Title

Utilization of Machine Learning for the Objective Assessment of Rhinoplasty Outcomes

Authors
Topsakal, Oguzhan
Dobratz, Eric J
Akbas, Mustafa Ilhan
et al.

Publication Date
2023

DOI
10.1109/access.2023.3270438

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 14 March 2023, accepted 23 April 2023, date of publication 26 April 2023, date of current version 3 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3270438

RESEARCH ARTICLE

Utilization of Machine Learning for the Objective Assessment of Rhinoplasty Outcomes

OGUZHAN TOPSAKAL[1], (Senior Member, IEEE), ERIC J. DOBRATZ[2], MUSTAFA ILHAN AKBAS[3], (Member, IEEE), WILLIAM M. DOUGHERTY[2], TAHIR CETIN AKINCI[4,5], (Senior Member, IEEE), AND MEHMET MAZHAR CELIKOYAR[6]

[1]Department of Computer Science, Florida Polytechnic University, Lakeland, FL 33805, USA
[2]Eastern Virginia Medical School, Norfolk, VA 23507, USA
[3]Electrical Engineering and Computer Science Department, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA
[4]Winston Chung Global Energy Center (WCGEC), University of California at Riverside, Riverside, CA 92521, USA
[5]Electrical Engineering Department, Istanbul Technical University, 34469 Istanbul, Turkey
[6]Department of Otolaryngology, School of Medicine, Demiroğlu Bilim University, 34394 Istanbul, Turkey

Corresponding author: Tahir Cetin Akinci (tahircetin.akinci@ucr.edu)

ABSTRACT Machine Learning started to provide solutions to various challenges in many fields, including medicine. The objective assessment of rhinoplasty results has been a challenge since the assessment of beauty is subjective in nature. This study explores if Machine Learning can be used to accomplish the complex task of objective evaluating the outcome evaluation and automated scoring for rhinoplasty. We introduce a methodology to map the aesthetics of visual appearance to the quantified measurements of pre-surgery, planned outcome, and post-surgery using machine learning. To develop the methodology, we generated synthetic 3D models utilizing artificial intelligence tools and applied various nasal deformities to simulate the pre-surgery, planned outcome, and post-surgery scans of rhinoplasty patients. The simulated outcomes were scored by reviewing the 3D visuals and corresponding measurements to prepare the training data for machine learning models. AutoGluon AutoML framework is used to generate the best-performing machine learning model. Machine learning models performed with 82% to 88% accuracy depending on the scoring method. We also identified the measurements that are highly influential in determining the scores. This is the first study that correlates the visual appearance and quantitative facial measurements of simulated rhinoplasty outcomes. The results suggest that an AI-based objective rhinoplasty outcome scoring tool is possible when machine learning algorithms are trained using consensus scores along with patients' pre-surgery, planned, and post-surgery measurements. This study introduces a methodology regarding how to map the aesthetics of visual appearance to the quantified measurements of pre-surgery, planned outcome, and post-surgery using machine learning.

INDEX TERMS Artificial intelligence, evaluation, machine learning, plastic surgery, rhinoplasty.

I. INTRODUCTION

The success of rhinoplasty has largely been determined by patient satisfaction and objective evaluation methods of aesthetic outcomes have been lacking. Objective evaluation is a challenging task since the assessment of beauty is subjective in nature. An objective, valid, reliable, agreeable, and effective evaluation tool for the assessment of rhinoplasty would be valuable for both surgeons and patients in evaluating rhinoplasty outcomes [1].

Rhinoplasty outcomes have generally been assessed with semi-quantitative questionnaires such as FACE-Q, ROE, [2], [3], [4], or through facial analyses in a quantitative fashion [5], [6]. There are studies that aimed to evaluate rhinoplasty outcomes utilizing questionnaires for subjective evaluation and facial measurements for objective evaluation [1], [7], [8], [9], [10], [11]. However, there was poor correlation and inconsistencies between subjective and objective evaluations, and it was not possible to

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine.

utilize objective (quantitative) scores to predict the success of outcomes [12].

In a recent systematic review study, Zijl et al. analyzed Patient Reported Outcome Measures (PROMs) to evaluate functional or aesthetic symptoms of patients undergoing rhinoplasty [13], [14]. and pointed out the importance of standardized outcomes for both patients and clinicians in rhinoplasty [13].

Researchers have utilized Artificial Intelligence (AI) to try and solve problems in many types of plastic surgeries [15], but there are only a handful of studies that utilize AI for rhinoplasty [16]. In one of the studies, a propriety Machine Learning tool was utilized to compare the predicted ages of pre-surgery and post-surgery to analyze the reversing effect of rhinoplasty on facial aging [17]. Another study also used proprietary AI software to predict and compare the age and attractiveness of pre-surgery and post-surgery photos to judge if the aesthetic surgery had a positive impact [18]. Other researchers used deep neural networks to make predictions on whether or not a rhinoplasty was performed [19]. In a recent study, Chandaliya and Nain proposed the PlasticGAN framework that is based on the Generative Adversarial Networks (GAN) 20, to generate images of post-surgery faces [21].

Stepanek et al. aimed to reveal correlations between facial attractiveness and facial measurements by analyzing the measurements taken on 2D photos. The study described which measurements could be changed surgically in order to achieve higher attractiveness scores [22], [23]. Our study has similarities as we are reporting on measurements that affect the perceived success of rhinoplasty, using machine learning to determine this information.

It has been suggested AI could assist surgeons in performing an objective assessment of rhinoplasty results [24], however, this has not been studied previously. In this study, we utilized machine learning to objectively evaluate rhinoplasty results by learning from the scores given to rhinoplasty outcomes by experts and non-experts. The machine then developed a model to score the outcome of rhinoplasty and determine the facial measurements most influential in affecting a successful outcome.

II. MATERIAL AND METHODS

Machine learning algorithms are trained to develop a model that encompasses a mathematical equation to map the input data to its output data. The developed model is then used to infer the corresponding output when new input data is given to the model. Machine learning requires quantifiable input data that affect, correlate, and cause the output.

For our machine learning solution, the target output (prediction) is the assessment score for rhinoplasty surgery. We accept that the planned outcome is an agreement between the patient and the surgeon. Therefore, a rhinoplasty assessment score should be high (indicating success) if the planned and post-surgery outputs are close (close to the output agreed outcome with the patient) and should be low if they are far apart (far from what was agreed on).

Machine learning algorithms require quantifiable input data that affect, correlate, and cause the output. We consider facial measurements as our quantifiable inputs (i.e., features of the machine learning algorithms). We use the pre-surgery, planned, and post-surgery measurements and use these as the input data for our ML algorithms. We compute the differences and ratios of the measurements of planned and post-surgery and the differences between and ratios of planned and pre-surgery measurements and use these computed data as the input data for our machine learning algorithms. We utilize the measurements of the pre-surgery state as well since new indicators can be generated using the pre-surgery measurement, such as the ratio of the difference between post and pre-surgery measurements and the difference between planned and pre-surgery measurements. The pre-surgery, planned, and post-surgery measurements were taken on the 3D models instead of 2D images due to the advantages of working with 3D models such as having higher fidelity, being more accurate, having the ability to rotate the model to examine the point of interest from various angles, and the possibility of calculating volumes and topographic distances on the 3D models [25], [26], [27], [28].

