



Artificial Intelligence Automation of Proptosis Measurement: An Indicator for Pediatric Orbital Abscess Surgery

Roxana Fu · Andriy Bandos · Joseph K. Leader · Samyuktha Melachuri ·
Tejus Pradeep · Aashim Bhatia · Srikala Narayanan · Ashley A. Campbell ·
Matthew Zhang · José-Alain Sahel · Jiantao Pu

Received: April 5, 2023 / Accepted: June 8, 2023 / Published online: June 23, 2023
© The Author(s) 2023

ABSTRACT

Introduction: To evaluate the ability of artificial intelligence (AI) software to quantify proptosis for identifying patients who need surgical drainage.

Methods: We pursued a retrospective study including 56 subjects with a clinical diagnosis of subperiosteal orbital abscess (SPOA) secondary to sinusitis at a tertiary pediatric hospital from 2002 to 2016. AI computer software was developed to perform 3D visualization and quantitative assessment of proptosis from computed

tomography (CT) images acquired at the time of hospital admission. The AI software automatically computed linear and volume metrics of proptosis to provide more practice-consistent and informative measures. Two experienced physicians independently measured proptosis using the interzygomatic line method on axial CT images. The AI software and physician proptosis assessments were evaluated for association with eventual treatment procedures as standalone markers and in combination with the standard predictors.

Results: To treat the SPOA, 31 of 56 (55%) children underwent surgical intervention,

R. Fu · J.-A. Sahel
Department of Ophthalmology, University of Pittsburgh, Pittsburgh, PA, USA

A. Bandos
Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, USA

J. K. Leader · J. Pu
Department of Radiology, University of Pittsburgh, Pittsburgh, PA, USA

J. K. Leader · J. Pu
Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA

S. Melachuri (✉)
School of Medicine, University of Pittsburgh, 1400 Locust St, Suite 500, Pittsburgh, PA 15219, USA
e-mail: melachurisr2@upmc.edu

T. Pradeep
Department of Ophthalmology, University of Pennsylvania, Philadelphia, PA, USA

A. Bhatia
Department of Radiology, Children's Hospital of Philadelphia, Philadelphia, PA, USA

S. Narayanan
Department of Radiology, Children's Hospital of Pittsburgh, Pittsburgh, PA, USA

A. A. Campbell
Department of Ophthalmology, Wilmer Eye Institute, Johns Hopkins, Baltimore, MD, USA

M. Zhang
Department of Ophthalmology, University of Washington, Seattle, WA, USA

including 18 early surgeries (performed within 24 h of admission), and 25 (45%) were managed medically. The physician measurements of proptosis were strongly correlated (Spearman $r = 0.89$, 95% CI 0.82–0.93) with 95% limits of agreement of ± 1.8 mm. The AI linear measurement was on average 1.2 mm larger ($p = 0.007$) and only moderately correlated with the average physicians' measurements ($r = 0.53$, 95% CI 0.31–0.69). Increased proptosis of both AI volumetric and linear measurements were moderately predictive of surgery (AUCs of 0.79, 95% CI 0.68–0.91, and 0.78, 95% CI 0.65–0.90, respectively) with the average physician measurement being poorly to fairly predictive (AUC of 0.70, 95% CI 0.56–0.84). The AI proptosis measures were also significantly greater in the early as compared to the late surgery groups ($p = 0.02$, and $p = 0.04$, respectively). The surgical and medical groups showed a substantial difference in the abscess volume ($p < 0.001$).

Conclusion: AI proptosis measures significantly differed from physician assessments and showed a good overall ability to predict the eventual treatment. The volumetric AI proptosis measurement significantly improved the ability to predict the likelihood of surgery compared to abscess volume alone. Further studies are needed to better characterize and incorporate the AI proptosis measurements for assisting in clinical decision-making.

Keywords: Orbital cellulitis; Subperiosteal orbital abscess; Proptosis; Computed tomography (CT); Automation

Key Summary Points

Surgical indications for pediatric orbital abscess drainage are vague or equivocal in literature.

Clinical measurement of proptosis in an acutely ill child is challenging, leading to incomplete ophthalmic exams.

This study seeks to automate CT proptosis measurements and evaluate any association with the need for eventual surgery.

Automated proptosis measurements are associated with the need for surgery, and its predictive ability is improved when combined with orbital abscess measurements.

Automation provides for the precise and accurate measurement of proptosis, calling for the need of an integrative tool that can provide these metrics in the clinical setting to inform decision-making.

INTRODUCTION

Orbital cellulitis is the most common sequela of sinusitis [1] and can lead to serious complications. Occurring predominately in the pediatric population, orbital cellulitis may lead to blindness, cavernous sinus thrombosis, meningitis, subdural empyema, and brain abscess [1–7]. Ocular examinations in the pediatric population are limited by their cooperation, mental status, and acute illness, leading to computed tomography (CT) as the most accurate modality for the diagnosis and extent of infection [8, 9]. Proptosis, gaze restriction, and pain with eye movement are common indicators of orbital extension, with their presence often prompting the need for CT scanning. While the presence of proptosis is a consistent sign of orbital cellulitis, distinguishing it from preseptal cellulitis, its ability to predict the presence of a subperiosteal orbital abscess (SPOA), or the subsequent need for abscess drainage have been equivocal in the literature (Table 1) [10–14].

