# The Feasibility of An Automatical Facial Evaluation System Providing Objective and Reliable Results for Facial Palsy

Yongli Zhang, *Member, IEEE*, Li Ding, Zhimin Xu, Huanghong Zha, Xinyi Tang, Chong Li, Shuo Xu, Zhijie Yan, and Jie Jia, *Member, IEEE*

***Abstract*—Facial palsy would lead to a series of physical and mental problems, as facial function plays an important role in various aspects of daily life. However, the current strategies for evaluating facial function relied heavily on raters and the results varied from the experience of raters. Thus, an objective and accurate facial evaluation system is always claimed. In this study, a customized automatical facial evaluation system (AFES) was proposed, which might have the potential to be employed as an adjunctive and efficient assessing method in clinic. In order to investigate the feasibility of AFES, ninety-two participants with facial palsy were recruited and received scale-based subjective manual evaluation (including mHBGS and mSFGS) and objective automatical evaluation of AFES (including aHBGS, aSFGS and indicators of facial regional features) at enrollment and after two weeks. The correlations between the results of the two methods were analyzed and the participants were stratified according to the severity of facial function for further analyses. Strong positive correlations between manual and automatical HBGS and SFGS were observed and higher correlations were reported in the participants with normal-mild and moderate facial palsy. Significant improvements in clinical scales and indicator of eye synkinesis were found in forty-two participants in two weeks. Furthermore, some of the indicators were correlated with scale scores (I4, I7) and one of them presented a significant change between the baseline evaluation and follow-up evaluation (I7). According to the results, AFES could be considered as a viable method to perform objective and reliable evaluation for patients with facial palsy and provide clarified results for prognosis.**

***Index Terms*—Automatical evaluation system, facial palsy, facial feature, HBGS, SFGS.**

Manuscript received 2 June 2022; revised 22 January 2023; accepted 6 February 2023. Date of current version 20 March 2023. This work was supported in part by the National Key Research and Development Program Project of China under Grant 2018YFC2002300, in part by the National Natural Science Foundation of China under Grant 82002385, in part by the National Natural Innovation Research Group Project of China under Grant 82021002, in part by the National Natural Integration Project under Grant 91948302, in part by the Key Discipline Construction Fund of Shanghai Jing'an District Health System under Grant 2021ZB02, and in part by the Shanghai Sailing Program under Grant 20YF1403400. *(Yongli Zhang and Li Ding contributed equally to this work.) (Corresponding author: Jie Jia.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Huashan Hospital.

Yongli Zhang is with the Department of Rehabilitation Medicine, Huashan Hospital, Fudan University, Shanghai 200040, China, and also with the Department of Rehabilitation Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou 350122, China (e-mail: islilyong@163.com).

Li Ding, Xinyi Tang, Chong Li, Shuo Xu, and Zhijie Yan are with the Department of Rehabilitation Medicine, Huashan Hospital, Fudan University, Shanghai 200040, China.

Zhimin Xu is with Sharpsight Ltd., Hong Kong, China.

Huanghong Zha is with the Department of Rehabilitation Medicine, Yuebei People's Hospital, Shaoguan, Guangdong 512027, China.

Jie Jia is with the National Clinical Research Center for Aging and Medicine, Department of Rehabilitation Medicine, Huashan Hospital, Fudan University, Shanghai 200040, China, and also with National Center for Neurological Disorders, Shanghai 200040, China (e-mail: shannonjj@126.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSRE.2023.3244563, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2023.3244563

## I. Introduction

HUMAN face plays an important role in feelings, emotional expressions, and social interactions [1], [2]. When facial palsy (FP) occurs, individuals might go through discomforts continuously, such as dry eyes and taste disorders [3]. As one of the most severe problems, sudden losing facial motor function after FP would make daily life inconvenient for functional and aesthetic deficits [4]. Individuals who suffer from FP might have difficulty to make coordinated expressions, eating disorder and increasing risk of ocular inflammation [5]. Moreover, those people with severe FP or in acute phase present facial asymmetry, which might hinder their abilities of communication and social cognition, even lead to psychosomatic problems [6]. It is reported that patients with FP often experienced negative emotions such as anxiety, depression and low self-esteem and their quality of life would also be greatly affected [7]. Therefore, individuals pay a lot of attention on the states or alterations of facial function during recovery stage.

Clinical scales are usually employed to evaluate the recovery of FP. The House-Brackmann grading system (HBGS) [8], the Sunnybrook Facial Grading System (SFGS) [9] and the Sydney facial grading system [10] are widely used. These scales grade the severity of facial motor dysfunction based on the static or dynamic facial states. However, the results of these current scales are mainly relied on the raters and vary from the experience and abilities of the raters [11]. Besides, the subjective selection of these scales might also bring difficulties in comparing the recovery process among individuals [12]. Neuroelectrophysiological examination, such as electromyography, electroneuronography [13], [14], the infrared thermal [15], [16] and laser speckle contrast imaging [17] are more objective ways to evaluate the facial function via physiological indicators, of which the relevant analyses might assist in assessing severity and predicting recovery. However, required equipment, strict implementation, and high costs limit the regular application of these evaluation tools in clinic or home use [18], [19]. Thus, an objective and convenient approach is necessary.

