Facial Morphology Analysis of Subjects Affected by Osteogenesis Imperfecta Type I, III and IV Using Computer Vision

Maxime Rousseau[a], Javier Vargas[bc], Frank Rauch[d,] Juliana Marulanda[d], Members of the BBDC*, Jean-Marc Retrouvey[e]

a: McGill University, Faculty of Dentistry, maxime.rousseau2@mail.mcgill.ca

b: Departamento de Óptica, Universidad Complutense de Madrid, Plaza de Ciencias 1, Madrid, 28040, Spain

c: Department of Anatomy and Cell Biology, McGill University 3640 Rue University, Montréal, QC H3A 0C7, Canada

d: Montreal Shriners Hospital

e: University of Missouri Kansas City

**: Brendan Lee, V. Reid Sutton, Sandesh CS Nagamani, Francis Glorieux, Janice Lee, Paul Esposito, Maegen Wallace, Michael Bober, David Eyre, Danielle Gomez, Gerald Harris, Tracy Hart, Mahim Jain, Deborah Krakow, Jeffrey Krischer, Eric Orwoll, Lindsey Nicol, Cathleen Raggio, Peter Smith, Laura Tosi.*

Abstract

Objective

Individuals affected by Osteogenesis Imperfecta (OI) type I, III and IV, an autosomal dominant genetic disease affecting mainly the COL1A1/A2 genes, have been reported to demonstrate characteristic facial manifestations of the disease. This study aims to assess morphological characteristics of patients affected by OI in a quantitative manner.

Materials and Methods

A retrospective case-control study involving 306 individuals (males:145, females:161) was conducted. This study uses automatic facial annotation and statistical shape analysis to compare the facial photographs of individuals affected by OI types I, III and IV to a normocephalic control group. Four facial ratios were used to compare the facial proportions. Additionally, the authors proposed a novel approach for facial analysis using 68 landmarks and statistical shape analysis to compare morphological features. A predictive model (PCALog) was trained to detect if a subject

was affected by OI based on facial landmarks.

Results

Our findings correlate with previous reports of facial characteristics of OI with type III subjects being the most severely affected. Our novel approach facilitated interpretation and comparison of morphological changes. Moreover, we were able to successfully train our PCALog model to automatically detect OI based on landmark features.

Conclusions

Facial manifestations of patients affected by OI were found to be greater at the level of the eyes and temples. A morphological approach facilitated the comparison of various groups and could be considered for future craniofacial analysis studies. Machine learning models can be trained using facial landmarks to detect presence of conditions affecting facial morphology.

Introduction

Osteogenesis Imperfecta (OI) is a rare, heritable bone disease caused. in most cases, by dominant mutations in *COL1A1* or *COL1A2*, the genes encoding collagen type I. However, mutations in about 20 other genes can also lead to OI[1]. The main clinical feature of OI is increased bone fragility and growth is often impaired[2]. The range of phenotypic severity is wide and is captured in the Sillence clinical classification that distinguishes four major OI types: type I ('mild'), type IV ('moderate'), type III ('severe') and type II ('lethal')[3]. There is no cure for OI and treatment usually involves the use of intravenous bisphosphonates to strengthen bones and thus decrease fracture rates[4].

The genetic alterations from this disease can result in significant craniofacial implications[5].

Qualitative assessments of facial morphology include triangular facial shape, low-set ears, and frontal bossing[6–8]. To our knowledge, no quantitative study on the facial morphology of individuals with OI has been published.

Quantitative facial analysis can be performed using anthropomorphic measurements or facial proportions. However, phenotypical manifestations are difficult to describe using these methods, hence facial manifestations of genetic diseases are often described in a qualitative manner[9]. This approach is highly prone to bias and is time consuming.

The goals of this study are to establish a reproducible method of assessment of facial morphology, compare the facial morphology of individuals with OI to a normocephalic control group, define the facial characteristics of OI types I, III and IV in a quantitative manner and train a supervised machine learning model to be able to detect if a subject is affected by OI based on their facial morphology. The null hypothesis is that there are no statistically significant differences in the facial morphology of subjects affected by different types of OI when compared to a normocephalic control group. This study was supported by the Brittle Bone Disorder Consortium (BBDC). The BBDC is part of Rare Diseases Clinical Research Network (RDCRN), an initiative of the Office of Rare Diseases Research (ORDR), the National Institute of Arthritis and Musculoskeletal and Skin (NCATS).