We prepared sets of 3D models representing rhinoplasty patients' pre-surgery and post-surgery 3D facial scans synthetically. The researchers have shown that synthetic images can be used successfully to achieve similar performance in machine learning studies [29]. The details of our methodology are given in the following subsections.

A. CREATION OF 3D MODELS

The 3D models for the study are created synthetically, therefore there were no human subjects in the study that would necessitate an Institutional Review Board (IRB) approval. For synthetic 3D model creation, we started by creating photos of people using the synthetic face generation website https://thispersondoesnotexist.com/ [30].

This website generates a photo of a person that does not really exist. It utilizes StyleGAN2 Face Generator developed by NVIDIA (NVIDIA Inc. Santa Clara, SA, USA) researchers [31], [32] which is a type of GAN (Generative Adversarial Network), a specialized architecture of neural networks [20].

A recent study that utilizes the same technology shows that synthesized photos are indistinguishable from real photos [33]. Another recent study was performed to generate photo images using the same technology and to offer them for facial aesthetic research [34].

We generated 24 (13 female, 11 male) front-facing and neutral (no smile) photos. Since the photos were generated synthetically using an AI tool, ethnicity cannot be determined. However, according to the looks of the photos, the race distribution can be determined as 80% White, 15% Asian, and 5% Black. We wanted to have photos representing various adult ages therefore we checked the ages of the photos using online tools at facialage.com and facelytics.io and selected

TABLE 1. Measurement statistics (mean +/− std) used from facebase.org statistics for 20 and 30 years olds: Cranial base width (CBW), Nasal width (NW), Nasal height (NH), Nasal protrusion (NP), Maximum cranial width (MCW), Morphological facial height (MFH), and Lower facial height (LFH).

Gender	Age	MCW	CBW	MFH	LFH	NW	NH
female	20	146.98 +/- 5.03	139.53 +/- 5.02	119.41 +/- 6.49	67.49 +/- 5.68	32.29 +/- 2.36	54.81 +/- 3.66
female	30	146.9 +/- 4.95	139.87 +/- 5.24	117.45 +/- 5.4	66.47 +/- 4.56	32.91 +/- 2.58	53.4 +/- 3.63
male	20	153.78 +/- 4	147.78 +/- 4.63	128.4 +/- 6.77	73.5 +/- 6.01	35.51 +/- 2.14	57.88 +/- 3.72
male	30	152.89 +/- 4.51	148.88 +/- 5.44	129.85 +/- 6.74	75.18 +/- 4.69	36.5 +/- 2.49	57.32 +/- 3.95

the photos that were predicted to be between 23 and 59. We used Reallusion Inc's Character Creator Software's Headshot plugin to generate 3D models from the photos [35].

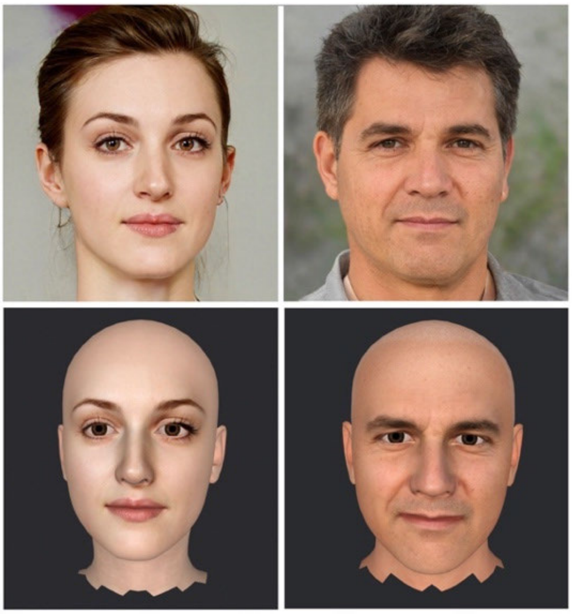

FIGURE 1. The example synthetic photos are in the top row. The bottom row shows a snapshot of the 3D models that correspond to the photos shown in the top row.

Examples of the generated photos and 3D models are shown in Figure 1. The images generated at https://thispersondoesnotexists.com are presented in the top row. The female on the left is assumed to be 29 years old, and the male on the right is assumed to be 45 years old. The bottom row shows a snapshot of the 3D models that correspond to the photos shown in the top row, generated using the Reallusion Character Creator Software.

We made sure that the 3D models had measurement values within realistic limits. We used the Face Analyzer tool [36], [37] to measure the 3D models and then compare them with the statistics from the 3D Facial Norms Database [38] listed at Facebase.org [39]. We checked the 3D models against the following measurements: 1) Maximum Cranial Width, 2) Cranial Base Width, 3) Morphological Facial Height, 4) Lower Facial Height, 5) Nasal Width, 6) Nasal Height, and 7) Nasal Protrusion; to make sure each measurement of a 3D model is within one standard deviation of the mean value for the supposed age of the 3D model. The measurement statistics for 20 and 30 years are given in Table 1. If the measurement is not within the range, the 3D model is resized to bring its measurement within the range using Blender 3D Software [40]. At the end of the processes explained above, we had 24 (13 male and 11 female) synthetically generated 3D models that can substitute a rhinoplasty patient's facial 3D model.

B. TAKING THE MEASUREMENTS

Ten measurements (6 distances, two angles, and two ratios) were selected as the most important based on the experienced rhinoplasty surgeon rankings and the literature [1], [28], [41], [42]. These measurements are shown in Figure 2 from top left to bottom right as 1) Alar Base Width, 2) Columella Length, 3) Columella Width, 4) Interalar Distance, 5) Nasal Bridge Length, 6) Nasal Tip Projection, 7) Interalar Angle, 8) Nasal Tip Angle, 9) Nasal Tip Projection: Goode, and 10) Nasal Width-Length Ratio. We marked the 22 landmarks on the 3D models that are used to calculate these ten important measurements using the Face Analyzer tool. Figure 3 shows the facial landmarks marked during this process. The landmarks are Alar_Base_Junction - left (ac_l), Alar_Base_Junction - right (ac_r), Alar Flare - left (al_l), Alar Flare - right (al_r), Columellar Break Point (cb), Endocanthion - left (en_l), Endocanthion - right (en_r), Maxilloanteriorale - left (ma_l), Maxilloanteriorale - right (ma_r), Maxillofrontale - left (mf_l), Maxillofrontale - right (mf_r), Nasal Parenthesis - left (np_l), Nasal Parenthesis - right (np_r), Nasion/Radix (n), Pronasale/Tip (prn), Rhinion (r), Subnasale - left (sn_l), Subnasale - right (sn_r), Subnasale (sn), Supratip Break Point (s), Tip Defining Point - left (td_l), Tip Defining Point - right (td_r).

