Artificial Intelligence-Based Facial Palsy Evaluation: A Survey

Yating Zhang, Weixiang Gao, Hui Yu, *Senior Member, IEEE*, Junyu Dong, *Member, IEEE*, and Yifan Xia

***Abstract*— Facial palsy evaluation (FPE) aims to assess facial palsy severity of patients, which plays a vital role in facial functional treatment and rehabilitation. The traditional manners of FPE are based on subjective judgment by clinicians, which may ultimately depend on individual experience. Compared with subjective and manual evaluation, objective and automated evaluation using artificial intelligence (AI) has shown great promise in improving traditional manners and recently received significant attention. The motivation of this survey paper is mainly to provide a systemic review that would guide researchers in conducting their future research work and thus make automatic FPE applicable in real-life situations. In this survey, we comprehensively review the state-of-the-art development of AI-based FPE. First, we summarize the general pipeline of FPE systems with the related background introduction. Following this pipeline, we introduce the existing public databases and give the widely used objective evaluation metrics of FPE. In addition, the preprocessing methods in FPE are described. Then, we provide an overview of selected key publications from 2008 and summarize the state-of-the-art methods of FPE that are designed based on AI techniques. Finally, we extensively discuss the current research challenges faced by FPE and provide insights about potential future directions for advancing state-of-the-art research in this field.**

***Index Terms*— Facial palsy evaluation, facial nerve function, artificial intelligence, survey.**

I. Introduction

FACIAL palsy is the most common and frequently occurring neuromuscular disorder among humans aged 15 to 50 worldwide. Despite enormous medical progress, facial palsy has still affected 11 to 40 persons per 100,000 worldwide, and its incidence has increased over the years [1], [2], [3]. The typical symptom of facial palsy is the remarked facial asymmetry. Due to trauma or acquired diseases (stroke, Bell's palsy, etc.), the patient suffers nerve damage, which makes it difficult to move the muscles of the face for some facial movements or expressions such as swallowing saliva and blinking [4], [5]. In addition to making daily life inconvenient and affecting the physical health of the patient, the functional disability or impairment from facial palsy can further drastically erode the patient's psychosocial well-being [6], [7], [8]. Facial movements or expressions are important in visual communication and contain much nonverbal information such as intent, feelings, meanings, emotions, and social interactions [9], [10]. Thus, the inability to perform facial movements or expressions and the generation of aesthetic deficits can significantly decrease the quality of a patient's daily life [11], [12], [13], which often leads to patients with negative emotions such as anxiety, depression, and low self-esteem [14].

Facial palsy evaluation (FPE) is the primary step for facial functional treatment and rehabilitation, ranging from self-resolution to surgical intervention [15], [16]. The purpose of evaluating facial palsy is to clearly understand the progression of the disease in clinical practice and provide a common language for healthcare professionals about the severity of facial palsy [17]. By using patients' facial data, it is possible to evaluate the treatment outcomes and analyze which rehabilitation plan has a higher likelihood of success or effect on the physical improvements of patients to facilitate decision-making and make an appropriate plan for enhancing the overall quality of care [18]. In addition, early evaluation of facial palsy can help people prevent the development of the disease and take proper treatment as early as possible, which can significantly reduce morbidity.

Traditionally, to conduct FPE, certain facial expressions of patients are judged in a visual manner by clinicians based on sophisticated grading scales [3], [19]. The main clinician grading scales for FPE include the House-Brackmann [20], Sunnybrook [21], eFACE [22], etc. In general, these traditional manual evaluation methods have been utilized by most hospitals and rehabilitation institutions. There is no doubt that the clinician's individual experience and perception are the most critical factors in the process of facial palsy evaluation. However, a significant issue with this traditional evaluation is the overly subjective judgments from clinicians in some cases, which might yield potential bias and low reproducible results [11], [20], [23]. Moreover, there are also other limitations, such as the time cost to train clinicians, the labour intensity to evaluate patients, and the need for

Manuscript received 5 June 2024; revised 26 July 2024; accepted 13 August 2024. Date of publication 22 August 2024; date of current version 3 September 2024. *(Corresponding author: Yifan Xia.)*

Yating Zhang, Weixiang Gao, and Yifan Xia are with the School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China (e-mail: xiayifan@sdu.edu.cn).

Hui Yu is with the cSCAN Centre, University of Glasgow, G12 8QB Glasgow, U.K. (e-mail: hui.yu@glasgow.ac.uk).

Junyu Dong is with the Faculty of Information Science and Engineering, Ocean University of China, Qingdao 266100, China (e-mail: dongjunyu@ouc.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2024.3447881