Methods

Study Population

A case-control study was conducted on a total of 306 individuals (males:145, females:161) aged 3 to 55 years (Table 1). 173 subjects with a diagnosis of OI were participants of the BBDC 7701

study at the Shriners Hospital for Children – Canada. These study participants were grouped according to OI type. The study groups were similar in age and sex distribution (Table 1).

The control group was composed of 133 patients recruited from the McGill Undergraduate Dental Clinic for orthodontic treatment who accepted to participate in this study. Control subjects were randomly selected to approximate the age and gender of the OI groups and represent 75% of the total number of OI affected subjects. We aimed at a control group size equal or greater to 70% of our OI sample subgroups combined. This group size was determined given the limitations in the number of control available and the time required to extract the data. An orthodontist reviewed the patient facial photographs to ensure that the control group subjects were of normocephalic facial type and did not have major skeletal discrepancies or malocclusions. All patients from the clinic signed an informed consent form. Additionally, the patients whose photographs were used in figures signed a release form. This study was approved by the McGill University Faculty of Medicine Institutional Review Board.

Facial Photographs

Antero-posterior facial photographs were used for this analysis. These pictures where all taken by the same operator using a Canon rebel T5™ camera with a Tamson™ SpDi macro 90 mm lens and a Canon Macro Ring Lite MR-14 EXII flash™. The camera settings were the following: AV mode, aperture - f/11, ISO -125, picture style – faithful, white balance - flash mode, image quality - JPEG L, metering mode – evaluative, drive mode - single shooting, focus - auto (distance subject to lens of camera = 1.5 m), orientation of the camera - portrait. Subjects had to face the camera head on, they could not rotate or tilt their head. The entire head and neck of the subject had to be included

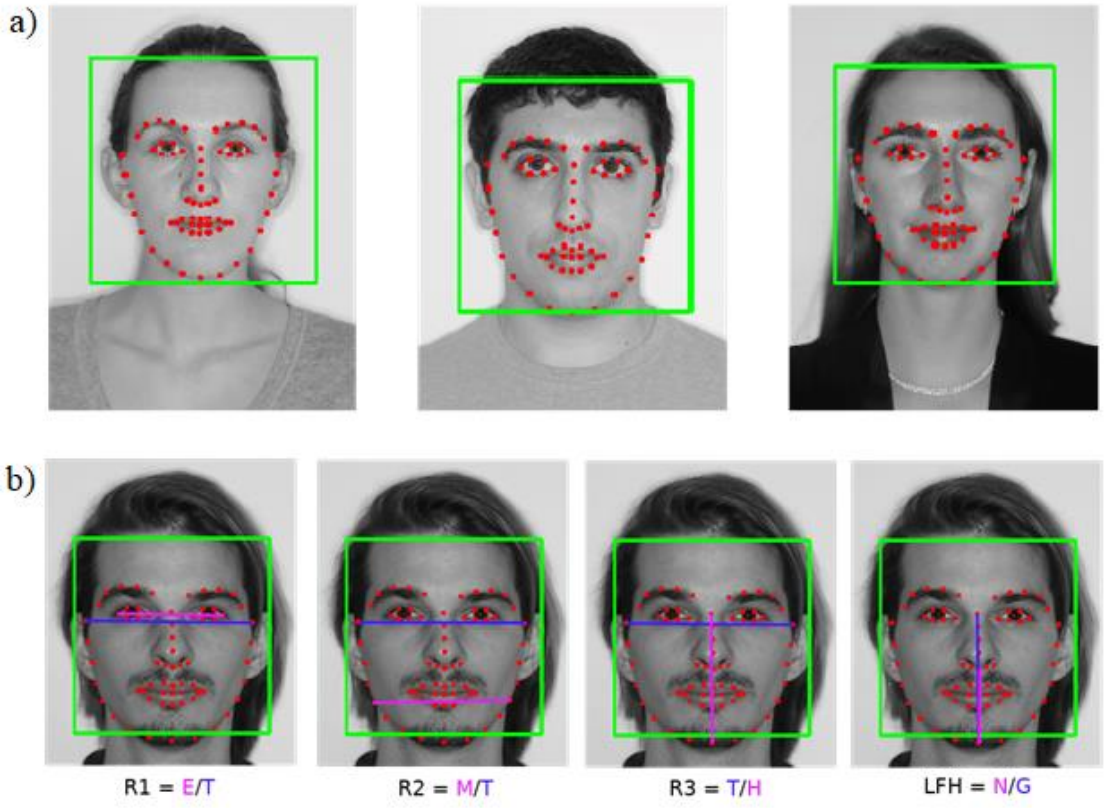
in the pictures. All photographs were taken with a white background. The subjects needed to clear their faces of any hair. Pictures that did not meet these requirements for a proper extra-oral orthodontic photograph were excluded from the study.

