Quantifying the Severity of Metopic Craniosynostosis: A Pilot Study Application of Machine Learning in Craniofacial Surgery

Riddhish Bhalodia, BS,[*] Lucas A. Dvoracek, MD,[†] Ali M. Ayyash, MD,[†] Ladislav Kavan, PhD,[*] Ross Whitaker, PhD,[*] and Jesse A. Goldstein, MD[†]

Abstract: The standard for diagnosing metopic craniosynostosis (CS) utilizes computed tomography (CT) imaging and physical exam, but there is no standardized method for determining disease severity. Previous studies using interfrontal angles have evaluated differences in specific skull landmarks; however, these measurements are difficult to readily ascertain in clinical practice and fail to assess the complete skull contour. This pilot project employs machine learning algorithms to combine statistical shape information with expert ratings to generate a novel objective method of measuring the severity of metopic CS.

Expert ratings of normal and metopic skull CT images were collected. Skull-shape analysis was conducted using ShapeWorks software. Machine-learning was used to combine the expert ratings with our shape analysis model to predict the severity of metopic CS using CT images. Our model was then compared to the gold standard using interfrontal angles.

Seventeen metopic skull CT images of patients 5 to 15 months old were assigned a severity by 18 craniofacial surgeons, and 65 nonaffected controls were included with a 0 severity. Our model accurately correlated the level of skull deformity with severity ($P < 0.10$) and predicted the severity of metopic CS more often than models using interfrontal angles ($\chi^2 = 5.46$, $P = 0.019$).

This is the first study that combines shape information with expert ratings to generate an objective measure of severity for metopic CS. This method may help clinicians easily quantify the severity and perform robust longitudinal assessments of the condition.

Key Words: Craniosynostosis severity, interfrontal angle, machine learning, metopic craniosynostosis

(*J Craniofac Surg* 2020;31: 697–701)

Metopic craniosynostosis (MC) refers to early fusion of the metopic suture and affects approximately 1 in every 6000 children born in the United States. Unlike other forms of craniosynostosis, MC represents a diagnostic challenge, especially in more moderate cases, because the metopic suture normally closes well before 1 year of age.[1,2] While both physical examination and computed tomography (CT) imaging are used to detect MC, the primary indicator of disease is the presence of an abnormal head shape, referred to as trigonocephaly. In this condition, the head is characterized by an abnormal triangular shape with bifrontal narrowing, biparietal widening, a metopic ridge, as well as lateral supra-orbital rim retrusion and hypotelorism.[3]

There is a wide variation in surgical approaches to patients with MC depending on the severity of the deformity and the age of the patient. Currently determining the extent of disease and timing of operative intervention is largely dependent on a qualitative evaluation of the degree of deformity. The subjective nature of this method can result in significant practice variation across different treatment centers and within centers from patient to patient. This is especially true in mild cases of MC where subtle differences in head shape may not result in a clearly stigmatizing deformity. This has created an opportunity in the literature to identify objective measures of severity for MC. While many attempts have been made to quantify the deformity in the past, most studies rely on measurements taken in 2-dimensions (2D) even though MC is inherently a 3D deformity. In this study, we generate a 3D shape model to describe the severity of the deformity in patients with MC. We then compare this 3D model to the predictive accuracy of the widely utilized interfrontal angle (IFA) introduced by Kellog et al.[4] This novel model will ultimately help eliminate subjectivity when determining the severity of MC and provide a unified platform for further outcome studies between different surgeons and institutions.

METHODS

Patients and Controls

This was an institutional review board approved multi-institutional pilot study. Computed tomography scans were obtained for subjects 5 to 15 months old seen at the University of Pittsburgh Medical Center Children's Hospital between 2002 and 2016. All images were obtained using a General Electric VCT 64 Slice CT scanner (General Electric, Fairfield, CT) with a standard low dose fine cut (0.25 mm) protocol. Patients with MC were diagnosed in clinic by a board-certified craniofacial plastic surgeon using CT imaging and physical exam. Controls were chosen from patients presenting to the

From the *School of Computing, University of Utah, Salt Lake City, UT; and †Department of Plastic Surgery, UPMC Children's Hospital, University of Pittsburgh Medical Center, Pittsburgh, PA.
Received August 28, 2019.
Accepted for publication October 30, 2019.
Address correspondence and reprint requests to Jesse A. Goldstein, MD, One Children's Hospital Drive, 4401 Penn Avenue Floor 3 Pittsburgh, PA 15224; E-mail: Jesse.goldstein@chp.edu
Presented at the Robert H. Ivy Society 2018 Plastic Surgery meeting and the Ohio Valley Society of Plastic Surgeons 2018 Meeting.
Accepted to the 18th Congress of International Society of Craniofacial Surgery for presentation September 16–19th, 2019 in Paris, France.
The authors report no conflicts of interest.
Supplemental digital contents are available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.jcraniofacialsurgery.com).