While abscess volume continues to be the most established risk factor for surgery, proptosis has been both negatively [15–17] and positively associated with surgical intervention [12, 18–20]. This discrepancy in the literature is likely secondary to small sample size, inconsistent measurement of proptosis, and inability to record a complete pediatric ophthalmic exam, which may happen in 38–43% of patients [15, 21].

Table 1 Studies of pediatric SPOA with proptosis analyzed as a variable for surgery

Study (year)	N = SPOAs, (patients with proptosis measured)	Method of measurement	Threshold for surgery (statistics)	Predictive of surgery? Statistics
Rahbar et al. [18]	19, (N/A)	Not reported	> 2 mm (descriptive only, reported as 92% chance of surgery within their cohort)	Yes. Multivariate analysis ($p = 0.003$)
Brown et al. [15]	42, (not reported, authors noted 43% of patients had incomplete ophthalmic exams)	Hertel in an unreported number of patients	N/A	No. Univariate analysis ($p = 0.3997$)
Oxford and McClay [19]	43, (17)	17 patients with Hertel measurement, 26 patients only noted as presence or absence of proptosis	> 5 mm (phi coefficient = 0.468)	Yes. Univariate analysis ($p = 0.008$)
Tabarino et al. [20]	32, (32)	Physician CT measurements using interzygomatic line (to center of eye), two readers	> 4 mm (not reported)	Yes. Univariate analysis ($p = 0.02$)
Nation et al. [16]	48, (N/A)	Patients only noted as presence or absence of proptosis	N/A	No. Multivariate analysis
Quintanilla-Dieck [17]	40, (40)	Physician CT measurements using interzygomatic line (to posterior globe)	N/A	No. Multivariate analysis (p value not reported)
Murphy et al. [12]	27, (N/A)	Patients only noted as presence or absence of proptosis	Not reported	Yes. Multivariate analysis ($p < 0.001$)
Fu et al. [31]	56, (56)	Physician CT measurements using interzygomatic line (to posterior globe), two readers. Automated linear and volumetric proptosis measurements	Not recommended	Yes. Multivariate analysis ($p < 0.01$, for all measures of proptosis)

N/A not applicable

Proptosis, anterior displacement of the globe, has been difficult to accurately and consistently measure in diseased states. While it is an important indicator of orbital pathology, common errors in measurement occur with vertical globe displacement, soft tissue or orbital rim changes, lack of uniform technique, and

parallax [22]. Hertel exophthalmometry (Hertel) is the most widely used measurement of proptosis [23, 24], but it was reported to have only moderate inter-reader agreement in a study of normal Chinese adults [25]. To clinically assess proptosis, the apex of the cornea is measured after seating the foot plates of the

Table 2 Non-proptosis characteristics of patients at the time of admission by the eventual treatment group

Patients' characteristics at the time of admission	Medical (<i>n</i> = 25)	Late surgery (<i>n</i> = 13)	Early surgery (<i>n</i> = 18)	<i>p</i> value
Age, mean (std)	4.84 (3.9)	6.77 (3.7)	6.44 (4.4)	0.165
Days on previous IV or pills, mean (std)	0.96 (1.0)	0.69 (0.9)	1.17 (1.2)	0.501
Abscess volume, mean (std)	334 (351)	822 (854)	1124 (544)	< 0.001
Sex (female), <i>n</i> (% in treatment group)	5 (20%)	3 (23%)	3 (17%)	0.901
Nasal steroids (yes), <i>n</i> (% in treatment group)	8 (32%)	7 (54%)	10 (56%)	0.258

Hertel instrument at the lateral canthus. Such requirements are complicated in the pediatric patient with orbital cellulitis, who routinely presents with eyelid edema, pain, and globe deviation or displacement. CT scans have been used as a substitute for the measurement of globe position. Physician measurement of proptosis on CT scans has been found to have good agreement with Hertel in normal populations and in patients with orbital pathology (e.g., orbital fractures, thyroid eye disease) [23, 26–29].

We developed and tested artificial intelligence (AI) software to perform 3D visualization and quantitative assessment of proptosis depicted on CT images. Our goal is to significantly improve the accuracy and consistency of quantifying globe protrusion. The influence of proptosis in relation to abscess volume and its role as a risk factor for surgical intervention was investigated in this retrospective study.

METHODS

Study Population

The study was approved by the Institutional Review Board (IRB) of University of Pittsburgh and adhered to the tenets of the Declaration of Helsinki. Informed consent was deferred with IRB approval. A single-institution, retrospective chart review was performed to identify patients with CT-confirmed SPOA from 2002 to 2016 at UPMC Children's Hospital of Pittsburgh. Patients with orbital cellulitis were identified

using ICD9 and ICD10 codes for preseptal cellulitis (373.13) and orbital cellulitis (376.01). Demographic characteristics including age, sex, use of nasal steroids, and treatment modalities including use of IV antibiotics and surgical drainage were collected for all subjects in the study cohort (Table 2).

The results of the chart review were narrowed to identify cases that met the inclusion criteria: (1) the presence of SPOA on CT scan, (2) access to the CT images for analysis, and (3) axial CT images were not severely degraded by motion artifact or patient positioning to preclude clinician planar measurement of proptosis. Cases not meeting these criteria were excluded. Ultimately, 56 cases met our inclusion criteria, with 31 (55%) undergoing surgery and 25 (45%) managed with medical therapies alone. Abscess volumes for each subject were computed by manually outlining each abscess with lesion-mapping software in a previous study, which was published in the literature [30].

