Automated Grading Hemifacial Spasm Using Smartphone Cameras

Ka I Chan[1*], Bo Hei[2*], Linghao Meng[3*], Ruen Liu[2†], Yuntao Wang[3 4 5†], Chang Chen[3], Qingpei Hao[2], Yuanchun Shi[3 5]
[1]*Global Innovation Exchange Institute, Tsinghua University, Beijing 100084, China*
[2]*Department of Neurosurgery, Peking University People's Hospital, Beijing 100044, China*
[3]*Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*
[4]*National Key Laboratory of Human Factors Engineering, Beijing 100094, China*
[5]*Qinghai University, Xining, Qinghai 810016, China*
Email: {chenjy22, menglh20, chenchan22}@mails.tsinghua.edu.cn, heibo12345@126.com,
{liuruen, haoqp}@pku.edu.cn, {yuntaowang, shiyc}@tsinghua.edu.cn

***Abstract*—Hemifacial spasm is a chronic neurological condition characterized by involuntary facial muscle contractions caused by nerve compression. While familiar to specialists, it is less known to the public and general practitioners, which can lead to difficulties in diagnosis and severity assessment, and even misdiagnosis. Consequently, patients are common to have a long medical history. However, long-term patients tend to have poorer outcomes following surgery, and one-third of patients experience a delayed cure during postoperative rehabilitation. Moreover, 4% of patients experience recurrence, highlighting the importance of early and accurate diagnosis as well as postoperative monitoring. In this paper, we collected a video dataset of 50 hemifacial spasm patients and 9 healthy adults. We identified three facial features from the videos to establish a novel grading system closely aligned with the medical standards, specifically the Cohen-Albert Grading System. We also developed algorithms capable of automatically grading hemifacial spasm using smartphone cameras based on facial keypoint detection. These algorithms were evaluated on the dataset, achieving an accuracy of 88% for detection and a mean absolute error of 0.42 for grading.**

***Keywords*-Hemifacial Spasm, Keypoint Detection, Cohen-Albert Grading System, AI-Assisted Diagnosis**

Asterisk indicates co-first authors. †Dagger indicates corresponding authors.

I. Introduction

Hemifacial spasm (HFS) is a chronic neurological disorder characterized by involuntary contractions of the facial muscles on one side of the face, which can significantly impact a patient's quality of life, leading to social discomfort, emotional distress, and functional limitations [1]. Although HFS predominantly affects middle-aged and elderly individuals, there is a trend towards younger age of onset [2].

HFS is uncommon due to its relatively low incidence and prevalence rates, and epidemiological data on the condition are limited. Studies conducted in 1990 and 2004 reported an average annual incidence rate of 0.78 per 100,000 [3], and a total prevalence rate of 9.8 per 100,000 [4]. However, there are patients have a medical history lasting up to 8.2 years [5], and in some cases, up to 15 years [6]. For untreated consecutive HFS patients, the natural history can extend up to 42 years (mean 12 years) from the onset of symptoms [7].

The reasons for the long duration from onset to surgery in treated patients are multifaceted:

1. **Lack of Awareness and Slow Disorder Progression**: Due to the low incidence rate and the rarity of encountering others with similar symptoms, many individuals may not recognize it as a disorder. The condition progresses relatively slowly [8], often only prompting medical assistance when it significantly impacts daily life, which can take several years to over a decade. Consequently, patients do not receive timely and effective guidance.
2. **Difficulty in Accurate Diagnosis and Defining Severity**: HFS is relatively uncommon within the field of neurology. Misdiagnosis and lack of referral from general practitioners likely contribute to underestimating the true prevalence of HFS [4]. Even neurologists may find it challenging to diagnose and differentiate HFS accurately. Traditional medical grading methods are not quantifiable, making it difficult to scientifically and effectively define the severity of the disorder.
3. **Limited Treatment Recommendations and Options in Primary Hospitals**: Patients often initially consider non-invasive treatments such as medication or acupuncture, which are not curative. High cure-rate treatments like microvascular decompression (MVD) surgery are not feasible in primary hospitals, which mainly rely on non-surgical conservative treatments. In underdeveloped and rural regions, some neurologists may not be aware of treatment options for this condition, nor do they know where such treatments are available, leading to ineffective treatment recommendations [9].

Overall, the main challenges for HFS stem from unclear diagnostic standards and the difficulty in medical grading, which require experienced specialists and even rely on patients' subjective feelings. Therefore, developing a quantifiable method for diagnosing and grading HFS is crucial. Such a method could reduce medical costs and the waste of

2993-396X/24/$31.00 ©2024 IEEE
DOI 10.1109/SWC62898.2024.00070

medical resources, as well as provide an efficient solution for distant patients to minimize their time and economic costs. Additionally, it could be used for postoperative follow-up in telehealth, enabling remote and low-cost medical care.

In this paper, we aim to design an AI-assisted diagnostic and severity grading algorithm for HFS, addressing the issue of diagnosis and grading. This solution enables autmomated diagnosis and grading anytime and anywhere, without requiring additional prior knowledge, and provides crucial suggestions through preliminary diagnosis and grading, determining whether it is necessary to seek consultation and treatment from an experienced specialist.