ISSN: 1049-2275
DOI: 10.1097/SCS.0000000000006215

hospital for trauma with no abnormalities on CT imaging. Descriptive statistics were calculated using Stata/SE version 15.1.

Expert Rater System

An expert panel of 10 craniofacial plastic surgeons and 8 pediatric neurosurgeons from the Synostosis Research Group were individually e-mailed the link to a web-based application called CranioRate to assign ratings to 3D metopic head shapes remotely (Fig. 1). Raters were able to rotate and manipulate the 3D reconstructed CT scan in any direction for visualization and assessment in any preferred viewpoint. Ratings were assigned on a 10-point Likert scale (1-least severe, 10-most severe). Normal controls were assumed to have a uniform severity rating of zero. Intraclass correlation was performed to quantify the level of agreement among surgeon scores.

Three-Dimensional Shape Modeling and Analysis

Computed tomography processing is summarized in Figure 2. Three dimensional skull CT's were cut parallel to the Frankfort Horizontal plane at the level of the zygomatic-frontal suture[5,6] in preparation for analysis by the ShapeWorks software.[7] ShapeWorks is an open-source software created by the University of Utah, Scientific Computing and Imaging Institute that utilizes an algorithm to automatically identify and place thousands of consistent 3D points (correspondences) on a cohort of shapes (Fig. 2). Principal component analysis (PCA) was used to reduce the 3D point representation for each skull into 8 dimensions (which capture the essence of the entire 3D skull shape) so that each skull could be quantified by a single shape descriptor x_i (Fig. 3).

It was assumed that each skull had a latent severity measure (l_i) which represents the actual severity of the pathology, and was dependent on the shape of the skull (shape descriptor x_i) and a regression coefficient (β_j) in a linear manner as given by $l_i = \sum_{\{j=1\}}^{8} \beta_j x_{ij}$. Rasch modeling[8–10] was used with the expert ratings to account for the internal bias of each rater with the latent severity of the scan (Fig. 4). Rater biases (θ_j) based on individual biases in the survey were incorporated into the model.

MPLUS software was used to create a maximum likelihood estimation machine-learning algorithm[11] with data collected from the expert ratings and shape descriptors in Shapeworks to approximate the unknowns: latent severity measure (l_i), rater biases (θ_j), and regression coefficients (β_j). A Pearson correlation was conducted to assess the accuracy of fit of the model with significance set at $\alpha = 0.05$. These regression values were used to generate a severity for each scan.

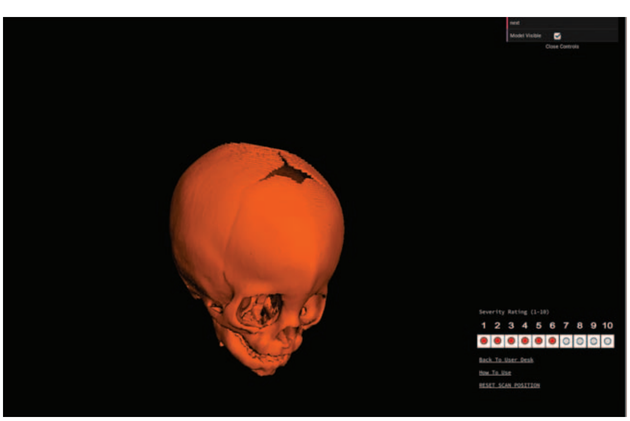

FIGURE 1. Example of a 3D skull shape sent to expert raters in the CranioRate application. Experts could rotate the 3D reconstruction in any direction to view the entirety of the skull, and then click on a number to record a severity rating. 3D, three dimensional.