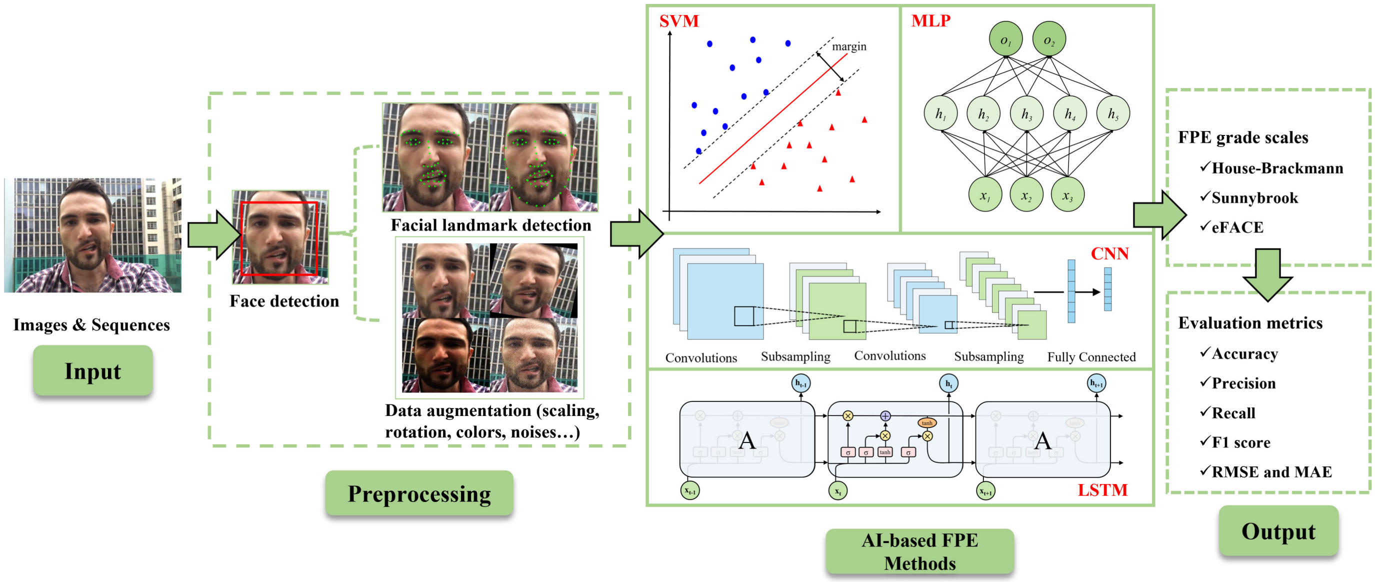


Fig. 1. The general pipeline of facial palsy evaluation systems: 1) databases, 2) data preprocessing, 3) AI-based FPE methods, and 4) evaluation metrics.

face-to-face evaluations. Over the years, disease diagnosis based on artificial intelligence (AI) has been one of the fastest-growing directions and the most promising application areas in healthcare [24], [25], [26], [27], [28], [29]. The machine with AI can mimic human intelligence and make decisions automatically like a human [30], [31], [32], [33], [34] and may cause a paradigm shift in healthcare by helping clinicians make decisions and identify diseases better and quickly [35], [36], [37], [38].

Currently, technological advances in AI have provided an automated and objective way for FPE to quantify the severity of facial palsy. Most current proposed AI-based approaches for evaluating facial palsy are based on facial images or videos. In the early period, most researchers manually computed the distance between the detected facial landmarks for FPE [39]. With the development of AI technologies, many hand-crafted features have been utilized to represent facial asymmetry and were adopted by machine learning algorithms to evaluate facial palsy severity [40]. However, since 2016, deep learning methods have increasingly been implemented in FPE by researchers because of the increase in computer hardware abilities, and have shown impressive performance [17], [41], [42], [43], [44]. In addition, some researchers have published public and relatively sufficient training data [45], [46], [47], which further promotes the transition of FPE from the period of machine learning to deep learning.

However, the FPE based on AI technologies has not been reviewed in previous survey papers, which is still an open challenge. Recently, automated facial nerve function assessment was surveyed in [3] but focused on the perspective of visual face capture. In contrast with [3], this survey conducts more specific and detailed research on the main AI-based FPE. The aim of this survey is to provide a systematic review of AI-based FPE for the researchers. We observed that most of these recent state-of-the-art studies mainly used AI techniques, more specifically, machine learning and deep learning, on the facial images of patients to assess facial palsy severity, which is because image or video data is more widely accessible in clinics or daily life than data from obtrusive physical interventions such as electromyography (EMG) and electroneuronography (ENoG) [48], [49], [50], [51]. Acquiring image or video data requires no special equipment or complex operation except for cameras. Thus, in this paper, we organise our survey from four parts involved in the general pipeline of the FPE systems based on facial images or videos (see Fig. 1): 1) databases, 2) data preprocessing, 3) AI-based FPE methods, and 4) evaluation metrics. Following this pipeline, this survey reviews the existing AI-based research from 2008 conducted to evaluate facial palsy severity. Our study not only reviews the state-of-the-art development of AI-based FPE but also discusses the existing research challenges and provides some guidelines about future directions for researchers regarding the possible application of AI in evaluating facial palsy severity.

The remainder of this paper is structured as follows. Section II introduces the existing public facial palsy databases and briefly reviews the evaluation metrics of FPE. Section III provides a summary of data preprocessing required in FPE. Section IV contains a detailed review of the state-of-the-art work on FPE based on AI techniques, including machine learning and deep learning. Section V discusses some of the research challenges in this field and Section VI identifies potential future directions. Finally, conclusions are drawn in Section VII.

II. Databases and Metrics

This section introduces the databases and evaluation metrics used in AI-based FPE methods.

