Article published online: 2021-05-29

An Artificial Intelligence Tool for Image Simulation in Rhinoplasty

Hernan Chinski, MD[1,2] Ricardo Lerch, Btech[3] Damián Tournour, MD[2] Luis Chinski, MD[1,2]
Diego Caruso, MD, MSc[4]

[1] Otolaryngology Center, Buenos Aires, CABA, Argentina
[2] Otolaryngology Department, Hospital de Clínicas José de San Martín, University of Buenos Aires, CABA, Argentina
[3] Algorithmic Laboratories Artificial Intelligence, Buenos Aires, Argentina
[4] Clinical Research Department, Hospital Dr. César Milstein, affiliated with the University of Buenos Aires, Buenos Aires, Argentina

Address for correspondence Hernan Chinski, MD, Department of ENT, Dr Chinski Otolaryngology Centre, Buenos Aires, Charcas 2777, Caba 1425, Argentina (e-mail: hernanch@gmail.com).

Facial Plast Surg 2022;38:201–206.

Abstract

During rhinoplasty consultations, surgeons typically create a computer simulation of the expected result. An artificial intelligence model (AIM) can learn a surgeon's style and criteria and generate the simulation automatically. The objective of this study is to determine if an AIM is capable of imitating a surgeon's criteria to generate simulated images of an aesthetic rhinoplasty surgery. This is a cross-sectional survey study of resident and specialist doctors in otolaryngology conducted in the month of November 2019 during a rhinoplasty conference. Sequential images of rhinoplasty simulations created by a surgeon and by an AIM were shown at random. Participants used a seven-point Likert scale to evaluate their level of agreement with the simulation images they were shown, with 1 indicating total disagreement and 7 total agreement. Ninety-seven of 122 doctors agreed to participate in the survey. The median level of agreement between the participant and the surgeon was 6 (interquartile range or IQR 5–7); between the participant and the AIM it was 5 (IQR 4–6), *p*-value < 0.0001. The evaluators were in total or partial agreement with the results of the AIM's simulation 68.4% of the time (95% confidence interval or CI 64.9–71.7). They were in total or partial agreement with the surgeon's simulation 77.3% of the time (95% CI 74.2–80.3). An AIM can emulate a surgeon's aesthetic criteria to generate a computer-simulated image of rhinoplasty. This can allow patients to have a realistic approximation of the possible results of a rhinoplasty ahead of an in-person consultation. The level of evidence of the study is 4.

Keywords
- artificial intelligence
- rhinoplasty
- simulation

Patients generally expect to know the probable results of a rhinoplasty.[1–4] A good surgical result generates an improved perception of facial beauty and bolsters a patient's self-esteem.[5] For more than 30 years, photographic simulations have been used to predict the possible results of a rhinoplasty surgery.[6]

A 2002 study of 56 patients showed that 64% of patients found agreement between their postoperative results and a prior simulation, and were more satisfied with the results than patients who did not receive a prior simulation.[7] Recently, patients were asked their opinions about a preoperative computer simulation; 70% indicated they would be interested in seeing a simulation prior to undergoing surgery and 68% indicated they would not undergo surgery without it.[4] Presently, creating such a simulation requires that the surgeon use photo editing software during a consultation.[3] This requires skill in the use of editing software and consumes valuable time during the consultation. Although there exist general criteria as to what the result of a rhinoplasty should be, there also exist small variations due to the individual preferences of professionals.[5]

Currently, there are smartphone applications that allow users to modify images in ways that are similar to that of a professional editing program.[8,9] However, patients lack the

published online May 29, 2021

Issue Theme Nonsurgical Facial Treatments; Guest Editor: Dario Bertossi, MD, PhD

DOI https://doi.org/10.1055/s-0041-1729911.
ISSN 0736-6825.

criteria to know if these simulations can be reproduced in the operating room.[10] If patients know their possible results beforehand, with a high degree of approximation of a particular surgeon's criteria, they are better able to decide whether to have the surgery or even consult with that particular professional.

In the decade of the sixties, with the advent of the computer age, humankind started developing artificial intelligence. The progressive increase in computing power and parallel data processing made exponential advances in the field possible.[11,12] Medicine has taken advantage of these advances to predict the presence of illnesses in the areas of dermatology and diagnostic imaging. A recent study used 4,216 three-dimensional images to train an artificial intelligence model (AIM) to classify and predict the best aesthetic result possible for every patient that was to undergo orthognathic surgery.[13]

Today, an AIM can be trained to generate a photographic image of the possible nasal profile that could result from surgery based on a photo of the patient's profile and the surgeon's criteria. In order for the AIM to acquire the surgeon's criteria, it must be provided with the simulations the surgeon created previously over the course of his or her professional career.