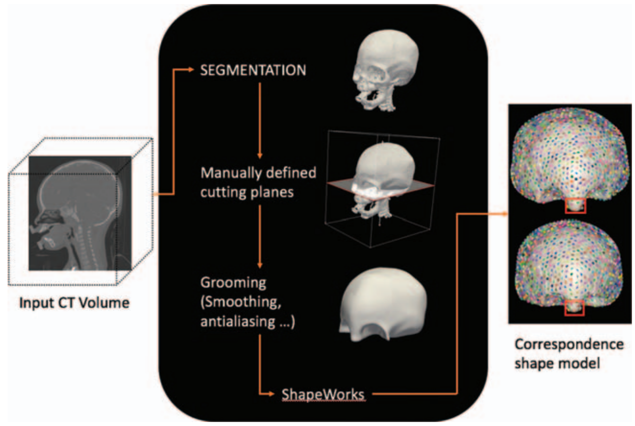

FIGURE 2. Computational process to prepare CT for processing by the Shapeworks Software. CTs were segmented into 3D models and manually cropped at previously identified planes. Models underwent advanced computational processing (smoothing and antialiasing) before being automatically converted into correspondence shape models (correspondence region depicted by the highlighted red box). 3D, three dimensional; CT, computerized tomography.

Validation and Comparison to the Interfrontal Angle

The efficacy of the shape descriptor was validated by comparison with the IFA, a widely adopted shape descriptor. The same machine learning algorithm was applied to predict severity with IFA in place of the shape descriptor. Severity ratings for each CT scan were then reconstructed using our new 3D model and the IFA model, as previously described in the literature.[3,4] Newly reconstructed ratings were compared with the ratings assigned by the experts using a Pearson Chi-squared test. A leave-one-out analysis[12] was performed after removing a single CT scan from the statistical model and after removing the ratings of a single expert to determine the influence of each additional rater on the model. The average error was estimated by mean squares difference between the latent severity prediction from the original model with that from the leave-one-out-analysis using MPLUS software.

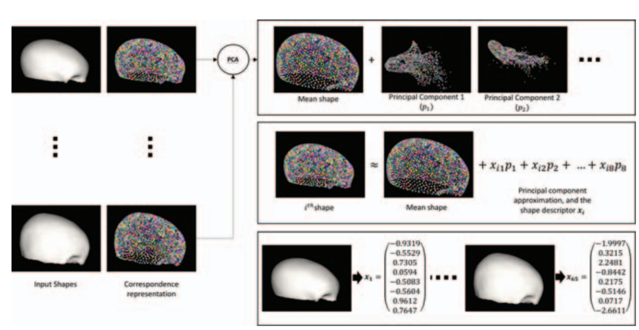

FIGURE 3. Process of principal component analysis. Each 3D skull was input into the ShapeWorks software, mapped with thousands of different points, and statistically analyzed in several different planes to generate a shape descriptor (x_i) which identifies the geometric variations in each skull. Each scan is then represented using 8 scalar values and these form the proposed shape descriptor. 3D, three dimensional.

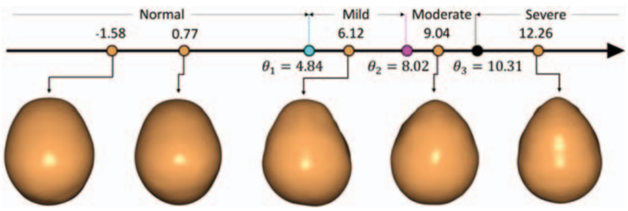

FIGURE 4. Representative top-down view of patients with varying degrees of metopic craniosynostosis severity.

RESULTS

Patients and Expert Ratings

A total of 17 metopic skull CT images and 65 nonaffected skull CT's were included (Supplementary Digital Content, Table 1, http://links.lww.com/SCS/B139). Each of the 18 experts assigned a severity ranking to the 17 metopic skull CTs for a total of 306 rankings. The calculated intraclass correlation coefficient was 0.716 (95% confidence interval 0.57–0.86), indicating a high level of agreement among surgeon rankings (values closer to 1 indicate more agreement).

Three-dimensional Shape Modeling and Severity Calculation

A total of 2048 3D correspondence points were analyzed for each skull. The correlation coefficient constants (β) for the skull shapes were calculated by the machine learning algorithm. There was a significant relationship between latent severity and the shape descriptor in 7 of the 8 dimensions of analysis ($P < 0.05$) (Supplementary Digital Content, Table 2, http://links.lww.com/SCS/B139). In the nonstatistically significant dimension, P-value was 0.056. The calculated regression coefficients from above and the results of the PCA were used to generate a map indicating the anticipated malformations as severity of MC increases (Fig. 5).