A. Facial Palsy Databases

Generally, AI can be regarded as a kind of data-driven technology. In principle, some complicated machine learning or deep learning algorithms always require much training data to train a good model. Therefore, the need for large, labelled,

TABLE I
AN OVERVIEW OF THE FACIAL PALSY DATABASES

Database	Samples	Subject	Source	Facial movement distribution	Annotation
YFP [47]	32 videos	21 patients	YouTube	Random facial movements: at rest, open mouth, close the eyes lightly, elevation of eyebrows, pursing lips, etc.	Palsy facial region (eyes or mouth) in each video
MEEI [45]	60 videos and 480 images	9 healthy subjects and 51 patients	Lab	8 facial movements: face at rest, eyebrows raised, gentle eye closure, forceful eye closure, gentle smile, forceful smile, pucker, and showing bottom teeth.	Each subject was labelled with an eFACE score
AFLFP [46]	5,632 images	88 subjects	Lab	16 facial movements: brow raise, close smile, frown, funny, gentle eye closure, left eyebrow, left smile, left snarl, left wink, open smile, right eyebrow, right smile, right snarl, right wink, snarl, and tight eye closure.	Each image was labelled with 68 facial landmarks

available databases for training, evaluating, and benchmarking has been widely acknowledged. This section only introduces the existing public databases used in our reviewed papers for AI-based FPE. Since facial images or videos are the most commonly used data in FPE, the scope of this survey is restricted to automatic FPE using 2D facial images or videos. Fig. 2 exhibits some examples of facial palsy images from these public databases. For completeness, we also provide a summary of existing public facial palsy databases in Table I, which provides an overview of these databases.

1) YFP: The YouTube Facial Palsy (YFP) [47] database is the most extensively used unconstrained database in FPE systems, which contains 32 YouTube videos from 21 patients. Some patients have multiple videos in the YFP database. Each video records the random facial movements and is converted into an image sequence with 6FPS. In YFP, three independent clinicians manually labelled the palsy facial regions, eyes, or mouth when the deformation intensity was considered sufficiently high.

2) MEEI: The Massachusetts Eye and Ear Infirmary (MEEI) [45] database is a laboratory-controlled facial palsy dataset with facial videos and images. The database is composed of 60 videos from 9 healthy subjects and 51 patients, and each video contains 8 different facial movements. In addition, this database provides images of 8 facial movements for each participant, which contains 480 high-resolution images in total. In this database, subjects were categorized by eFACE [22] score: normal (96–100), near-normal (91–95), mild (80–90), moderate (70–79), severe (60–69), and complete flaccid or nonflaccid facial palsy (< 60).

3) AFLFP: The Annotated Facial Landmarks for Facial Palsy (AFLFP) [46] database is a diverse and laboratory-controlled database that contains facial images from 88 subjects. For each subject, the AFLFP database collects 16 expression videos (such as brow raise, close smile, gentle eye closure, open smile, etc.). The database contains keyframes of each facial expression video to express the four key states (neutral, onset, a mid-state between onset and peak, and peak). This database contains 5,632 facial images, 1,408 samples for each key state, and 64 samples for each subject. Each facial image was independently and manually annotated with 68 facial landmarks.

B. Evaluation Metrics

To achieve the AI-based evaluation of facial palsy, a clinical practitioner must evaluate and label the facial data. Various evaluation grade scales have been used to evaluate facial palsy severity for more uniform and accurate results. These scales divide the degree of facial nerve damage into discrete levels or continuous scores based on rigorously validated measures. The House-Brackmann (HB) [20] is the widely used discrete grade scale for evaluating facial palsy, which grades facial palsy into six discrete classes from normal to total paralysis according to the functional performance of facial muscles (see Table II). In addition, the discrete grade scales could be a binary value indicating whether the subject has facial palsy [52], [53], several customized categories of severity [17], [44], or if a specific face region is paralyzed [47]. Compared with discrete grade scales, continuous evaluation grade scales such as Sunnybrook [21] and eFACE [22] comprehensively consider factors such as the symmetry of the patient's face in a stationary state, symmetry during movement, coordinated movement, and movement amplitude. Continuous values ranging from 0 to 100 are used to score the severity of facial palsy, with higher scores indicating better facial nerve function. From the perspective of AI, the task is called classification when the prediction assigns the facial data of patients into the predefined discrete classes of facial palsy severity. But when the prediction output is continuous, the task is called regression. Therefore, according to the evaluation results of facial palsy severity, there are two categories of evaluation metrics: classification metrics and regression metrics.

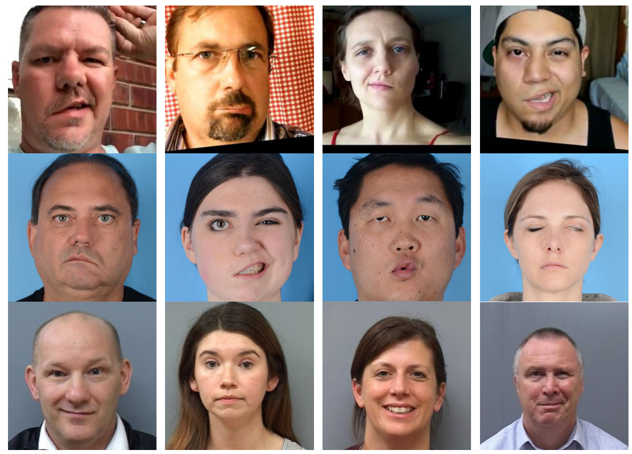


Fig. 2. Sample images with facial palsy from different databases. The images of per row from YFP [47], MEEI [45], AFLFP [46].