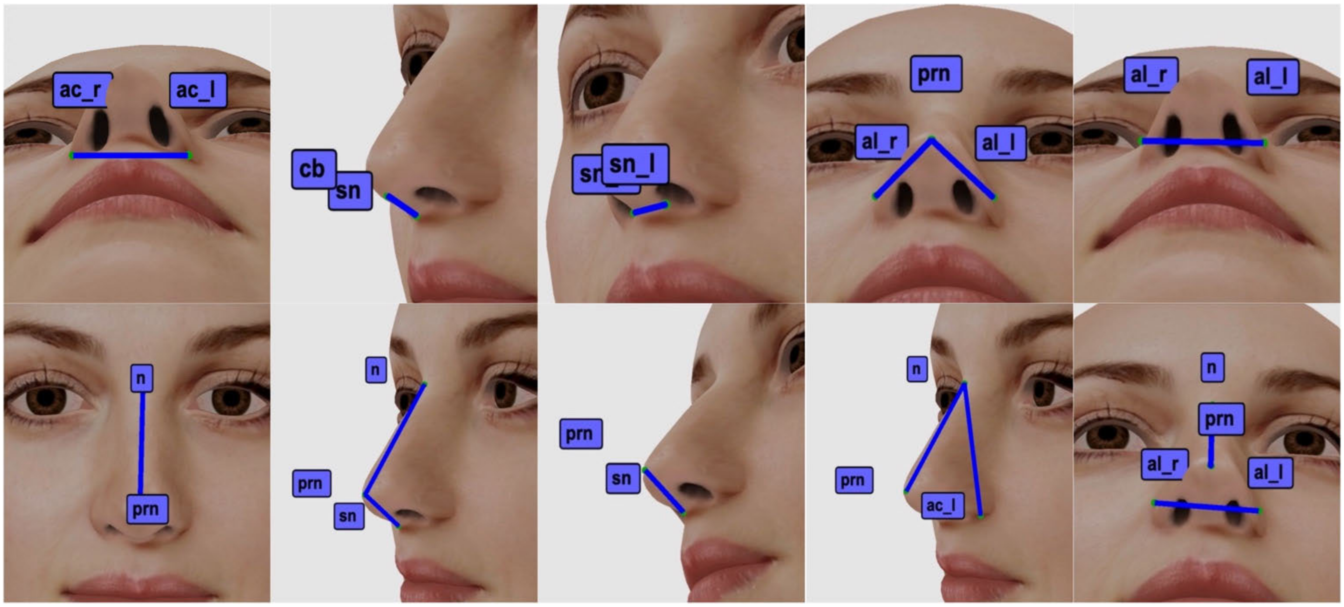

FIGURE 2. Measurements selected, from top left to bottom right: 1) Alar base width, 2) Columella length, 3) Columella width, 4) Interalar distance, 5) Nasal bridge length, 6) Nasal tip projection, 7) Interalar angle, 8) Nasal tip angle, 9) Nasal tip projection: Goode, and 10) Nasal width-length ratio.

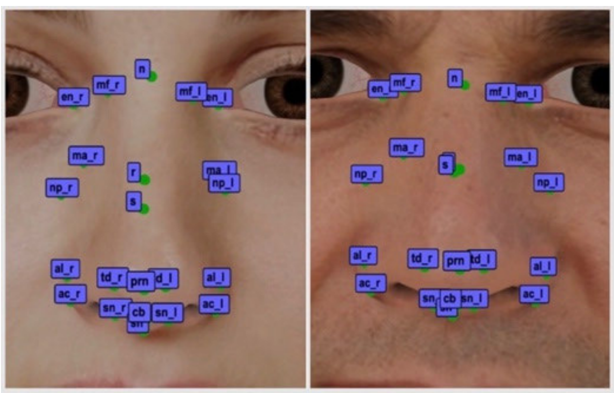

FIGURE 3. The 22 marked facial feature points are shown on 3D models. The landmarks are Alar_Base_Junction - left (ac_l), Alar_Base_Junction - right (ac_r), Alar Flare - left (al_l), Alar Flare - right (al_r), Columellar Break Point (cb), Endocanthion - left (en_l), Endocanthion - right (en_r), Maxilloanteriorale - left (ma_l), Maxilloanteriorale - right (ma_r), Maxillofrontale - left (mf_l), Maxillofrontale - right (mf_r), Nasal Parenthesis - left (np_l), Nasal Parenthesis - right (np_r), Nasion/Radix (n), Pronasale/Tip (prn), Rhinion (r), Subnasale - left (sn_l), Subnasale - right (sn_r), Subnasale (sn), Supratip Break Point (s), Tip Defining Point - left (td_l), Tip Defining Point - right (td_r).

C. GENERATING DEFORMITIES ON 3D MODELS

We determined nine common deformities for which patients typically desire a rhinoplasty based on the experience of the authors who perform rhinoplasties. These nose types are Crooked, Drooping Tip, Hooked, Hump, Large, Pinched, Saddle, Snub, and Wide. Figure 4 shows these nine deformities. We developed a morphing software application and used it to morph the 3D models to have new 3D models with the selected deformities [43], [44]. It is important to note that the methodology can be extended with additional types of deformities.

The morphing software reads the current locations of landmarks from an input file and then reads a configuration file describing how much each landmark location needs to be shifted. Then the morphing software outputs the morphed 3D model along with the new locations of the landmarks. For example, the location of the pronasale (prn) landmark was $x = 2.58$, $y = -7.64$, $z = 111.5$ before morphing the model for Drooping Tip deformity, and it is $x = 2.17$, $y = -18.28$, $z = 115.19$ after the application of the deformity using the program. That means there is a 0.31, 10.64, and 3.69-millimeter shift in the x, y, and z coordinates, respectively, for the pronasale (prn) landmark. These updated landmark locations are used to calculate the measurements for the morphed 3D model.

D. ORGANIZING THE 3D MODELS TO FORM THE PRE, PLANNED, AND POST-SURGERY-SETS

We applied the nine deformities at various magnitudes to each 3D model representing 24 patients and generated 19 different 3D models for each patient (a total of $456 = 24 \times 19$). We organized the resulting 3D models to form a set of pre/planned/post-surgery 3D models as follows:

- We selected a deformed 3D model to represent the pre-surgery,
- We selected a less deformed or a not deformed model to represent the planned, and
- We again selected a less deformed or not deformed model for representing post-surgery outcome.

For example, we used a 3D model having a crooked deformity as its pre-surgery 3D model, then used the normal (not deformed) 3D model as its planned 3D model, and then used the slightly deformed model as its post-surgery outcome.

FIGURE 4. Nose types used. From top left to bottom right in order: Crooked, Drooping Tip, Hooked, Hump, Large, Pinched, Saddle, Snub, and Wide.

We formed a total of 648 sets of 3D models representing the pre-surgery, planned, and post-surgery facial scans of 24 rhinoplasty patients.

E. GETTING SCORES TO TRAIN MACHINE LEARNING ALGORITHMS

324 sets of 3D models were randomly selected from the 648 sets and scored to prepare the training data for the machine learning models. The set of 3d models was presented to raters in random order, and each rater gave an outcome score between 1 to 10. We grouped the raters into two groups: expert and non-expert. The expert group consisted of two rhinoplasty surgeons. Non-expert raters were junior college students pursuing computer science degrees without medical knowledge.

We developed a web-based software to collect scores for each set of 3D models. Figure 5 is a screenshot of this web-based software showing pre/planned/post-surgery 3D models side by side. The software enabled the raters to rotate, zoom in/out and move the 3D models in x, y, and z coordinates to analyze the pre, planned, and post 3D models in detail.

