ADITUM
Enrich your Research

Clinical Case Reports and Clinical Study

Research Article

Automated eFACE Grading Of Facial Function After Facial Nerve Decompression: Agreement With Clinician eFACE And The Sunnybrook System

Ghouzlanı Imane, A. Chraibi, M. Lahjaouj, M. Loudghırı, W. Bıjou, Y. Oukessou, S. Rouadı, R.L. Abada, M. Roubal, M. Mahtar

Service D'ORL Et Chirurgie Cervico-Faciale, Hopital 20 Aout Casablanca.

Article Information

Received: July 15, 2026
Accepted: July 27, 2026
Published: August 13, 2026

***Corresponding author:** Ghouzlanı Imane, Service D'ORL Et Chirurgie Cervico-Faciale, Hopital 20 Aout Casablanca.

Citation: Imane G, A. Chraibi, M. Lahjaouj, M. Loudghırı, W. Bıjou, Y. Oukessou, S. Rouadı, R.L. Abada, M. Roubal, M. Mahtar. (2026) "Automated eFACE Grading Of Facial Function After Facial Nerve Decompression: Agreement With Clinician eFACE And The Sunnybrook System" Clinical Case Reports and Clinical Study, 13(4); DOI: 10.61148/2766-8614/JCCRCS/246.

Abstract

Objective: To evaluate the agreement between an automated, smartphone-application–based eFACE, clinician-graded (manual) eFACE, and the Sunnybrook Facial Grading System (Sunnybrook) for grading facial motor function in patients who had undergone facial nerve decompression, and to determine whether an automated tool can reliably stand in for expert clinician assessment.

Methods: Sixteen consecutive patients operated on for peripheral facial palsy were assessed. At the post-operative visit, facial function was graded with the House-Brackmann (HB) scale, the Sunnybrook, a clinician-graded eFACE, and an automated eFACE generated by the mobile application from a standardized photo/video set. Agreement was analyzed with Pearson and Spearman correlations, Bland-Altman limits of agreement, and the intraclass correlation coefficient (ICC).

Results: Complete grading with all three instruments was feasible in 13 of 16 patients (81%; the remaining three had bilateral or non-gradable palsy). Scores correlated strongly across instruments (automated vs. clinician eFACE $r = 0.89$; clinician eFACE vs. Sunnybrook $r = 0.93$; automated eFACE vs. Sunnybrook $r = 0.90$; all $p < 0.001$). Bland-Altman analysis showed that the automated eFACE systematically scored 8.6 points lower than the clinician eFACE (95% limits of agreement −16.7 to −0.6), while the ICC between the two was moderate (0.60).

Conclusion: An automated smartphone eFACE tracks clinician grading and the Sunnybrook closely but systematically underestimates facial function by roughly nine points. It is a promising, accessible screening and monitoring tool, yet clinician grading should remain the reference until the systematic bias is corrected.

Keywords: facial nerve decompression; facial palsy; eFACE; Sunnybrook Facial Grading System; House-Brackmann; automated facial grading; smartphone application

Introduction

Few things are read as instinctively as a human face, and few losses are felt as acutely as the loss of its movement. Severe peripheral facial palsy reshapes how a person eats, speaks, closes an eye, and is seen by others, and it carries a measurable psychological burden. Facial nerve decompression remains a valuable option in this setting—most clearly after temporal bone trauma, when electrophysiological testing points to a poor chance of spontaneous recovery[1,2]—and modern middle-fossa and transmastoid techniques have made the intratemporal course of the nerve reliably accessible.[3] The operation itself has a long lineage, from the first descriptions of complete decompression[4] to the systematic rehabilitation of the traumatically injured nerve.[5]

Yet an operation is only as meaningful as our ability to measure what it achieves, and here otology has never quite settled on a single language. The House-Brackmann scale is still the most widely spoken[6,7], prized for its simplicity, but its six coarse grades flatten the very differences—resting tone, voluntary excursion, synkinesis—that matter most to a recovering patient. The Sunnybrook Facial Grading System was designed to be more sensitive, weighting these components into a single composite score[8,9], and the clinician-graded electronic facial function scale (eFACE) went further still, capturing static, dynamic, and synkinetic domains on continuous 0–100 scales.[10,11] Repeated calls for uniformity[12] and dedicated synkinesis metrics[13] reflect how unsettled the field remains.