The current situation with coronavirus disease 2019 (COVID-19) pandemic has modified the way medical consultation is done, namely through video visits. And rhinoplasty is not an exception. Having an automatic facial simulation AIM that recreates the surgeon's aesthetic concept could facilitate this way of consulting.

The objective of this study is to determine whether an AIM is capable of generating images that simulate the result of an aesthetic rhinoplasty by imitating a surgeon's criteria.

Materials and Methods

Participants

Between November 1, 2019 and November 21, 2019, specialists and residents in otolaryngology from the Autonomous City of Buenos Aires who attended a scientific conference were recruited to complete a web-based survey through a weblink. The sampling was nonprobabilistic and the main exclusion criteria was refusal to participate or visual impairment.

The study was conducted in accordance with the care of clinical research participants as per the Declaration of Helsinki and as approved by the Hospital Dr. César Milstein's Institutional Review Board, including written informed consent. The survey data was anonymous. Patients signed an informed consent form to authorize the use of their images.

Survey Instrument

A cell phone application was created to display pairs of photographs sequentially. Each photograph pair included the original right-profile photograph of the patient on the superior half of the screen, and the simulation of the proposed rhinoplasty on the inferior half (►**Fig. 1**).

It is important to underscore that the inferior image displayed by the app could be either the surgeon's or the AIM's simulation, as randomly selected by the app.

Fig. 1 Screenshot from the mobile app used in our survey showing a simulated image.

Participants were asked to indicate their level of agreement with the proposed simulation using a seven-point Likert scale, with 1 indicating total disagreement and 7 total agreement.

Other participant data was also collected, such as years of experience in the practice of medicine, number of rhinoplasties performed, and whether simulations were used prior to surgery. Each participant was shown a maximum of 20 pairs of images. An example of the app can be found at: http://ain.algolabs.ai.

Machine Learning and Deep Learning

Machine Learning is a type of Generative Adversarial Network model that consists of abstraction layers of images, known as convolution and deconvolution layers, and two dense neural networks that function as discriminators. The images are codified in matrices of numbers based on the position and RGB value of each pixel, providing the neural network with two images—(1) the original image of the patient, and (2) the image simulated by the surgeon. The AIM compares the images, detecting differences in a way that enables it to compose an approximation function. Applying this function, the system takes an image and generates a new one. It then verifies the result and revises the function to tailor it even further. This process is repeated over and over again, constantly correcting the function, thus resulting in machine learning.

In this study, the neural network was provided with 1,200 pairs of images of patients who attended a consultation for an aesthetic primary rhinoplasty. The images were generated by the lead author over the course of his long professional career.

The images were taken with a Canon camera, model EOS 50D, with a fixed 85 mm lens, shutter speed 1/125, F 6.3, ISO 1200, and using a uniform chromakey blue backdrop.

Images of highly complex cases—such as congenital nasal deformities, saddle nose, revision surgery, or a history of facial trauma surgery—were excluded. Simulations that included changes to other parts of the face, especially simulations of chin augmentations, were also excluded.

The model was created in the Python programming language using the open-source machine learning library known as TensorFlow.

The cellphone application's interface was also developed in Python, and executed in a cloud computing service.

Equipment

The AIM was developed on a high-performance computer with the following characteristics: 8-core CPU i7, 16GB of RAM, 1TB hard drive, GPU Nvidia 1050 Ti with 4GB video card.

Training the Model

The model was trained with the hardware described above for 800 hours, consuming 100% of the computing power and GPU memory with an approximate average consumption of 75 Watt/h per GPU. Afterward, the AIM was able to simulate images that were deemed adequate by the authors. Once the training was completed, the software was able to take a side profile image of a patient and return a simulated image in a matter of milliseconds. The generated images were 512×512 pixels (►**Fig. 2**).

Fig. 2 Examples of simulations created by the AIM. AIM, artificial intelligence model.

Photo Selection for the Survey

The selection was made with a python library that generates random numbers within a range, with the minimum being 1 and the maximum being the number of available images, thus generating a list of images. Individual and different lists were generated for each user. The application did not indicate whether the image displayed was generated by a surgeon or the AIM.