The automatical evaluation systems based on computers, websites or mobile phones developed recent years have the high potential to be useful tools in hospital and for remote evaluation [20]. Specific facial features in the static and the dynamic conditions are captured to classify the severity of FP in these systems [5], [21], [22], [23]. Guarin et al. [24] developed a software platform for automated facial measurements, called Emotrics. It was based on a database of normal faces and established to provide automatical facial displacements computation of FP faces according to frontal-view photographs. Lee et al. [25] demonstrated a computer-based facial asymmetry assessment program, which could score the asymmetry of palsied face or even a certain part. Hsu et al. [26] proposed a deep-learning algorithm to evaluate FP patients quantitatively, named the deep hierarchical network (DHN). Based on the line segmentation strategy, the DHN increased the accuracy of facial landmark localization which played an important role in improving the overall performance of the algorithm. A trained automatical evaluation system was provided by Mothes et al. [27], which output the scores of the HBGS, the SFGS and the Stennert index. However, photographs were used as input materials in these above studies, which might limit the accuracy of the general evaluation for neglecting the impact of dynamic status in facial symmetry, like synkinesis. Therefore, some studies [28], [29], [30], [31] suggested video-based evaluation system. Monini et al. [32]showed a novel digital video-analysis system, which recorded videos through smartphones and output automatical HBGS scores. Zhao et al. [31] presented a videography-based system, which combined the SFGS and three-dimensional dynamic quantitative analysis system for prognostic use. However, the bulky size of the system limited the clinical applications. Although these evaluation systems provided more objective and accurate results, most of them remained in the preclinical stage [33] which lacked clinical verification and they were unable to provide finer grading scores or demonstrate quantified scale-based facial feature alterations. Another main problem was many studies constructed new evaluation indicators [34], [35], [36], which were not easily to combine with clinical practice. Moreover, few studies [5], [18] have presented tentative evidence for the follow-up evaluation based on regional facial features to indicate the specific therapeutic alterations via video evaluation systems.

Therefore, we propose a novel customized apparatus, automatical facial evaluation system (AFES), to provide objective outcome measurements and establish a comparable assessment method. In this study, we aimed to explore the clinical feasibility of AFES in different severity individuals with FP.

TABLE I
DEMOGRAPHICS OF PARTICIPANTS.
Y, YEAR. SD, STANDARD DEVIATION

| **Demographics** | N | % |
|---|---|---|
| **Gender** | | |
| Male | 59 | 35.87 |
| Female | 33 | 64.13 |
| **Affected side** | | |
| Right | 46 | 50.00 |
| Left | 46 | 50.00 |
| **Etiology** | | |
| Idiopathic | 57 | 61.96 |
| Neoplastic | 6 | 6.52 |
| Infectious | 17 | 18.48 |
| Traumatic | 4 | 4.35 |
| Central | 6 | 6.52 |
| Iatrogenic | 2 | 2.17 |
| **Age (y)**/Mean ± SD | 43.82±16.14 | |
| **BMI**/Mean ± SD | 23.02±3.81 | |
| **Duration(m)**/Mean ± SD | 13.26±16.31 | |

## II. METHODOLOGY

### A. Participants

This study was conducted in Huashan Hospital, Fudan University from November 2021 to March 2022. Ninety-two participants with FP were recruited (Table I). The inclusion criteria were: (1) aged between 18-80 years; (2) unilateral facial palsy; (3) volunteer to participate. The exclusion criteria were: (1) obvious wounds on faces; (2) facial movements are affected by other reasons except facial palsy (eg. pain or swelling); (3) without enough cognitive ability to complete the study. The study was approved by the Ethics Committee of Huashan Hospital (Ethic number: KY2021-814) and performed according to the Declaration of Helsinki. All the participants agreed to take part in the study and were given a written informed consent.

### B. Facial Motor Function Evaluation

The evaluation included scale-based subjective evaluation and the AFES-based evaluation. It was taken out on the first day after the participants enrollment (baseline evaluation) and two weeks later (follow-up evaluation).These two methods were performed in random order and with a fifteen-minute break for participants in between.

*1) Scale-Based Subjective Evaluation:* In the study, the scales of HBGS and the SFGS were employed. The HBGS was a widely used and accepted system, which was a gross scale with six grades [37]. Grade I represented the normal status and the grade IV represented the most severe facial palsy [38]. The SFGS, with a total score of 100, combined static and dynamic evaluation, and synkinesis of the paralyzed face [39]. Higher scores stand for better facial motor function. As a


Fig. 1. **AFES Evaluation flowchart.** During AFES evaluation process, patients watching the evaluation guiding page (A), the calibration stage (B), the instructional videos (C) and the results presented page (D). Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images included in this article.

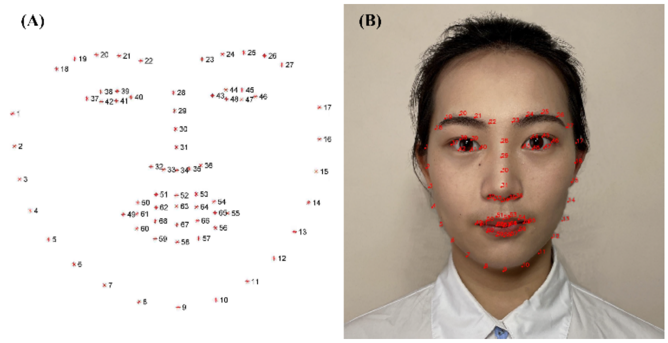


Fig. 2. **The 68 annotated landmarks.** Illustration of the position of 68 landmarks (A) and the 68 facial landmarks (red dots) automatically placed by AFES (B). Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images included in this article.

sensitive, reliable, and validated grading system, the SFGS was recommended to assess facial function [40].