For the classification task, the performance is usually reported using four distinct accuracy metrics (the bigger is

TABLE II
HOUSE-BRACKMANN FACIAL PALSY EVALUATION SCALE

Grade	Severity	Gross description	At rest	Motion		
				Forehead	Eye	Mouth
1	Normal	Normal facial function in all facial areas	Symmetry	Normal facial function		
2	Mild dysfunction	Slight weakness noticeable on close inspection, may have very slight synkinesis	Normal symmetry	Moderate to good function	Complete closure with minimum effort	Slight asymmetry
3	Moderate dysfunction	Obvious but not disfiguring difference between the two sides, noticeable but not severe synkinesis	Normal symmetry	Slight to moderate movement	Complete closure with effort	Slightly weak with maximum effort
4	Moderately severe dysfunction	Obvious weakness and/or disfiguring asymmetry	Normal symmetry	None	Incomplete closure	Asymmetric with maximum effort
5	Severe dysfunction	Only barely perceptible motion	Asymmetry	None	Incomplete closure	Slight movement
6	Total paralysis	No perceptible motion	Asymmetry	No movement		

better), including accuracy, precision, recall, and F1 score, which are formulated as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively. In general, to evaluate the performance of a given AI model, multiple evaluation metrics are usually used because of the presence of some imbalanced classes in the database [54].

For the regression task, the performance is usually reported using two error metrics (the smaller is better), including the root mean square error (RMSE) and mean absolute error (MAE). These metrics are formulated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - g_i)^2} \tag{5}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|p_i - g_i| \tag{6}$$

where N is the number of samples in the evaluation set; p_i and g_i are the prediction and ground truth of the i th sample.

III. DATA PREPROCESSING

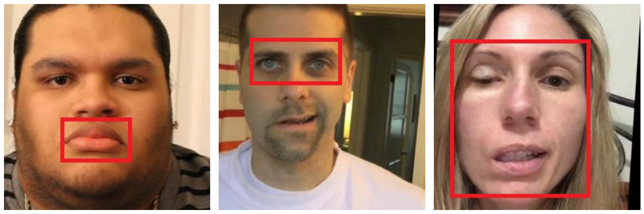
In this section, we briefly summarize the common data preprocessing steps and their methods in AI-based FPE according to the reviewed papers, which can improve the data quality and are beneficial to the utilization of the semantic information conveyed by the face during training of the AI model.

A. Face Detection

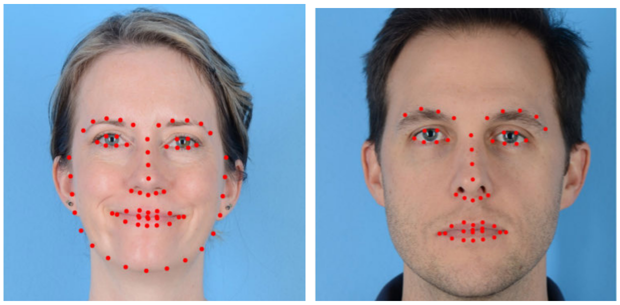
Given a series of training data, detecting the whole face or facial regions is usually the first procedure for FPE [55], [56], which aims to remove background and non-face areas. Fig. 3 (a) exhibits some examples of face or facial region detection. A classic implementation for face detection is the Viola & Jones (V&J) detector [57], which has been widely used for detecting near-frontal faces in FPE [53], [58], [59] because of its robustness and computational simplicity. Other FPE methods [4], [60], [61], [62], [63] utilize the Dlib library [64], another popular open-source face detector. Some recent FPE works have successfully employed object detection algorithms, such as Faster R-CNN and YOLO, to perform face or facial region detection [42], [43], [47], [65], [66], [67], [68]. For a method with higher detection performance and lower computational cost, see a recent survey [69] for the state-of-the-art development of face detection.

B. Facial Landmark Detection

After face detection, facial landmark detection is an indispensable procedure and plays a vital role in FPE to enhance performance substantially. Currently, most existing FPE methods represent asymmetry and shape features of facial images based on 68 or 49 facial landmarks (as shown in Fig. 3 (b) and (c)). Moreover, accurate facial landmarks can allow for better facial region detection and focus on the key facial regions for FPE, thus reducing interference from unrelated facial areas and improving performance [17], [42], [43]. Many research studies have performed facial landmark detection before FPE [4], [43], [52], [58], [63], [70], [71], [72], [73], [74], [75], [76], [77], [78]. Among these works, the most widely used facial landmark detection methods in this field are the active shape model (ASM) [79], active appearance model (AAM) [80], supervised descent method (SDM) [81], iPAR-CLR [82] and the ensemble of regression trees (ERT) [83], which usually trained on the normal facial database such as LFPW [84], HELEN [85], AFW [86], and 300-W [87]. In addition, other works [17], [62], [88], [89], [90], [91], [92], [93] used open-source tools for facial landmark detection, such as Dlib [64], IntraFace [94], Emotrics [95], and Auto-eFACE [96], which are also widely used in FPE. Recently, some researchers have used deep learning-based methods of facial landmark detection in FPE [42], [44], [47], [61], [66],



(a) The detection results of the facial regions or whole face



(b) 68 facial landmarks (c) 49 facial landmarks

Fig. 3. Some face detection and facial landmark detection examples.

