Article published online: 2022-03-21

APS
Archives of Plastic Surgery

Pediatric/Craniomaxillofacial/Head&Neck
Original Article

Computerized Sunnybrook facial grading scale (SBface) application for facial paralysis evaluation

Supasid Jirawatnotai[1], Pojanan Jomkoh[1], Tsz Yin Voravitvet[1], Wuttipong Tirakotai[2], Natthawut Somboonsap[3]

[1]*Plastic and Reconstructive Surgery Unit, Department of Surgery, Lerdsin Hospital, College of Medicine, Rangsit University, Bangkok;* [2]*Department of Neurosurgery, Prasat Neurological Institute, Bangkok;* [3]*Department of Otorhinolaryngology, Buddhasothorn Hospital, Chachoengsao, Thailand*

Background The Sunnybrook facial grading scale is a comprehensive scale for the evaluation of facial paralysis patients. Its results greatly depend on subjective input. This study aimed to develop and validate an automated Sunnybrook facial grading scale (SBface) to more objectively assess disfigurement due to facial paralysis.

Methods An application compatible with iOS version 11.0 and up was developed. The software automatically detected facial features in standardized photographs and generated scores following the Sunnybrook facial grading scale. Photographic data from 30 unilateral facial paralysis patients were randomly sampled for validation. Intrarater reliability was tested by conducting two identical tests at a 2-week interval. Interrater reliability was tested between the software and three facial nerve clinicians.

Results A beta version of the SBface application was tested. Intrarater reliability showed excellent congruence between the two tests. Moderate to strong positive correlations were found between the software and an otolaryngologist, including the total scores of the three individual software domains and composite scores. However, 74.4% (29/39) of the subdomain items showed low to zero correlation with the human raters ($\kappa < 0.2$). The correlations between the human raters showed good congruence for most of the total and composite scores, with 10.3% (4/39) of the subdomain items failing to correspond ($\kappa < 0.2$).

Conclusions The SBface application is efficient and accurate for evaluating the degree of facial paralysis based on the Sunnybrook facial grading scale. However, correlations of the software-derived results with those of human raters are limited by the software algorithm and the raters' inconsistency.

Keywords Facial paralysis / Scale / Mobile application / Photographic / Software validation

Correspondence:
Supasid Jirawatnotai
Plastic and Reconstructive Surgery Unit, Department of Surgery, Lerdsin Hospital, 190 Silom Road, Bangrak, Bangkok 10500, Thailand
Tel: +66-2-353-9801
Fax: +66-2-353-9756
E-mail: add345@gmail.com

Received: September 26, 2020 • Revised: January 5, 2021 • Accepted: February 24, 2021
pISSN: 2234-6163 • eISSN: 2234-6171 • https://doi.org/10.5999/aps.2020.01844 • Arch Plast Surg 2021;48:269-277

INTRODUCTION

Facial paralysis remains a major challenge to reconstructive surgeons due to its complex etiology, disease course, and variation in severity. Assessment of the severity of facial paralysis and evaluation of its progression are crucial and require a dependable quantitative grading scale. Various scoring systems have been proposed to measure the degree of disfigurement as nu-

Copyright © 2021 The Korean Society of Plastic and Reconstructive Surgeons
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

merical data [1-5]. Simple and user-friendly systems are commonly accepted, but may not capture certain critical details. Complicated systems yield elaborate information, but they are unpopular due to being time- and labor-intensive.

The conventional systems most used by facial paralysis specialists are the House-Brackmann, Yanagihara, and Sunnybrook facial grading scale. Researchers have reported good consistency across these three systems [6]. However, the Sunnybrook scale is considered the current standard in evaluating outcomes and synkinesis due to its comprehensive scope, ease of use, and rapid results interpretation [7]. We have been using the Sunnybrook facial grading scale in our facial paralysis clinic since 2014. The Sunnybrook facial grading scale, unlike others, systematically focuses on each subunit of facial movement (eyebrows, eyelids, nasal base, upper lip, and lower lip), while subjects are instructed to make six simple facial expressions. Furthermore, the Sunnybrook facial grading scale globally evaluates resting symmetry, symmetry of voluntary movement, and the degree of synkinesis, which are vital for physicians to clearly understand the progression and improvement of the disease [4].

A new clinician-graded electronic scale has been proposed to address the above drawbacks. However, similar to the Sunnybrook facial grading scale, it still requires experienced evaluators to subjectively assess and score the degree of disfigurement and synkinesis using a form [5]. Either system may require a patient to perform each facial movement multiple times, causing fatigue, and the input data can be affected by human inconsistency as a result of a subjective visual interpretation. Our objective in this study was to develop a software application based on the Sunnybrook facial grading scale to evaluate, calculate, and interpret the result while minimizing subjective human input. It was designed to be fast and easy-to-use by inexperienced observers.