Two experienced therapists were selected as raters and they all received training from a senior professor. The evaluation was conducted in a separate room under normal fluorescent lighting conditions. The raters would demonstrate the evaluation actions for the participants. Then the participants were asked to perform the corresponding facial movements after they heard the instructions and maintain it for a few seconds. To ensure the accuracy of the evaluation, each action would be repeated two or three times. At the same time, the two raters assessed the subject's facial function with the HBGS and SFGS. For the manual grades of HBGS (mHBGS), if the grading results were different, the two raters would re-evaluate and discuss with each other to reach a consensus. As for the manual scores of SFGS (mSFGS), the mean scores were calculated as the final results.

*2) Automatical Evaluation of AFES:* The AFES was employed in this study to perform objective evaluation, which was a customized and marker-free evaluation system. It consisted of a camera, a screen, a computer, and two LED lamps, and was mounted on a movable supporting holder.

The AFES included three steps for standard evaluation process, evaluation guide, calibration and evaluation. At the beginning, the evaluation guiding page would be presented to the patients with some specific requirements, such as minimizing body movements and removing items that might block the face (Fig.1A). In the calibration stage (Fig.1B), the participants were required to keep his/her face tracked in a proper area. As for the evaluation stage, participants were instructed to perform facial motor tasks following instructional videos and voice prompts (Fig.1C). They were asked to relax their faces and remain resting for 15 seconds. Then five predetermined facial motor tasks were presented, including raising eyebrows, closing eyes gently, wrinkling the nose, pouting, grinning, and each task was repeated three times and maintained for five seconds. After completing a series of facial movements, the AFES would display the results on the screen within a few seconds (Fig.1D).

During the automatical evaluation, the performance of participants was recorded by the AFES. The processing of the video data includes rough video interception, video stabilization, key frame extraction, image geometric normalization and grayscale normalization. The conduction of key frame extraction was one of the important steps, which aimed to select the images of no facial expression state and maximum facial expression state. Sixty-eight points [41] which represented the facial landmarks (Fig.3) were marked automatically by using Ensemble of Regression Trees algorithm and were used to split and select the static and dynamic features. Thirty-two static features that described the symmetry of the human face were extracted. Subsequently, the region of interest of the images were divided according to these points and six key regions (both sides of eyebrows, eyes and mouth) were extracted based on the Horn-Schunck optical flow method. Then the dynamic characteristics of the optical flow difference between the left and right corresponding key regions were calculated according to the symmetry of the face. Finally, the data of the dynamic and static feature classifications were weighted and comprehensively analyzed to gain ratings of the participants. The algorithm of AFES has been trained and tested with video frames from more than 100 patients and the 80%, 20% of the video frames were used as training sets and testing sets, respectively. In addition, the five-fold cross-validation method [26] was used to optimize the Support Vector Machine kernel function and parameters on the training set: divided the data into 5 randomly, and four of them were taken as the training set for obtaining the classifier, and then used the learned classifier to test the remaining data.

The AFES presented the results of auto-scored HBGS (aHBGS), auto-scored SFGS (aSFGS), and seven regional facial features (see the supplementary for details). Five out of seven facial features (I1, I2, I3, I4, I6) were mainly used to calculate the differences of facial regions between the healthy

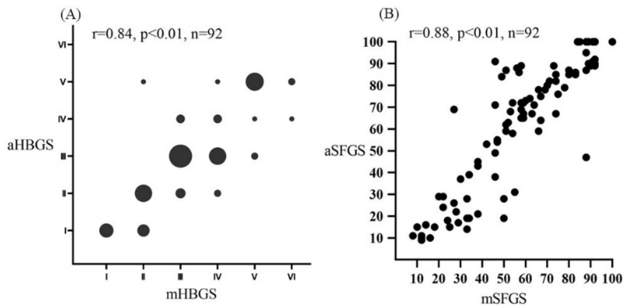


Fig. 3. **Correlations between the subjective evaluation and AFES.** (A) The correlation between aHBGS and mHBGS. The size of dots is proportional to the number of samples. The larger the area of dots, the more samples; The smaller the dot area, the smaller the sample. (B) The correlation between aSFGS and mSFGS. AFES, automatical facial evaluation system; HBGS, the House-Brackmann grading system; SFGS, the Sunnybrook Facial Grading System; aHBGS, auto-scored HBGS; mHBGS, manually scored HBGS; aSFGS, auto-scored SFGS; mSFGS, manually scored SFGS.

and the affected side, contributing to assess the severity. Two of them (I5, I7) were used for the evaluation of synkinesis, which was one of the major sequelae of facial palsy and an important item for assessing.

### C. Statistical Analysis

All statistical analyses were performed using IBM SPSS version 25. To verify the reliability of AFES, we analyzed the correlation between automated ratings and manual scorings. The Kolmogorov-Sminov test was used as the normality test for mSFGS and aSFGS. If the data did not conform to the normal distribution, the Spearman's correlation univariate analysis method was then employed. The Spearman's correlation univariate analysis method was used to analyze the correlations between mHBGS and aHBGS. To observe the accuracy of the aSFGS for participants with different severities, subgroup analyses based on the mHBGS (normal-mild, HBGSI-II; moderate, HBGS III-IV; severe, HBGSV-VI) were done. The Wilkerson rank-sum tests were applied to verify whether the change in forty-two participants of HBGS, SFGS, indicators between two evaluation was statistically significant. The Spearman's correlation univariate analysis method was used to analyze the correlations between seven regional facial features and the two scales (the scales of SFGS and HBGS). All the analysis was two-tailed t test. When P value was less than 0.05, we considered it was significant.