Analysis of Facial Photographs

The extraction of landmark coordinates from the images was performed using the pfla program[10]. The program uses deep learning models[11,12] to detect and annotate the faces with 68 landmarks (Figure 1a). The coordinates from these landmarks were used to perform our statistical analyses using the R programming language. The predictive machine learning models were implemented in Python 3 and Matlab.

Figure 1: Face Detection and Landmark Placement

a) *Automatic face detection (green rectangle) and 68 landmark annotation b) Four facial ratios computed using the landmark coordinates from the PFLA program*

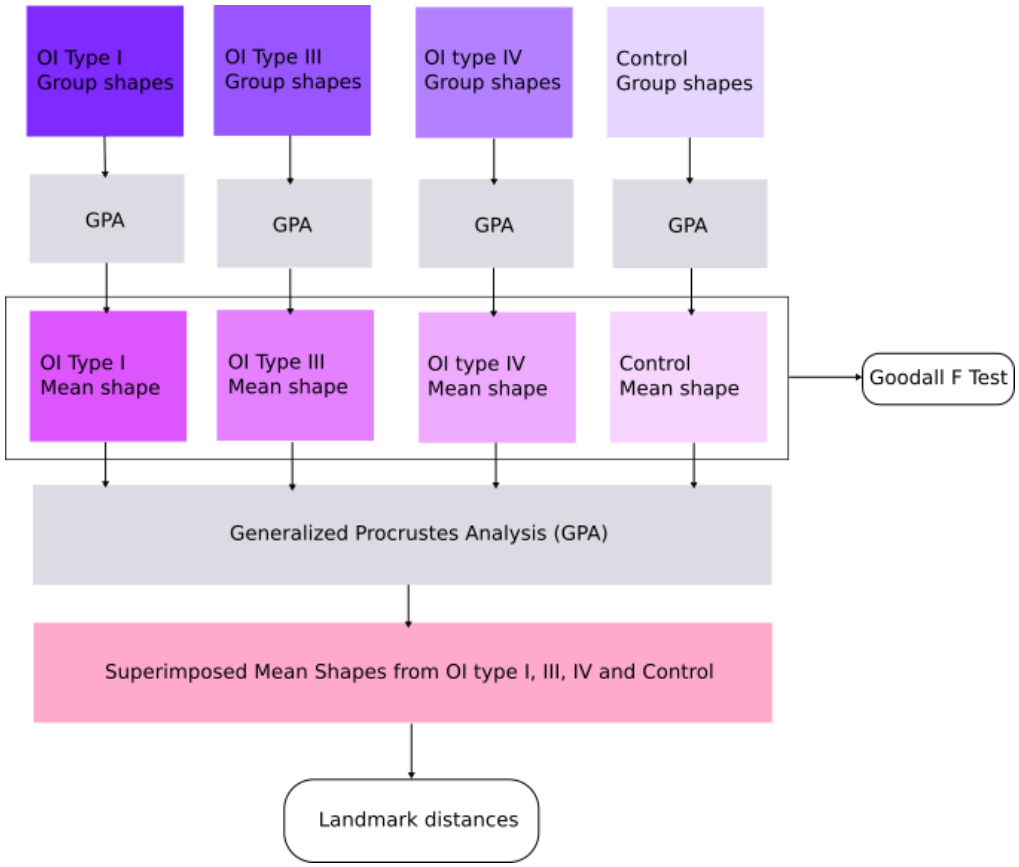


The proposed morphological method for facial analysis used all 68 automatically detected landmarks. This methodology is not tied to any set of landmarks and can be applied to other sets of facial annotations. The authors selected these landmarks as they were pertinent to identify the previously reported craniofacial manifestations of the disease[7].

The morphological approach to understanding the manifestations of OI has two objectives: determining if there is a difference between the group and if present, determining what these differences are. The shape processing and analysis methodology is illustrated in Figure 2.

Figure 2: Morphological Analysis Method

The study population was subdivided into subgroups (OI type I, III, IV and control). Generalized Procrustes Analysis is performed for each group and outputs the group mean shapes. The four mean shapes are compared using the Goodall F-Test. The four mean shapes are then aligned using GPA and landmark distances are computed to compare morphological differences.