We collected facial video datasets from 50 HFS patients and 9 healthy adults. Using qualitative methods and the expertise of an experienced specialist, we quantified three facial features for grading HFS and proposed a novel HFS grading system based on them. We then designed and developed detection and grading algorithms for HFS. The algorithms were tested on the dataset, achieving a detection accuracy of 88% and a mean absolute error (MAE) of grading of 0.42. Our contributions are threefold:

1. We identified measurable three facial features for assessing and grading HFS and proposed a novel HFS grading system, which addresses the vagueness and non-quantifiability of traditional medical grading systems that rely mostly on the accumulated experience of doctors.
2. We designed and developed algorithms capable of quantifying HFS for detection and grading.
3. The detection algorithm's accuracy and the grading algorithm's MAE were evaluated using collected dataset, showing results that are close to medical grading system.

II. Preliminaries and Related Work

A. Hemifacial Spasm

Hemifacial spasm (HFS) is a chronic disorder characterized by unilateral facial muscle contractions due to abnormal compression of the facial nerve, often from vascular anomalies in the cerebellopontine angle region [10], [11]. These contractions can cause involuntary eyelid closure, eyebrow elevation, visual interference, and social embarrassment, significantly impacting quality of life [5], [11]–[13]. Limited epidemiological data report an incidence rate of 0.78 per 100,000 in Olmsted County, Minnesota [3], and a prevalence rate of 9.8 per 100,000 in Oslo, Norway, increasing with age [4]. Interestingly, HFS appears more common in some Asian populations [14].

The efficacy of oral medication and botulinum neurotoxin injections for HFS is generally limited and transient [15], [16]. MVD surgery is the most effective approach, addressing the underlying cause with a high success rate ranging from 80% to 88% [16]–[19]. However, about one-third of patients may experience a delayed cure during postoperative rehabilitation, which can range from 7 days to 12 months [20]–[27], and 4% of patients experience recurrence [5]. Moreover, long-term HFS patients often have poorer surgical outcomes [5]. Therefore, early diagnosis and intervention are crucial for optimizing treatment results, along with effective postoperative follow-up to monitor the condition.

A grading system standardizes HFS severity assessment, aiding clinical diagnosis, treatment planning, and monitoring. The most common is Cohen-Albert Grading System [28], which categorizes spasticity as follows:

0: None.
I: Increased blinking caused by external stimuli.
II: Mild, noticeable fluttering, not incapacitating.
III: Moderate, very noticeable spasm, mildly incapacitating.
IV: Severely incapacitating (unable to drive, read, etc.).

However, HFS diagnosis must be made by a professional doctor, as the grading system contains non-quantifiable descriptions. Moreover, HFS has a relatively low incidence and prevalence rate and shares similarities with other conditions causing facial twitches, such as psychogenic facial spasm, facial nerve tic, facial myokymia, blepharospasm, and tardive dyskinesia. These often leads to misdiagnosis and lack of referral from general practitioners [4], necessitating specialist evaluation.

B. Facial Keypoint Detection

Facial keypoint detection technology is a crucial technique in computer vision, enabling the automatic identification and localization of key feature points on facial images or videos. These distinguishable characteristic points on the face, known as facial keypoints, include the corners and center of the mouth, nose, and eyes. Typical applications include facial expression analysis, face recognition, face tracking, and head gesture understanding [29], [30]. Facial keypoint detection algorithms can be broadly divided into traditional machine learning algorithms, such as the holistic method, constrained local method, and regression-based method, as well as deep learning-based algorithms [31], [32].

Our work selected two classic facial keypoint detection models: the 68-point annotation model in dlib and the Face Mesh model in MediaPipe. Dlib's model uses an Ensemble of Regression Trees to predict and refine keypoint positions [33]. In contrast, Google's MediaPipe provides the Face Mesh model, which detects and tracks 468 facial keypoints using a deep convolutional neural network [34], [35]. While both models are effective for facial keypoint detection, dlib's model offers high accuracy but limited detail with its 68 keypoints, whereas MediaPipe's model provides very high accuracy and detailed feature detection with its 468 keypoints, as well as real-time capabilities, making it popular for applications such as virtual try-ons, facial animation capture, and augmented reality filters.

C. AI-Assisted Diagnosis From Faces

Most neurological conditions manifest on the face, making facial information valuable for diagnosis. These conditions include facial paralysis, Parkinson's disease, Alzheimer's disease, amyotrophic lateral sclerosis, and epilepsy [36]–[38]. Among these, facial paralysis presents symptoms most similar to HFS. Objective methods for measuring face paralysis mainly include linear measurement method and image subtraction method, both of which use quantifiable indicators to reduce diagnostic errors [39]. Based on these objective and quantifiable metrics, AI can be employed to detect facial paralysis. Traditional machine learning approaches typically rely on extracting asymmetries between the two sides of the face, achieving an accuracy rate of 60.7% [40], [41]. A deep learning approach has also been proposed with a classification accuracy of 91.25% [42]. However, the developed tools are still rarely used in clinical practice due to the reliability issues, stemming from a lack of clinical validation and insufficient applicability.

III. DESIGN A QUANTIFIABLE GRADING SYSTEM

In this section, we mainly describe our data collection procedure. Then, we quantified three facial features through qualitative approach based on the data, identifying their relevant patterns for automated grading HFS. Subsequently, we proposed a novel quantifiable grading system for HFS and validated it with Kappa coefficient, Spearman correlation coefficient and MAE to confirm its consistency with the widely adopted Cohen-Albert Grading System in medical field.