## III. Result

In this study, ninety-two participants were recruited and all of them completed the first evaluation. However, fifty participants withdrew from the second evaluation after two-week rehabilitation due to work arrangement, logistical problems, and quarantine policy of COVID-19. No adverse reactions were reported.

### A. Correlations Between Subjective Evaluation and AFES

Fig.3 showed the correlations of the results from manual evaluation and automatic evaluation. Very strong positive correlations, where the correlation coefficient located between 0.7 and 0.99 as suggested by Abdullah [42], were observed, which suggested the accuracy of evaluation via AFES (mHBGS and aHBGS, $r=0.84$, $p<0.01$, $n=92$; mSFGS and aSFGS, $r=0.88$, $p<0.01$, $n=92$).

### B. Subgroup Analyses According to Severity

Strong correlations ($0.4 \leq r \leq 0.69$) were showed in the subgroup analyses. A correlation coefficient of 0.655 was reported in the participants with normal-mild facial impairment ($p<0.01$, $n=27$). A correlation coefficient of 0.67 was reported in the participants with moderate facial impairment ($p<0.01$, $n=46$). A correlation coefficient of 0.57 was reported in the participants with severe FP patients ($p=0.01$, $n=19$). These results showed that the AFES might be more suitable for participants with normal-mild and moderate FP, comparing to severe ones. In addition, the accuracy of aHBGS in patients with different grades was calculated (see the supplementary for details).

### C. Correlations of Indicators and Subjective Scales

In order to explore potential associations of indicators of facial regional features with clinical scales, additional analyses were conducted and showed a strong negative correlation between I4 and mHBGS ($r=-0.64$, $p<0.01$, $n=92$) and a strong positive correlation between I4 and mSFGS ($r=0.64$, $p<0.01$, $n=92$). A weak positive correlation between I7 and aHBGS ($r=0.29$, $p=0.005$, $n=92$) and a moderate negative correlation between I7 and aSFGS ($r=-0.35$, $p=0.001$, $n=92$) were observed. These results suggested better performance in eyebrow raising and less symptom of synkinesis present a better recovery in facial function, which is consistent with the clinical observation.

### D. Analyses on Indicators of Seven Regional Facial Features

There were statistical differences between the baseline evaluation and follow-up evaluation in both AFES and subjective evaluations in forty-two participants (pre-mSFGS: $57.79\pm22.39$, post-mSFGS: $68.36\pm21.05$, $p<0.01$; pre-mHBGS: $3.21\pm1.16$, post-mHBGS: $2.69\pm1.24$, $p<0.01$; pre-aSFGS: $65.69\pm24.65$; post-aSFGS: $78.74\pm20.67$, $p<0.01$;pre-aHBGS: $2.62\pm0.94$; post-aHBGS: $2.14\pm0.98$, $p=0.005$). According to the significant improvements in facial motor function, further analyses on the indicators of seven regional facial features between the baseline evaluation and follow-up evaluation were conducted. It showed significant difference only on I7 (pre-I7: $1.05\pm0.11$; post-I7: $1.01\pm0.07$, $p=0.003$, $n=42$), which represented the synkinesis of FP and suggested the synkinesis relieved in the participants of our study. No obvious changes were found in the other indicators (pre-I1: $0.89\pm0.68$; post-I1: $0.76\pm0.58$; pre-I2: $1.01\pm0.12$; post-I2: $1.00\pm0.12$; pre-I3: $1.44\pm1.54$; post-I3: $1.24\pm1.21$; pre-I4: $0.65\pm0.28$; post-I4: $0.72\pm0.31$; pre-I5: $0.99\pm0.05$; post-I5: $0.98\pm0.05$; pre-I6: $1.00\pm0.10$; post-I6: $1.00\pm0.09$).

## IV. Discussion

Objective quantification of facial function plays an important role in the progression, recovery, treatment response of FP, and inter-clinician communication during rehabilitation [31], [43]. The present study proposed a novel evaluation apparatus and provided tentative evidence for its feasibility in evaluating facial motor function as an objective and reliable grading system.

### A. Feasibility of AFES in Clinic

A lot of algorithms and strategies were proposed in previous studies for automatical evaluation of FP. A quantitative evaluation system that combined parallel hierarchy convolutional neural network (PHCNN) with Long Short-Term Memory (LSTM) was proposed by Liu et al. [44]. The system could distinguish the images of FN patients from the healthy one and grade FN patients with three level. Zhuang et al. [45] suggested the facial detection framework based on a neurologists-verified dataset and they found it was feasible in clinical use, as it reached the evaluation level of residents and equivalent to the paramedics. Azoulay et al. [30] demonstrated their self-developed mobile application and found it could be a useful tool for grading FP patients with a very strong correlation of HBGS between their system and clinicians. In line with these studies, a very strong association of aHBGS and mHBGS was observed in our study, which suggested the accuracy of AFES. In addition, the detection of facial landmark is a crucial part of the facial palsy evaluation [46]. Recently, many scholars [47], [48], [49] put immense effort into exploring the balance among speed, accuracy and stability of technology of facial landmark detection. Xia et al. [50] suggested it was necessary to train the facial landmark detection with palsy-faces, which had significant differences from those trained with normal-faces. Comparing with similar studies used open access repositories of healthy people [24], [35], [45], [51], the AFES used real patients to train the algorithm which meant a better performance in landmark localization [43], [52]. Additionally, the simple and smooth evaluation process of AFES improve the patients' experience and cooperation, which might be beneficial to assess the real situation of the patients.