Statistical Analysis

The categorical variables were described as absolute frequency and proportions, and the continuous variables as mean and standard deviation or median and interquartile range based on the distribution of the data.

The evaluator's degree of agreement with the simulation shown was measured using a seven-point Likert scale, with 1 indicating total disagreement and 7 total agreement.

Agreement was considered total or partial for scores of 5, 6, and 7.

A Bland Altman plot was used to establish the level of agreement between the survey respondent and the simulated image, and the Spearman's correlation coefficient was used to determine agreement between the average score obtained for the surgeon's and AIM's images in each evaluation strata. In other words, the frequency count at each evaluation level between 1 and 7, was expressed as a percentage of total evaluations. Python version 3.7 was used for the calculations.

Sample Size

The calculation was based on the difference between means. Considering that a prior ad hoc pilot study presented a mean on the Likert scale of 6.23 with a variance of 0.67 for images generated by the surgeon and a mean of 5.92 with a variance of 1.27 for the AIM's simulations, it was estimated that at least 318 images needed to be evaluated to have a significance of 0.05 and a power of 80.

Table 1 Basal demographic data from the study's 97 participants

	$n = 97$	%
Years of professional medical experience, mean (SD)	AVG 12.47	SD 9.8
Rhinoplasties performed		
Less than 10	30	30.93%
10–50	25	25.77%
50–100	9	9.28%
More than 100	33	34.02%
Use of a simulation		
Always	33	34.02%
Almost always	15	15.46%
Sometimes	13	13.40%
Almost never	12	12.37%
Never	24	24.74%

Abbreviations: AVG, average; SD, standard deviations.

Results

An invitation to participate was extended to 122 medical specialists, and 97 accepted, completing at least one evaluation of image included in the analysis.

The mean time of experience in medical practice was 12.4 years, 49.5% responded that they created simulations almost always or always, and 34.0% had performed more than 100 aesthetic rhinoplasties (►**Table 1**). A total of 1,436 images were evaluated.

Level of Agreement

The median final score obtained by the surgeon's simulation was 6 (interquartile range or IQR 5–7), while for the AIM it was 5 (IQR 4–6), *p*-value <.0001.

The evaluators were in total or partial agreement with the result of the AIM's simulation 68.4% of the time (95% confidence interval or CI 64.9–71.7) and 77.3% of the time with the surgeon's simulation (95% CI 74.2–80.3).

The evaluation obtained by the AIM and the surgeon at each strata of agreement can be seen in ►**Fig. 3**.

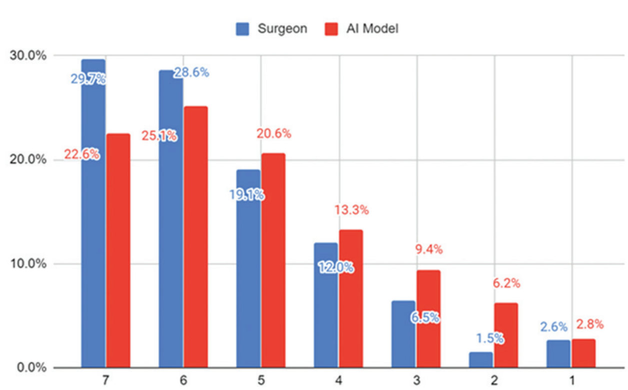

Fig. 3 Graphic showing proportion of evaluations at each Likert strata (1–7) scored by the surgeon and by the artificial intelligence model.

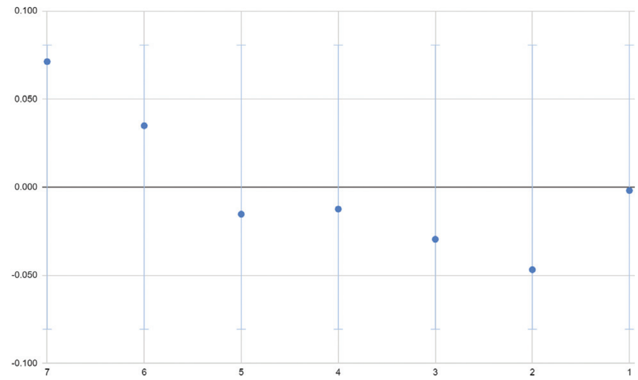

Fig. 4 Bland Altman plot showing the correlation between the scores received by the artificial intelligence model and the surgeon. For all of the evaluation strata, the difference in the Likert scale score (1 to 7) does not exceed two standard deviations.