Into this landscape has come the smartphone. Automated, application-based grading promises to put a reproducible eFACE in the hands of any clinician—or patient—with a camera, but its trustworthiness against expert assessment is not yet established. Recent work has shown that even the manual eFACE and Sunnybrook behave differently in the hands of trained versus untrained raters[14], which makes it all the more important to ask a simpler, practical question before an app enters routine use: when an algorithm grades the face, does it agree with the surgeon? We addressed this question in a cohort of patients followed after facial nerve decompression, comparing an automated eFACE against clinician-graded eFACE and the Sunnybrook.

Materials And Methods

Study design and patients

We conducted a single-center, cross-sectional study at a tertiary referral center. Sixteen consecutive patients who underwent facial nerve decompression for peripheral facial palsy and attended structured post-operative facial-function assessment were included. Patients with incomplete follow-up or without a recorded post-operative grading set were not analyzed. The study followed the principles of the Declaration of Helsinki, and institutional approval and patient consent were obtained.

Grading instruments

At the post-operative visit, each patient was graded with four instruments. The House-Brackmann scale provided a global six-grade severity anchor[6,7]. The Sunnybrook composite score (0–100) was computed from resting symmetry, voluntary movement, and synkinesis[8,9]. A clinician-graded (manual) eFACE was completed by an experienced examiner across its static, dynamic, and synkinetic subdomains, yielding a total score of 0–100.[10] Finally, an automated eFACE total was generated by the mobile application from a standardized photographic and video set acquired at the same visit, under the same lighting and instructions used for the clinician assessment—an approach consistent with reported in-person versus video equivalence work for the eFACE.[15]

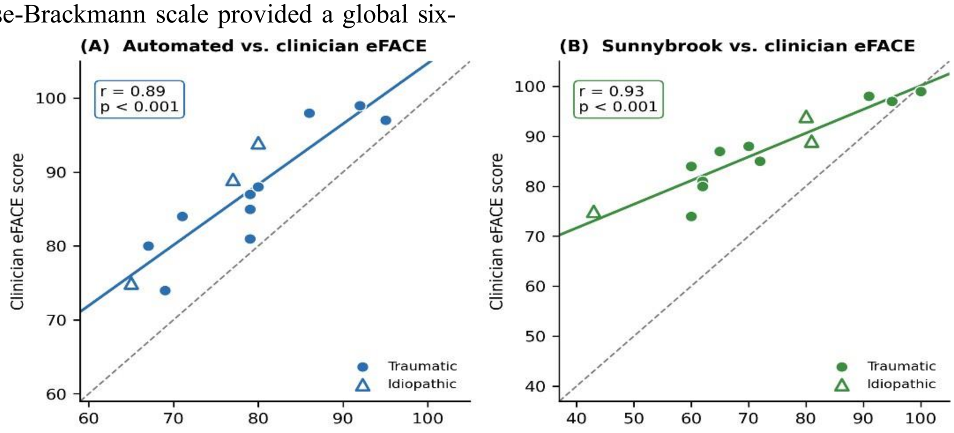
Statistical analysis

Continuous scores are reported as mean ± standard deviation with ranges. Relationships between instruments were quantified with Pearson and Spearman correlation coefficients. Agreement was assessed with Bland-Altman analysis—reporting the mean bias and 95% limits of agreement (bias ± 1.96 SD)—and with the two-way random-effects, absolute-agreement intraclass correlation coefficient (ICC).[22] A two-sided $p < 0.05$ was considered significant. Analyses were performed in Python (SciPy).

Results

The cohort comprised 16 operated patients—13 with a traumatic and 3 with an idiopathic etiology—with palsy of the right side in 8 and the left in 8, and pre-operative House-Brackmann grades ranging from I to IV. Post-operative delay at assessment ranged from the immediate post-operative period to 66 months. Complete grading with all three quantitative instruments (Sunnybrook, clinician eFACE, automated eFACE) was feasible in 13 of the 16 patients (81%); the remaining three could not be graded with symmetry-based instruments because of bilateral or profoundly severe palsy and were therefore excluded from the agreement analysis. Patient characteristics and scores are summarized in Table I.