Relationship to Interfrontal Angle

The recreated severity rankings from the new model and relationship to the IFA model are shown in Figure 6. The number of times the reconstructed ratings differed from the expert ratings, the sum of these values, and the relative average of the differences between the 2 methods are displayed in Supplementary Digital Content, Table 3, http://links.lww.com/SCS/B139. A Pearson Chi-squared test statistic revealed a statistically significant lower proportion of misclassifications in the method using this newly proposed 3D model versus the IFA model ($\chi^2 = 5.46$, $P = 0.019$).

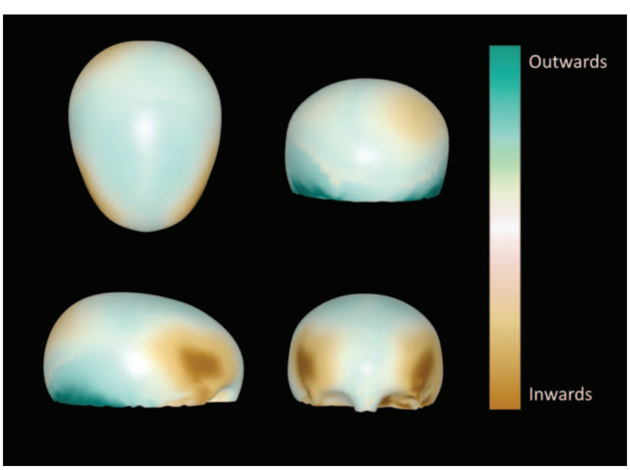

FIGURE 5. Normal skull with colorimetric overlay generated through the statistical model indicating the regions of malformation in the metopic CS skull, with color intensity proportional to severity. CS, craniosynostosis.

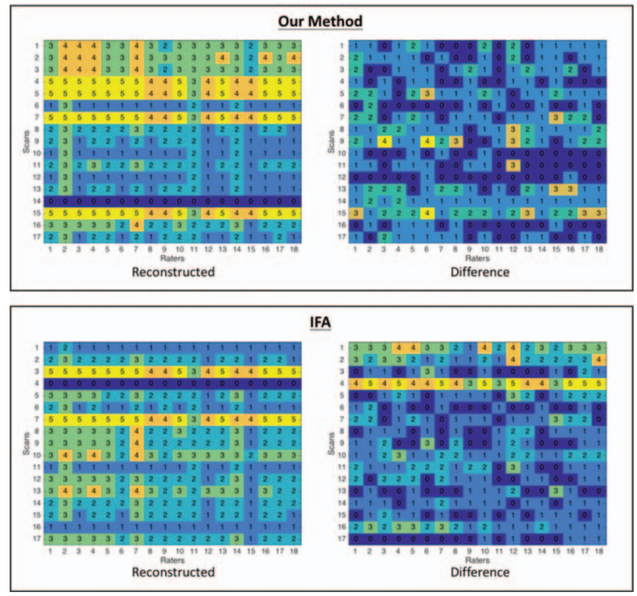

FIGURE 6. Comparison of our model to traditional IFA model of severity. Top Left: The ratings of each skull were constructed using our hypothesized statistical method (using a downsampled Likert scale range 1 to 5 and using the estimated latent severity and the rater bias to bin into the categorical values). Top Right: These recreated ratings were then compared to the ratings for the skulls provided by the experts to determine the accuracy of our model. Bottom Left: The ratings of each skull were constructed using the measured interfrontal angles. Bottom Right: These recreated ratings were then compared to the ratings for the skulls provided by the experts to determine the accuracy of the interfrontal angle model. The regression coefficient (λ) relating the latent severity measure (l_i with the interfrontal angle (A_i) was calculated to be -0.496 in the relationship $l_i = \lambda A_i$. IFA, interfrontal angle.

Model Strength Evaluation

A leave-one-out analysis was performed after removing a single CT scan from this model, yielding a leave-one-out error of 9.61, variance of 87.37, and predicted r^2 of 89%. Similarly, the leave-one-out analysis after removing a single rater yielded a leave-one-out error of 2.81, variance of 87.37, and predicted r^2 of 97%. These errors and variances are on the unitless latent severity scale and represent how robustly the statistical model describes the given set of CT scans and raters respectively.