Generalized Procrustes analysis (GPA) were used to normalize and align the shapes of each group which will allow for accurate comparison[13]. The change in morphology was deemed more important than measurements as they offer more information about the phenotypic manifestations of OI. GPA was performed for both rotation and scaling for optimal superimposition. A second GPA was performed on the mean shapes of the groups. This allowed for comparisons with landmark distance calculations.

Goodall's F-Test was used to determine whether a group's facial morphology was different from

another one. This test identifies statistically significant differences between the group mean shapes[14].

To explore and interpret the differences the Euclidean distances were computed between corresponding landmarks from each mean shape to the baseline mean shape from the control group. The resulting landmark distances were grouped according to their anatomical areas to improve the interpretability of the results (Figure 3b). We will refer to this as anatomical shape discrepancy (Equation 1).

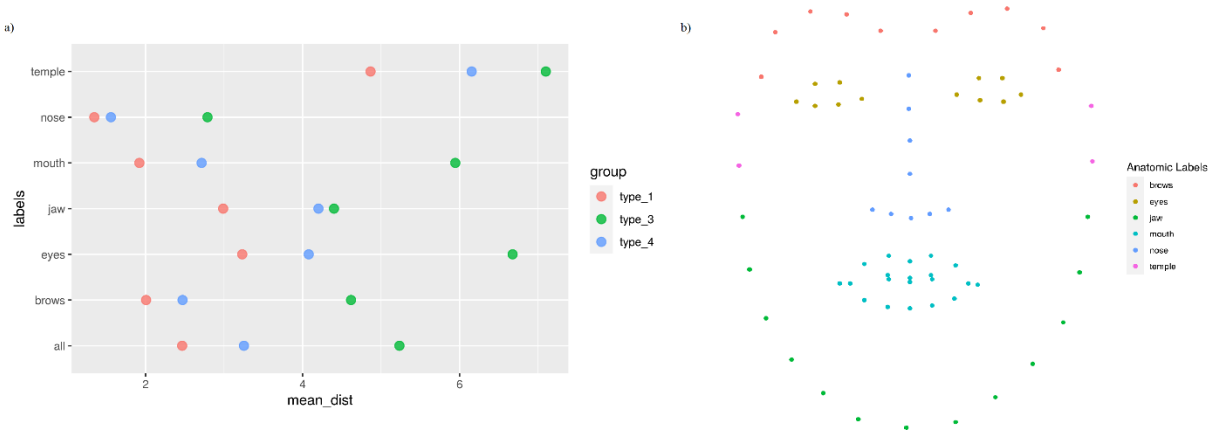
Equation 1: Anatomical Shape Discrepancy

$$Ld = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(a_x - b_x)^2 + (a_y - b_y)^2}$$

Figure 3: Landmark Labels

a) *Landmark distances of OI mean shapes compared to the control group mean shape. The distances are compiled by anatomic regions to enhance interpretability of results. The distances are relative, and units are not measured in absolute values as the shapes have been superimposed by GPA.*

b) *Anatomic labeling of the facial landmarks*



To validate our novel approach, the results of the proposed morphological analysis were compared to facial ratios analyses commonly used in orthodontics[15]. Four facial ratios were measured (Equation 2): biocular width /bitemporal width (R1), bimandibular width / bitemporal width (R2), bitemporal width / facial height (R3), and the lower face height (LFH) (Figure 1b). The groups were compared using ANOVA with Bonferroni post-hoc tests.