The raters first reviewed the 3D models visually and gave a 'look' score. The software also displayed the measurements of each 3D model in a table at the bottom of the page. The table included one row for each measurement and the columns showed the pre, planned, and post measurements along with the differences between these measurements and the ratios of the measurements. The last column presents the ratio of the differences between post and pre-surgery measurements over the differences between planned and pre-surgery measurements. If this ratio is 1, that means the post-surgery outcome is exactly the same as planned. After reviewing the data, the raters gave a 'data' score. The average score of raters was computed for each pre/planned/post set and considered as the consensus.

The consensus scores are used as labels (output) for training the machine learning algorithms and the input data included values of the ten measurements for the pre, planned, and post-surgery 3D models, along with other data generated by subtracting or finding the ratio of pre, planned, and post-surgery measurement values such as the ratio of differences between post and pre-surgery and differences between planned and pre-surgery. Some of these features are shown as column headers in the data table presented to the raters in Figure 5. The complete list of features for each measurement used as an input to the machine learning training algorithms are listed in Table 2. We had nine values for each of the measurements and therefore we had a total of 90 features for training the machine learning algorithms.

F. RELIABILITY OF EXPERT AND NON-EXPERT SCORES

We utilized the intraclass correlation coefficient (ICC) method to analyze the inter-rater reliability of scores given

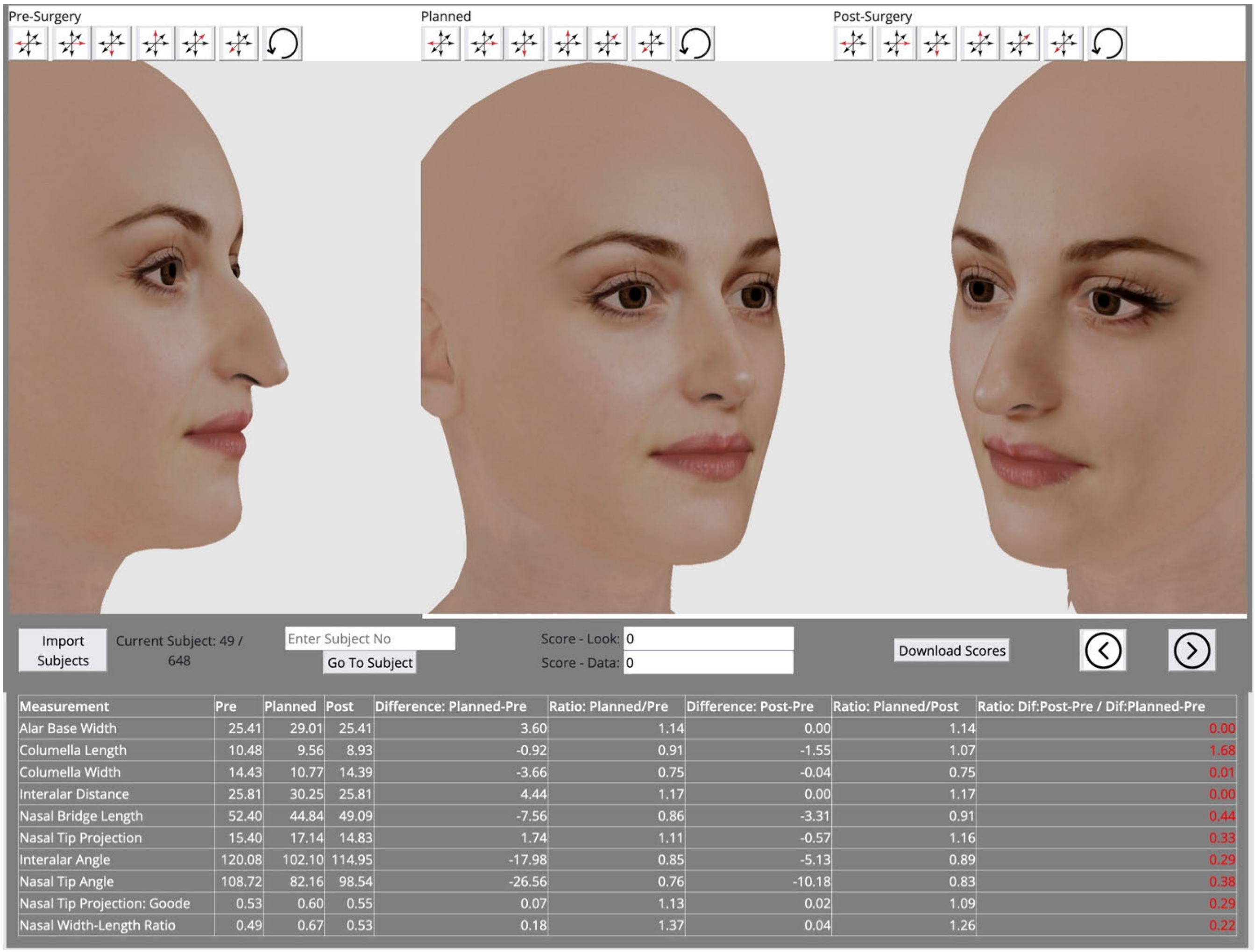

FIGURE 5. A screenshot of this web-based software for scoring rhinoplasty outcomes.

TABLE 2. Features (attributes) used for each measurement for training machine learning models.

Feature Name	Description
pre	The measurement taken on pre-surgery 3D model
planned	The measurement taken on planning 3D model
post	The measurement taken on post 3D model
diff:planned-pre	The difference between planned and pre-surgery
diff:post-pre	The difference between post and pre-surgery
ratio:planned/pre	The ratio of planned and pre-surgery
ratio:post/pre	The ratio of post and pre-surgery
ratio:planned/post	The ratio of planned and post-surgery
ratio:diff:post-pre/diff:planned-pre	The ratio of differences between post and pre-surgery and differences between planned and pre-surgery

TABLE 3. The intraclass correlation coefficient (ICC) scores for inter-rater reliability (N = 324).

Expert versus non-expert, scores based on the measurements	0.979
Expert versus non-expert, scores based on the look	0.970
Two expert raters, scores based on the measurements	0.975
Two expert raters, scores based on the look	0.901
Two non-expert raters, scores based on the measurements	0.984
Two non-expert raters, scores based on the look	0.963

by the expert and non-expert raters. The inter-rater reliability was calculated using a two-rater, consistency, two-way mixed-effects model [45]. The inter-rater reliability scores were calculated for both scores given based on the look and measurements and tested the reliability of scores from two expert raters, two non-expert raters, and the average of expert versus non-expert raters. An ICC of less than 0.5 is considered as poor, 0.50 to 0.75 as fair, 0.75 to 0.90 as good, and 0.90 to 1.00 as excellent reliability [46]. All the inter-rater reliability scores are excellent (>0.9) as presented in Table 3.

G. SEARCHING FOR THE BEST MACHINE LEARNING MODEL USING AUTOGLUON

We utilized the AutoGluon (version 0.5.2) Auto ML framework and Python language (version 3.7.13) on the Colab environment (colab.research.google.com) for finding the best machine learning model. Auto Gluon is an open-source AutoML framework by Amazon (https://auto.gluon.ai/) [47].