Subjects were also grouped into “early” and “late” surgical intervention. The criteria to be assigned to the early surgery group was surgery performed within 24 h of admission.

Proptosis Assessment

Maxillofacial CT scans were performed to evaluate the presence of orbital abnormalities. The CT scans were performed on a 16-detector (*n* = 12) or 64-detector (*n* = 51) CT scanner (GE Healthcare, Waukesha, WI, USA). The scans

were acquired using a helical technique with radiopaque contrast. Images were reconstructed to encompass the entire orbit using the GE “bone” at an image thickness of 0.625–1.25 mm within an interval from 0.325 to 1.25 mm. In-house software was used to assess proptosis depicted on the CT images.

Physician Planar Measurement of Proptosis

Proptosis was measured by two independent expert reviewers: a neuroradiologist (AB) and an oculoplastics surgeon (RF). The readers were blind to subject history, demographics, hospital course, and radiology reports. Proptosis was measured on a single axial CT image (planar measurement) by measuring posterior globe perpendicular to the interzygomatic line. This was performed for both eyes, and the absolute difference between the two eyes represented relative proptosis (Fig. 1) [17].

Automation of Globe Displacement

In-house convolutional neural network (CNN) software was used to automatically identify the globe and postseptal region depicted on CT images (Fig. 2a–e), which was trained and validated on the same dataset [31]. The software



Fig. 1 Axial CT image of a representative subject with left-sided proptosis, periorbital edema, ethmoid sinusitis, and trace medial subperiosteal abscess. Physician measurement of relative proptosis by measuring the posterior globe perpendicular from the interzygomatic line

was initially developed to visualize and quantify SPOA depicted on CT images. Once the globe and the postseptal regions are segmented, the AI software computed two different 3D metrics of proptosis that quantified the displacement of the globe from the orbit: (1) linear displacement and (2) volume displacement. The displacement distance “ d ” was computed by first identifying the intersection plane between the globe and the postseptal region (Fig. 3b). Then, the point on the globe surface that was outside the extraconal region and had the farthest anterior distance (i.e., the distance d) to the intersection plane was identified. This represents the maximum anterior projection of the globe relative to the postseptal region. The linear proptosis measurement was defined at the difference between the distance “ d ” between each eye. The volume displacement was computed as the difference between the globe volume anterior to the postseptal region of each eye (Fig. 3a–c).

Statistical Analysis

The statistical analysis was based on non-parametric methods implemented using SAS statistical software (SAS, v.9.4, SAS Institute, Cary NC). The proptosis measurements (relative protrusion in millimeters) were correlated between physicians and compared with the AI (using rank-based Spearman correlation and exact Wilcoxon test at a two-sided significance level of 0.05); the level of inter-reader agreement was quantified using the 95% limits of agreement [32]. To assess possible imbalances, demographic and diagnosis-related parameters were compared across the three treatment groups (using the overall Kruskal–Wallis test). Individual predictiveness of the proptosis measurements was evaluated on the basis of the empirical receiver operating characteristics (ROC) curves for medical versus surgical groups combined as well as for early versus late surgical subgroups (the 95% confidence intervals for the area under the ROC curve, AUC, were constructed using the method of DeLong et al. [33]). The added significance of the proptosis measurements for predicting the likelihood of surgery was evaluated within the logistic