Equations 2: Facial Ratios

$$a)\ R1 = \frac{(l_{46} - l_{37})}{(l_{17} - l_1)}$$

$$b)\ R2 = \frac{(l_{13} - l_5)}{(l_{17} - l_1)}$$

$$b)\ R3 = \frac{(l_{17} - l_1)}{(l_{28} - l_9)}$$

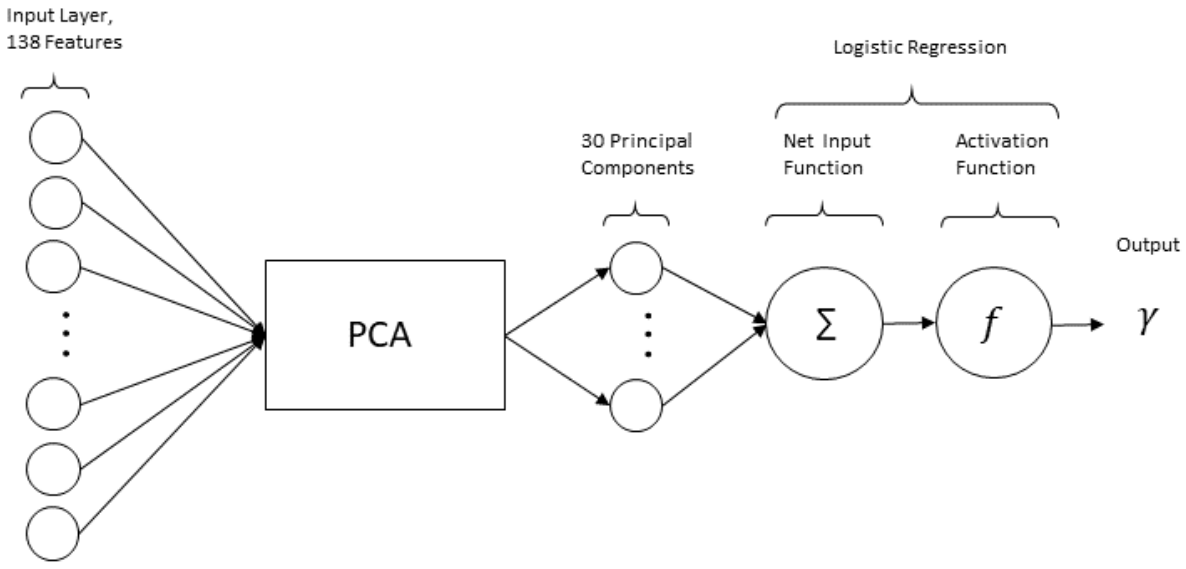
$$c)\ LFH = \frac{(l_{34} - l_9)}{(l_{28} - l_9)}$$

Predictive models were trained for the purpose of OI type detection. Three binary classifiers were trained. Training and test set were created from an 80/20 split of the dataset and were used for all models.

The first model named PCALog, (Figure 4) consisted of 138 input features. 68 x and 68 y coordinates as well as the age (1) and sex (1) of the subject. For this model, a principal component analysis (PCA) was used to reduce the dimensions to 30 features which were then passed through a logistic regression layer (Figure 4).

The second and third models were pre-trained AlexNet[16] and VGG16[17] convolutional neural networks. Both models used RGB images of fixed size of 224 pixels in height and width as inputs. Transfer learning was used for both deep learning architectures.

Figure 4: Schematic Representation of the PCALog Model



To assess the efficacy of the predictive models the precision recall and F1-score will be calculated for each class and for the total results of the test set (Equation 3).