A. Data Collection

To design our quantifiable grading system and algorithms for HFS, we collected a facial video dataset involving 50 HFS patients (31 females and 19 males, age ranging from 26 to 74, with a mean age of 54.48 and a standard deviation of 11.37) and 9 healthy adults (3 females and 6 males, age ranging from 22 to 30, with a mean age of 25.67 and a standard deviation of 2.96), with the approval by the Ethics Committee of Peking University People's Hospital. All HFS patients were scheduled to undergo MVD surgery, and the healthy adults were all college students who identified as not having any facial disorders. Informed consent was obtained from all participants prior to their inclusion in the study.

We recorded the facial videos using an iPhone 12 Pro camera at 1080P and 60 frames per second (FPS). All recordings were conducted in a controlled environment, either at the hospital or on campus. Participants sat in a comfortable chair with a frontal smartphone placed on a desk facing them. They were instructed to ensure that their face and neck were clearly visible, free from glasses and hair obstructions. As HFS can be triggered by relaxing the face after voluntary and forceful facial muscle contractions [11], participants were asked to perform two specific facial movements in the videos: forcefully closing their eyes and pulling the corners of their mouth to the sides (Figure 1).

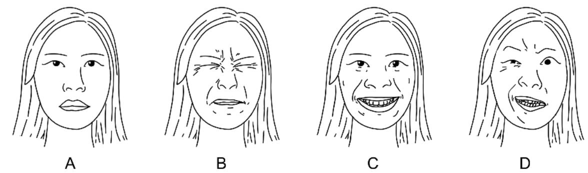

Figure 1. Participants were instructed to do two specific facial movements to trigger HFS symptoms: forcefully closing their eyes (B) and forcefully pulling the corners of their mouth to the sides (C). Natural status and the occurrence of HFS are indicated by (A) and (D), respectively.

The procedure includes obtaining consent from the participants and collecting the data, took five to ten minutes to complete. We recorded data from 59 participants, and 7 patients were recorded a second time three to six days after surgery. This resulted in a total of 9428 seconds of video footage, with an average length of 142.85 seconds for each video. All videos were graded and labeled by a specialist with 14 years of experience based on the Cohen-Albert Grading System [28].

B. Facial Features Extraction

By observing the HFS symptoms of patients in the videos and consulting with an experienced specialist, we found three facial features, each associated with specific HFS symptoms:

1. **Eye Twitching**: Symptoms include eyelid twitching or involuntary eyelid closure.
2. **Mouth Twitching**: Symptoms include twitching of the corners of the mouth.
3. **Platysma Twitching**: Symptoms include involuntary twitching of the cheeks.

We labeled the patients in the videos for unnatural twitching of the eyes, the mouth, and the platysma. For each grade of the Cohen-Albert Grading System, we counted the proportion of patients exhibiting eye twitching, mouth twitching, and platysma twitching. The proportions of each type of twitching are presented in Figure 2, revealing two distinct patterns:

1. **Sequence in the occurrence of twitches**: Patients with mouth twitches often also experience eye twitches, and patients with platysma twitches frequently exhibit simultaneous eye and mouth twitches.
2. **Variation in twitch locations among different grades**: According to the Cohen-Albert Grading System, patients rated as grade IV often have platysma twitches, grade III patients frequently exhibit mouth twitches, grade I and II patients predominantly have eye twitches, and grade 0 patients do not exhibit any facial twitches.

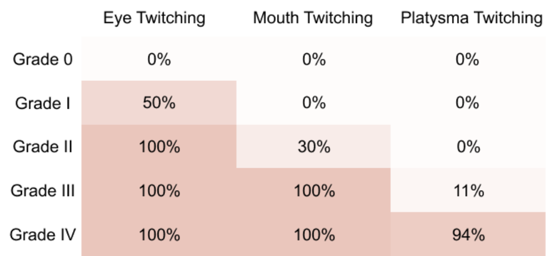

Figure 2. Proportion of patients with different twitching symptoms categorized by Cohen-Albert Grading System.

C. Validation of Proposed Grading System

The process of using Cohen-Albert Grading System to grade patients relies on the patients' subjective feelings and the specialists' clinical knowledge. Therefore, it is neccessary to establish a novel grading standard for HFS. This grading system should be clearly described, objective, and quantifiable, allowing for the implementation of an automated grading approach. Thus, we proposed a quantifiable grading system based on the three facial features.

Since we discovered a sequence in the occurrence of twitching and observed that the location of twitching vary among patients with different severity grades, it is possible to grade patients based on the location of twitching. We designed six different grading systems (S1 to S6). For example, according to the S1 system, level 0 patients have no twitching at all, level 1 patients have only eye twitching, level 2 patient have eye and mouth twitching, and level 3 patients have eye, mouth and platysma twitching. In other systems, the sequence of the location of twitching may be changed (Table I).

We graded each participant in the dataset based on these six grading systems and compared the results with the Cohen-Albert Grading System, calculating their Kappa coefficient, Spearman correlation coefficient and MAE respectively (Table I).

Table I
CONSISTENCY BETWEEN SIX PROPOSED GRADING STANDARDS AND COHEN-ALBERT GRADING SYSTEM

	Level 1	Level 2	Level 3	KC	SCC	MAE
S1	Eye	Mouth	Platysma	0.85	0.95	0.48
S2	Eye	Platysma	Mouth	0.72	0.88	0.76
S3	Mouth	Eye	Platysma	0.79	0.92	0.67
S4	Mouth	Platysma	Eye	0.48	0.77	1.14
S5	Platysma	Eye	Mouth	0.66	0.89	0.95
S6	Platysma	Mouth	Eye	0.48	0.77	1.14

Note: Level 0 is "None" for all standards. KC stands for Kappa coefficient, SCC stands for Spearman correlation coefficient, and MAE stands for mean absolute error.