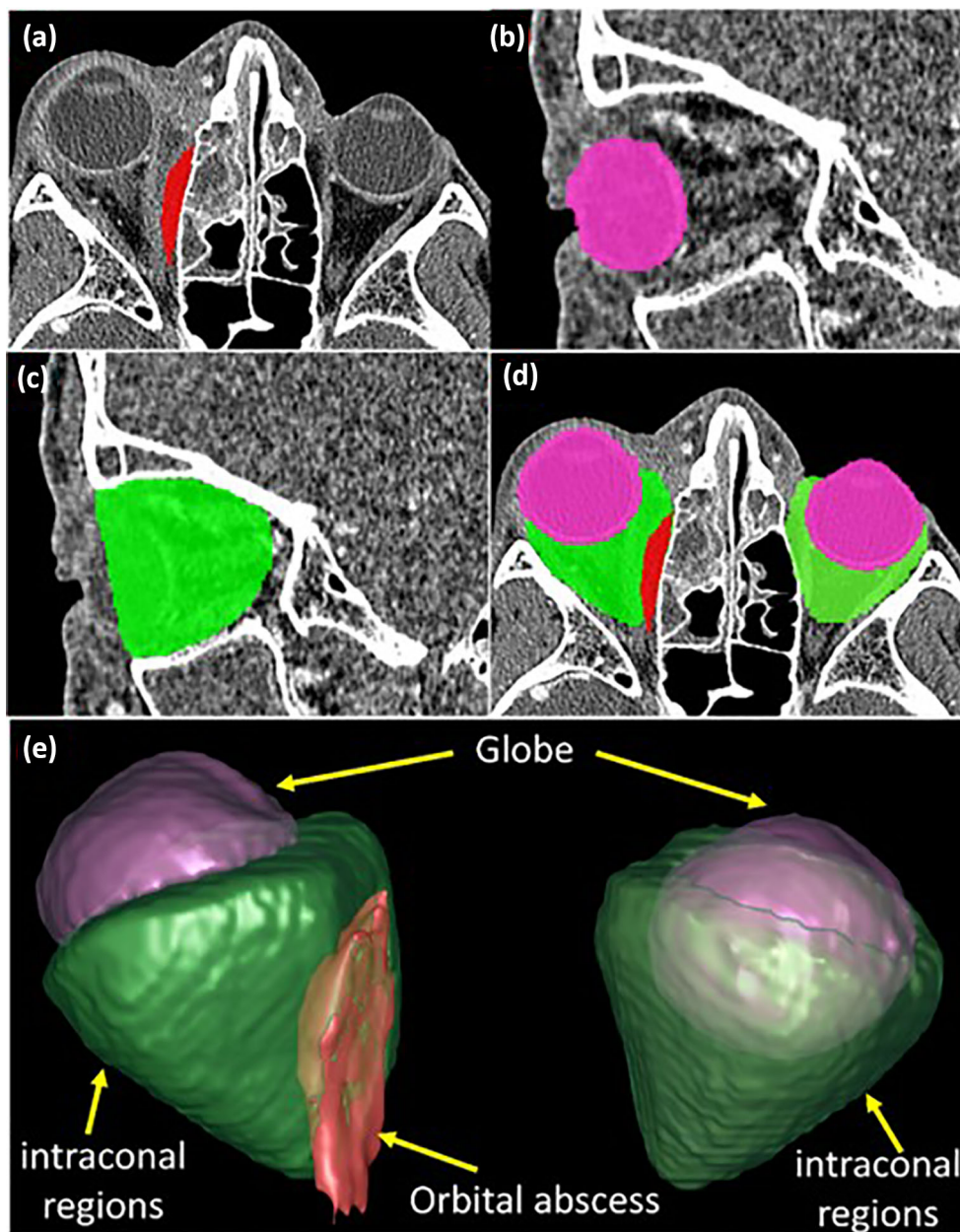


Fig. 2 Automated identification of orbital abscess (a), globe (b), and postseptal region (c) and their visualization (d, e)

regression model built on the basis of the demographic and diagnosis-related parameters. The predicted probability of future surgery was used to estimate the overall level of predictive ability for surgery and to illustrate possible decision rules for predicting respectively a medical treatment, late surgery, and early surgery group, with the accuracy of the resulting

classifications quantified with the Somers' D statistic.

RESULTS

Of 56 children with CT-confirmed SPOAs that met our study inclusion criteria 31 of 56 (55%) underwent surgical intervention, including 18

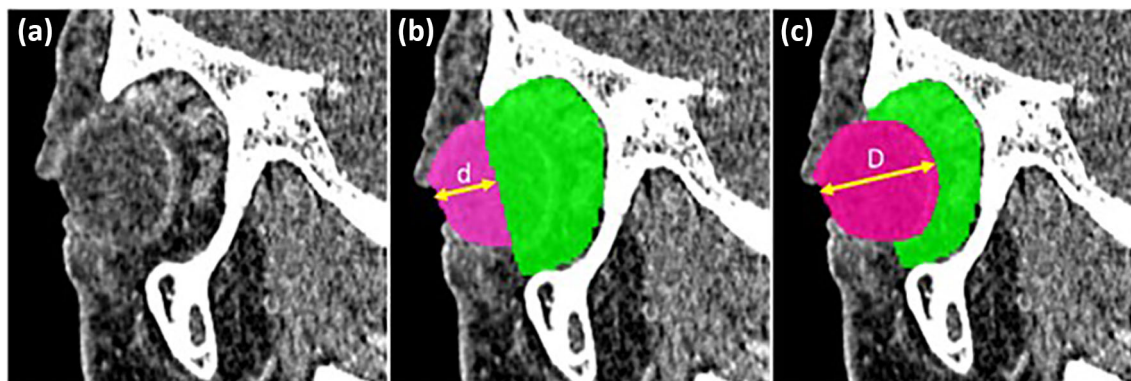


Fig. 3 Proptosis assessment by artificial intelligence (AI) software. **a** Original CT image, **b** “*d*” is the farthest 3D distance from the globe’s anterior surface to the intersection plane between the globe and the postseptal region. **c** “*D*” is the diameter of the globe. The AI software linear proptosis measurement is defined as the difference between

“*d*” of each eye. The AI volume measurement of proptosis is defined as the volume of the globe beyond the postseptal region, with the relative volume of proptosis calculated as the volume difference between each eye

early surgeries (performed within 24 h of admission), and 25 (45%) were managed medically. The median age of the cohort at presentation was 4 years old, with 45 male and 11 female patients (Table 2). There was no difference in gender, age, or prior treatment between the surgical and medical groups, but the abscess volume was significantly larger for patients who underwent surgery, especially early surgery ($p < 0.0001$).

The physician’s measurements of proptosis were strongly correlated (Spearman $r = 0.89$, 95% CI 0.82–0.93) with 95% limits of agreement of ± 1.8 mm (Fig. 4). Six physician measurements demonstrated a negative protrusion of the abscess-affected side (with globe protrusion measuring greater on the unaffected side).