TABLE 4. The performance of machine learning models that were trained using the scores from non-experts given based on the appearance "look" of 3D models and based on their measurements.

Based on Look			*Based on Measurements*		
ML Algorithm	*Test*	*Validation*	*ML Algorithm*	*Test*	*Validation*
KNeighborsUnif	0.86	0.84	RandomForestGini	0.88	0.75
KNeighborsDist	0.86	0.82	RandomForestEntr	0.88	0.75
ExtraTreesEntr	0.86	0.84	ExtraTreesGini	0.86	0.77
CatBoost	0.84	0.84	ExtraTreesEntr	0.86	0.73
RandomForestGini	0.84	0.80	CatBoost	0.84	0.80
RandomForestEntr	0.84	0.82	WeightedEnsemble_L2	0.84	0.80
WeightedEnsemble_L2	0.82	0.89	KNeighborsUnif	0.82	0.75
ExtraTreesGini	0.80	0.87	KNeighborsDist	0.82	0.77

TABLE 5. The performance of machine learning models that were trained using the scores from experts given based on the appearance of 3D models and based on their measurements.

Based on Look			Based on Measurements		
ML Algorithm	Test	Validation	ML Algorithm	Test	Validation
CatBoost	0.82	0.70	CatBoost	0.84	0.78
ExtraTreesGini	0.82	0.70	RandomForestEntr	0.84	0.82
ExtraTreesEntr	0.82	0.70	RandomForestGini	0.84	0.78
WeightedEnsemble_L2	0.82	0.70	WeightedEnsemble_L2	0.84	0.82
KNeighborsDist	0.80	0.68	KNeighborsUnif	0.82	0.71
RandomForestGini	0.80	0.70	KNeighborsDist	0.82	0.71
RandomForestEntr	0.80	0.70	ExtraTreesEntr	0.82	0.78
KNeighborsUnif	0.79	0.64	ExtraTreesGini	0.82	0.78

AutoGluon ensembles multiple models and stacks them in multiple layers to find the best-performing model along with its hyperparameters by experimenting with various machine learning algorithms such as neural networks, LightGBM boosted trees, CatBoost boosted trees, and Random Forests. AutoGluon does not perform cross-validation by default. We selected AutoGluon since it is shown to perform better than five widely used AutoML frameworks such as Auto-WEKA, auto-sklearn, The Tree-based Pipeline Optimization Tool (TPOT), H2O AutoML, Google Cloud Platform AutoML Tables [47] and would fine-tune hyperparameters better compared to manually tweaking them [48], [49].

We partitioned our dataset into training and test dataset. 80% of the data is used for training, and the rest is used for testing. Since each rater can assign scores between 1 to 10, this machine learning problem can be considered a classification problem where we assign each rhinoplasty outcome into one of ten classes. We utilized the accuracy metric for the performance evaluation of the machine learning models.

To avoid overfitting during the training, we applied the early-stopping technique and compared the validation and test scores to identify the number of training instances when the models started to overfit. We trained the machine learning algorithms with expert scores and non-expert scores separately. We also trained them using the scores given based on the look and based on the data separately. Hence, we performed four sets of training utilizing the AutoGluon AutoML framework. To find out which measurements affect the outcome score the most, we utilized the feature importance function that works based on permutation importance algorithms [50].

III. RESULTS

We first ran AutoGluon using the non-expert scores given based on the 'look' (appearance) of the 3D models and given based on the 'data' (measurement values). The model WeightedEnsemble_L2 performed the best on the validation data (89% accuracy), and models KNeighborsUnif, KNeighborsDist, and ExtraTreesEntr performed the best on the test data (all 86% accuracy) when scores based on 'look' were used as presented in Table 4. The models WeightedEnsemble_L2 and *CatBoost* performed the best on the validation data (80% accuracy), and models RandomForestGini and RandomForestEntr performed the best on the test data (88% accuracy) when scores based on 'measurements' were used as presented in Table 4.

We then ran the feature importance function of AutoGluon to see which features contribute the most to the best-performing model's inference. While the models utilized nine different input data for each measurement, the most important input data were consistently the ratio of two differences (the difference between post and pre-surgery and the difference between planned and pre-surgery). We list the top 3 measurements that this ratio affects the model performance the most in Table 6.

Next, we ran AutoGluon using the expert scores given based on the 'look' and 'measurements'. Several models performed 70% accuracy on the validation data and several models such as CatBoost, ExtraTreesGini, etc. performed the

TABLE 6. The top 3 measurements that are affecting the performance of machine learning models trained with the scores from non-experts.

When Scores Given Based on Look	When Scores Given Based on Measurements
Interalar Distance	Alar Base Width
Columella Width	Interalar Angle
Nasal Bridge Length	Nasal Width-Length Ratio

best on the test data with 82% accuracy when scores based on 'look' were used. The models WeightedEnsemble_L2 and RandomForestEntr performed the best on validation data (82% accuracy), and models such as CatBoost, RandomForestGini and RandomForestEntr performed the best on the test data (84% accuracy) when scores based on measurements were used.

TABLE 7. The top 3 measurements that are affecting the performance of machine learning models trained with the scores from experts.

When Scores Given Based on Look	When Scores Given Based on Measurements
Interalar Distance	Nasal Tip Projection
Alar Base Width	Alar Base Width
Columella Width	Nasal Width-Length Ratio

Next, we ran the feature importance function of AutoGluon to see which features contribute the most to the best-performing model's inference. The top 3 most important measurement data are listed in Table 7.

IV. DISCUSSION

The literature points out the poor correlation between the subjective scores that are based on surveys and objective scores that are based on measurements and suggests that objective scores could not be used to predict successful outcomes of nasal surgeries due to these inconsistencies [12]. We believe our study fills the gap between the objective and subjective scores by utilizing a visual input (3D images) for scoring and by figuring out the mapping between the scores and the measurements utilizing machine learning.

A machine learning algorithm utilizes the training data and creates a model that (mathematically) defines the relationship (equation) between the features (attributes such as the difference between planned and post measurements) and the labels (outcome scores). A machine learning model can, at most, be as accurate as the scores in the training data. If the scores given by the raters in the training data are not objective, the scores predicted by the model will not be objective. Personal opinions, especially on a subjective concept like beauty, will always exist. We believe we can mitigate subjectivity by aggregating many opinions to find a consensus. In order to achieve a more objective score, we computed the consensus score among the raters by averaging the scores given by them.

The intraclass correlation coefficient (ICC) reliability indicators for the scores were between 0.901 and 0.984, corresponding to the scores from two expert raters based on the look and two non-expert raters based on the measurements, respectively. We noticed that the lowest predictive performance was achieved for the training data that has the lowest reliability. The researchers have pointed out that the reliability of the data correlates with predictive performance [51].

We have pre-surgery, planned, and post-surgery values for each of the ten selected measurements making it 30 input parameters for machine learning models. However, we utilized feature engineering techniques [52] to develop new input data that might correlate better with the output label. For each measurement, we added six new types of data by computing the differences or rations of the pre/planned/post measurements, bringing the total feature number to 90. Our results show that the input data computed by getting the ratio of differences between post and pre-surgery over the differences between planned and pre-surgery correlates well with the output which seems logical.