Because our grading systems divide patients into four levels, while the Cohen-Albert Grading System divides patients into five grades, it is difficult to judge the consistency of the two systems from the directly calculated Kappa coefficient and MAE. However, we found that the symptoms of grade I and grade II patients in Cohen-Albert Grading System were both mild HFS. Therefore, to facilitate comparison, we performed additional processing with when calculating the Kappa coefficient and MAE: grade I and grade II patients were uniformly recorded as level 1, and grade III and grade IV patients were recorded as level 2 and level 3 respectively.

As can be seen from Table I, S1 has the highest Kappa coefficient and Spearman correlation coefficient, and the lowest MAE. Therefore, S1 is the closest and most consistent grading system to the Cohen-Albert Grading System.

Therefore, the novel grading system categorizes patients into four major levels based on the location of twitching:

0: No symptom of HFS.
1: Unnatural twitching only in the eyes.
2: Unnatural twitching in the eyes and mouth.
3: Unnatural twitching in the eyes, mouth and platysma.

IV. AUTOMATED GRADING ALGORITHM

We aim to implement a video-based automated HFS grading algorithm based on our proposed grading system (Figure 3). The algorithm consists of three sub-algorithms: an eye twitching detection algorithm, a mouth twitching detection algorithm, and a platysma twitching detection algorithm. Each sub-algorithm independently detects the presence of twitching in the corresponding locations. Ultimately, we combine the detection results to achieve automated grading for patients.

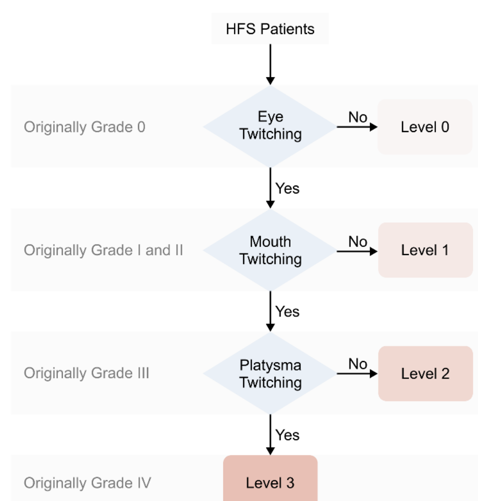

Figure 3. Flowchart of our proposed quantifiable HFS grading system.

A. Eye Twitching Detection

Generally speaking, when a patient's eye twitches, the size of the eye changes significantly within a short period of time.

To calculate the eye size, we use Mediapipe's FaceMesh model to detect five key points on the upper and lower eyelids of both the left and right eyes, totaling 20 key points. We then calculate the average y-coordinates of the key points on the upper and lower eyelids of each eye respectively, and define the eye size as the difference between these average y-coordinates.

However, different face shapes may lead to different eye sizes, and different video recording resolutions may also lead to different eye sizes. In order to eliminate the errors caused by different face shapes and video ratios, we calculate the percentage of eye size as follows:

$$P_{\text{eye}} = \frac{S_i}{\frac{1}{n}\sum_{i=1}^{n} S_i} \tag{1}$$

where P_{eye} is the percentage of eye size, S_i is the eye size in a certain frame i, n is the total number of frames.

When patients blink and close their eyes in the video, the eye size also changes significantly in a short period of time. In order to eliminate the errors caused by blinking and closing eyes, when the eye size is less than or equal to a certain blink and close eye threshold, we uniformly record the subsequent eye size as the threshold until the eye size is greater than the threshold again.

We find that the eye size of a patient with eye twitching rapidly decrease and then increase within a short period of time. Based on this pattern, we designed an algorithm to detect eye twitching. Additionally, since an eye twitch is short-lived, we choose one-third of a second as the maximum frame interval of a twitch. The pseudo code of the eye twitching detection algorithm is provided below (Algorithm 1).

Algorithm 1 Eye Twitching Detection.

```
extremes ← FindLocalExtremes(ratio_recorder)
for each extreme in extremes do
    if extreme is 'Local Minimum' then
        extreme_l ← FindLeftExtreme(extreme)
        extreme_r ← FindRightExtreme(extreme)
        amplitude ← extreme_l+extreme_r−2∗extreme
        if amplitude ≥ TWITCH_THRESHOLD then
            twitch_recorder ← eye, frame, amplitude
        end if
    end if
end for
```

B. Mouth Twitching Detection

Similar to eye twitching, when a patient experiences mouth twitching, the distance between the corners of the mouth changes significantly within a short period of time. To calculate this distance, we use Mediapipe's FaceMesh model to detect three key points at the left and right corners of the mouth, totaling six key points. We define the mouth corner distance as the Euclidean distance between the central points of the left and right mouth corners. To eliminate the errors caused by face shape and video resolution, we calculate the percentage of mouth corner distance as follows:

$$P_{\text{mouth}} = \frac{D_i}{\frac{1}{n}\sum_{i=1}^{n} D_i} \tag{2}$$

where P_{mouth} is the percentage of mouth corner distance, D_i is the mouth corner distance in a certain frame i, n is the total number of frames.