The physicians’ AI linear measure of proptosis was on average 1.2 mm larger ($p = 0.007$) than the average of two physicians’ average, with the difference prevailing across the treatment groups (Table 3). The AI linear proptosis measurement was positive in all six cases when the average physicians’ measure was negative. There was only one instance when the AI linear measurement was negative (-1.1 mm, with physicians’ measurements of 5.2 mm, and 5.9 mm). The larger values for AI linear measurements prevailed for the subset of cases with

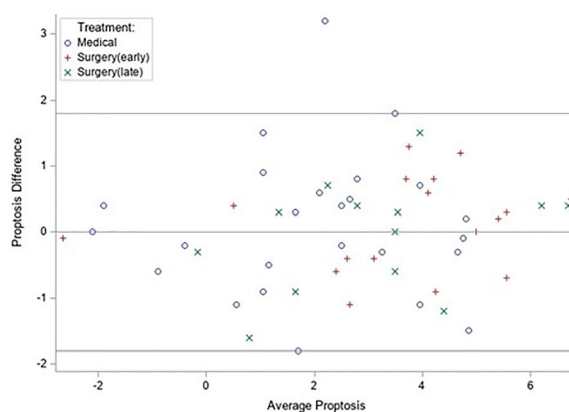


Fig. 4 Bland–Altman plot of the difference between by the average of the proptosis measurements of two physicians (the 95% limits of agreement are ± 1.8 mm, based on the standard deviation of the difference of 0.9 mm)

only positive physicians’ measurements (overall and within treatment groups).

A potentially more informative AI volumetric proptosis modality had positive proptosis values in all cases. Both linear and volumetric AI measurements were only moderately correlated with the average physicians’ measurement of proptosis (Spearman correlations of 0.53, 95% CI 0.31–0.69, and 0.57, 95% CI 0.35–0.72, respectively).

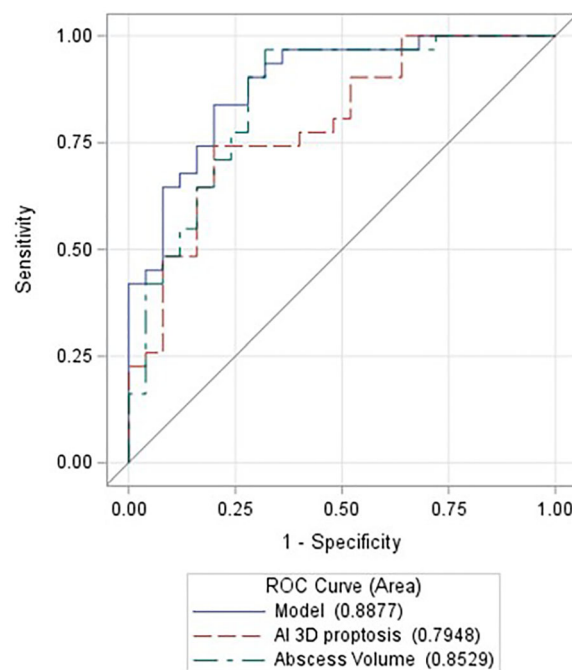
Table 3 Distribution of possible predictors in treatment groups and their AUCs for differentiating between medical and surgery group as well as between early and late surgery

Potential predictors	Surgery ($n = 31$) versus medical ($n = 25$)			Early ($n = 18$) versus late ($n = 13$) surgery		
	Medical Mean (std)	Surgery Mean (std)	AUC (Wilcoxon p value)	Late Mean (std)	Early Mean (std)	AUC (Wilcoxon p value)
Clinical proptosis (reader 1)	2.1 (2.1)	3.5 (2.2)	0.69 (0.02)	3.1 (2.2)	3.8 (2.3)	0.63 (0.22)
Clinical proptosis (reader 2)	2 (2.1)	3.5 (2)	0.71 (0.01)	3.1 (1.9)	3.7 (2.2)	0.62 (0.27)
Clinical proptosis (average)	2.1 (2.0)	3.5 (2.1)	0.70 (0.01)	3.1 (2)	3.7 (2.2)	0.64 (0.21)
AI proptosis 2D	3.1 (1.6)	4.8 (2.0)	0.78 (< 0.001)	4.2 (1.9)	5.2 (2.0)	0.72 (0.04)
AI proptosis 3D	1.2 (0.8)	2.2 (0.9)	0.79 (< 0.001)	1.8 (1)	2.5 (0.7)	0.76 (0.02)

Individually, all proptosis measurements varied significantly across the three treatment groups ($p < 0.028$) being significantly lower for the medical than surgery group and somewhat lower for late than early surgery subgroups (Table 3). The AI volumetric proptosis measurement was the most predictive, achieving the AUC of 0.79 (95% CI 0.68–0.91) for discriminating between the medical and surgery groups and the AUC of 0.76 (95% CI 0.55–0.97) for discriminating between the early and late surgical subgroups (Table 3).

Adjusting for the abscess volume (a known indicator for surgery and the only non-proptosis factor significantly varying across the treatment groups), only the AI volumetric proptosis method was statistically significant ($p = 0.03$), indicating the potential to improve the prediction of the likelihood of surgery on the basis of the abscess volume alone. The small sample size of the current dataset does not allow statistically supporting the specific aspects of improvements in predictive ability beyond the implications of the overall improvement in risk prediction indicated by the multivariable logistic regression. However, the data indicate the potential of the AI volumetric proptosis to improve the overall discrimination between the surgical and medical groups (combined marker's empirical AUC of 0.89) and improve the ability to confidently predict a substantial fraction of surgeries

while maintaining the ability to confidently identify a substantial fraction of the medical group (Fig. 5).