Across the 13 gradable patients, mean scores were 72.4 ± 16.3 for the Sunnybrook, 78.4 ± 9.0 for the automated eFACE, and 87.0 ± 8.3 for the clinician eFACE. The three instruments correlated strongly with one another (Fig. 1): the automated eFACE tracked the clinician eFACE with a Pearson r of 0.89 (Spearman ρ = 0.89), the clinician eFACE and the Sunnybrook agreed most closely (r = 0.93; ρ = 0.94), and the automated eFACE correlated with the Sunnybrook at r = 0.90 (all $p < 0.001$). Correlation was preserved across both traumatic and idiopathic patients.



Figuee 1. Correlation between grading instruments in the 13 gradable patients. (A) Automated versus clinician-graded eFACE. (B) Sunnybrook versus clinician-graded eFACE. Solid lines are least-squares regression fits; dashed grey lines are lines of identity. Circles denote traumatic and open triangles idiopathic etiology. Pearson r is annotated; all $p < 0.001$.

Correlation, however, is not agreement, and the Bland-Altman analysis told the more clinically important story (Fig. 2). The automated eFACE scored, on average, 8.6 points lower than the clinician eFACE (95% limits of agreement −16.7 to −0.6), a consistent, one-directional underestimation across the range of severity; the corresponding ICC was moderate (0.60). The clinician eFACE in turn ran above the Sunnybrook by 14.6 points (95% limits of agreement −3.2 to 32.5), reflecting the known scaling difference between the two instruments rather than disagreement about rank order. Representative automated eFACE outputs for two of our patients are shown in Figure 3.

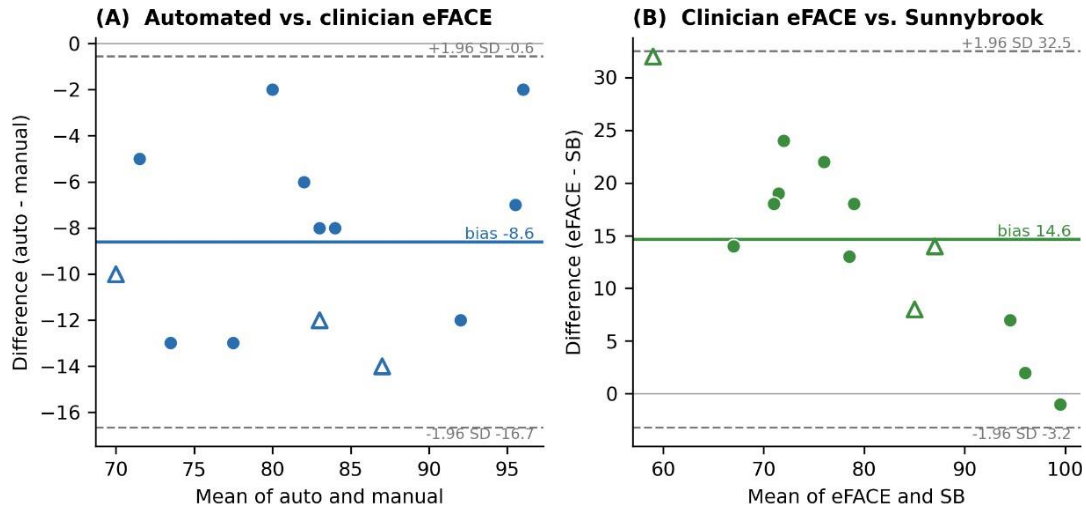


Figure 2. Bland-Altman agreement plots. (A) Automated versus clinician eFACE: the automated score is systematically lower (mean bias −8.6). (B) Clinician eFACE versus Sunnybrook. Solid coloured lines mark the mean bias; dashed grey lines mark the 95% limits of agreement (bias ± 1.96 SD). Symbols as in Figure 1.