We developed the "SBface" application using image processing technology to automatically detect and mark facial features. The software functionality spans three domains: resting symmetry, symmetry of voluntary movement, and synkinesis assessment. Comparisons are then made between the paralyzed and the unaffected sides of the face. Intrarater and interrater consistency testing between the scores generated by the software and three practicing facial nerve physicians were performed to validate the software.

METHODS

The study was approved by the Ethics Committee of Lerdsin Hospital (EC No. 611038/2561). The photographs of 30 unilateral facial paralysis patients with varying severities were randomly selected from our facial clinic photographic database between July 2018 to July 2019. Photographs were included in this study with patients' consent.

Design of the application

Software developers designed an iOS-based mobile application with the researchers' guidance. Photos of patients making the six standardized facial expressions from the Sunnybrook scale were taken using the device's built-in camera. Photographic data could also be retrieved from the device storage. An on-screen guiding marker was displayed over the patient's face to ensure a correct face position and angle (Fig. 1).

After all photos were verified by the user, the basic information of the patient, including the date of assessment, identification number, and side of the face affected, was recorded. All records and photos were automatically stored on the device.

The VNFaceObservation image analysis technology (Vision Framework, iOS version 11.0+, 2017; Apple Inc., Cupertino, CA, USA) automatically detected each facial feature on both the paralyzed and non-paralyzed sides. Comparisons of the resting positions of each facial feature were made between the paralyzed and non-paralyzed sides. Differences in position for each facial expression were recorded. The movement of facial mark-

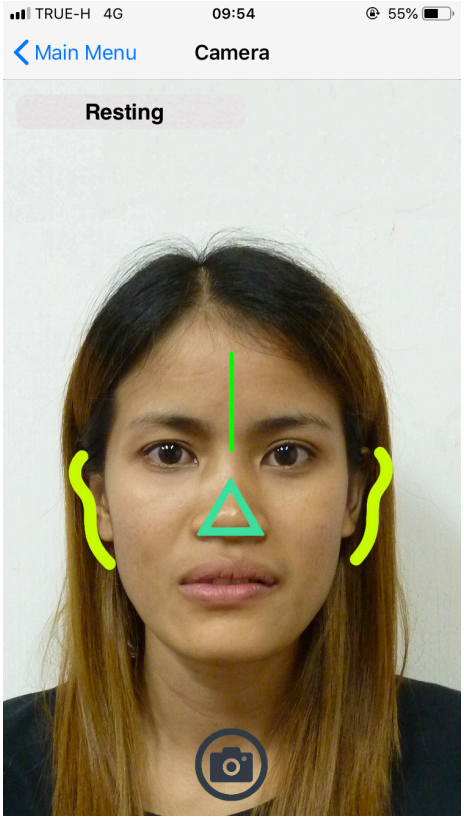

Fig. 1. An on-screen guiding marker. Throughout the photograph-taking steps, a guide was displayed over the patient's face to ensure a correct face position and angle. The face must be aligned with the guiding marker and in the center of the frame.

ers on the paralyzed side was compared to that of the unaffected side, and the difference was calculated as a percentage. Unwanted movements were detected as synkinesis (Fig. 2).

Locations of markers, movement determination, and calculations

Forehead wrinkle (frontalis muscle)
The midpoints of each eyebrow were marked. The amount of vertical displacement of the marker on the paralyzed side was compared to that of the unaffected side and calculated as a percentage.

Gentle eye closure (orbicularis oculi muscle)
The distance between the upper and lower eyelids was measured at the midpupillary point. The amount of vertical distance reduction on the paralyzed side was compared to that of the unaffected side and calculated as a percentage.

Open-mouth smile (zygomaticus and risorius muscles)
The amount of commissure excursion from the resting position to maximal smile on the paralyzed side was compared to that of the normal side and calculated as a percentage.

Snarl (levator labii alaeque nasi and levator labii superioris muscle)
The widest points of the alar base were marked. The amount of excursion from the resting position to maximal snarl on the paralyzed side was compared that of the normal side and calculated as a percentage.

Lip pucker (orbicularis oris muscle)
The horizontal distances from each side of the patient's oral commissures to the lateral edges of the face were measured. The amount of medial displacement from the resting position to maximal puckering on the paralyzed side was compared to that of the normal side and calculated as a percentage.

Score generation

Image analysis protocol was organized into three domains based on the Sunnybrook facial grading system.