The Spearman's correlation of scores between the strata 1 to 7 was 0.92, *p*-value 0.0025.

In the Bland Altman plot, it can be observed that, at each stratum, the difference in the evaluations obtained for the surgeon and the AIM did not exceed two standard deviations (►**Fig. 4**).

Discussion

This study demonstrated that an AIM can simulate an image of a rhinoplasty result in a way that is similar to that of a surgeon, with 92% concordance, according to the criteria of the doctors surveyed.

Participants indicated a high degree of agreement with the simulations created by both the surgeon and the AIM.[14]

The high levels of agreement attained in this study between the participants and the displayed images suggest that the AIM is designing images similar to what a surgeon would generate. Had the agreement been low, we would not be able to confirm this premise. For instance, it could have turned out that the surgeon generates simulations that are "too upturned" and the AIM generates simulations that are "too droopy." In both cases there would be disagreement, but for different reasons.

Hamet and Tremblay described the use of artificial intelligence in medicine, particularly the use of deep learning and machine learning as used in this study.[11] Kanevsky et al described different applications of artificial intelligence in the field of plastic surgery using machine learning concepts.[12] Yeong described the use of an AIM to prognosticate healing time for a burn wound.[14] However, these types of models are trained to select from a predetermined number of options. The generation of an image formed by thousands of pixels is a much more complex process for an AIM. Knoops et al used supervised machine learning to generate 3D simulations and potential results of orthognathic surgery [13].

There are many studies that describe the usefulness of a facial simulation in aesthetic rhinoplasty, but until now there has not been any research on the application of AI for this purpose.[1–4,7]

The AIM was provided only with preoperative photographs. This study is not based on postoperative results. The purpose of the simulations is to allow the patient and the surgeon to arrive at an agreement regarding the aesthetic criteria to be used, and is not intended as a guaranteed result.

The AIM generated simulations of 512×512 pixels. This resolution is useful for viewing the image on a cell phone but can be suboptimal for the larger monitors usually used during a consultation. The AIM developed for this study is limited in that it cannot simulate complex cases of nasal deformities or surgeries for patients with thick skin.

The AIM was provided with images that had a uniform backdrop and was not trained to edit photos with nonuniform backdrops. This means that if an image with a nonuniform backdrop is inputted, the AIM will return a distorted image. The cropping of an image to eliminate a nonuniform backdrop requires either manual retouching of the image or the use of high-cost licensed software. The taking of the photograph can pose a challenge for patients trying to do it by themselves.

Further, selfie-type photos taken at home tend to produce deformed images due to the large angular effect of a cell phone lens.[8,9] Both of these situations could represent a limitation if one attempts to use the AIM with homemade photographs.[10] Additionally, the AIM that we developed will always return a simulation using the surgeon's criteria, and also taking into account the feasibility of making the change in the operating room; however, it does not consider the patient's opinion. As presently programmed, the AIM is unable to satisfy a patient's request to see a pointier or smaller nose. *During consultation, the surgeon can retouch the simulated image according to the patient's wishes and technical possibilities. For example, some patients prefer their nose to be more or less rotated or projected.*

Another limitation worth mentioning is that the professionals who agreed to respond to the survey were possibly more motivated than those who refused, which could introduce some type of selection bias.

In this study the AIM was given photographs produced by a single surgeon of patients who had a primary rhinoplasty consultation. The results cannot be extrapolated to cases of greater complexity, such as a revision rhinoplasty, or cases of a combined rhinoplasty and chin augmentation. The result returned by the AIM is based on the experience of prior simulations.

To train a model, it is necessary to provide the system with several thousands of photograph pairs of patients. Consequently, only those surgeons with several years of professional experience have enough material to provide to the AIM to imitate surgeon's own aesthetic criteria.

One of the strengths of this study is that the evaluators did not know which simulations were created by the AIM and which were done by the surgeon, thus preventing social desirability bias. Another of its strengths is that the evaluators were otolaryngology doctors who were trained and had knowledge of aesthetic rhinoplasty; moreover, half of the participants always or almost always used simulations in their daily practice. This made them more demanding than other populations.

The results suggest that using the simulation created by the AIM could free up consulting time, making the surgeon's job easier. An application could be developed and made available to surgeons, who could take a photo of a patient and process it with the AIM, instantly producing a simulated result.