In this study, the AFES was a customized apparatus to provide objective and accurate evaluations for FP via automatically grading and scoring algorithm. The main evaluation results of AFES were presented as classical clinical scales, HBGS and SFGS, rather than a novel and unfamiliar indicator, which might contribute to the clinical promotion. As a finer grading system, SFGS demonstrated enhanced repeatability and good sensitivity to clinical changes rather than HBGS, which was proposed in few studies [27], [53]. Mothes et al. [27] attempted to verify the feasibility of different scales as classification methods and the results showed the automated Sunnybrook grading would be a good choice. There was a very strong correlation of aSFGS and mSFGS in Mothes' study. Similarly, a very strong correlation was also found in our study and it seemed that AFES might provide more accurate result of SFGS. The difference of input materials might lead to the different findings. Comparing to the photographs that Mothes used, the videos we applied might perform better in reflexing the real function of patients.

Furthermore, as suggested by the subgroup analyses, participants with normal-mild and moderate FP were found to be more suitable for AFES. The low accuracy in severe participants might be due to the excessive gap in facial muscle strength between affected and unaffected face of the severe ones. Thus, they were more likely to exhibit large facial movement displacement, which may affect the results. Furthermore, another reason for the difference of conclusions might be the sample size and distribution [44]. It should be noted that the small sample size may lead to data bias [54] as the sample size of severe participants was smallest in this study.

### B. Indicators of Facial Regional Features

For patients with facial palsy, taking more concern on the facial regional areas is particularly necessary for evaluating their facial function accurately. Liu et al. [44] improved the overall performance of facial evaluation system by considering the correlations between facial regions and reducing the contribution of unrepresentative organs. Parra-dominguez et al. [55] reaffirmed the importance of regional information as specific facial regions play a decisive role in the extraction of facial key features. In our study, seven indicators of facial regional features were provided to evaluate the local asymmetry and to show the specific improvements of facial function. As a series of social and aesthetic deficits secondary to the loss of facial motor function [56], patients with facial palsy are prone to have emotional problems. Studies [57], [58] showed that FP patients have a high degree of psychological distress, which was 3 to 5 times that of the normal population. One of the important reasons was they always emphasized on the improvements of facial motor function. However, clinical scales for FP could only present a general grade or score for the present status. Besides, the ceiling effect of these scales hindered the accurate evaluation for almost recovered people. It was suggested that the ideal facial nerve grading instrument should have the scores reflecting regional facial function [59]. A growing body of studies have focused on local asymmetries, which included regional asymmetry and angular asymmetry [25], [60]. However, most studies used it as important data in algorithm processing, and only a few studies input local asymmetry as a result for clinical reference [25], [61], [62]. In our study, indicators reflecting the movements of the patient's facial regions (eyebrows, eyes, and lips) and the severity of synkinesis in FP patients were suggested. With reference to the seven indicators, we can evaluate the regional efficacy of the treatments better and adjust the training plan accurately. Besides, the indicators might also be able to make supplementary instruction to the changes of clinical scales. As suggested by the correlations of indicators with clinical scales, an increase in I4 and a decrease in I7 both indicated the increase of SFGS scores and the decrease of HBGS, which showed the improvement of patients' facial motor function.

Several studies [59], [63], [64] have demonstrated the sensitivity in evaluation of their devices via follow-up evaluation. Similarly, our study also demonstrated the ability of AFES in clarifying the changes of FP via indicators, which confirmed the sensitivity of AFES. Furthermore, a statistically significant change in I7 was observed, which might suggest that the relief of eye synkinesis (I7) might contribute to the enhancements of SFGS and HBGS in the present study.

These indicators can help us to specifically interpret the therapeutic effects of exercise and provide precise guidance for adjustment of clinical treatment.

AFES provided objective and reliable results for FP, including auto-scoring clinical scales and complementary quantified facial features, which showed the feasibility in clinic. Besides, with the simple evaluation process, patients can complete AFES evaluation independently without relying on clinicians which suggested the possibility of applying AFES for remote medical evaluation and treatment. There is potential for AFES to be deployed onto portable devices like smartphones, which will expand its applications [65]. With the extra indicators offered by AFES, the patients can monitor the specific changes of facial motor function in real time and at any time, which might help patients establish a correct perception of their facial function and alleviate their psychological problems when they stay at home [66], [67].

There were several limitations of our study. One was the sample size, which might hinder the power of statistical analyses. Another limitation was the different participant distribution of training sets and testing sets, which might affect the performance of the algorithm. Thus, in the future study, we will recruit feasible number of participants with different types and severities.

## V. Conclusion

This study verified the feasibility of an automatical evaluation system providing objective and accurate evaluation results of HBGS and SFGS for FP. Moreover, it was the first to demonstrate the change of facial regional indicators which might contribute to reveal the specific changes of facial function and interpret the prognosis. This study provided tentative evidence for the clinical application of AFES and suggested the potential to use at home.