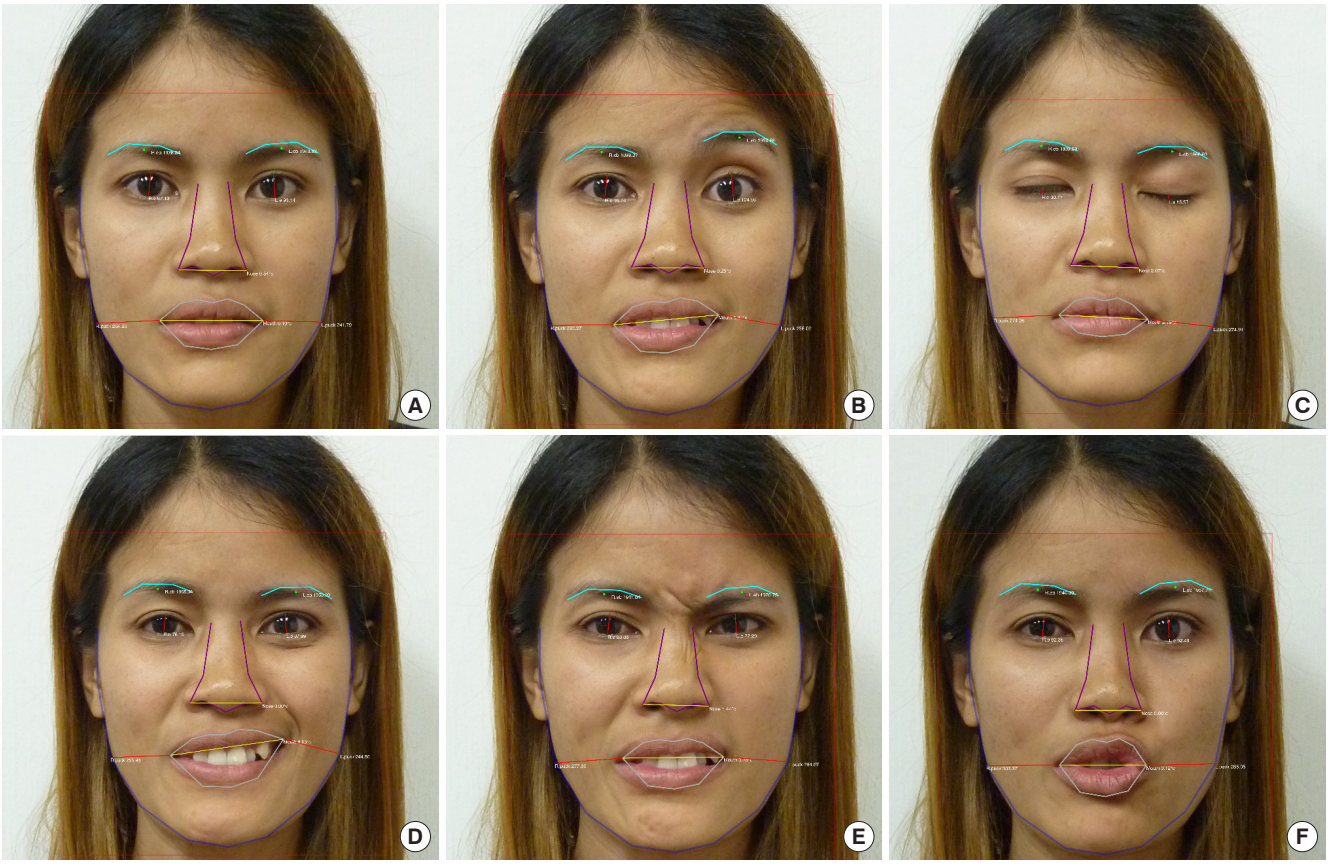

Fig. 2. Locations of markers. The image analysis technology automatically detected each facial feature. The resting positions of facial features were compared between paralyzed and non-paralyzed sides. The movements of each marker on the paralyzed side were also compared to ones on the unaffected side during expression. (A) Resting, (B) forehead wrinkle, (C) gentle eye closure, (D) open-mouth smile, (E) snarl, (F) lip pucker.

Resting symmetry domain

At rest, the differences between the normal side and the affected side were compared in three regions: the size of the palpebral openings, the presence of the nasolabial fold, and the degree of a patient's oral commissures. Using a cutoff point at 20%, scores were given as 0 or 1, with 0 signifying a difference of 20% or less and 1 signifying a difference of more than 20%. A score was not given to measure the nasolabial fold since the software could not detect it.

Symmetry of voluntary movement domain

The Sunnybrook scale's five subjective descriptions of movement, which are "no movement," "initiates slight movement," "movement with mild excursion," "movement almost complete," and "complete movement," were quantified as objective scores of 1 to 5. Every 20 percentage point difference was equal to 1 point, with 20% difference being 1 point and 100% being 5 points.

Synkinesis domain

The software detected abnormal movements in other parts of the face beyond the scope of each assigned facial expression. For example, if brow lifting occurred in the smiling photo, it was detected as synkinesis. The subjective criteria for synkinesis in the Sunnybrook scale are "no synkinesis," "slight synkinesis," "obvious synkinesis," and "disfiguring synkinesis/gross movement," which were transformed into scores of 0 to 3. Since the movements during synkinesis were usually not as strong as voluntary movements, the percentage tiers were lowered so that a difference exceeding 61% was considered "disfiguring." Only the most intense unwanted movements were calculated.

After the six photographs were taken, the software calculated the scores and presented the results. The automated score of each parameter could always be overridden by users. The last display screen showed the calculated sum of each parameter and the composite score (Fig. 3).