Since patients can be induced HFS symptoms by forcefully pulling the corners of their mouth to the sides when recording videos, conscious pulling behavior needs to be excluded when detecting patients' mouth twitching. Therefore, when the mouth corner distance is greater than or equal to a certain pulling threshold, the subsequent mouth corner distances can be uniformly recorded as the threshold until the mouth corner distance is less than the threshold again.

We discover that the mouth corner distance of a participant with mouth twitching rapidly increase and then decrease in a short period of time. Based on this pattern, we design an algorithm similar to eye twitching detection. The difference is that when detecting mouth twitching, we use the local maximum instead of the local minimum.

C. Platysma Twitching Detection

The eye and mouth twitching detection algorithms are very similar in principle and method. However, it is difficult to use the same principle and method to detect the platysma twitching for two main reasons: the facial key point detection technology cannot detect the key points near the platysma, and platysma twitching may be difficult to observe directly due to light, clothing, hair and other factors in the video.

Therefore, we aim to use six possible feature vectors to determine whether the patient in the video has platysma twitching. These six feature vectors are:

1. Frequency of eye twitching
2. Average twitch amplitude of eye twitching
3. Maximum twitch amplitude of eye twitching
4. Frequency of mouth twitching
5. Average twitch amplitude of mouth twitching
6. Maximum twitch amplitude of mouth twitching

In terms of model selection, we chose the decision tree. It is a white box model, so we can understand the decision process of the model by looking at the tree structure. However, due to the size of the current dataset, we need to prevent overfitting when using decision tree. We used five-fold cross-validation to calculate the average accuracy of decision tree of different depths in predicting platysma twitching.

V. Results

A. Thresholds for Eye and Mouth Twitching Detection

We selected two videos, one featuring a participant with eye twitching and one without, calculated the eye size in the videos frame by frame, performed the above two steps of preprocessing, and plotted the curve of eye size changing over time. In Figure 4, the blue and purple curve is the curve of eye size changing over time of the participant without eye twitching, while the red and orange curve is the participant with eye twitching.

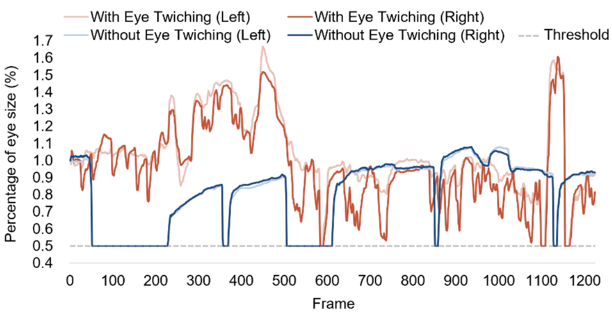

Figure 4. The eye size changes were compared between participants with eye twitching (red) and without eye twitching (blue), with a 0.5 threshold of eye blinking and closing (grey).

We also selected two videos, one featuring a participant with mouth twitching and one without, calculated the mouth corner distance in the videos frame by frame. In Figure 5, the blue curve is the participant without mouth twitching, while the red curve is the participant with mouth twitching.

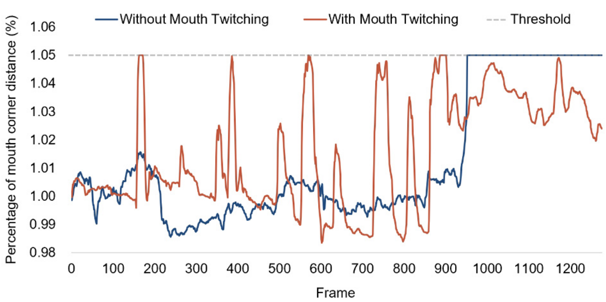

Figure 5. The mouth twitching curve changes were compared between participants with mouth twitching (red) and without mouth twitching (blue), with a 1.05 pulling threshold (grey).

It was observed that when the eye blinking and closing threshold is 0.5, it effectively distinguishes blinking and closing eyes without interfering with the detection of eye twitching. Meanwhile, when the pulling threshold is 1.05, it effectively identifies pulling behavior without interfering with the detection of unnatural twitching of the mouth corners.

In order to select the best twitch thresholds for eye and mouth, we used different twitch thresholds to detect eye twitching and mouth twitching. However, the time cost of annotating each eye and mouth twitch of the patient in the video is relatively large. Therefore, when calculating the algorithm metrics, we only compare whether the patient in the video actually has eye twitching and mouth twitching with the detected eye twitching and mouth twitching. Table II is the algorithm metrics when using different eye and mouth twitch thresholds.

Table II
ALGORITHM METRICS FOR DIFFERENT EYE AND MOUTH TWITCH THRESHOLDS

Threshold		Accuracy	Precision	Recall	F1 Score
Eye	0.03	83.33%	87.18%	94.44%	0.91
	0.04	88.1%	91.89%	94.44%	0.93
	0.05	85.71%	94.12%	88.89%	0.91
	0.06	88.1%	96.97%	88.89%	0.93
	0.07	88.1%	96.97%	88.89%	0.93
	0.08	85.71%	96.88%	86.11%	0.91
Mouth	0.015	78.57%	85.19%	82.14%	0.84
	0.016	76.19%	84.62%	78.57%	0.81
	0.017	76.19%	87.5%	75%	0.81
	0.018	71.43%	86.36%	67.86%	0.76
	0.019	66.67%	85%	60.71%	0.71
	0.020	61.9%	83.33%	53.57%	0.65

We found that the algorithm had the highest accuracy and precision when the eye twitch threshold was 0.06, which indicates that when the eye size changes by more than 6% in a short period of time, it often sigifies the eye twitching. Similarly, we found that the algorithm had the highest precision and relatively high accuracy when the mouth twitch threshold was 0.017, which indicates that when the mouth corner distance changes by more than 1.7% in a short period of time, it often signifies the mouth twitching.