DISCUSSION

Assessing the severity of patients with MC is a critical but underappreciated facet of the care of these complex patients. Traditional methods to gauge severity include plain radiographs, physical exam, and CT imaging; however, each of these modalities ultimately rely on the subjective assessment of the treating physician. This can introduce unintended treatment variability and inconsistency between patients, surgeons, and treatment centers and highlights a need for objective measures of severity for MC. Such a measure should be reproducible and lead to meaningful stratification. Ultimately, such measures could dictate the ideal circumstances for operative intervention, aid in predicting patients with poor postoperative outcomes (intracranial hypertension or aesthetic concerns), help compare longitudinal outcomes of MC between differing institutions and surgeons, and even inform patient-specific interventions.

The effort to define objective measures to quantify the severity of craniosynostosis has been a long undertaking among craniofacial surgeons. Perhaps the most obvious indication of severe craniosynostosis (and clear indication for surgery) are signs or symptoms of intracranial hypertension. Indeed, papilledema has been shown to be a highly specific (98%) measurable marker for intracranial pathology in patients with craniosynostosis.[13] Other signs such as headaches, and vision changes may also indicate the need for operative intervention. However, such physical symptoms are unreliable since the incidence of intracranial hypertension is exceedingly low at presentation and the vast majority of patients are asymptomatic initially.[14–16] Indeed, the indication for surgery is rarely to treat elevated pressure. Rather, surgery is performed both to minimize the potential for future intracranial hypertension and to normalize the head shape, and much study is focused on anthropometric measurements and skull landmarks to represent skull variations in metopic CS patients.

The trigonocephalic head shape present in metopic CS results from bilateral constriction of the frontal bones with an associated parieto-occipital bossing.[17] Weinzweig et al[2] recently observed an endocranial metopic notch in 97% of metopic synostosis patients, helping to distinguish abnormal from normal suture fusion. While helpful as a diagnostic marker, the presence of an endocranial notch provides limited information on the exact timing of fusion and potential contribution of such a fusion on the degree of skull dysmorphology. Therefore, the effort to define objective measures for identifying the presence and severity of MC has evolved to the study of trigonocephalic components using anthropometric measurements. For example, Bottero et al[18] measured the ratio of the interparietal to intercoronal distance to describe the degree of frontal stenosis present. Similar methods to associate trigonocephalic severity with skull measurements have been conducted using the IFA, angle of the nasopterion,[3] interobital distance and bitemporal width,[19] intercanthal distance to midfacial width ratio and endocranial bifrontal angle,[20] interzygomaticofrontal suture and interdacryon distance,[21] and a trigonocephaly severity index calculated from outlined cranial shapes (Supplementary Digital Content, Table 4, http://links.lww.com/SCS/B139).[22]

While these measurements have proven helpful in describing the differences in trigonocephalic head shapes from normal, their influence over clinical decision making has been relatively limited. For example, many of these studies determined the influence of these variables on functional outcomes rather than morphological properties.[3,18] Similarly, much of the data from these studies shows a great amount of overlap between control and metopic groups. Beckett et al[21] found a significant degree of overlap in the endocranial bifrontal angle between metopic and control patients where angles in metopic patients ranged from 100 to 148 degrees and normal controls ranged from 134 to 160 degrees. Despite these factors, perhaps the biggest limitation of these studies is that they rely on measurements taken in 2D when MC is inherently a 3D deformity. With the advent of modern advanced analytical software and machine learning, the subtleties of the 3D head shape changes in craniosynostosis no longer need to be ignored so that the shape can be distilled into simple angles and dimensions which are measured by hand. More powerful 3D analytical techniques can evaluate magnitudes more data and help identify patterns which would otherwise be overlooked by simpler methods. This factor advocates strongly for our method of analysis using PCA and the Shapeworks Software. Our methodology analyzed the geometric relationships of 2048 different points on the 3D skull capturing 95% of the variability in head shapes. To our knowledge, there has been no other study that has been able to describe the morphologic aberrations in metopic head shapes using such a complete 3D map of the skull.

In the described model, we hypothesized that the severity seen on a CT scan of the skull could be predicted by the shape descriptor variable, generated from an analysis of the thousands of correspondence points located on the 3D skull, and correlated with the expert ratings from 18 craniofacial and neurosurgeons. The values from this analysis were then utilized in a modernized, machine learning algorithm called maximum likelihood estimation to calculate unknown parameters in our model that would complete the predicted severity for each skull.

