, Rabie AB (2007) Chinese norms of McNamara's cephalometric analysis. Angle Orthodontist 77(1): 12–20.
14. Raberin M, Laumon B, Martin J, Brunner F(1993) Dimensions and form of dental arches in subjects with normal occlusions. American Journal of Orthodontics and Dentofacial Orthopedics 104: 67–72.
15. Lee SJ, Lee S, Lim J, Ahn SJ, Kim TW (2007) Cluster analysis of tooth size in subjects with normal occlusion. American Journal of Orthodontics and Dentofacial Orthopedics, 132: 796–800.
16. Hwang HS, Youn IS, Lee KH, Lim HJ (2007) Classification of facial asymmetry by cluster analysis. American Journal of Orthodontics and Dentofacial Orthopedics 132: 279 e1–6.
17. De Veld DC, Skurichina M, Witjes MJ, Duin RP, Sterenborg DJ, et al. (2003) Autofluorescence Characteristics of Healthy Oral Mucosa at Different Anatomical Sites. Lasers in Surgery and Medicine 32: 367–376.
18. Xu R, Wunsch D (2005) Survey of clustering algorithms. IEEE Transactions on Neural Network 16(3): 645–678.
19. Rand WM (1971) Objective Criteria for the Evaluation of Clustering Methods. Journal of the American Statistical Association 66(336): 846–850.
20. Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. Royal Statistical Society 63(2): 411–423.
21. Kaufman L, Rousseeuw PJ (1990) Find Groups in Data: An Introduction to Cluster Analysis, John Wiley & Sons. New York.
22. Hubert L, Arabie P (1985) Comparing partitions. Journal of Classification, 2: 193–218.
23. Gordon AD (1999) Classification, Chapman& Hall. New York.
24. Jolliffe IT (2002) Principal Component Analysis. Springer-Verlag, New York.
25. Nguyen HP, Ortiz IP, Temiyasathit C, Kim SB, Schug KA (2008) LDI–MS fingerprinting of complex hydrocarbon mixture: application to crude oils using data mining techniques. Rapid Commun. Mass Spectrom, 22: 2220–2226.
26. Johnson RA, Wichern DW (2002) Applied Multivariate Statistical Analysis. New Jersey: Prentice-Hall.
27. Mei Y, Kim SB, Tsui KL (2009) Linear mixed effects models for feature selection in high-dimensional NMR spectra. Expert Systems with Applications 36: 4703–4708.
28. Kim SB, Chen VCP, Park Y, Ziegler TR, Jones DP (2008) Controlling the false discovery rate for feature selection in high-resolution NMR spectra. Statistical Analysis and Data Mining, 1: 58–66.
29. Kim SB, Tsui KL, Borodovsky M (2006) Multiple testing in large-scale contingency tables: inferring patterns of pair-wise amino acid association in b-sheets. International Journal of Bioinformatics Research and Applications 2: 193–217.
30. Park T, Yi SG, Lee S, Lee SY, Yoo DH, et al. (2003) Statistical tests for identifying differentially expressed genes in time-course microarray experiments. Bioinformatics 19: 694–703.
31. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. J.R. Statist. Soc.B, 57: 289–300.
32. Benjamini Y, Yekutieli D (2001)The control of the false discovery rate in multiple testing under dependency. Annals of Statistics 29(4): 1165–1188.
33. Mitchell TM (1997) Machine Learning. New York: McGraw-Hill.