B. *Model Visualization for Platysma Twitching Detection*

We used five-fold cross-validation to calculate the average accuracy of decision tree of different depths in predicting platysma twitching, and found that the decision tree with a depth of three had the highest average accuracy.

We plotted the rule graph of the decision tree (Figure 6) and calculated the importance of different feature vectors (Table III) in the model. Due to the current size of the dataset, the decision tree may not necessarily be suitable for all HFS patients. However, we found that the rules of the decision tree are consistent with our expectation, which is patients with platysma twitching often exibit the following characteristics: eye and mouth twitching, high frequency of eye twitching, and large amplitude of eye twitching and mouth twitching.

C. *Algorithm Performance*

For patients and healthy individuals in our video dataset, we extracted a 20-second video clip for each, ensuring to include the HFS symptoms and instructed facial movements (such as blinking, and pulling the corners of mouth to each side) as much as possible. We then manually grade each clip based on the presence of eye twitching, mouth twitching, or platysma twitching. Ultimately, our test set comprised

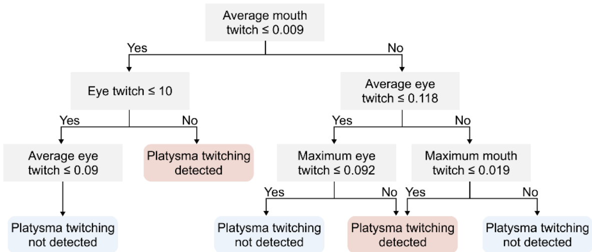

Figure 6. Rule graph of the decision tree for platysma twitching detection algorithm.

Table III
IMPORTANCE OF DIFFERENT FEATURE VECTORS

Feature	Importance
Frequency of Eye Twitching	0.13
Avg Amplitude of Eye Twitching	0.24
Max Amplitude of Eye Twitching	0.17
Frequency of Mouth Twitching	0
Avg Amplitude of Mouth Twitching	0.34
Max Amplitude of Mouth Twitching	0.12

50 video clips, each with a duration of 20 seconds. We compared the results of automated grading with manual grading and generated the corresponding confusion matrix (Figure 7).

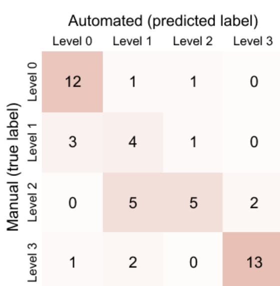

Figure 7. Confusion matrix between automated and manual grading.

The accuracy of the detection algorithm is 88%. By observing the confusion matrix, we found that only one patient without HFS was misdiagnosed as having HFS, and four patients with HFS were misdiagnosed as not having HFS.

The accuracy of the grading algorithm is 68%. By observing the confusion matrix, we found that the algorithm performed well in grading patients at level 0 and level 3, but had lower accuracy for patients at level 1 and level 2. This discrepancy is primarily due to the relatively low recall of the mouth twitching detection algorithm. The MAE of the grading algorithm is 0.42, indicating that the majority of patients are classified as the correct grading or do not differ by more than one grade from the correct grading.

VI. DISCUSSION

Although many existing studies have applied facial keypoint and localization detection in assistive medical diagnosis, these typically address common disorders (e.g. facial paralysis) that people and doctors can easily identify, while our work focuses on an uncommon disorder, hemifacial spasm. Patients suffering from HFS often go undiagnosed, misdiagnosed and untreated for long periods. Additionally, the slow progression of HFS and the possibility of delayed cure necessitate frequent doctor visits to monitor the severity of the condition. However, most patients are unable to do so due to the costs of time, distance, and medical fees. Therefore, our work provides a novel approach for people, particularly distant patients and primary care doctors, to assist in diagnosis in a cost-effective and relatively accurate way.

By analyzing videos of patients exhibiting HFS symptoms and consulting with an experienced specialist, we identified three primary facial features: eye twitching, mouth twitching, and platysma twitching. We observed a sequential pattern in the occurrence of these twitches, with variations in twitch locations corresponding to different severity grades of the condition. Therefore, we designed a grading system based on the specific locations of these twitches. To detect eye and mouth twitching, we utilize Mediapipe's FaceMesh model to capture 26 key facial points. We define two critical feature values: eye size and mouth corner distance. Our findings indicate that an eye size change exceeding 6% within a short period typically signifies eye twitching, while a mouth corner distance change exceeding 1.7% often indicates mouth twitching.

However, there are several limitations in our work. First, detecting platysma twitching is challenging, so the algorithms were designed based on metrics of eye and mouth movements to substitute for platysma twitching detection. Second, while triggering spasms is possible, it is impossible to control the exact timing of their occurrence. Therefore, we cannot ensure that the most severe spasm happens when data are collected from each patient. Our algorithms can only provide results based on the input videos.

Our work established AI-assisted diagnostic algorithms based on facial keypoint detection using a collected facial video dataset, marking the first medical-engineering application research related to HFS. Our algorithms provide a significant foundation for subsequent app development, enabling the detection and grading of the severity of the condition. This offers people a convenient, fast, and low-cost preliminary diagnosis without requiring any medical knowledge. Additionally, if we can deploy the algorithm to a mobile app and promote it to the public, it will facilitate data growth. Eventually, using deep learning approaches may enable continuous improvement in the accuracy of diagnosis and grading. The ultimate goal is to develop a self-learning AI-assisted diagnostic algorithm that can achieve a level of expertise comparable to that of an experienced specialist.