Software reliability test

Interrater and intrarater reliability were tested. Interrater reliability between the software and three clinicians, which included a plastic surgeon, otolaryngologist, and plastic surgery resident, was tested using photographic data from 30 randomly selected patients with unilateral facial paralysis from our practice. Cohen's kappa coefficient (κ), the weighted kappa, and Pearson correlation coefficient (r) analyses were applied to the assessment results. Total domain-specific and composite test scores were used to measure correlation between raters. Pairs of data sets that showed a strong correlation ($r > 0.7$) were further tested for agreement using a Bland-Altman plot. To determine software repeatability, seven patients with varying degrees of unilateral facial paralysis were selected from the study group and underwent evaluation using the software a second time after a 2-week interval.

RESULTS

A beta version of the SBface application was tested in this study. Thirty sets of photographs of unilateral facial paralysis patients were randomly chosen for assessments of interrater and intrarat-

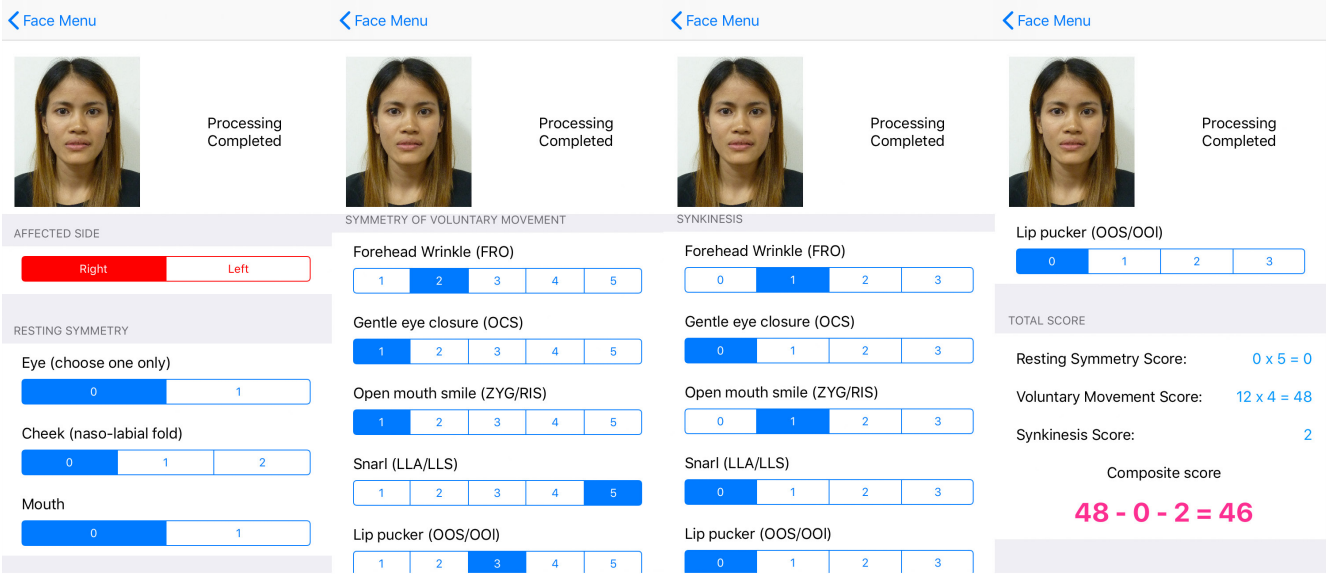

Fig. 3. Interface of the application. Scores of each subdomain item, which can be manually changed by users, are displayed. The overall scores of each domain and the final composite score are calculated and shown in the last screen.

er reliability. Fifty-four percent of the patients were female. The mean age was 45.9 ± 16.7 years (range, 12–72 years). The most common cause of paralysis was vestibular schwannoma, followed by Bell's palsy and trauma (Table 1).

The analysis of software repeatability showed good congruence after the 2-week test interval. Ten of the 13 categorical items in the three software domains showed moderate to almost perfect agreement after the 2-week test interval ($\kappa > 0.5$, $P < 0.05$). The total domain-specific and composite scores after the 2-week interval showed strong positive correlations ($r > 0.8$, $P < 0.05$) with good agreement on the Bland-Altman plot. The mean differences of the total domain-specific and composite scores after the 2-week test intervals were –2.14, 1.71, 1.00, and 2.85, respectively. More than 95% of all data points fell within the limits of agreement (Table 2, Fig. 4).

Software reliability was also tested against the three human raters. Strong positive correlations were found between the software and the otolaryngologist, including the total scores regarding the symmetry of voluntary movement ($r = 0.72$, $P < 0.001$) and the composite scores ($r = 0.70$, $P < 0.001$), whereas only a moderate correlation was found in the resting symmetry domain ($r = 0.43$, $P = 0.029$). In the symmetry of voluntary movement domain, the software also showed moderate correlation ith the plastic surgeon and a fair correlation with the plastic surgery resident ($r = 0.53$, $P = 0.014$ and $r = 0.35$, $P = 0.041$, respectively). Other domain-specific and composite scores were not correlated between the software and the human raters. Furthermore, 74.4% (29/39) of subdomain items showed only low to zero correlation with the human raters ($\kappa < 0.2$) (Table 3, Fig. 5).