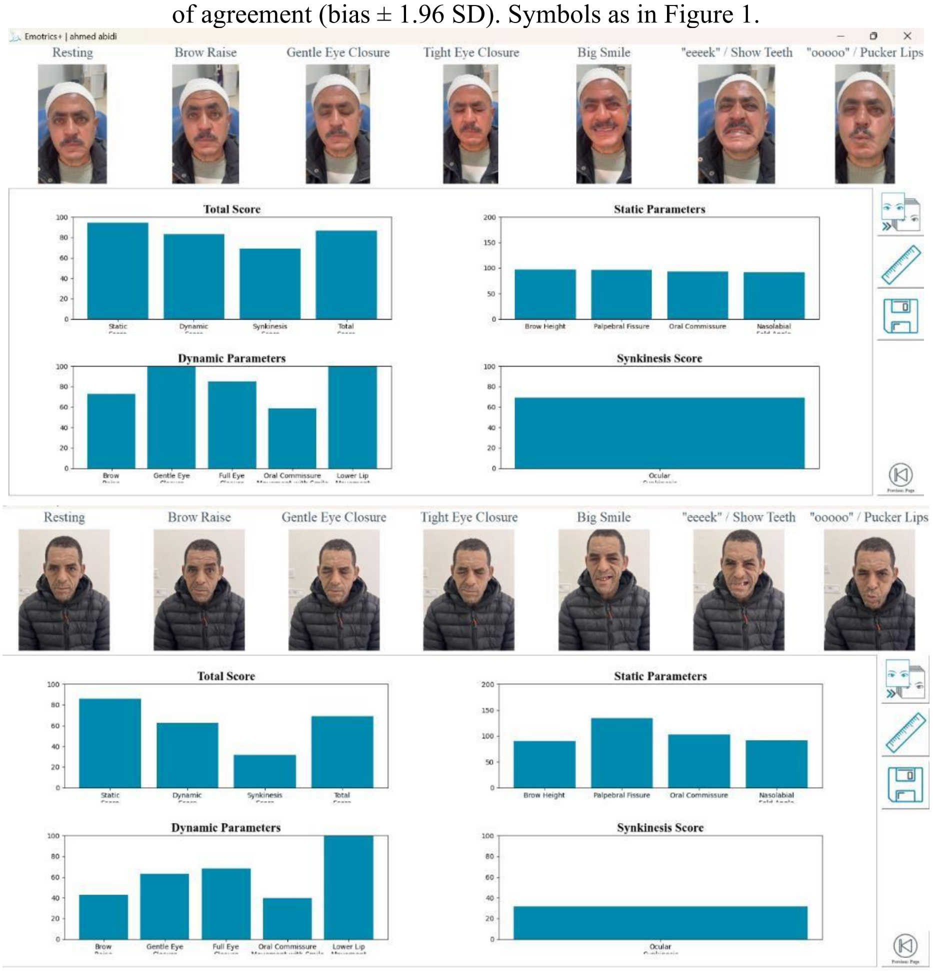


Figure 3. Representative automated eFACE application outputs for two study patients.

Table I. Facial-function scores of the 13 patients graded with all three instruments.

Patient	Etiology	HB	Sunnybrook	Auto eFACE	Clin. eFACE	Side / delay
P1	Traumatic	II	62	79	81	Right / 66 mo
P2	Traumatic	II	62	67	80	Right / 55 mo
P3	Traumatic	III	65	79	87	Left / 46 mo
P4	Traumatic	III	60	69	74	Right / 31 mo
P5	Traumatic	III	72	79	85	Left / 32 mo
P6	Traumatic	I	95	95	97	Left / 36 mo
P7	Traumatic	II	91	86	98	Right / 16 mo
P8	Idiopathic	III	81	77	89	Right / 9 mo
P9	Traumatic	II	70	80	88	Left / 8 mo
P10	Idiopathic	II	80	80	94	Left / 6 mo
P11	Traumatic	I	100	92	99	Right / 9 mo
P12	Idiopathic	III	43	65	75	Right / post-op
P13	Traumatic	III	60	71	84	Left / post-op
Mean ± SD	—	—	**72.4 ± 16.3**	**78.4 ± 9.0**	**87.0 ± 8.3**	—

HB = House-Brackmann grade; mo = months; post-op = immediate post-operative assessment. Delays are approximate.