VII. Conclusion

Hemifacial spasm has a relatively low incidence and prevalence rate, and its similarity to other facial disorders makes it prone to misdiagnosis. The grading system for HFS is complex and non-quantifiable, posing a challenge for non-specialists such as general practitioners to diagnose HFS. In this paper, we presented an approach to quantify facial features and indicators for detecting and grading HFS. We collected facial videos of HFS patients and healthy adults to design and develop algorithms capable of detecting and grading HFS. Our algorithms were evaluated on the dataset, achieving an accuracy of 88% for detection and a mean absolute error of 0.42 for grading, closely aligning with the medical grading system.

Acknowledgment

This work is supported by the foundation of National Key Laboratory of Human Factors Engineering under Grant No. HENKL.2024W06, Natural Science Foundation of China under Grant No. 62472244, No. 62132010 and No. 62102221, Tsinghua University Initiative Scientific Research Program, Beijing Key Lab of Networked Multimedia, and the Beijing Municipal Health Commission under Grant No. SHOUFA-202224085. We thank all participants and Shiqi Dai for assistance in data collection.

References

[1] J. Reimer, K. Gilg, A. Karow, J. Esser, and G. Franke, "Health-related quality of life in blepharospasm or hemifacial spasm," *Acta Neurologica Scandinavica*, vol. 111, no. 1, pp. 64–70, 2005.

[2] E. Tan and L. Chan, "Young onset hemifacial spasm," *Acta neurologica scandinavica*, vol. 114, no. 1, pp. 59–62, 2006.

[3] R. G. Auger and J. P. Whisnant, "Hemifacial spasm in rochester and olmsted county, minnesota, 1960 to 1984," *Archives of Neurology*, vol. 47, no. 11, pp. 1233–1234, 1990.

[4] B. Nilsen, K.-D. Le, and E. Dietrichs, "Prevalence of hemifacial spasm in oslo, norway," *Neurology*, vol. 63, no. 8, pp. 1532–1533, 2004.

[5] C. Rosenstengel, M. Matthes, J. Baldauf, S. Fleck, and H. Schroeder, "Hemifacial spasm: conservative and surgical treatment options," *Deutsches Ärzteblatt International*, vol. 109, no. 41, p. 667, 2012.

[6] H. Miwa, Y. Mizuno, and T. Kondo, "Familial hemifacial spasm: report of cases and review of literature," *Journal of the neurological sciences*, vol. 193, no. 2, pp. 97–102, 2002.

[7] J.-A. Lee, K.-H. Kim, and K. Park, "Natural history of untreated hemifacial spasm: a study of 104 consecutive patients over 5 years," *Stereotactic and functional neurosurgery*, vol. 95, no. 1, pp. 21–25, 2017.

[8] G. Ehni and H. W. WOLTMAN, "Hemifacial spasm: review of one hundred and six cases," *Archives of Neurology & Psychiatry*, vol. 53, no. 3, pp. 205–211, 1945.

[9] K. I. Chan, Y. Wang, S. Hu, B. Hei, Z. Lu, P.-L. P. Rau, and Y. Shi, "The odyssey journey: Hemifacial spasm patients' top-tier medical resource seeking in china from an actor-network perspective," 2024.

[10] N.-C. Tan, L.-L. Chan, and E.-K. Tan, "Hemifacial spasm and involuntary facial movements," *Qjm*, vol. 95, no. 8, pp. 493–500, 2002.

[11] G. Abbruzzese, A. Berardelli, and G. Defazio, "Hemifacial spasm," *Handbook of clinical neurology*, vol. 100, pp. 675–680, 2011.

[12] R. H. Wilkins, "Hemifacial spasm: a review," *Surgical neurology*, vol. 36, no. 4, pp. 251–277, 1991.

[13] J. Greenwood, "The surgical treatment of hemifacial spasm," *Journal of Neurosurgery*, vol. 3, no. 6, pp. 506–510, 1946.

[14] N. Poungvarin, V. Devahastin, and A. Viriyavejakul, "Treatment of various movement disorders with botulinum a toxin injection: an experience of 900 patients." *Journal of the Medical Association of Thailand= Chotmaihet thangphaet*, vol. 78, no. 6, pp. 281–288, 1995.

[15] A. Wang and J. Jankovic, "Hemifacial spasm: clinical findings and treatment," *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, vol. 21, no. 12, pp. 1740–1747, 1998.

[16] F. G. Barker, P. J. Jannetta, D. J. Bissonette, P. T. Shields, M. V. Larkins, and H. D. Jho, "Microvascular decompression for hemifacial spasm," *Journal of neurosurgery*, vol. 82, no. 2, pp. 201–210, 1995.

[17] M. Montava, V. Rossi, C. C. Fais, J. Mancini, and J.-P. Lavieille, "Long-term surgical results in microvascular decompression for hemifacial spasm: efficacy, morbidity and quality of life," *Acta Otorhinolaryngologica Italica*, vol. 36, no. 3, p. 220, 2016.

[18] Y. Yuan, Y. Wang, S.-x. Zhang, L. Zhang, R. Li, and J. Guo, "Microvascular decompression in patients with hemifacial spasm: report of 1200 cases," *Chinese medical journal*, vol. 118, no. 10, pp. 833–836, 2005.