[67], [97], [98], [99], such as face alignment network (FAN), and demonstrated higher localization accuracy than traditional methods. Landmark detection algorithms are trained using real patients rather than healthy people, which may perform better [46], [75], [77]. Please see the survey [100] for a more exhaustive and state-of-the-art development of facial landmark detection.

C. Data Augmentation

The models of AI generally require sufficient training data to ensure better performance and generalization. If there are small volumes of data, it may run the risk of overfitting. However, most publicly available databases for facial palsy do not have enough images for training, and the data imbalance commonly appears in most existing databases. Some works [17], [42], [47], [62], [66], [67], [71], [92], [97], [101] incorporated images from facial expression databases, such as the CK+ database [102], to balance healthy and paralyzed facial images. The more common strategy is data augmentation, which can help overcome the problem caused by datasets with imbalanced classes or small datasets. The idea of sample augmentation is to increase the variability of the actual sample dataset by altering the existing samples in a preset way to simulate variability naturally encountered. Therefore, data augmentation is a vital step for preparing a high-quality dataset for training an accurate FPE model, which ensures an equal number of images in each facial palsy class, increases the diversity of the training data, and improves the model's generalization ability. As shown in Fig. 4, the widely used data augmentation operations during the training step in FPE [17], [41], [58], [59], [61], [62], [63], [67], [78], [101], [103], [104], [105] include random flipping, adding noise, random scaling, random shearing, etc. Combinations of multiple operations can generate more unseen training samples, which can result in a database that is much larger than the original training data. Please see [106] for more details about data augmentation.

Fig. 4. Illustration of the dataset augmentation over a sample image. The images of each row from left to right: raw image, rotating, adding noise, cropping, shearing, flipping, brightness, and scaling.

IV. AI-Based Facial Palsy Evaluation

FPE has witnessed promising progress in recent years with the development of AI techniques, which provide a highly efficient and cost-effective means to quantify facial palsy severity automatically and objectively. The evolution process of AI-based FPE is shown in Fig. 5. In this section, according to the AI methods used in the literature, we broadly divide the current FPE approaches into two main categories: machine learning approaches and deep learning approaches, and provide an overview of the FPE approaches in these categories.

A. Machine Learning Approaches for FPE

Machine Learning (ML), as a branch of AI, can build a predictive model trained on labelled image data to predict the facial palsy severity grade on new data. Recent studies have shown significant advancements in the field of FPE using machine learning techniques. For the machine learning approaches, the facial image data of patients with facial palsy are firstly preprocessed, and then key features aiding in the identification of facial palsy are extracted. These hand-crafted features are subsequently utilized in general machine learning algorithms, such as support vector machine (SVM), multilayer perceptron (MLP), random forests (RF), etc., to assess the facial palsy by mapping the features to the facial palsy severity grades [107], [108]. Therefore, according to feature types, we categorize existing machine learning approaches for FPE into two main categories: appearance feature-based approaches and geometric feature-based approaches.

1) Appearance Feature-based Approaches: The appearance features, including facial area texture and colour, are the most commonly used in FPE. They primarily focus on the movement capability and symmetry of facial muscles, providing intuitive assessment indicators that can quickly reflect the severity of facial palsy and the extent of facial nerve function damage in patients. LBP is an operator commonly used to describe local texture features in images and is the most frequently used feature extraction method in facial palsy evaluation systems. Many works successfully used partitioned LBP features to evaluate facial palsy (e.g., [40], [71], [73], [99], [109], [110]). For example, He et al. [40] extended LBP to the spatiotemporal domain and used the similarity between multiresolution LBP (MLBP) features on the left and right sides of the face to evaluate facial palsy in image sequences. Other works [109], [110], [111], [112] used filter features