## References

[1] I. J. Kleiss, C. H. G. Beurskens, P. F. M. Stalmeier, K. J. A. O. Ingels, and H. A. M. Marres, "Quality of life assessment in facial palsy: Validation of the Dutch facial clinimetric evaluation scale," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 272, no. 8, pp. 2055–2061, Aug. 2015.

[2] A. L. Ho, A. M. Scott, A. F. Klassen, S. J. Cano, A. L. Pusic, and N. Van Laeken, "Measuring quality of life and patient satisfaction in facial paralysis patients: A systematic review of patient-reported outcome measures," *Plastic Reconstructive Surg.*, vol. 130, no. 1, pp. 91–99, Jul. 2012.

[3] M. J. Warner, J. Hutchison, M. Varacallo, and T. H. Busby, *Bell Palsy (Nursing)*. Treasure Island, FL, USA: StatPearls, 2022.

[4] H. S. Kim, S. Y. Kim, Y. H. Kim, and K. S. Park, "A smartphone-based automatic diagnosis system for facial nerve palsy," *Sensors*, vol. 15, no. 10, pp. 26756–26768, Oct. 2015.

[5] T. A. Hadlock and L. S. Urban, "Toward a universal, automated facial measurement tool in facial reanimation," *Arch. Facial Plastic Surg.*, vol. 14, no. 4, p. 277, Jul. 2012.

[6] E. Chávez-Serna, J. E. Telich-Tarriba, C. Altamirano-Arcos, L. Nahas-Combina, and A. Cárdenas-Mejía, "Facial paralysis, etiology and surgical treatment in a tertiary care center in plastic and reconstructive surgery in Mexico," *Cirugia Cirujanos*, vol. 89, no. 6, pp. 718–727, Jan. 2021.

[7] S. E. Coulson, N. J. O'Dwyer, R. D. Adams, and G. R. Croxson, "Expression of emotion and quality of life after facial nerve paralysis," *Otol. Neurotol.*, vol. 25, no. 6, pp. 1014–1019, Nov. 2004.

[8] J. W. House and D. E. Brackmann, "Facial nerve grading system," *Otolaryngol.-Head Neck Surg.*, vol. 93, no. 2, pp. 146–147, Apr. 1985.

[9] J. G. Neely, N. G. Cherian, C. B. Dickerson, and J. M. Nedzelski, "Sunnybrook facial grading system: Reliability and criteria for grading," *Laryngoscope*, vol. 120, no. 5, pp. 1038–1045, May 2010.

[10] C. G. Se, "Assessing physiotherapy rehabilitation outcomes following facial nerve paralysis," *Austral. J. Otolaryngol.*, vol. 2, no. 1, pp. 20–24, 1995.

[11] L. N. Lee, S. M. Susarla, M. H. Hohman, D. K. Henstrom, M. L. Cheney, and T. A. Hadlock, "A comparison of facial nerve grading systems," *Ann. Plastic Surg.*, vol. 70, no. 3, pp. 313–316, 2013.

[12] M. W. Robinson and J. Baiungo, "Facial rehabilitation: Evaluation and treatment strategies for the patient with facial palsy," *Otolaryngolog. Clinics North Amer.*, vol. 51, no. 6, pp. 1151–1167, 2018.

[13] E. D. Toffola, D. Bossi, M. Buonocore, C. Montomoli, L. Petrucci, and E. Alfonsi, "Usefulness of BFB/EMG in facial palsy rehabilitation," *Disab. Rehabil.*, vol. 27, no. 14, pp. 809–815, Jul. 2005.

[14] N. P. Schumann, K. Bongers, O. Guntinas-Lichius, and H. C. Scholle, "Facial muscle activation patterns in healthy male humans: A multi-channel surface EMG study," *J. Neurosci. Methods*, vol. 187, no. 1, pp. 120–128, Mar. 2010.

[15] X. L. Liu, B. R. Fu, L. W. Xu, N. Lu, C. Y. Yu, and L. Y. Bai, "Automatic assessment of facial nerve function based on infrared thermal imaging," *Guang Pu Xue Yu Guang Pu Fen Xi*, vol. 36, no. 5, pp. 1445–1450, May 2016.

[16] X. Liu, S. Dong, M. An, L. Bai, and J. Luan, "Quantitative assessment of facial paralysis using infrared thermal imaging," in *Proc. 8th Int. Conf. Biomed. Eng. Informat. (BMEI)*, Oct. 2015, pp. 106–110.

[17] C. Jiang et al., "Automatic facial paralysis assessment via computational image analysis," *J. Healthcare Eng.*, vol. 2020, pp. 1–10, Feb. 2020.

[18] A. Gerós, R. Horta, and P. Aguiar, "*Facegram*—Objective quantitative analysis in facial reconstructive surgery," *J. Biomed. Informat.*, vol. 61, pp. 1–9, Jun. 2016.

[19] E. Yurdakurban, G. S. Duran, and S. Görgülü, "Evaluation of an automated approach for facial midline detection and asymmetry assessment: A preliminary study," *Orthodontics Craniofacial Res.*, vol. 24, no. S2, pp. 84–91, Dec. 2021.

[20] J. R. Dusseldorp, M. M. van Veen, S. Mohan, and T. A. Hadlock, "Outcome tracking in facial palsy," *Otolaryngolog. Clinics North Amer.*, vol. 51, no. 6, pp. 1033–1050, Dec. 2018.