Equations 3: Prediction Model Accuracy Metrics

$$a)\ Precision = \frac{TP}{TP + FP}$$

$$b)\ Recall = \frac{TP}{TP + FN}$$

$$c)\ F1 - score = \frac{Precision \cdot Recall}{Precision + Recall}$$

Results

Mean shape comparisons between OI types and with the control group were all statistically significant except for the comparison between individuals with OI type I and those with OI type IV. Figure 3a summarizes the morphological differences between OI types when compared to control groups according to their anatomical regions. OI type III shows considerably greater morphological differences compared to types I and IV. OI type I is closest to the control group. Overall, the temple and eye landmarks showed the greatest discrepancies when compared to the baseline shape.

Table 1: Results from the Measurements and Statistical Testing.

	Control	OI-I	OI-III	OI-IV	P Value
N (M/F)	133 (68/65)	88 (42/46)	28 (9/19)	57 (26/31)	P (Chi Square) = 0.32

Mean Age (SD)	17.7 (7.7)	19.7 (14)	15.4 (8.4)	17.1 (8.4)	P(ANOVA) = 0.19
Mean Ratio 1 (SD)	0.612[a] (0.026)	0.595 (0.026)	0.604 (0.036)	0.602 (0.031)	P(ANOVA) = <0.05
Mean Ratio 2 (SD)	0.799[ab] (0.027)	0.784 (0.038)	0.774 (0.030)	0.791 (0.031)	P(ANOVA) = <0.05
Mean Ratio 3 (SD)	1.26[abc] (0.075)	1.33[b] (0.106)	1.40[c] (0.124)	1.33 (0.097)	P(ANOVA) = <0.05
Mean LFH (SD)	0.569(0.022)	0.571 (0.027)	0.562 (0.035)	0.567 (0.026)	P(ANOVA) = 0.47

Bonferroni Post-Hoc Analysis a: $p < 0.05$ in comparison to OI-I b: $p < 0.05$ in comparison to OI-III c: $p < 0.05$ in comparison to OI-IV

Ratios are summarized in Table 1. Between-group comparisons were statistically significant for all facial metrics except for lower face height (LFH). Ratio 1 differed between the control group and OI type I. Ratio 2 differed between the control group and OI types I and III. Ratio 3 results were statistically different for the control group and all OI types, as well as between OI types I and III and types III and IV.

Table 2: Accuracy Metrics from the Classifiers Test Set

Model	Class	Precision	Recall	F1-Score
PCALog	Control	1.00	1.00	1.00
	OI Subject	1.00	1.00	1.00
AlexNet	Control	0.88	0.78	0.82
	OI Subject	0.84	0.91	0.88
VGG16	Control	0.86	0.89	0.87
	OI Subject	0.91	0.89	0.90

The accuracy metrics for the different models are summarized in Table 2. Precision is the positive predictive value and recall is the sensitivity. F1-score is a metric that tests the accuracy of a predictive model considering both precision and recall. The model based on facial morphology classifiers had perfect performance on our test set. AlexNet and VGG showed decreased accuracy in comparison.

Discussion

In this study we used an automated facial morphology analysis based on computer vision to objectively assess the facial characteristics of OI. This novel approach allowed us to conduct face detection, labeling and shape analysis easily and efficiently without human annotation. We found that the morphological differences between individuals with OI and controls are greatest at the level of the eyes and temples. Morphological differences across all face labels were consistently more severe in OI type III. Using GPA and other shape analysis methods proved to be reliable in our comparison of the facial morphology of the groups. This approach simplified the interpretation of our findings by using the semantic distance graph. Moreover, we were able to train a model that was able to classify OI and controls based on their morphological features.

The OI and control samples were not matched according to ethnicity given the amount of variation found in the OI subjects. The heterogenicity of ethnicity of the sample presents a limitation with respect to the mean shape and predictive models. The facial shape analyses were conducted on groups which included male and female, additionally we did not further subdivide the groups according to their age. Further studies could be conducted to investigate the morphological

differences according to gender and their changes with age.