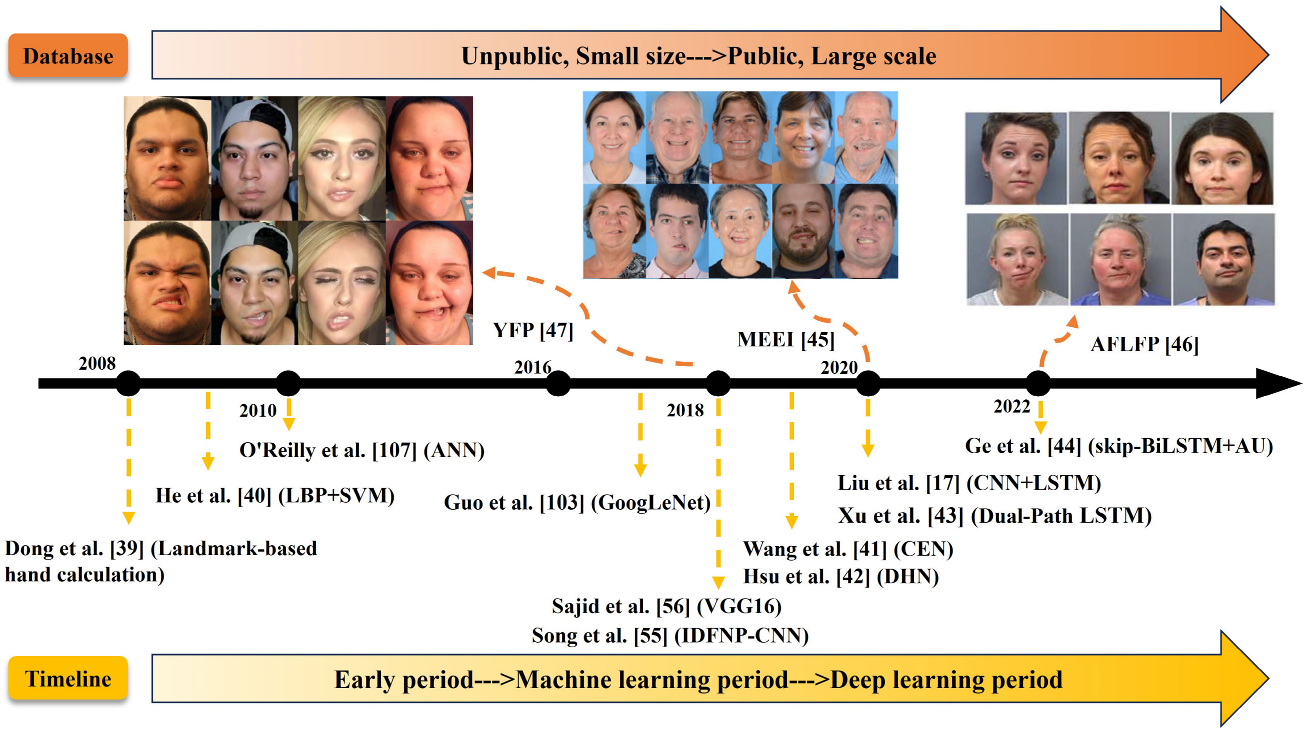


Fig. 5. The evolution of facial palsy evaluation. From the perspective of the database, there is a shift from an unpublic and small size to a public and large scale. In terms of methods, the development of facial palsy evaluation can be divided into three periods: the early period, the machine learning period, and the deep learning period.

such as Gabor and wavelet decomposition to extract specific frequency or texture information from facial images to analyze facial nerve function abnormalities.

Some works also tried to combine the above two appearance features to promote the overall improvement of FPE. In [110], a method was proposed using Gabor filters to remove noise and redundant information from LBP images of face regions (e.g., eyebrows, eyes, nose, and mouth) of facial palsy patients. Similarly, Ngo et al. [109] compared traditional methods' facial palsy recognition rates based on single appearance features with methods based on LBP images processed with Gabor filters (GBLBP) and LBP images decomposed by wavelet transformation (WLLBP). GBLBP and WLLPB exhibited higher evaluation rates for facial palsy than previous traditional methods. However, the effectiveness of GBLBP and WLLPB varies in detecting facial palsy during different facial movements. For instance, GBLBP performs better in expressing eyebrow movements, while WLLBP performs better in expressions involving tightly closed eyes and teeth movements.

Additionally, in terms of extraction strategies, most research works extracted the appearance features of local facial regions (e.g., [40], [71], [73], [99], [109], [110], [111], [112]). Typical local feature extraction involves dividing the face into non-overlapping small segments and applying feature descriptors to each segment to obtain block-based local features. The local features can be used directly to assess facial palsy, or all local features can be concatenated to form a feature vector. For example, Li et al. [99] cropped several segments from each frame of the video, extracted the LBP features of all segments, and concatenated them into a single local area motion feature vector to describe the dynamic changes in the local area. For FPE, local features can better capture key facial information and focus on detailed descriptions of specific facial areas.

2) Geometric Feature-based Approaches: The calculation of geometric features is based on a set of facial reference points known as landmarks, which capture statistics derived from the location of facial landmarks. The machine learning methods based on geometric features first detect the facial area in the image and then extract facial landmarks to calculate other metrics such as distances, angles, and areas between facial landmarks [93], [113]. In the process of facial palsy evaluation, these landmark-based measurements are input into classifiers or regressors for training to determine the degree of facial asymmetry and facial nerve dysfunction.

Researchers have delved into the facial characteristics of patients with facial palsy and designed various geometric features based on facial landmarks that reflect facial asymmetry. The most common geometric feature analysis involves measuring the distances between key facial landmarks, which is used as a key basis for classifying the severity of facial palsy [114]. In [115], Anguraj et al. proposed a method that uses the salient point selection algorithm (SPSA) to calculate the distances between salient points in different expression states and assesses the degree of facial palsy by comparing the ratio of distances between salient points on the affected and unaffected sides. Arora et al. [91] extracted 10 distance-related landmark features and identified the three most important features in facial palsy detection. These features were then input into SVM and Logistic Regression models to identify facial palsy.

In addition to measuring the straight-line distances between facial landmarks, researchers have developed a series of landmark features based on the full face range, including angles