**Fig. 5** ROC curve for discriminating between the medical and surgical groups on the basis of the abscess volume, AI volumetric proptosis, and their linear combination (model)

DISCUSSION

Outside of emergent indications, such as vision loss or CNS involvement, the decision to intervene surgically remains complex for orbital abscesses. Historically, criteria for initial close observation with medical management were described by Garcia [9]. Since then, many publications have both disputed and supported these criteria. Abscess volume appears to be the most significant predictive risk factor for surgery [34]. The “large” abscess as described by Garcia has been more recently described with precise volume thresholds [16, 17, 20, 30, 35–37] culminating in the development of lesion mapping software and automation to measure abscess volumes [30, 31]. However, while accurate, the clinical significance of these automated abscess volumes has yet to be applied in larger or prospective studies.

Proptosis was not included in the indications for surgery by Garcia [9]. Several recent relatively large studies reported no association between proptosis and the need for surgical intervention (Table 1) [16, 17]. The predictive value of proptosis for surgery has been equivocal, likely related to small sample sizes and the difficulties in obtaining exophthalmometry measurements in the pediatric population. In the acutely ill child, factors such as age, cooperation, changes in globe position, and presence of eyelid edema and pain can lead to the lack of or inaccurate proptosis measurements [26]. CT scans can serve as a substitute in these conditions. Measurement of proptosis on a CT image can be much more straightforward, with clear, highly contrasted areas available for both physician estimation and computer automation.

The physicians in our study measured proptosis using the interzygomatic line method, which is defined as the distance from lateral rims to apex of the globe. It was chosen because of its high correlation with Hertel readings and ease of use [17]. While physician measurements had moderate agreement with the AI software linear measurements, six of the physician measurements indicated no proptosis on the

affected side. While relative proptosis may be negative in cases of preexisting orbital pathology, none was seen radiographically within our cohort. Given that the AI linear had only one negative measurement and AI volumetric had none, the authors extrapolate that the AI measurements were more reflective of the pathology seen on CT imaging. This trend is likely due to the limitations of the physician measurements using only 2D anatomy with a singular axial plane. The relationship of the lateral orbital rims to globe is influenced by patient positioning such as head tilt, and orbital pathology causing additional, non-axial globe displacement relative to the unaffected globe. These changes limit the accuracy of the 2D physician measurements but are more accounted for when using 3D landmarks to measure AI linear and volumetric proptosis.

The AI software used 3D reconstruction of the globe and postseptal regions and compared their maximal anterior distance to measure of proptosis, reporting a linear assessment measuring from the septal plane to the anterior apex of the globe and a volumetric assessment, measuring the entire volume of globe protrusion past the septal plane. Compared to physicians' linear method, both automated linear and volumetric methods achieved substantially higher levels of the overall predictive ability, although the current dataset is too small to provide the corresponding statistical support (AUCs of 0.777 and 0.795 versus AUC of 0.699, with p of 0.07 and 0.04, respectively). In the current study, the AI measures were significantly greater for the early compared to late surgical subgroups with physician measurements unable to meaningfully differentiate between these subgroups.

AI volumetric measurements had the strongest predictive ability for surgery overall and early versus late surgery among the other proptosis modalities and was the only proptosis measurement to improve the predictive ability of abscess volume ($p = 0.03$). The consideration of AI volumetric measurement in addition to the abscess volume fractionally improved the empirical ROC curve overall and especially in the region of high specificity (indicating the potential for more confident prediction of

surgery). Yet, the small sample size of the current dataset does not allow statistically supporting the specific aspects of improvements in predictive ability beyond the implications of the overall improvement in prediction of the likelihood of surgery indicated by the multivariable logistic regression (for surgery versus medical groups as well as for the three treatment groups simultaneously). Volumetric analysis measures the entire globe displacement relative to each respective orbital rim, which decreases the errors related to the assumption of normal surrounding orbital and periorbital tissue with orthogonal planes and uniform symmetry.

To our knowledge, this is the first use of computer software to quantify proptosis in a volumetric fashion to help improve the prediction of the likelihood of surgery. Willaert et al. [38] described using CT imaging technology to measure proptosis in control and thyroid eye disease patients before and after surgical decompression, but the clinicians were still required to choose maximum diameters of the globes for the creation of a sphere. Chaganti and colleagues [39] used computer software to measure orbital landmarks in patients with thyroid eye disease, including globe position; however, it was unclear if relative proptosis was defined in terms of length or volume and no surgical threshold was defined.

We have developed automated measures of proptosis which are significantly different from the physician assessments and, at the same time, provide a strong marker of future surgery. The volumetric automated measures also show the potential to provide substantial improvements to the predictions based on the abscess volume. Larger studies are needed to better define these radiographic markers to assess patients with SPOA and to better characterize and incorporate the AI proptosis measurements for assisting in clinical decision-making.

LIMITATIONS

This is a retrospective, single-institution study of SPOAs with limited sample size. Small data size prevents meaningful assessment of the

specific characteristics of improvement in predicting the eventual treatment from adding AI volumetric measure of proptosis (i.e., AUC, sensitivity at high specificity or specificity at high sensitivity). External validation or a larger study is needed to obtain reliable estimates for the decision rule and the related characteristics. In this same vein, a surgical threshold for proptosis is not recommended by the authors given the small sample size, despite our study being the largest for the analysis of proptosis as an indication for surgery (Table 1).