To our knowledge, machine learning has never been applied to predict the severity of MC using 3D head shapes. The technique originates from advancements in computer science describing the ability of a computer to improve performance of a task without explicit programming for that task, essentially "learning" by pattern recognition and data. In our model, machine learning was utilized to provide more accurate predictions for severity after understanding the patterns and assigned rankings from expert raters. The effectiveness of this methodology is summarized by the correlation coefficients of our model, which were statistically significant at the 0.05 level in 7 of the 8 dimensions of analysis (with P-value of the eighth approaching significance at 0.056). A leave-one-out analysis yielded strong positive correlations after correction ($r^2 = 0.89$ and 0.97, respectively) and indicate that the shape descriptor utilized in this model was accurately predictive of a linear relationship to the overall severity of each CT scan.

While the majority of surgeons (95.3%) still rely heavily on the clinical assessment and results from CT imaging to determine the degree of trigonocephaly,[14] the IFA has recently been utilized in conjunction with these modalities to help determine severity. The IFA measures the angle formed when 2 lines are connected from the coronal sutures to the most anterior portion of the mid-line forehead on axial CT imaging.[23] This objective measurement was validated in 2012 by Kellogg et al[4] who observed a significantly more narrow-angle in metopic head shapes vs. normal controls (117.78 versus 144.8 degrees, $P < 0.0001$). Anolik et al[24] later built upon this model by using IFAs with expert rankings to classify the severity of MC. "Severe" synostosis corresponded to a ranking of 4 on the Likert scale with angles between 92.3 and 114.3 degrees, while normal head shapes corresponded to a median score of 1 with angles between 136.1 to 140.6 degrees. The authors found a significant level of agreement ($P < 0.0001$) among expert ratings for head shapes falling into the severe and minor strata; however, expert ratings varied significantly for IFAs between 114.3 and 136.1 degrees. Ultimately, the authors concluded that the IFA could help as an operative threshold, especially for angles greater than 118.2 degrees. As part of our analysis, the IFA was calculated for the 17 metopic skull shapes and then compared to our model. When compared to the severities assigned by our 18 craniofacial experts, our model misclassified the severity rankings of trigonocephalic head shapes statistically significantly less than the model utilizing IFA ($P = 0.019$). These results indicate that our model, built from an aggregate of the entire 3D skull, may more accurately predict the severity of metopic head shapes than previously utilized models, including the widely adopted IFA.

One of the limitations of this study includes the relatively small sample size for both metopic and normal control CT images. Furthermore, our model relied upon the ratings of craniofacial experts from both neurosurgical and plastic surgery disciplines potentially attributing to differences in the ratings assigned by each surgeon. Also, the 10-point Likert scale used by the surgeons in our Craniorate application was downsampled to a 5-point scale mathematically. Our analysis assumes that experts would rate skulls equally given a smaller scale, although this may not be accurate in practice. Another limitation arises from our assumption that the

IFA was linearly predictive of head shape severity; however, it is possible that the relationship between IFA and head shape is correlated but not totally linear. It should also be noted that our correlation coefficient was not statistically significant ($P = 0.056$) in 1 of the 8 dimensions. This undoubtedly skews the impact of our analysis and accuracy of our model. Similarly, the use of expert raters may re-introduce subjectivity into the severity ratings. Finally, the machine learning tool used for this study may not be easily usable in clinical practice as it requires powerful machinery and skilled personnel. Currently, work is underway to expand our dataset to validate our results and to develop an easy to use, automated interface to take advantage of this novel tool as well as to use deep machine learning without the use of expert raters.

Objective determination of the severity of metopic CS may help clinicians chose when to operate and when to manage the condition conservatively, determine predictors of poor surgical outcomes including the development of intracranial hypertension and aesthetic deformities, and allow investigators to compare outcomes for MC between institutions and surgeons. We present a new model for predicting MC severity, built using guided machine learning, incorporating shape models visualizing the entire 3D skull with ratings correlation from 18 expert craniofacial surgeons. This is the first study that has adopted machine learning techniques to enhance the predictive ability of a severity stratification model in MC, and the results indicate that such a model can potentially predict the severity of MC more accurately than previously adopted methods. These results will serve as the infrastructure on which a user-friendly technical application can be built for clinicians to use in the clinical and nonclinical setting for determining the severity of MC.