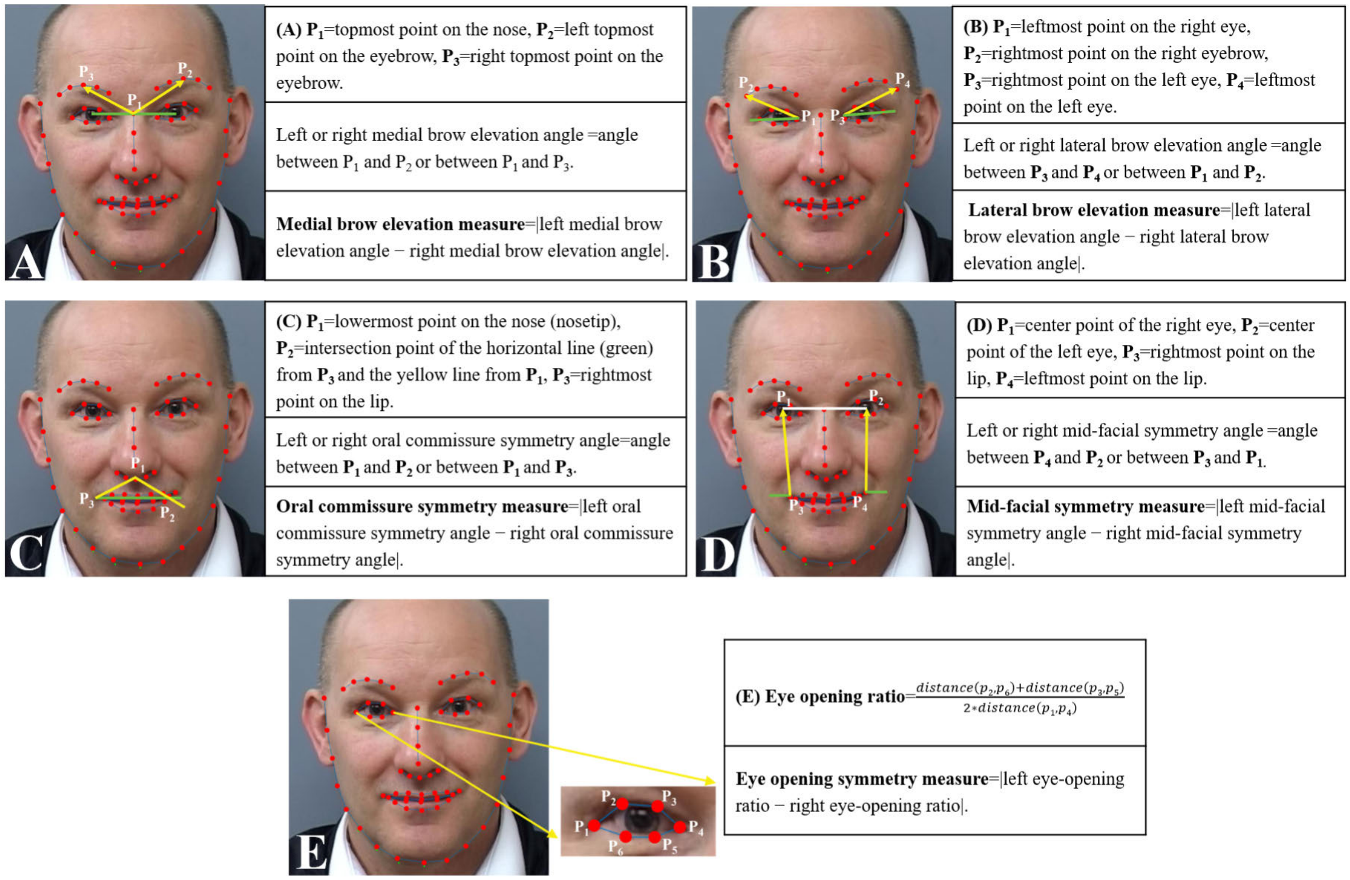
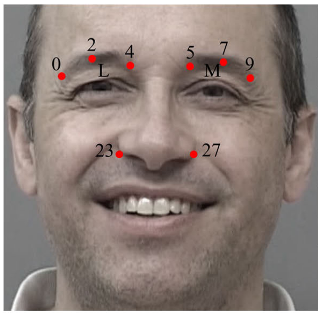


Fig. 6. The five landmark features proposed in [90] for FPE are utilized to derive the differences of the left–right (L-R) parts of the face for eye opening symmetry ratios, brow elevation, mid-facial symmetry, and oral commissure symmetry.

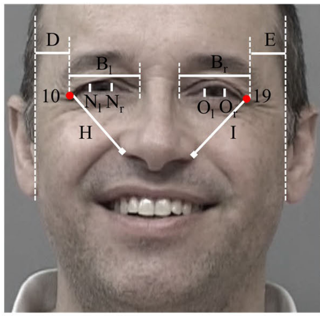
between facial landmarks, ratios of distances, etc., which are crucial for revealing the asymmetry of facial muscles [58], [116]. The L-FAM method was proposed in [90], which includes five landmark features, as shown in Fig. 6. These features were used to measure the differences in eye-opening symmetry ratio, eyebrow height, facial midline symmetry, and oral connection symmetry between the left and right sides of the face, making it more convenient to repeat measurements of facial asymmetry and severity. In [4], researchers quantified the degree of asymmetry between the two sides of the face through a set of landmark-based features, which used simple mathematical operations to measure the extracted facial landmarks and could achieve the binary classification of healthy and facial palsy states without performing specific facial movements. An efficient classification was achieved by calculating 29 symmetric measurements (as shown in Fig. 7), such as angles, inclinations, and distance ratios between facial landmarks, and inputting these features into the MLP classifier.