The 68 landmarks were used to compare the mean shape of the groups and to evaluate differences in facial features. The mean shape was computed using all landmarks at once. Facial feature comparison was performed by grouping landmarks according to their anatomic labels (Figure 3b). Mean shape comparison allows to determine if patients have similar facial morphology. The landmark distances of anatomic regions detail how the groups differ. Figure 3a illustrates the differences in anatomic regions between each group. A statistically different facial morphology associated with deviations of anatomic regions compared to the control could signify underlying craniofacial malformations and should be considered during the diagnostic phase of orthodontics. Individuals with OI type I, III and IV present with a different facial morphology when compared to a control group. All three ratios which incorporated the bi-temporal width were significantly different between groups. The LFH measurement, which only takes the vertical aspect of the face into account was the only metric that showed no intergroup differences. Thus, differences in facial width are a key characteristic of facial morphology in OI. These findings are in accordance with the triangular facial morphology described in previous articles[7,9,18].

Our results also indicate that OI is associated with a different morphology at the level of the temples. The difference in the mean shapes can be easily visualized and interpreted using the semantic distance plot. In a separate study using cone beam computed tomography we had also observed prominent temporal bones in severe OI[19]. Our present analysis suggests that these manifestations vary with OI type, and that the OI type III group is the more severely and consistently affected.

The morphological approach is a more reliable method of comparing facial morphology than any

single measurement or metric such as facial ratios. A more intuitive interpretation of the results that encompasses all the facial landmarks allows to find out where differences lie. Mean shape testing appeared to be the most efficient way to detect morphological differences between groups, as it incorporates the complete set of landmarks at once. Moreover, semantic distance analysis provides us with more precise information about the differences and is less cumbersome in its interpretation. This type of analysis could be extended to conduct studies correlating morphological changes with genetic mutations. This technique can also be adapted to automate the analysis of other types of medical imaging such as CT scans, radiographs, and 3D models. A deeper understanding of the manifestations of various genetic mutations may be achieved. In orthodontics, these techniques have the potential to become a reliable method of assessment and tracking of the effects of orthodontic and dentofacial orthopedic treatment, orthognathic surgery, or cleft palate management on soft tissue facial morphology.

Given the small dataset of subjects affected with OI, there was insufficient data to train and evaluate a multi-class classifiers model. Too few type III and IV patients would create a skewed training set which would be prone to overfitting. Moreover, such a small sample makes it difficult to assess the accuracy of the model. For these reasons, we aimed to train a simple binary classifier to illustrate the use of machine learning models with minimal data.

Results from the trained models are promising especially for the detection and classification of OI. The efficiency of logistic regression of landmark data compared to pre-trained deep neural networks is observed. The PCALog model outperformed common deep learning architectures which is probably due to the small sample size. The ability to train successful classifiers using landmarks and general subject data as input features creates the potential for new diagnostics aids.

As OI is a rare disease, the main limitation to the sophistication and accuracy of the models is the size of the dataset. Simple models can still be effective for preprocessed data such as landmarks. The PCALog model was only trained on controls and patients affected with OI. At its current stage of development, this model cannot be considered a diagnostic model or screening tool. OI patients may share some morphological features with other syndromes which could lead to misdiagnosis. The efficacy of our PCALog model trained on small amounts of data demonstrates the potential in clinical applications of machine learning methods for the assessment and classification of facial features of rare diseases and syndromes. These models and methods may be applied for other conditions such as Marfan, Crouzon, Ehler-Danlos and Treacher Collins syndromes that are associated with changes in facial morphology.

Conclusion

Facial morphology in OI presents specific features mostly at the level of the temples and the eyes. Using a small dataset, we were able to train the PCALog model to distinguish individuals with OI from controls based on their morphological landmarks.

Acknowledgements

The Brittle Bone Disease Consortium (1U54AR068069-0) is a part of the National Center for Advancing Translational Sciences (NCATS) Rare Diseases Clinical Research Network (RDCRN), and is funded through a collaboration between the Office of Rare Diseases Research (ORDR), NCATS, the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), the National Institute of Dental and Craniofacial Research (NIDCR), and the Eunice Kennedy Shriver

National Institutes of Child Health and Development (NICHD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The Brittle Bone Disease Consortium is also supported by the Osteogenesis Imperfecta Foundation.