The validity test of the three clinicians evaluating patients using the conventional Sunnybrook facial grading scale showed strong correlations ($r > 0.70$, $P < 0.05$) in 83.3% (10/12) of total and composite scores, and 70% of the correlated scores had good agreement. Four of the 39 (10.3%) of the subdomain items failed to correspond ($\kappa < 0.2$) (Table 4).

Table 1. Demographic data and patient characteristics

Characteristic	Value
Age (yr)	
Mean ± SD	45.9 ± 16.7
Range	12–72
Sex, No. (%)	
Female	16 (54)
Male	14 (46)
Cause of facial paralysis, No. (%)	
Vestibular schwannoma	21 (70)
Bell's palsy	8 (26)
Trauma	1 (4)

Table 2. Intrarater reliability

Variable	App 1 vs. App 2	
	κ	P-value
Resting symmetry		
Eye	0.36	0.106
Cheek	0.81	0.002*
Mouth	0.41	0.135
Total score	0.83[a]	0.018*
	–2.14 (–10.011 to 5.725)[b]	0[c]
Symmetry of voluntary movement		
Forehead wrinkle	0.68	0.003*
Gentle eye closure	0.71	0.003*
Open mouth smile	0.76	0.009*
Snarl	0.60	0.007*
Lip pucker	–0.17	0.670
Total score	0.95[a]	< 0.001*
	1.71 (–9.464 to 12.892)[b]	0[c]
Synkinesis		
Forehead wrinkle	0.50	0.009*
Gentle eye closure	1.00	0.004*
Open mouth smile	0.63	0.010*
Snarl	0.53	0.026*
Lip pucker	0.58	0.043*
Total score	0.91[a]	0.004*
	1.00 (–1.000 to 3.000)[b]	0[c]
Composite score	0.95[a]	0.001*
	2.85 (–10.972 to 16.686)[b]	0[c]

App 1, first application test; App 2, second application test; LOA, limits of agreement.
[a]Pearson correlation coefficient (r); [b]Mean difference (LOA); [c]Percentage of data lying outside LOA.
*Statistically significant ($P < 0.05$).

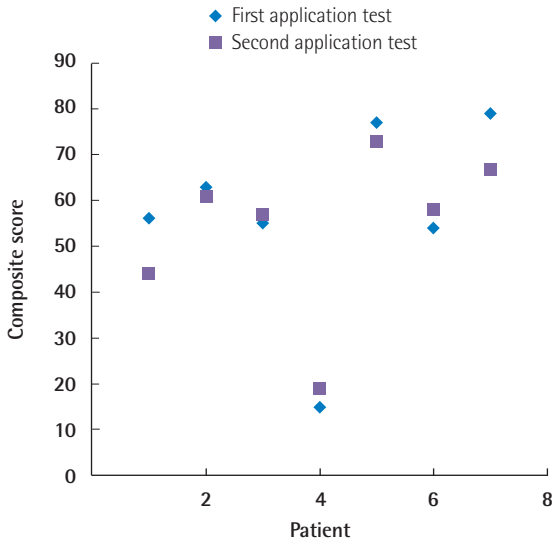

Fig. 4. Intrarater reliability. The diagram demonstrates composite scores calculated from seven subjects rated by the computerized Sunnybrook facial grading system (SBface application) after a 2-week interval. The composite scores between the two tests showed a strong positive correlation ($r = 0.95$, $P = 0.001$) with good agreement on the Bland-Altman plot.

Table 3. Interrater reliability of the software against the three human raters

Variable	SBface vs. plastic surgeon		SBface vs. otolaryngologist		SBface vs. resident	
	κ	P-value	κ	P-value	κ	P-value
Resting symmetry						
Eye	–0.03	0.556	0.05	0.351	0.17	0.151
Cheek	–0.01	0.526	0.29	0.009*	0.03	0.387
Mouth	0.70	<0.001*	0.45	0.007*	0.40	0.009*
Total score	0.13[a]	0.580	0.43[a]	0.029*	0.26[a]	0.123
Symmetry of voluntary muscle						
Forehead wrinkle	0.00	0.500	0.03	0.303	0.01	0.390
Gentle eye closure	–0.02	0.611	0.00	0.500	–0.05	0.740
Open mouth smile	0.55	<0.001*	0.63	<0.001*	0.40	<0.001*
Snarl	–0.09	0.797	0.05	0.231	0.13	0.945
Lip pucker	0.21	0.050*	0.26	0.019*	0.12	0.116
Total score	0.53[a]	0.014*	0.72[a]	<0.001*	0.35[a]	0.041*
Synkinesis						
Forehead wrinkle	–0.00	0.500	0.04	0.155	0.01	0.256
Gentle eye closure	0.01	0.440	–0.17	0.894	-0.02	0.681
Open mouth smile	0.27	0.030*	0.21	0.027*	0.01	0.465
Snarl	0.17	0.057	0.14	0.050*	0.14	0.007*
Lip pucker	–0.07	0.651	–0.13	0.822	–0.10	0.799
Total score	0.15[a]	0.524	0.33[a]	0.104	0.05[a]	0.769
Composite score	0.37[a]	0.099	0.70[a]	<0.001*	0.25[a]	0.141

SBface, computerized Sunnybrook facial grading system.
[a]Pearson correlation coefficient (r).
*Statistically significant (P<0.05).