The measurements Interalar Distance and Columella Width are highly influential in determining the scores in ML algorithms based on appearance, and Alar base Width and Columella Width are influential when measurements are considered. While tables 1 and 2, which list the most influential measurements based on expert and non-expert training data, have some common measurements, the measurements listed are not the same. This is expected since expert and non-expert scores are different. At this point, it is difficult to determine if the specific measurements noted as being most influential in this study actually correlate clinically since the planned and post-operative images were only simulations. In general, columella width is less clinically important than many of the other measurements listed, and the influence may be more influential in simulated images than in actual patients. However, it is helpful to know that it is possible to determine which measurements are most influential in determining scores for future studies.

V. LIMITATIONS AND FUTURE DIRECTIONS

We simulated three types of post-surgery outcomes: the perfect outcome, the total failure, and somewhere in between. However, in practice, it is often that we have patients that are very close to the planned image except for a small deviation in one or two areas. Future studies may look to utilize cases with more minor deviations from the perfect outcome and utilize real-life simulations and postoperative outcomes.

Utilizing actual patients may correlate with the most influential measurements in score generation and help surgeons to focus more on these areas while performing surgery. Generating the algorithm requires a lot of data, so performing this study with clinical images may require collaboration.

To mitigate the issue of subjectivity and to reach more objective results, we aimed to find a consensus on the outcome scores. Therefore, we aggregated the scores given by raters. However, the number of raters we had in this study is not enough to reach a common, publicly agreeable consensus. A larger group of raters should be utilized to reach a better consensus on outcome scores.

This study utilized ten common measurements for rhinoplasty. Rhinoplasty aims to achieve harmony with the whole face. Therefore, other anatomic relationships between the nose and the face should be included. The study can be repeated using facial measurements and even more distance, angle, and ratio measurements around the nose.

While this study utilizes 3D models for visualization, it does not benefit from some of the measurements that can be performed on 3D models. The study can be extended to incorporate new measurements such as area and volume.

Finally, the need for new assessment tools that encompass functional, psycho-relational, and aesthetic aspects has been pointed out in the literature [52]. The objective assessment solution we introduce in this study only evaluates the visual (cosmetics) outcomes of rhinoplasty and cannot be used to assess an outcome score for functional aspects of rhinoplasty.

VI. CONCLUSION

This study shows that an AI-based objective rhinoplasty outcome scoring tool is possible when an adequate number and variety of patients' pre-surgery, planned outcome, and post-surgery 3D models are used to train machine learning algorithms with the consensus scores given by an adequate number of raters. Once the machine learning algorithms are trained and a mathematical model is produced, this model can be utilized to make predictions (assess an outcome score for rhinoplasty). With the estimated accuracy, a prediction by the trained model would answer the question: "How would the raters score the new rhinoplasty outcome when the input data (measurements of pre, planned, and post-surgery) are given?"

This is the first study that forms the connection between the visual appearance of the outcome and quantitative facial measurements of the outcome. Machine learning algorithms learn how to map satisfaction, which mostly depends on aesthetics of visual appearance, to the quantified measurements of pre-surgery, planned outcome, and post-surgery.

Moreover, this is the first study to analyze each measurement's contribution to the outcome scores. As the study repeated with more raters and measurements, it may suggest which measurements are more important for a successful outcome.

Objective assessment tools need to be easy to use, cost-effective, accurate, reliable and provide repeatable results. We believe our machine learning study is one of the major initial steps in having such an evaluation tool for rhinoplasty and all facial plastic surgeries.

REFERENCES

[1] M. Eloteify, M. El-Shazly, W. Saadeldeen, and M. El-Sonbaty, "Subjective psychic perception versus objective nasofacial measures in assessment of rhinoplasty results: A clinical report," *Eur. J. Plastic Surgery*, vol. 38, no. 2, pp. 109–116, Apr. 2015, doi: 10.1007/s00238-015-1070-9.

[2] A. Kalaaji, S. Dreyer, J. Schnegg, L. Sanosyan, T. Radovic, and I. Maric, "Assessment of rhinoplasty outcomes with FACE-Q rhinoplasty module: Norwegian linguistic validation and clinical application in 243 patients," *Plastic Reconstructive Surg. Global Open*, vol. 7, p. 9, Sep. 2019, doi: 10.1097/GOX.0000000000002448.

[3] S. C. Izu, E. M. Kosugi, K. V. Brandao, A. S. Lopes, L. B. Garcia, V. M. Suguri, and L. C. Gregorio, "Normal values for the Rhinoplasty Outcome Evaluation (ROE) questionnaire," *Brazilian a Otorhinolaryngol.*, vol. 78, no. 4, pp. 9–76, Jul. 2012, doi: 10.1590/s1808-86942012000400015.

[4] I. Khansa, L. Khansa, and G. D. Pearson, "Patient satisfaction after rhinoplasty: A social media analysis," *Aesthetic Surg. J.*, vol. 36, no. 1, pp. 1–5, Jan. 2016, doi: 10.1093/asj/sjv095.

[5] S. Sporri, D. Simmen, H. R. Briner, and N. Jones, "Objective assessment of tip projection and the nasolabial angle in rhinoplasty," *Arch. Facial Plast Surg.*, vol. 6, no. 5, pp. 8–295, 2004, doi: 10.1001/archfaci.6.5.295.

[6] P. Chatrath, J. De Cordova, S. A. R. Nouraei, J. Ahmed, and H. A. Saleh, "Objective assessment of facial asymmetry in rhinoplasty patients," *Arch. Facial Plastic Surg.*, vol. 9, no. 3, pp. 184–187, May 2007, doi: 10.1001/archfaci.9.3.184.

[7] A. Bazazz, A. Bazazz, M. Ficms, M. Dawood, A. Ficms, and K. Khammas, "Subjective and objective assessment of rhinoplasty outcome," *Iraqi Med. J.*, vol. 63, no. 1, pp. 17–28, 2017.

[8] E. Chisholm and N. Jallali, "Rhinoplasty and septorhinoplasty outcome evaluation," *Ear, Nose Throat J.*, vol. 91, no. 3, pp. 10–14, Mar. 2012, doi: 10.1177/014556131209100315.

[9] S. H. Abianeh, S. M. Bajestani, J. Rahmati, M. A. Shahrbaf, and A. F. Meybodi, "Evaluation of aesthetic and functional outcomes after open rhinoplasty: A quasi-experimental study by the aid of ROE and RHINO questionnaires," *Aesthetic Plastic Surg.*, vol. 45, no. 2, pp. 663–669, Apr. 2021, doi: 10.1007/s00266-020-01905-w.

[10] M. W. Saadeldeen and M. M. El-Shazly, "Objective versus subjective assessment for rhinoplasty," *Egypt. J. Plastic Reconstr. Surg.*, vol. 37, no. 1, pp. 73–79, 2013.

[11] S. S. Esteves, M. G. Ferreira, J. C. Almeida, J. Abrunhosa, and C. A. E. Sousa, "Evaluation of aesthetic and functional outcomes in rhinoplasty surgery: A prospective study," *Brazilian J. Otorhinolaryngology*, vol. 83, no. 5, pp. 552–557, Sep. 2017, doi: 10.1016/j.bjorl.2016.06.010.