[21] M. G. Calvo, A. Fernández-Martín, G. Recio, and D. Lundqvist, "Human observers and automated assessment of dynamic emotional facial expressions: KDEF-dyn database validation," *Frontiers Psychol.*, vol. 9, p. 2052, Oct. 2018.

[22] S. He, J. J. Soraghan, B. F. O'Reilly, and D. Xing, "Quantitative analysis of facial paralysis using local binary patterns in biomedical videos," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 7, pp. 1864–1870, Jul. 2009.

[23] M. J. Javitt, E. A. Vanner, A. L. Grajewski, and T. C. Chang, "Evaluation of a computer-based facial dysmorphology analysis algorithm (Face2Gene) using standardized textbook photos," *Eye*, vol. 36, no. 4, pp. 859–861, Apr. 2022.

[24] D. L. Guarin, J. Dusseldorp, T. A. Hadlock, and N. Jowett, "A machine learning approach for automated facial measurements in facial palsy," *JAMA Facial Plastic Surg.*, vol. 20, no. 4, pp. 335–337, 2018.

[25] D. Y. Lee, H. S. Kim, S. Y. Kim, K. S. Park, and Y. H. Kim, "Comparison between subjective scoring and computer-based asymmetry assessment in facial nerve palsy," *J. Audiol. Otol.*, vol. 23, no. 1, pp. 53–58, Jan. 2019.

[26] G.-S. J. Hsu, J.-H. Kang, and W.-F. Huang, "Deep hierarchical network with line segment learning for quantitative analysis of facial palsy," *IEEE Access*, vol. 7, pp. 4833–4842, 2019.

[27] O. Mothes et al., "Automated objective and marker-free facial grading using photographs of patients with facial palsy," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 276, no. 12, pp. 3335–3343, 2019.

[28] G.-S.-J. Hsu, W.-F. Huang, and J.-H. Kang, "Hierarchical network for facial palsy detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 580–586.

[29] T. H. Ngo, Y. Chen, M. Seo, N. Matsushiro, and W. Xiong, "Quantitative analysis of facial paralysis based on three-dimensional features," in *Proc. Int. Conf. Image Process.*, 2016, pp. 1319–1323.

[30] O. Azoulay, Y. Ater, L. Gersi, Y. Glassner, and D. Halperin, "Mobile application for diagnosis of facial palsy," in *Proc. MOBMED*, 2014, pp. 1–14.

[31] Y. Zhao et al., "Prognostic value of a three-dimensional dynamic quantitative analysis system to measure facial motion in acute facial paralysis patients," *Head Face Med.*, vol. 16, no. 1, pp. 1–10, 2020.

[32] S. Monini et al., "Validation of the objective assessment of facial movement with a new software-based system," *Acta Oto-Laryngol.*, vol. 139, no. 5, pp. 456–460, May 2019.

[33] J. Lou, H. Yu, and F.-Y. Wang, "A review on automated facial nerve function assessment from visual face capture," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 488–497, Feb. 2020.

[34] M. M. van Veen, J. H. A. T. Berge, P. M. N. Werker, and P. U. Dijkstra, "Three-dimensional stereophotogrammetry assessment of facial asymmetry in facial palsy," *J. Craniofacial Surg.*, vol. 31, no. 4, pp. 893–897, 2020.

[35] J. Taeger, S. Bischoff, R. Hagen, and K. Rak, "Utilization of smartphone depth mapping cameras for app-based grading of facial movement disorders: Development and feasibility study," *JMIR mHealth uHealth*, vol. 9, no. 1, Jan. 2021, Art. no. e19346.

[36] B. Hontanilla and C. Aubá, "Automatic three-dimensional quantitative analysis for evaluation of facial movement," *J. Plastic, Reconstructive Aesthetic Surg.*, vol. 61, no. 1, pp. 18–30, 2008.

[37] B. R. Ross, G. Fradet, and J. M. Nedzelski, "Development of a sensitive clinical facial grading system," in *The Facial Nerve*, E. R. Stennert, G. W. Kreutzberg, O. Michel, and M. Jungehülsing, Eds. Berlin, Germany: Springer, 1994, pp. 180–181.

[38] Y. Satoh, J. Kanzaki, and S. Yoshihara, "A comparison and conversion table of 'the House–Brackmann facial nerve grading system' and 'the Yanagihara grading system,'" *Auris Nasus Larynx*, vol. 27, no. 3, pp. 207–212, Jul. 2000.

[39] E. E. Sanches, F. W. de Jongh, K. J. A. O. Ingels, and S. Pouwels, "E-health and telemedicine implementation in facial paralysis: Challenges and pitfalls," *Eur. J. Plastic Surg.*, vol. 44, no. 4, pp. 429–434, Aug. 2021.

[40] A. Y. Fattah et al., "Survey of methods of facial palsy documentation in use by members of the Sir Charles Bell Society," *Laryngoscope*, vol. 124, no. 10, pp. 2247–2251, Oct. 2014.

[41] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Jun. 2013, pp. 397–403.

[42] C. L. Henriquez, T. B. Croat, P. Poczai, and I. Ahmed, "Mutational dynamics of aroid chloroplast genomes II," *Frontiers Genet.*, vol. 11, Jan. 2021, Art. no. 610838.

[43] D. L. Guarin et al., "Toward an automatic system for computer-aided assessment in facial palsy," *Facial Plastic Surg. Aesthetic Med.*, vol. 22, no. 1, pp. 42–49, Feb. 2020.