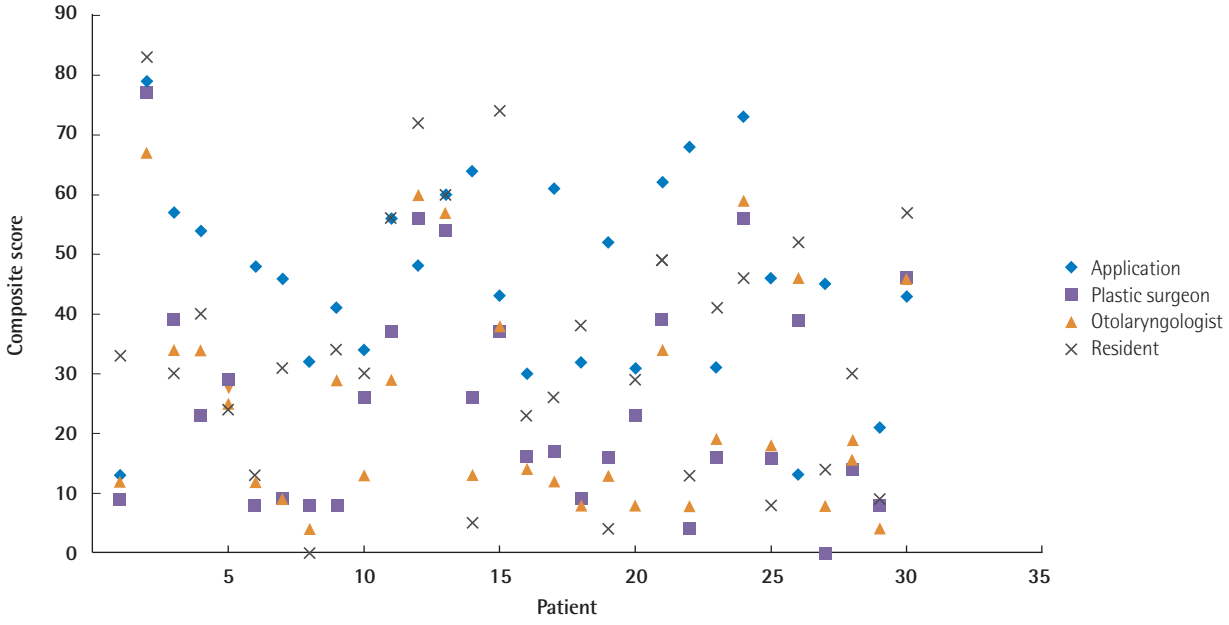

Fig. 5. Interrater reliability. The diagram demonstrates composite scores calculated from 30 subjects rated by the computerized Sunnybrook facial grading system (SBface application) and the three human raters. There was only a strong correlation between composite scores rated by the SBface application and the otolaryngologist (r=0.70, P<0.01) with a mean difference of 20.48 (limits of agreement, –8.11 to 49.07), and 4% of the data lay outside the limits of agreement. The composite scoring generated by the application was not correlated with that of the other raters.

DISCUSSION

An objective grading system for patients with facial paralysis is crucial for physicians to evaluate severity and monitor treatment results, as well as for patients to conduct self-evaluations [8-10]. A large number of reconstructive surgeons prefer the Sunnybrook scale over the House-Brackmann scale due to its detailed assessment of specific regions of the face and their dynamics.