Discussion

Our central finding is reassuring and cautionary in equal measure: an automated smartphone eFACE agrees with expert grading in rank and trend but not in absolute value. The strong correlations we observed—up to r = 0.93 between clinician eFACE and Sunnybrook—sit comfortably within the range reported when these instruments are compared head to head[14], and they confirm that all three scales are measuring the same underlying construct in operated patients. The clinically decisive result, however, emerged only from the Bland-Altman analysis: the automated eFACE underscored the clinician eFACE by a consistent 8.6 points. A high correlation with a fixed offset is the signature of a tool that ranks patients correctly while miscalibrating the scale—precisely the pattern one wants to detect before an app is used to make longitudinal decisions.

This distinction matters because the different instruments were built with different sensitivities. Novice and expert raters are known to converge on the Sunnybrook only after dozens of gradings[16], and cross-cultural validations have repeatedly confirmed its robustness once learned[17]. Comparative studies of the House-Brackmann, Facial Nerve Grading System 2.0, and Sunnybrook have similarly shown high agreement in rank with meaningful differences in granularity[18,19], and even the House-Brackmann–Sunnybrook relationship, though strong overall, weakens at baseline severity[20]. Our results extend this literature to an automated tool and to a decompression cohort specifically, where accurate serial measurement is essential to judge surgical benefit and to counsel patients whose quality of life tracks closely with their grade[21].

The practical implication is nuanced. An automated eFACE that reliably underestimates by a known margin is not useless—far from it. Once its systematic bias is characterized and corrected, it could democratize follow-up, allowing a patient to capture a standardized recording at home and a clinician to trend recovery without a clinic visit. What it should not yet do is replace an expert grade in a treatment decision, nor be compared naively against a manual eFACE recorded on a different day. The moderate ICC of 0.60 between automated and clinician eFACE underscores this: adequate for screening and monitoring, insufficient as a stand-alone reference.

Our study has clear limitations. The sample is small and single-center, three of sixteen patients could not be graded with symmetry-based instruments because of bilateral involvement, and assessments were cross-sectional rather than longitudinal, so we could not yet test whether the automated bias is stable over time within a patient. Etiology and post-operative delay were heterogeneous, and a single clinician performed the manual grading, leaving inter-rater reliability of the manual eFACE unexamined in this cohort. Larger, longitudinal, multi-rater studies—ideally incorporating intra-operative facial nerve monitoring data[24] and standardized endoscopic or middle-fossa operative documentation[23]—will be needed to calibrate the automated tool and confirm its place in routine practice.

Conclusion

In patients followed after facial nerve decompression, an automated smartphone eFACE correlated strongly with clinician-graded eFACE and the Sunnybrook system but systematically underestimated facial function by roughly nine points. The tool is a promising, accessible way to screen and monitor recovery, yet clinician grading should remain the reference standard until this reproducible bias is corrected. Automated facial grading is close to clinical usefulness—provided we read its numbers with the offset in mind.

References

1. Darrouzet V, Duclos JY, Liguoro D, Truilhe Y, De Bonfils C, Bébéar JP. Management of facial paralysis resulting from temporal bone fractures: our experience in 115 cases. Otolaryngol Head Neck Surg. 2001;125(1):77-84.