[12] S. S. Pawar, G. J. Garcia, J. S. Kimbell, and M. Rowland, "A systematic review of patient satisfaction following rhinoplasty," *J. Plast Reconstr. Aesthet. Surg.*, vol. 68, no. 5, pp. 547–561, 2015, doi: 10.1016/j.bjps.2014.09.017.

[13] F. V. W. J. van Zijl, L. B. Mokkink, J. A. Haagsma, and F. R. Datema, "Evaluation of measurement properties of patient-reported outcome measures after rhinoplasty: A systematic review," *JAMA Facial Plastic Surg.*, vol. 21, no. 2, pp. 152–162, Mar. 2019, doi: 10.1001/jamafacial.2018.1639.

[14] J. J. Gagnier, J. Lai, L. B. Mokkink, and C. B. Terwee, "COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures," *Qual. Life Res.*, vol. 30, no. 8, pp. 2197–2218, Aug. 2021, doi: 10.1007/s11136-021-02822-4.

[15] T. Jarvis, D. Thornburg, A. M. Rebecca, and C. M. Teven, "Artificial intelligence in plastic surgery: Current applications, future directions, and ethical implications," *Plastic Reconstructive Surg. Global Open*, vol. 8, no. 10, p. e3200, Oct. 2020, doi: 10.1097/GOX.0000000000003200.

[16] A. S. Eldaly, F. R. Avila, R. A. Torres-Guzman, K. Maita, J. P. Garcia, L. P. Serrano, and A. J. Forte, "Simulation and artificial intelligence in rhinoplasty: A systematic review," *Aesthetic Plastic Surg.*, vol. 46, no. 5, pp. 2368–2377, Oct. 2022, doi: 10.1007/s00266-022-02883-x.

[17] R. Dorfman, I. Chang, S. Saadat, and J. Roostaeian, "Making the subjective objective: Machine learning and rhinoplasty," *Aesthetic Surg. J.*, vol. 40, no. 5, pp. 493–498, Apr. 2020, doi: 10.1093/asj/sjz259.

[18] C. J. Peck, V. K. Patel, Y. Parsaei, N. Pourtaheri, O. Allam, J. Lopez, and D. Steinbacher, "Commercial artificial intelligence software as a tool for assessing facial attractiveness: A proof-of-concept study in an orthognathic surgery cohort," *Aesthetic Plastic Surg.*, vol. 46, no. 2, pp. 1013–1016, Apr. 2022, doi: 10.1007/s00266-021-02537-4.

[19] E. Borsting, R. DeSimone, M. Ascha, and M. Ascha, "Applied deep learning in plastic surgery: Classifying rhinoplasty with a mobile app," *J. Craniofacial Surgery*, vol. 31, no. 1, pp. 102–106, 2020, doi: 10.1097/SCS.0000000000005905.

[20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[21] P. K. Chandaliya and N. Nain, "PlasticGAN: Holistic generative adversarial network on face plastic and aesthetic surgery," *Multimedia Tools Appl.*, vol. 81, no. 22, pp. 32139–32160, Sep. 2022, doi: 10.1007/s11042-022-12865-5.

[22] L. Stepanek, P. Kasal, and J. Mestak, "Evaluation of facial attractiveness for purposes of plastic surgery using machine-learning methods and image analysis," in *Proc. IEEE 20th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Sep. 2018, pp. 1–6, doi: 10.1109/HealthCom.2018.8531195.

[23] L. Stepanek, P. Kasal, and J. Mest'ak, "Evaluation of facial attractiveness after undergoing rhinoplasty using tree-based and regression methods," in *Proc. E-Health Bioeng. Conf. (EHB)*, Nov. 2019, pp. 1–4.

[24] A. Chandawarkar, C. Chartier, J. Kanevsky, and P. E. Cress, "A practical approach to artificial intelligence in plastic surgery," *Aesthetic Surg. J. Open Forum*, vol. 2, no. 1, Jan. 2020, Art. no. ojaa001, doi: 10.1093/asjof/ojaa001.

[25] O. Topsakal, M. I. Akbas, B. S. Smith, M. F. Perez, E. C. Guden, and M. M. Celikoyar, "Evaluating the agreement and reliability of a web-based facial analysis tool for rhinoplasty," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 16, no. 8, pp. 1381–1391, Jun. 2021, doi: 10.1007/s11548-021-02423-z.

[26] G. Lekakis, G. Hens, P. Claes, and P. W. Hellings, "Three-dimensional morphing and its added value in the rhinoplasty consult," *Plastic Reconstructive Surg. Global Open*, vol. 7, no. 1, p. e2063, Jan. 2019, doi: 10.1097/GOX.0000000000002063.

[27] D. M. Toriumi and T. K. Dixon, "Assessment of rhinoplasty techniques by overlay of before-and-after 3D images," *Facial Plastic Surg. Clinics North Amer.*, vol. 19, no. 4, pp. 711–723, Nov. 2011, doi: 10.1016/j.fsc.2011.07.011.

[28] O. Topsakal, E. Dobratz, and M. M. Celikoyar, "A web-based interactive three-dimensional (3D) tool for theeducation of landmarks and measurements of facial surface anatomy," *Facial Plastic Surgery Aesthetic Med.*, vol. 25, no. 3, pp. 1–3, 2009, doi: 10.1089/fpsam.2023.0019.

[29] E. Wood, T. Baltrusaitis, C. Hewitt, S. Dziadzio, T. J. Cashman, and J. Shotton, "Fake it till you make it: Face analysis in the wild using synthetic data alone," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3661–3671, doi: 10.1109/ICCV48922.2021.00366.

[30] *This Person Does not Exist*. Accessed: Jan. 4, 2023. [Online]. Available: https://thispersondoesnotexist.com

[31] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4217–4228, Dec. 2021, doi: 10.1109/TPAMI.2020.2970919.

[32] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8107–8116, doi: 10.1109/CVPR42600.2020.00813.

[33] S. J. Nightingale and H. Farid, "AI-synthesized faces are indistinguishable from real faces and more trustworthy," *Proc. Nat. Acad. Sci. USA*, vol. 119, no. 8, Feb. 2022, Art. no. e2120481119, doi: 10.1073/pnas.2120481119.

[34] J. P. Flynn, E. Cha, T. J. Flynn, J. D. Kriet, and C. D. Humphrey, "Machine learning generated synthetic faces for use in facial aesthetic research," *Facial Plastic Surg. Aesthetic Med.*, vol. 23, no. 5, pp. 339–343, Oct. 2021, doi: 10.1089/fpsam.2020.0328.

[35] *Reallusion Character Creator*. Accessed: Jan. 4, 2023. [Online]. Available: https://www.reallusion.com/character-creator/headshot/

[36] Digitization of Rhinoplasty Project. *Face Analyzer*. Accessed: Jan. 4, 2023. [Online]. Available: https://www.digitized-rhinoplasty.com/

[37] O. Topsakal, M. I. Akbas, D. Demirel, R. Nunez, B. S. Smith, M. F. Perez, and M. M. Celikoyar, "Digitizing rhinoplasty: A web application with three-dimensional preoperative evaluation to assist rhinoplasty surgeons with surgical planning," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 15, no. 11, pp. 1941–1950, Nov. 2020, doi: 10.1007/s11548-020-02251-7.