In a retrospective study, it is always difficult to determine if the clinical chosen treatment approach is indeed the correct choice. This can confound analyses for determining appropriate cutoff values for when or when not to select the appropriate treatment. It can be argued that increased proptosis may be associated more so with decision to pursue surgery, wherein increased proptosis decreases the surgeon's threshold for surgery, rather than being a true indicator of severe disease recalcitrant to medical management. The influence of proptosis on surgeons compared to abscess volume is likely decreased given the inherent difficulties in accurately quantifying proptosis. The lack of a clinical gold standard for comparison is notable, but as discussed, the accuracy of Hertel exophthalmometry in this acute setting would be subject to common errors leading to inaccurate measurements. There may be local clinical factors that influence degree of proptosis, such as microbiologic profile; a robust comparison of clinical markers associated with proptosis is beyond the scope of this study. As a single-institution study, the results may or may not generalize to other institutions.

CONCLUSION

Physicians using orbital landmarks to measure proptosis in a 2D plane agreed moderately with 3D AI software to compute similar linear measurements. All three approaches to compute proptosis (i.e., physicians' manual, AI linear, and AI volume) were found to be markers associated with the need for surgical drainage of SPOAs. The AI volumetric of proptosis was the

most robust performing metric in terms of identifying which patients ultimately underwent surgery, predicting early vs late surgery, and was the only measure of proptosis to improve the ability of abscess volume to discriminate surgical vs medical patients. The AI-estimated volumetric measure of proptosis can improve the prediction of the eventual treatment approach. Larger studies are needed to better estimate the specific improvement characteristics and optimal surgical vs medical management thresholds. An integrative, automated tool is necessary to improve early treatment decision-making and be available in an accessible clinical application.

ACKNOWLEDGEMENTS

Funding. This work is supported in part by the National Institutes of Health (NIH) (Grant No. R01CA237277) and the UPMC Hillman Developmental Pilot Program. No funding or sponsorship was received for the publication of this article.

Author Contributions. Roxana Fu MD: concept, design, manuscript preparation and review. Andriy Bandos PhD: statistical analysis. Joseph K. Leader PhD: statistical analysis. Samyuktha Melachuri MD: manuscript preparation and submission. Tejus Pradeep MD: manuscript preparation, statistical analysis. Aashim Bhatia MD: manuscript preparation and review. Srikala Narayanan MD: radiology review and guidance. Ashley A. Campbell MD: review and evaluation of radiology. Matthew Zhang MD: manuscript preparation and review. José-Alain Sahel MD: project overview and guidance. Jiantao Pu PhD: statistical analysis.

Disclosures. All named authors confirm that no conflicting relationship exists.

Compliance with Ethics Guidelines. The study was approved by the Institutional Review Board of University of Pittsburgh and adhered to the tenets of the Declaration of Helsinki.

Informed consent was deferred with IRB approval.

Data Availability. Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Wald ER, Applegate KE, Bordley C, et al. Clinical practice guideline for the diagnosis and management of acute bacterial sinusitis in children aged 1 to 18 years. *Pediatrics*. 2013;132:e262–80.
2. Liu IT, Kao SC, Wang AG, Tsai CC, Liang CK, Hsu WM. Preseptal and orbital cellulitis: a 10-year review of hospitalized patients. *J Chin Med Assoc*. 2006;69:415–22.
3. Davis JP, Stearns MP. Orbital complications of sinusitis: avoid delays in diagnosis. *Postgrad Med J*. 1994;70:108–10.
4. Wagenmann M, Naclerio RM. Complications of sinusitis. *J Allergy Clin Immunol*. 1992;90:552–4.
5. Swift AC, Charlton G. Sinusitis and the acute orbit in children. *J Laryngol Otol*. 1990;104:213–6.
6. Moloney JR, Badham NJ, McRae A. The acute orbit. Preseptal (periorbital) cellulitis, subperiosteal