Table 4. Interrater reliability between the three human raters

Variable	Resident vs. plastic surgeon		Resident vs. otolaryngologist		Plastic surgeon vs. otolaryngologist	
	κ	P-value	κ	P-value	κ	P-value
Resting symmetry						
Eye	0.45	0.008*	0.11	0.283	0.62	0.013*
Cheek	0.58	<0.001*	0.31	0.014*	0.67	0.001*
Mouth	0.51	0.008*	0.58	0.001*	0.64	0.011*
Total score	0.79[a]	<0.001*	0.58[a]	0.002*	0.92[a]	<0.001*
Symmetry of voluntary muscle						
Forehead wrinkle	0.27	0.001*	0.52	<0.001*	0.77	0.001*
Gentle eye closure	0.45	<0.001*	0.39	<0.001*	0.60	<0.001*
Open mouth smile	0.58	<0.001*	0.54	<0.001*	0.85	<0.001*
Snarl	0.08	0.264	0.40	0.001*	0.82	<0.001*
Lip pucker	0.50	<0.001*	0.25	0.038*	0.70	<0.001*
Total score	0.82[a]	<0.001*	0.75[a]	<0.001*	0.96[a]	<0.001*
Synkinesis						
Forehead wrinkle	0.47	0.001*	0.64	<0.001*	0.62	0.013*
Gentle eye closure	[b]	[b]	0.32	0.005*	0.72	0.001*
Open mouth smile	0.20	0.037*	0.46	0.003*	0.45	0.013*
Snarl	0.27	0.069	–0.03	0.672	0.45	0.010*
Lip pucker	0.32	0.011*	0.78	<0.001*	0.49	0.005*
Total score	0.72[a]	<0.001*	0.65[a]	0.001*	0.92[a]	<0.001*
Composite score	0.88[a]	<0.001*	0.80[a]	<0.001*	0.95[a]	<0.001*

[a]Pearson correlation coefficient (r); [b]Kappa (κ) was not calculated for this data set because observed concordance was smaller than mean-chance concordance.
*Statistically significant (P < 0.05).

However, the main shortcoming of the Sunnybrook facial grading scale itself is the inconsistency of user-given scores. Biases in the Sunnybrook facial grading scale have been reported, which can affect the results [11-13].

Evaluation with this system can be time-consuming, especially for inexperienced users, and some may even require the use of the Sunnybrook handout while examining patients, which is unfeasible. In response to this dilemma, we developed an electronic mobile application based on the Sunnybrook facial grading scale. The software provides exact calculations of grading parameters, reducing human error resulting from subjective visual assessments. We designed the SBface application to improve the existing well-designed scale rather than create a new system altogether. This makes it easier for familiar users of the Sunnybrook scale to adopt our own automated version. Furthermore, it was designed to be used by inexperienced raters. Thus, anyone capable of taking the required photographs would be able to utilize the application without a significant learning curve.

A different group of researchers developed a software known as electronic facial paralysis assessment (eFACE) in 2015 [5]. The program assesses recorded video clips of patients to grade static, dynamic, and synkinesis disfigurements. The use of the eFACE software has become popular due to its ease of use and compatibility with the Sunnybrook facial grading scale [14,15]. The eFACE software's visual analog scale, as opposed to a simple ordinal scale, optimizes sensitivity for detecting differences; however, it can be quite tedious to use. The requirement of using video recordings to assess patients, necessitating scrolling and pausing of the video, can be a time-consuming task for users. Evaluating different variables through different frames of a video can also cause deviation in the results, even from the same raters.

The facogram and the OSCAR system are other software programs developed for the same purpose, neither of which need facial markers. The facogram was based on the House-Brackmann grading system, while the OSCAR system used its own method of analysis. However, both programs also require video recordings for assessments [16,17].

As opposed to clinician-graded scales, image processing software identifies distinct locations of human facial features. Patients with facial paralysis exhibit abnormal facial movements within the image processor's parameters compared to the unaffected side. The software very precisely calculated the differences in each specific location; however, it failed to locate some facial features, particularly in extremely disfigured subjects. Another drawback we encountered during the development of the application was the presence of the nasolabial fold, which is a very subjective parameter. It was inaccurately identified by the program, and the depth of the nasolabial fold could not be easily determined using a two-dimensional representation [18]. We addressed this issue by enabling users to manually mark a loca-

tion that the system then used to calculate the results. This led to the next phase of software development, in which we applied a machine learning algorithm to the image processing steps to improve feature recognition and analysis.

The reproducibility of the application was tested, and all parameters showed significant consistency on two separate tests, demonstrating the excellent accuracy of the SBface application. The interrater reliability test between the application and otolaryngologist showed a high level of agreement in the total and composite scores as well as between the three raters. However, the software-derived scores regarding the subdomain items did not correlate well with the scores from the clinicians, and 10.3% of the subdomain items failed to show correlations among the three clinicians themselves. This illustrates that the human group, despite similar total scores, was inconsistent in determining the levels of disfigurement. The accuracy of the subdomain scores is important since they measure the changes in static, dynamic, and synkinesis conditions along the treatment course. We hypothesized that, among the three specialists, the otolaryngologist was more familiar with the entire spectrum of facial paralysis symptoms, while plastic surgeons usually see more severe cases and are less frequently exposed to minor cases. Since the software equally divides its scores according to severity, ratings from clinicians who understand the whole spectrum of disfigurements are expected to correlate more closely with the ratings made by the software.

Another research group proposed automating an already-existing grading system using still photographs and no markers. Their results showed a high correlation between the automated Sunnybrook facial grading scale and ratings by clinicians, while there was low correlation between the House-Brackmann and Stennert grading systems [19].