The authors would also like to acknowledge the contributions of the following BBDC site coordinators: M Abrahamson (OHSU), S Alon (UCLA), M Azamian and A Turner (BCM), C Brown (Nemours Alfred I. duPont Hospital), E Carter and E Yonko (HSS), A Caudill (Chicago Shriners), K Dobose (KKI), M Durigova (Montreal Shriners Hospital for Children), A Giles and E Rajah (CNMC), M Gross-King (Tampa Shriners), E Strudthoff (UNMC).

References

1. Tauer JT, Robinson M-E, Rauch F. Osteogenesis Imperfecta: New Perspectives From Clinical and Translational Research. *JBMR Plus*. 2019;3(8):e10174.

2. Graff K, Syczewska M. Developmental charts for children with osteogenesis imperfecta, type I (body height, body weight and BMI). *Eur J Pediatr*. 2017;176(3):311-316.

3. Sillence DO, Rimoin DL. Classification of osteogenesis imperfect. *Lancet*. 1978;1(8072):1041-1042.

4. Rossi V, Lee B, Marom R. Osteogenesis imperfecta: advancements in genetics and treatment. *Curr Opin Pediatr*. 2019;31(6):708-715.

5. Shapiro JR. Osteogenesis imperfecta: a translational approach to brittle bone disease. Published online 2013.

6. Rousseau M, Retrouvey JM. Osteogenesis imperfecta: potential therapeutic approaches. *PeerJ*. 2018;6:e5464-e5464.

7. Chang PC, Lin SY, Hsu KH. The craniofacial characteristics of osteogenesis imperfecta patients. *Eur J Orthod*. 2007;29(3):232-237.

8. Rizkallah J, Schwartz S, Rauch F, et al. Evaluation of the severity of malocclusions in children affected by osteogenesis imperfecta with the peer assessment rating and discrepancy indexes. *Am J Orthod Dentofacial Orthop*. 2013;143(3):336-341.

9. Retrouvey J-M, Taqi D, Tamimi F, et al. Oro-dental and cranio-facial characteristics of osteogenesis imperfecta type V. *Eur J Med Genet*. 2019;62(12):103606.

10. Rousseau M, Retrouvey JM. pfla: A Python Package for Dental Facial Analysis using Computer Vision and Statistical Shape Analysis. *J Open Source Softw*. 2018;3(32):855-855.

11. Zhang K, Li Z, Qiao Y, Zhang Z. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Process Lett*. 2016;23(10):1499-1503.

12. Sagonas C, Tzimiropoulos G, Zafeiriou S, Pantic M. A Semi-automatic Methodology for Facial Landmark Annotation. Published online 2013:896-903.

13. Airey DC, Wu F, Guan M, Collins CE. Geometric morphometrics defines shape differences in the cortical area map of C57BL/6J and DBA/2J inbred mice. *BMC Neurosci*. 2006;7:63.

14. Goodall C. Procrustes Methods in the Statistical Analysis of Shape. *J R Stat Soc Ser B Methodol*. 1991;53(2):285-321.

15. Negi G, Ponnada S, Aravind NKS, Chitra P. Photogrammetric Correlation of Face with

Frontal Radiographs and Direct Measurements. *J Clin Diagn Res JCDR*. 2017;11(5):ZC79-ZC83.

16. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*.:1097-1105.

17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *ArXiv Prepr ArXiv14091556*. Published online 2014.

18. Foster BL, Ramnitz MS, Gafni RI, et al. Rare bone diseases and their dental, oral, and craniofacial manifestations. *J Dent Res*. 2014;93(7 Suppl):7S-19S.

19. Reznikov N, Dagdeviren D, Tamimi F, Glorieux F, Rauch F, Retrouvey J-M. Cone-Beam Computed Tomography of Osteogenesis Imperfecta Types III and IV: Three-Dimensional Evaluation of Craniofacial Features and Upper Airways. *JBMR Plus*. 2019;3(6):e10124.