The COVID-19 pandemic has shown us the importance of video visits. Having a tool like this could be really helpful for both patient and surgeon.

If the abovementioned limitations regarding the taking of homemade photos can be overcome, another possible use scenario would be for patients to take their own profile photos with the cell phone application and thus generate the anticipated surgery result that the surgeon would propose without even having an in-person consultation. Knowing the potential result a surgeon might propose would allow patients to determine if it meets their expectations. This would help patients decide whether or not to schedule an in-person consultation, saving their time and money, especially for patients who live hundreds of kilometers from the doctor's office. Further, by having access to a simulation prior to a medical consultation, the patient has more time to reflect on the possible aesthetic result and discuss the matter with family and friends, thereby making more detailed questions in the consultation.

Despite what we just said, we feel it would be important for the application to have a disclaimer which informed the patient that simulation is not a promise of result, it is just an estimation of a result and an objective to reach.

The simulation created by the AIM could give the patient a good general idea of the realistic expectations for a rhinoplasty result performed by other surgeons, as well. Additionally, each surgeon could train the AIM with their own set of simulations to create a personal style. For educational purposes, professionals in training could use an AIM trained by a more experienced professional and learn how that surgeon would resolve a certain rhinoplasty case.

An even more ambitious possibility is the creation of a single collaborative application that can store the criteria of several surgeons who contributed their images to the development of such an AIM.

Conclusion

An AIM can be trained to generate images of computer simulations of an aesthetic rhinoplasty result, emulating the aesthetic criteria of a surgeon. This allows patients to have a realistic approximation of the possible results of their rhinoplasties.

Conflict of Interest
None declared.

References

1 Lekakis G, Claes P, Hamilton GS III, Hellings PW. Evolution of preoperative rhinoplasty consult by computer imaging. Facial Plast Surg 2016;32(01):80–87
2 Persing S, Timberlake A, Madari S, Steinbacher D. Three-dimensional imaging in rhinoplasty: a comparison of the simulated versus actual Result. Aesthetic Plast Surg 2018;42(05):1331–1335
3 Vuyk HD, Stroomer J, Vinayak B. The role of computer imaging in facial plastic surgery consultation: a clinical study. Clin Otolaryngol Allied Sci 1998;23(03):235–243

4 Aksakal İA, Keles MK, Engin MS, Aydoğdu İO, Küçüker İ Preoperative simulation in planning rhinoplasty: evaluation from patients' and surgeons' perspectives. Facial Plast Surg 2017;33(03): 324–328
5 Chinski H, Chinski L, Armijos J, Arias JP. Rhinoplasty and its effects on the perception of beauty. Int Arch Otorhinolaryngol 2013;17 (01):47–50
6 Constantian MB. Interactive computer graphics: a new technology to improve judgment in rhinoplasty. Clin Plast Surg 1987;14 (04):623–630
7 Sharp HR, Tingay RS, Coman S, Mills V, Roberts DN. Computer imaging and patient satisfaction in rhinoplasty surgery. J Laryngol Otol 2002;116(12):1009–1013
8 Kayastha D, Vakharia KT. The evolving roles of computer-based technology and smartphone applications in facial plastic surgery. Curr Opin Otolaryngol Head Neck Surg 2019;27(04):267–273
9 Larrosa F, Dura MJ, Roura J, Hernandez A. Rhinoplasty planning with an iPhone app: analysis of otolaryngologists response. Eur Arch Otorhinolaryngol 2013;270(09):2473–2477
10 Henderson JL, Larrabee WF Jr, Krieger BD. Photographic standards for facial plastic surgery. Arch Facial Plast Surg 2005;7(05):331–333
11 Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism 2017;69S:S36–S40
12 Kanevsky J, Corban J, Gaster R, Kanevsky A, Lin S, Gilardino M. Big data and machine learning in plastic surgery: a new frontier in surgical innovation. Plast Reconstr Surg 2016;137(05):890e–897e
13 Knoops PGM, Papaioannou A, Borghi A, et al. A machine learning framework for automated diagnosis and computer-assisted planning in plastic and reconstructive surgery. Sci Rep 2019;9(01):13597
14 Yeong E-K, Hsiao T-C, Chiang HK, Lin C-W. Prediction of burn healing time using artificial neural networks and reflectance spectrometer. Burns 2005;31(04):415–420