We believe that artificial intelligence (AI) technology could benefit our application, as the integration of AI into the system would provide much more in-depth analysis and result in even higher accuracy. For example, AI could show the degree of improvement or worsening of paralysis and predict the prognosis after multiple follow-up visits. The development of this application is merely a preliminary study for its use in patients with unilateral facial paralysis. Its promising design shows an improvement over the conventional Sunnybrook facial grading scale in terms of both functionality and accuracy for inexperienced users.

NOTES

Conflict of interest

No potential conflict of interest relevant to this article was reported.

Ethical approval

The study was approved by the Ethics Committee of Lerdsin Hospital (EC No. 611038/2561) and performed in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained.

Patient consent

The patient provided written informed consent for the publication and the use of her images.

Author contribution

Conceptualization: S Jirawatnotai, W Tirakotai. Data curation: S Jirawatnotai, P Jomkoh, N Somboonsap. Formal analysis: P Jomkoh. Methodology: S Jirawatnotai, N Somboonsap. Project administration: S Jirawatnotai, W Tirakotai. Visualization: N Somboonsap. Writing - original draft: P Jomkoh. Writing - review & editing: S Jirawatnotai, TY Voravitvet, W Tirakotai.

ORCID

Supasid Jirawatnotai https://orcid.org/0000-0002-3831-473X
Pojanan Jomkoh https://orcid.org/0000-0002-9033-3591
Tsz Yin Voravitvet https://orcid.org/0000-0002-9755-2594
Wuttipong Tirakotai https://orcid.org/0000-0003-2553-6138
Natthawut Somboonsap https://orcid.org/0000-0003-3786-0111

REFERENCES

1. House JW. Facial nerve grading systems. Laryngoscope 1983;93:1056-69.
2. House JW, Brackmann DE. Facial nerve grading system. Otolaryngol Head Neck Surg 1985;93:146-7.
3. Yanagihara N. Grading of facial palsy. In: Fisch U, editor. Facial nerve surgery. Proceedings of the Third International Symposium on Facial Nerve Surgery, Zurich, 1976. Amstelveen: Kugler Medical Publications; 1977. p. 533-5.
4. Ross BG, Fradet G, Nedzelski JM. Development of a sensitive clinical facial grading system. Otolaryngol Head Neck Surg 1996;114:380-6.
5. Banks CA, Bhama PK, Park J, et al. Clinician-graded electronic facial paralysis assessment: the eFACE. Plast Reconstr Surg 2015;136:223e-230e.
6. Berg T, Jonsson L, Engstrom M. Agreement between the Sunnybrook, House-Brackmann, and Yanagihara facial nerve grading systems in Bell's palsy. Otol Neurotol 2004;25:1020-6.
7. Fattah AY, Gurusinghe ADR, Gavilan J, et al. Facial nerve grading instruments: systematic review of the literature and suggestion for uniformity. Plast Reconstr Surg 2015;135:569-79.

8. Neely JG, Cherian NG, Dickerson CB, et al. Sunnybrook facial grading system: reliability and criteria for grading. Laryngoscope 2010;120:1038-45.
9. Gaudin RA, Robinson M, Banks CA, et al. Emerging vs time-tested methods of facial grading among patients with facial paralysis. JAMA Facial Plast Surg 2016;18:251-7.
10. Chee GH, Nedzelski JM. Facial nerve grading systems. Facial Plast Surg 2000;16:315-24.
11. Burres S, Fisch U. The comparison of facial grading systems. Arch Otolaryngol Head Neck Surg 1986;112:755-8.
12. Lee LN, Susarla SM, H Hohman M, et al. A comparison of facial nerve grading systems. Ann Plast Surg 2013;70:313-6.
13. Kanerva M, Poussa T, Pitkaranta A. Sunnybrook and House-Brackmann facial grading systems: intrarater repeatability and interrater agreement. Otolaryngol Head Neck Surg 2006;135:865-71.
14. Chong LSH, Eviston TJ, Low TH, et al. Validation of the clinician-graded electronic facial paralysis assessment. Plast Reconstr Surg 2017;140:159-67.
15. Banks CA, Jowett N, Azizzadeh B, et al. Worldwide testing of the eFACE facial nerve clinician-graded scale. Plast Reconstr Surg 2017;139:491e-498e.
16. O'Reilly BF, Soraghan JJ, McGrenary S, et al. Objective method of assessing and presenting the House-Brackmann and regional grades of facial palsy by production of a facogram. Otol Neurotol 2010;31:486-91.
17. Meier-Gallati V, Scriba H. Objective assessment of the reliability of the House-Brackmann and Fisch grading of synkinesis. Eur Arch Otorhinolaryngol 2017;274:4217-23.
18. Baude M, Hutin E, Gracies JM. A bidimensional system of facial movement analysis conception and reliability in adults. Biomed Res Int 2015;2015:812961.
19. Mothes O, Modersohn L, Volk GF, et al. Automated objective and marker-free facial grading using photographs of patients with facial palsy. Eur Arch Otorhinolaryngol 2019;276:3335-43.