- abscess and orbital cellulitis due to sinusitis. *J Laryngol Otol Suppl.* 1987;12:1–18.
7. Jain A, Rubin PA. Orbital cellulitis in children. *Int Ophthalmol Clin.* 2001;41:71–86.
 8. Ryan JT, Preciado DA, Bauman N, et al. Management of pediatric orbital cellulitis in patients with radiographic findings of subperiosteal abscess. *Otolaryngol Head Neck Surg.* 2009;140:907–11.
 9. Garcia GH, Harris GJ. Criteria for nonsurgical management of subperiosteal abscess of the orbit: analysis of outcomes 1988–1998. *Ophthalmology.* 2000;107:1454–8.
 10. Santos JC, Pinto S, Ferreira S, Maia C, Alves S, da Silva V. Pediatric preseptal and orbital cellulitis: a 10-year experience. *Int J Pediatr Otorhinolaryngol.* 2019;120:82–8.
 11. Danishyar A, Sergent SR. Orbital cellulitis. Treasure Island: StatPearls; 2022
 12. Murphy DC, Meghji S, Alfiky M, Bath AP. Paediatric periorbital cellulitis: a 10-year retrospective case series review. *J Paediatr Child Health.* 2021;57:227–33.
 13. Jabarin B, Eviatar E, Israel O, Marom T, Gavriel H. Indicators for imaging in periorbital cellulitis secondary to rhinosinusitis. *Eur Arch Otorhinolaryngol.* 2018;275:943–8.
 14. Huang SF, Lee TJ, Lee YS, Chen CC, Chin SC, Wang NC. Acute rhinosinusitis-related orbital infection in pediatric patients: a retrospective analysis. *Ann Otol Rhinol Laryngol.* 2011;120:185–90.
 15. Brown CL, Graham SM, Griffin MC, et al. Pediatric medial subperiosteal orbital abscess: medical management where possible. *Am J Rhinol.* 2004;18:321–7.
 16. Nation J, Lopez A, Grover N, Carvalho D, Vinocur D, Jiang W. Management of large-volume subperiosteal abscesses of the orbit: medical vs surgical outcomes. *Otolaryngol Head Neck Surg.* 2017;157:891–7.
 17. Quintanilla-Dieck L, Chinnadurai S, Goudy SL, Virgin FW. Characteristics of superior orbital subperiosteal abscesses in children. *Laryngoscope.* 2017;127:735–40.
 18. Rahbar R, Robson CD, Petersen RA, et al. Management of orbital subperiosteal abscess in children. *Arch Otolaryngol Head Neck Surg.* 2001;127:281–6.
 19. Oxford LE, Mcclay J. Medical and surgical management of subperiosteal orbital abscess secondary to acute sinusitis in children. *Int J Pediatr Otorhinolaryngol.* 2006;70:1853–61.
 20. Tabarino F, Elmaleh-Bergès M, Quesnel S, Lorrot M, Van Den Abbeele T, Teissier N. Subperiosteal orbital abscess: volumetric criteria for surgical drainage. *Int J Pediatr Otorhinolaryngol.* 2015;79:131–5.
 21. Pereira KD, Mitchell RB, Younis RT, Lazar RH. Management of medial subperiosteal abscess of the orbit in children—a 5 year experience. *Int J Pediatr Otorhinolaryngol.* 1997;38:247–54.
 22. Davanger M. Principles and sources of error in exophthalmometry. A new exophthalmometer. *Acta Ophthalmol.* 1970;48:625–33.
 23. Kim IT, Choi JB. Normal range of exophthalmos values on orbit computerized tomography in Koreans. *Ophthalmol J.* 2001;215:156–62.
 24. Nkenke E, Maier T, Benz M, et al. Hertel exophthalmometry versus computed tomography and optical 3D imaging for the determination of the globe position in zygomatic fractures. *Int J Oral Maxillofac Surg.* 2004;33:125–33.
 25. Lam AK, Lam CF, Leung WK, Hung PK. Intra-observer and inter-observer variation of Hertel exophthalmometry. *Ophthalmol Physiol Optics.* 2009;29:472–6.
 26. Park NR, Moon JH, Lee JK. Hertel exophthalmometer versus computed tomography scan in proptosis estimation in thyroid-associated orbitopathy. *Clin Ophthalmol.* 2019;13:1461–7.
 27. Bingham CM, Sivak-Callcott JA, Gurka MJ, et al. Axial globe position measurement: a prospective multicenter study by the international thyroid eye disease society. *Ophthalmic Plast Reconstr Surg.* 2016;32:106–12.
 28. Tan NYQ, Leong Y, Lang SS, Htoon ZM, Young SM, Sundar G. Radiologic parameters of orbital bone remodeling in thyroid eye disease. *Invest Ophthalmol Vis Sci.* 2017;58:2527–33.
 29. Ramli N, Kala S, Samsudin A, Rahmat K, Abidin ZZ. Proptosis-correlation and agreement between Hertel exophthalmometry and computed tomography. *Orbit.* 2015;34:257–62.
 30. Pradeep T, Kalra G, Leader JK, et al. Lesion-mapping software for the measurement of subperiosteal abscess volume for pediatric orbital cellulitis: a retrospective chart review. *Ophthalmol Ther.* 2021;10:261–71.
 31. Fu R, Leader JK, Pradeep T, et al. Automated delineation of orbital abscess depicted on CT scan using deep learning. *Med Phys.* 2021;48:3721–9.

32. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8(2):135–60.
33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44:837–45.
34. Adil EA, Muir ME, Kawai K, Dombrowski ND, Cunningham MJ. Pediatric subperiosteal abscess secondary to acute sinusitis: a systematic review and meta-analysis. *Laryngoscope.* 2020;130:2906–12.
35. Todman MS, Enzer YR. Medical management versus surgical intervention of pediatric orbital cellulitis: the importance of subperiosteal abscess volume as a new criterion. *Ophthalmic Plast Reconstr Surg.* 2011;27:255–9.
36. Gavriel H, Yeheskeli E, Aviram E, Yehoshua L, Eviatar E. Dimension of subperiosteal orbital abscess as an indication for surgical management in children. *Otolaryngol Head Neck Surg.* 2011;145:823–7.
37. Le TD, Liu ES, Adatia FA, Buncic JR, Blaser S. The effect of adding orbital computed tomography findings to the Chandler criteria for classifying pediatric orbital cellulitis in predicting which patients will require surgical intervention. *J Am Assoc Pediatric Ophthalmol Strabismus.* 2014;18:271–7.
38. Willaert R, Shaheen E, Deferm J, Vermeersch H, Jacobs R, Mombaerts I. Three-dimensional characterisation of the globe position in the orbit. *Graefes Arch Clin Exp Ophthalmol.* 2020;258:1527–32.
39. Chaganti S, Mundy K, DeLisi MP, et al. Assessment of orbital computed tomography (CT) imaging biomarkers in patients with thyroid eye disease. *J Digit Imaging.* 2019;32:987–94.