2. Nash JJ, Friedland DR, Boorsma KJ, Rhee JS. Management and outcomes of facial paralysis from intratemporal blunt trauma: a systematic review. Laryngoscope. 2010;120(7):1397-1404.
3. Ebode D, Tuset MP, Michel J, Radulesco T, et al. How I do it: middle fossa approach of facial nerve decompression. Laryngoscope. 2026;136(4):1896-1899.
4. Pulec JL. Total decompression of the facial nerve. Laryngoscope. 1966;76(6):1015-1028.
5. Glasscock ME 3rd, Wiet RJ, Jackson CG, Dickins JRE. Rehabilitation of the face following traumatic injury to the facial nerve. Laryngoscope. 1979;89(9):1389-1404.
6. House JW. Facial nerve grading systems. Laryngoscope. 1983;93(8):1056-1069.
7. House JW, Brackmann DE. Facial nerve grading system. Otolaryngol Head Neck Surg. 1985;93(2):146-147.
8. Ross BG, Fradet G, Nedzelski JM. Development of a sensitive clinical facial grading system. Otolaryngol Head Neck Surg. 1996;114(3):380-386.
9. Neely JG, Cherian NG, Dickerson CB, Nedzelski JM. Sunnybrook facial grading system: reliability and criteria for grading. Laryngoscope. 2010;120(5):1038-1045.
10. Banks CA, Bhama PK, Park J, Hadlock CR, Hadlock TA. Clinician-graded electronic facial paralysis assessment: the eFACE. Plast Reconstr Surg. 2015;136(2):223e-230e.
11. Banks CA, Jowett N, Azizzadeh B, et al. Worldwide testing of the eFACE facial nerve clinician-graded scale. Plast Reconstr Surg. 2017;139(2):491e-498e.
12. Fattah AY, Gurusinghe AD, Gavilan J, et al. Facial nerve grading instruments: systematic review of the literature and suggestion for uniformity. Plast Reconstr Surg. 2015;135(2):569-579.
13. Berner JE, Kamalathevan P, Kyriazidis I, Nduka C. Facial synkinesis outcome measures: a systematic review of the available grading systems and a Delphi study. J Plast Reconstr Aesthet Surg. 2019;72(6):946-963.
14. Bruins TE, Lammens RF, van Veen MM, Tamási K, Dijkstra PU, Werker PMN, Broekstra DC. Assessing facial palsy: does feedback improve assessment using the eFACE and Sunnybrook Facial Grading System? Laryngoscope. 2024;134(7):3105-3111.
15. Banks CA, Jowett N, Hadlock TA. Test-retest reliability and agreement between in-person and video assessment of facial mimetic function using the eFACE facial grading system. JAMA Facial Plast Surg. 2017;19(3):206-211.
16. Hu WL, Ross B, Nedzelski J. Reliability of the Sunnybrook facial grading system by novice users. J Otolaryngol. 2001;30(4):208-211.
17. Pavese C, Tinelli C, Furini F, et al. Validation of the Italian version of the Sunnybrook Facial Grading System. Neurol Sci. 2013;34(4):457-463.
18. Mengi E, Orhan Kara C, Necdet Ardıç F, et al. Comparison of the reliability of the House-Brackmann, Facial Nerve Grading System 2.0, and Sunnybrook Facial Grading System. J Int Adv Otol. 2024;20(1):14-18.
19. Mat Lazim N, Ismail H, Abdul Halim S, Nik Othman NA, Haron A. Comparison of 3 grading systems (House-Brackmann, Sunnybrook, Sydney) for the assessment of facial nerve paralysis and prediction of neural recovery. Medeni Med J. 2023;38(2):111-119.
20. Williams A, Eapen N, Kochar A, et al. Agreement between House-Brackmann and Sunnybrook facial nerve grading systems in Bell's palsy in children: secondary analysis of a randomized, placebo-controlled multicenter trial. J Child Neurol. 2023;38(1-2):89-97.
21. Bylund N, Hultcrantz M, Jonsson L, Marsk E. Quality of life in Bell's palsy: correlation with Sunnybrook and House-Brackmann over time. Laryngoscope. 2021;131(2):E612-E618.
22. Killip S, Mahfoud Z, Pearce K. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. Ann Fam Med. 2004;2(3):204-208.
23. Marchioni D, Alicandri-Ciufelli M, Piccinini A, et al. Surgical anatomy of transcanal endoscopic approach to the tympanic facial nerve. Laryngoscope. 2011;121(7):1565-1573.
24. Kartush JM, Rice KS, Minahan RE, et al. Best practices in facial nerve monitoring. Laryngoscope. 2021;131(Suppl 4):S1-S42.