[19] M. Dannenbaum, B. C. Lega, D. Suki, R. L. Harper, and D. Yoshor, "Microvascular decompression for hemifacial spasm: long-term results from 114 operations performed without neurophysiological monitoring," *Journal of neurosurgery*, vol. 109, no. 3, pp. 410–415, 2008.

[20] M. Sindou and P. Mercier, "Microvascular decompression for hemifacial spasm: Outcome on spasm and complications. a review." *Neuro-chirurgie*, vol. 64, no. 2, pp. 106–116, 2018.

[21] K. Heuser, E. Kerty, P. Eide, M. Cvancarova, and E. Dietrichs, "Microvascular decompression for hemifacial spasm: postoperative neurologic follow-up and evaluation of life quality," *European journal of neurology*, vol. 14, no. 3, pp. 335–340, 2007.

[22] R. G. AUGER, D. G. PIEPGRAS, and E. R. LAWS Jr, "Hemifacial spasm: results of microvascular decompression of the facial nerve in 54 patients," in *Mayo Clinic Proceedings*, vol. 61, no. 8. Elsevier, 1986, pp. 640–644.

[23] T. Fukushima, "Microvascular decompression for hemifacial spasm: results in 2890 cases," *Neurovascular surgery*, pp. 1133–1145, 1955.

[24] R. Illingworth, D. Porter, and J. Jakubowski, "Hemifacial spasm: a prospective long-term follow up of 83 cases treated by microvascular decompression at two neurosurgical centres in the united kingdom." *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 60, no. 1, pp. 72–77, 1996.

[25] M. Ishikawa, T. Ohira, J. Namiki, M. Kobayashi, M. Takase, T. Kawase, and S. Toya, "Electrophysiological investigation of hemifacial spasm after microvascular decompression: F waves of the facial muscles, blink reflexes, and abnormal muscle responses," *Journal of neurosurgery*, vol. 86, no. 4, pp. 654–661, 1997.

[26] T. Iwakuma, A. Matsumoto, and N. Nakamura, "Hemifacial spasm: Comparison of three different operative procedures in 110 patients," *Journal of Neurosurgery*, vol. 57, no. 6, pp. 753–756, 1982.

[27] J. C. Shin, U. H. Chung, Y. C. Kim, and C. I. Park, "Prospective study of microvascular decompression in hemifacial spasm," *Neurosurgery*, vol. 40, no. 4, pp. 730–735, 1997.

[28] D. A. Cohen, P. J. Savino, M. B. Stern, and H. I. Hurtig, "Botulinum injection therapy for blepharospasm: a review and report of 75 patients," *Clinical neuropharmacology*, vol. 9, no. 5, pp. 415–429, 1986.

[29] O. Çeliktutan, S. Ulukaya, and B. Sankur, "A comparative study of face landmarking techniques," *EURASIP Journal on Image and Video Processing*, vol. 2013, pp. 1–27, 2013.

[30] D. Rathod, A. Vinay, S. Shylaja, and S. Natarajan, "Facial landmark localization-a literature survey," *Int J Current Eng Technol*, vol. 4, no. 3, pp. 1901–1907, 2014.

[31] Y. Wu and Q. Ji, "Facial landmark detection: A literature survey," *International Journal of Computer Vision*, vol. 127, no. 2, pp. 115–142, 2019.

[32] A. Bätz, A. D. M. Palacio, G. Mishra, I. L. Rodrıguez, M. R. Pranti, P. S. Mulet, and C. Lippert, "Facial landmarks detection: A brief chronological survey & practical implementation."

[33] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1867–1874.

[34] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[35] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3476–3483.

[36] J. Thevenot, M. B. López, and A. Hadid, "A survey on computer vision for assistive medical diagnosis from faces," *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1497–1511, 2017.

[37] J. Qiang, D. Wu, H. Du, H. Zhu, S. Chen, and H. Pan, "Review on facial-recognition-based applications in disease diagnosis," *Bioengineering*, vol. 9, no. 7, p. 273, 2022.

[38] A.-M. Tăuţan, B. Ionescu, and E. Santarnecchi, "Artificial intelligence in neurodegenerative diseases: A review of available tools with a focus on machine learning techniques," *Artificial Intelligence in Medicine*, vol. 117, p. 102081, 2021.

[39] P. Dulguerov, F. Marchal, D. Wang, and C. Gysin, "Review of objective topographic facial nerve evaluation methods," *Otology & Neurotology*, vol. 20, no. 5, pp. 672–678, 1999.

[40] T. H. Ngo, M. Seo, N. Matsushiro, and Y.-W. Chen, "Quantitative analysis of facial paralysis based on filters of concentric modulation," in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. IEEE, 2015, pp. 1758–1763.

[41] Y. Zhuang, M. McDonald, O. Uribe, X. Yin, D. Parikh, A. M. Southerland, and G. K. Rohde, "Facial weakness analysis and quantification of static images," *IEEE journal of biomedical and health informatics*, vol. 24, no. 8, pp. 2260–2267, 2020.

[42] Z. Guo, M. Shen, L. Duan, Y. Zhou, J. Xiang, H. Ding, S. Chen, O. Deussen, and G. Dan, "Deep assessment process: Objective assessment process for unilateral peripheral facial paralysis via deep convolutional neural network," in *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*. IEEE, 2017, pp. 135–138.