[38] S. Weinberg, Z. D. Raffensperger, M. J. Kesterke, C. L. Heike, M. L. Cunningham, J. T. Hecht, C. H. Kau, J. C. Murray, G. L. Wehby, L. M. Moreno, and M. L. Marazita, "The 3D facial norms database: Part 1. A web-based craniofacial anthropometric and image repository for the clinical and research community," *Cleft Palate-Craniofacial J.*, vol. 53, no. 6, pp. 185–197, 2016.

[39] *Facebase, A Resource for Craniofacial Researchers*. Accessed: Jan. 4, 2023. [Online]. Available: https://facebase.org

[40] *Blender. Free and Open-Source 3D Creation*. Accessed: Jan. 4, 2023. [Online]. Available: https://blender.org

[41] M. M. Celikoyar, M. F. Pérez, M. I. Akbas, and O. Topsakal, "Facial surface anthropometric features and measurements with an emphasis on rhinoplasty," *Aesthetic Surg. J.*, vol. 42, no. 2, pp. 133–148, Jan. 2022, doi: 10.1093/asj/sjab190.

[42] *Educator, A Web-Based 3D Tool for the Facial Anatomy*. Accessed: Jan. 4, 2023. [Online]. Available: https://digitized-rhinoplasty.com/app/educator.html

[43] O. Topsakal, J. Glinton, M. I. Akbas, and M. Celikoyar, "Generating morphed facial 3D models realisticallyusing facebase statistics," presented at the FaceBase Community Forum, Washington, DC, USA, Jun. 2022. [Online]. Available: https://www.facebase.org/events/2022-facebase-annual/

[44] *3D Face Morph Software*. Accessed: Jan. 4, 2023. [Online]. Available: https://github.com/research-digitized-rhinoplasty/3D-face-morph

[45] R. Landers, "Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS," *Winnower*, pp. 1–4, Jun. 2015, doi: 10.15200/winn.143518.81744.

[46] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *J. Chiropractic Med.*, vol. 15, no. 2, pp. 155–163, Jun. 2016.

[47] N. Erickson, J. Mueller, A. Shirkov, H. Zhang, P. Larroy, M. Li, and A. Smola, "AutoGluon-tabular: Robust and accurate AutoML for structured data," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1–28.

[48] *AutoGluon: AutoML for Text, Image, and Tabular Data*. Accessed: Feb. 3, 2023. [Online]. Available: https://auto.gluon.ai/

[49] *AutoML for Image, Text, Time Series, and Tabular Data*. Accessed: Feb. 3, 2023. [Online]. Available: https://github.com/awslabs/autogluon

[50] A. Altmann, L. Tolosi, O. Sander, and T. Lengauer, "Permutation importance: A corrected feature importance measure," *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, Jan. 2010.

[51] G. Duwe, "Why inter inter-rater reliability matters for recidivism risk assessment (policy brief number 2017 2017–03)," The Public Safety Risk Assessment Clearinghouse, Washington, DC, USA, 2017. Accessed: Feb. 3, 2023. [Online]. Available: https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/pb-interrater-reliability.pdf

[52] A. Casari and A. Zheng, *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. Berlin, Germany: O'Reilly Media, 2018.

[53] M. Barone, A. Cogliandro, N. Di Stefano, V. Tambone, and P. Persichetti, "A systematic review of patient-reported outcome measures after rhinoplasty," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 274, no. 4, pp. 1807–1811, Apr. 2017, doi: 10.1007/s00405-016-4359-9.

OGUZHAN TOPSAKAL (Senior Member, IEEE) received the B.S. degree in computer engineering from Istanbul Technical University, Turkey, in 1996, and the M.S. and Ph.D. degrees in computer science from the University of Florida, in 2003 and 2007, respectively.

After gaining extensive experience in the software industry, he is currently an Assistant Professor with the Department of Computer Science, Florida Polytechnic University. He teaches courses related to machine learning, algorithm design, databases, and mobile development. His research interests include developing innovative solutions utilizing machine learning and deep learning.

ERIC J. DOBRATZ received the medical degree from the Eastern Virginia Medical School. He completed his residency at the University of Virginia. He received specialized training in facial plastic and reconstructive surgery through a fellowship with the University of Minnesota. He has a highly specialized practice where he focuses exclusively on plastic surgery of the face, head, and neck. He is currently a Professor of otolaryngology and the Director of the Division of Facial Plastic and Reconstructive Surgery, Eastern Virginia Medical School. He has double board certifications from the American Board of Otolaryngology and the American Board of Facial Plastic and Reconstructive Surgery.

MUSTAFA ILHAN AKBAS (Member, IEEE) received the B.S. and M.S. degrees from the Department of Electrical and Electronics Engineering, Middle East Technical University, and the Ph.D. degree in computer engineering from the University of Central Florida. He is currently an Assistant Professor with the Electrical Engineering and Computer Science Department, Embry-Riddle Aeronautical University. He has industry experience in projects with multinational defense industry partners and large enterprises. His research interests include connected and autonomous cyber-physical systems, validation and verification, wireless networks, and mobile computing. He is a member of ACM, AAAI, and SAE.

WILLIAM M. DOUGHERTY received the medical degree from the University of Kentucky College of Medicine. He completed his residency at the University of Virginia. He received specialized training in facial plastic and reconstructive surgery through a fellowship with Oregon Health Sciences University. He is currently an Assistant Professor of otolaryngology and a Facial Plastic Surgeon with the Eastern Virginia Medical School, specializing exclusively in plastic surgery of the head, neck, and face. He holds board certification in otolaryngology-head and neck surgery.

TAHIR CETIN AKINCI (Senior Member, IEEE) received the bachelor's degree in electrical engineering, in 2000, and the master's and Ph.D. degrees, in 2005 and 2010, respectively. From 2003 to 2010, he was a Research Assistant with Marmara University, Istanbul, Turkey. He became an Associate Professor with Istanbul Technical University (ITU) and held the position until 2021, after which he was promoted to Full Professor with the Department of Electrical Engineering, ITU, in 2021. He assumed the role of a Visiting Scholar with the University of California at Riverside (UCR). His research interests include artificial neural networks, deep learning, machine learning, cognitive systems, signal processing, and data analysis.

MEHMET MAZHAR CELIKOYAR received the degree from the Hacettepe School of Medicine, Ankara, Turkey, in 1982. He specialized in otolaryngology-head and neck surgery with Sisli Children's Hospital and Marmara University Hospital, in 1989. He had his fellowship training on head and neck surgery and facial plastic and reconstructive surgery with the Methodist Hospital of Indiana, in 1996. Since 1996, he has been practicing in Istanbul, Turkey. He is currently an Otolaryngologist-Head and Neck Surgeon practicing with Istanbul Florence Nightingale Hospital. He is also one of the teaching members of the Department of Otolaryngology-Head and Neck Surgery, School of Medicine, Demiroğlu Bilim University, Istanbul. His research interest includes integrating rhinoplasty with computer science.