[44] X. Liu, Y. Xia, H. Yu, J. Dong, M. Jian, and T. D. Pham, "Region based parallel hierarchy convolutional neural network for automatic facial nerve paralysis evaluation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2325–2332, Oct. 2020.

[45] Y. Zhuang et al., "Video-based facial weakness analysis," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 9, pp. 2698–2705, Sep. 2021.

[46] O. O. Abayomi-Alli, R. Damaševičius, R. Maskeliūnas, and S. Misra, "Few-shot learning with a novel Voronoi tessellation-based image augmentation method for facial palsy detection," *Electronics*, vol. 10, no. 8, p. 978, Apr. 2021.

[47] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2D & 3D face alignment problem? (And a dataset of 230,000 3D facial landmarks)," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Jun. 2017, pp. 1021–1030.

[48] Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, "Joint 3D face reconstruction and dense alignment with position map regression network," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 534–551.

[49] J. Guo, X. Zhu, Y. Yang, F. Yang, Z. Lei, and S. Li, "Towards fast, accurate and stable 3D dense face alignment," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 152–168.

[50] Y. Xia, C. Nduka, R. Y. Kannan, E. Pescarini, J. E. Berner, and H. Yu, "AFLFP: A database with annotated facial landmarks for facial palsy," *IEEE Trans. Computat. Social Syst.*, early access, Jul. 15, 2022, doi: 10.1109/TCSS.2022.3187622.

[51] V. Lum, M. S. Goonewardene, A. Mian, and P. Eastwood, "Three-dimensional assessment of facial asymmetry using dense correspondence, symmetry, and midline analysis," *Amer. J. Orthodontics Dentofacial Orthopedics*, vol. 158, no. 1, pp. 134–146, Jul. 2020.

[52] B. A. Asgarian, "Biases in facial landmark with dementia," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2019, pp. 28–36.

[53] M. Tramontano et al., "Sunnybrook facial grading system reliability in subacute stroke patients," *Eur. J. Phys. Rehabil. Med.*, vol. 57, no. 5, pp. 685–690, 2021.

[54] Y. Zhang, T. Shao, L. Yao, H. Yue, and Z. Zhang, "Effects of tirofiban on stent thrombosis, Hs-CRP, IL-6 and sICAM-1 after PCI of acute myocardial infarction," *Exp. Therapeutic Med.*, vol. 16, no. 4, pp. 3383–3388, Aug. 2018.

[55] G. S. Parra-Dominguez, C. H. Garcia-Capulin, and R. E. Sanchez-Yanez, "Automatic facial palsy diagnosis as a classification problem using regional information extracted from a photograph," *Diagnostics*, vol. 12, no. 7, p. 1528, Jun. 2022.

[56] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychol. Sci. Public Interest*, vol. 20, no. 1, pp. 1–68, Jul. 2019.

[57] S. Brown, B. Isaacson, W. Kutz, S. Barnett, and S. M. Rozen, "Facial nerve trauma: Clinical evaluation and management strategies," *Plastic Reconstructive Surg.*, vol. 143, no. 5, pp. 1498–1512, May 2019.

[58] M. Sugiura et al., "An assessment of psychological stress in patients with facial palsy," *Nippon Jibiinkoka Gakkai Kaiho*, vol. 106, no. 5, pp. 491–498, May 2003.

[59] A. Y. Fattah et al., "Facial nerve grading instruments: Systematic review of the literature and suggestion for uniformity," *Plastic Reconstructive Surg.*, vol. 135, no. 2, pp. 569–579, 2015.

[60] A. Song, G. Xu, X. Ding, J. Song, G. Xu, and W. Zhang, "Assessment for facial nerve paralysis based on facial asymmetry," *Australas. Phys. Eng. Sci. Med.*, vol. 40, no. 4, pp. 851–860, 2017.

[61] J. Feng, Z. Guo, J. Wang, and G. Dan, "Using eye aspect ratio to enhance fast and objective assessment of facial paralysis," *Comput. Math. Methods Med.*, vol. 2020, pp. 1–11, Jan. 2020.

[62] A. Gaber, M. F. Taher, and M. A. Wahed, "Quantifying facial paralysis using the Kinect v2," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2497–2501.

[63] D. Haase, L. Minnigerode, G. F. Volk, J. Denzler, and O. Guntinas-Lichius, "Automated and objective action coding of facial expressions in patients with acute facial palsy," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 272, no. 5, pp. 1259–1267, May 2015.

[64] J. J. Greene, J. Tavares, D. L. Guarin, and T. Hadlock, "Clinician and automated assessments of facial function following eyelid weight placement," *JAMA Facial Plastic Surg.*, vol. 21, no. 5, pp. 387–392, Sep. 2019.

[65] T. Wang, S. Zhang, L. Liu, G. Wu, and J. Dong, "Automatic facial paralysis evaluation augmented by a cascaded encoder network structure," *IEEE Access*, vol. 7, pp. 135621–135631, 2019.

[66] F. C. Macgregor, "Facial disfigurement: Problems and management of social interaction and implications for mental health," *Aesthetic Plastic Surg.*, vol. 14, no. 1, pp. 249–257, Dec. 1990.

[67] S. R. Chaiet and S. M. Carpenter, "Importance of reporting duration of facial paralysis in studies of emotion and well-being," *JAMA Facial Plastic Surg.*, vol. 19, no. 5, pp. 437–438, Sep. 2017.