REFERENCES

1. Cornelissen M, Ottelander B, Rizopoulos D, et al. Increase of prevalence of craniosynostosis. *J Craniomaxillofac Surg* 2016;44:1273–1279
2. Weinzweig J, Kirschner RE, Farley A, et al. Metopic synostosis: defining the temporal sequence of normal suture fusion and differentiating it from synostosis on the basis of computed tomography images. *Plast Reconstr Surg* 2003;112:1211–1218
3. Oi S, Matsumoto S. Trigonocephaly (metopic synostosis). Clinical, surgical and anatomical concepts. *Childs Nerv Syst* 1987;3:259–265
4. Kellogg R, Allori AC, Rogers GF, et al. Interfrontal angle for characterization of trigonocephaly: part 1: development and validation of a tool for diagnosis of metopic synostosis. *J Craniofac Surg* 2012;23:799–804
5. Bayome M, Park JH, Kook Y-A. New three-dimensional cephalometric analyses among adults with a skeletal Class I pattern and normal occlusion. *Korean J Orthod* 2013;43:62–73
6. Lonic D, Sundoro A, Lin H-H, et al. Selection of a horizontal reference plane in 3D evaluation: identifying facial asymmetry and occlusal cant in orthognathic surgery planning. *Sci Rep* 2017;7:2157
7. Cates J, Fletcher TP, Martin S, et al. Shape modeling and analysis with entropy-based particle systems. *Inf Process Med Imaging* 2007;20:333–345
8. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Dankmarks Paedagogiske Institut; 1960
9. Uebersax JS, Grove WM. A latent trait finite mixture model for the analysis of rating agreement. *Biometrics* 1993:823–835
10. Lord FM, Novick MR, Brinbum A. *Statistical Theories of Mental Test Scores*. Oxford, England: Addison-Wesley; 1968
11. Muthén LK, Muthén BO. *Mplus User's Guide: Statistical Analysis with Latent Variables.* 3rd ed. Los Angeles, CA: Muthén & Muthén; 2007
12. James G, Witten D, Hastie T, et al. *An Introduction to Statistical Learning: With Applications in R*. Berlin, Germany: Springer; 2014
13. Tuite GF, Chong WK, Evanson J, et al. The effectiveness of papilledema as an indicator of raised intracranial pressure in children with craniosynostosis. *Neurosurgery* 1996;38:272–278
14. Yee ST, Fearon JA, Gosain AK, et al. Classification and management of metopic craniosynostosis. *J Craniofac Surg* 2015;26:1812–1817
15. Renier D, Sainte-Rose C, Marchac D, et al. Intracranial pressure in craniostenosis. *J Neurosurg* 1982;57:370–377
16. Cornelissen MJ, Loudon SE, van Doorn FE, et al. Very low prevalence of intracranial hypertension in trigonocephaly. *Plast Reconstr Surg* 2017;139:97e–104e
17. Kim HJ, Roh HG, Lee IW. Craniosynostosis: updates in radiologic diagnosis. *J Korean Neurosurg Soc* 2016;59:219–226
18. Bottero L, Lajeunie E, Arnaud E, et al. Functional outcome after surgery for trigonocephaly. *Plast Reconstr Surg* 1998;102: 952–958
19. Posnick JC, Lin KY, Chen P, et al. Metopic synostosis: quantitative assessment of presenting deformity and surgical results based on CT scans. *Plast Reconstr Surg* 1994;93:16–24
20. Diluna ML, Steinbacher DM. Simulated fronto-orbital advancement achieves reproducible results in metopic synostosis. *J Craniofac Surg* 2012;23:e231–e234
21. Beckett JS, Chadha P, Persing JA, et al. Classification of trigonocephaly in metopic synostosis. *Plast Reconstr Surg* 2012;130:442e–447e
22. Ruiz-Correa S, Starr JR, Lin HJ, et al. New severity indices for quantifying single-suture metopic craniosynostosis. *Neurosurgery* 2008;63:318–324
23. Wood BC, Mendoza CS, Oh AK, et al. What's in a name? Accurately diagnosing metopic craniosynostosis using a computational approach. *Plast Reconstr Surg* 2016;137:205–213
24. Anolik RA, Allori AC, Pourtaheri N, et al. Objective assessment of the interfrontal angle for severity grading and operative decision-making in metopic synostosis. *Plast Reconstr Surg* 2016;137:1548–1555