In general, full-face landmark features cover the entire face, providing more comprehensive facial information, but may require more data and computational resources; regional features typically require less data volume because they only focus on specific facial areas, which may make model training and computation more efficient, but may not capture overall facial changes comprehensively. For example, in [58], the researchers divided the face into key areas such as eyebrows, eyes, nose, and mouth using landmarks and extracted features highly related to the HB scoring system. These features were then input into the SVM model for classifying patients with facial palsy. To explore the different effects and usage scenarios of full-face and regional features in facial palsy evaluation, Parra-Dominguez et al. [62] further refined the facial features based on [4], dividing the facial features into four facial areas (eyebrows, eyes, nose, and mouth). The study comprehensively analyzed full-face features and compared the classification effects based on features from each region. This study indicates that although full-face analysis has a slight advantage in accuracy, regional features are also effective in facial palsy detection, especially when dealing with images with partial occlusion, still achieving satisfactory results. Considering the landmarks around the eyes are particularly useful for measuring the severity of the palsy, Barbosa et al. [53] used iris segmentation and key point detection based on localized active contour (LAC) to measure facial symmetry by the ratio of iris area to the vertical distance between two facial landmarks on the face, combined with a hybrid classifier based on rule-based and regularized logistic regression to analyze and predict the type and severity of facial palsy.

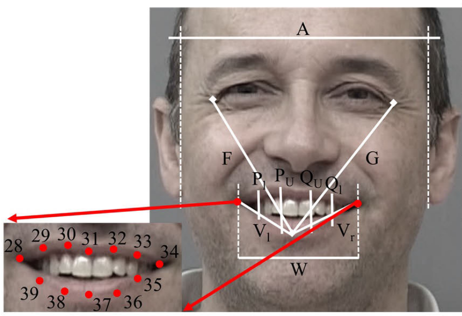
3) System Applications: In practical applications, researchers have developed a series of integrated automatic diagnostic systems using machine learning models to evaluate patients with facial palsy efficiently. In [117], a smartphone-based application was developed to track facial landmarks in real-time, complete video capture, and apply grading



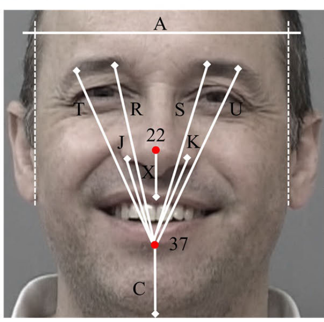
Feature	Face Element	Distance	Description	Formula	Equation
f_0	Eyebrows	**L,M** : Height average of the 5 points within each eyebrow.	Angle	$\lvert\angle(P0,P9)\rvert$	**Angle:** $\angle(Pa,Pb) = arctan2(\Delta x, \Delta y) \times 180/\pi$, $\Delta x = Pa_x - Pb_x, \Delta y = Pa_y - Pb_y$. **Slope:** $m(Pa,Pb) = \lvert\frac{Pa_y - Pb_y}{Pa_x - Pb_x}\rvert$
f_1			Angle	$\lvert\angle(P2,P7)\rvert$	
f_2			Angle	$\lvert\angle(P4,P5)\rvert$	
f_3			Maximum value	max(L/M, M/L)	
f_4			Slope	m(P0,P9)	
f_5			Slope	m(P2,P7)	
f_6			Slope	m(P4,P5)	
f_{22}	Nose		Angle	$\lvert\angle(P23,P27)\rvert$	



Feature	Face Element	Distance	Description	Formula	Equation
f_7	Eyes	**B_l,B_r** : Width of the eye(left and right). **D,E** : Distance from the outer eye corner to the head side. **H,I** : Distance from the outer eye corner to the nose edge point. **N_l,N_r,O_l,O_r** : Distance between the upper and lower eyelids (left and right) of both eyes. **N** : Average height of the eye(N_l+N_r)/2 **O** : Average height of the eye(O_l+O_r)/2	Angle	$\lvert\angle(P10,P19)\rvert$	**Euclidean distance:** $d(Pa,Pb) = \sqrt{(Pa_x - Pb_x)^2 + (Pa_y - Pb_y)^2}$
f_8			Maximum value	max(B_l/B_r,B_r/B_l)	
f_9			Maximum value	max(D/E,E/D)	
f_{10}			Maximum value	max(H/I,I/H)	
f_{11}			Maximum value	max(N/O,O/N)	
f_{12}			Maximum value	max(N_l/O_r,O_r/N_l)	
f_{13}			Maximum value	max(N_r/O_l,O_l/N_r)	



Feature	Face Element	Distance	Description	Formula	Equation
f_{14}	Mouth	**F,G** : Distance from the outer eye corner to the middle mouth point. **P_l,P_u,Q_l,Q_u** : Distance between the upper and lower lip of both sides of the mouth. **V_l,V_r** : Distance from the mouth corner to the outer middle point of the mouth(left and right) . **A** : Distance between the first and last jaw points. **W** : Distance from the mouth corners.	Angle	$\lvert\angle(P28,P34)\rvert$	$\boldsymbol{W_l} = \overline{S}(P28, P29, P30, P31, P37, P38, P39)$, $\boldsymbol{W_r} = \overline{S}(P31, P32, P33, P34, P35, P36, P37)$, **perimeter of a close segment:** $\overline{S}(P_s, \ldots, P_l) = \sum_{x=s}^{l-1} d(p_x, p_x + 1) + d(p_s, p_l)$
f_{15}			Maximum value	max(F/G, G/F)	
f_{16}			Maximum value	max(P_l/Q_l,Q_l/P_l)	
f_{17}			Maximum value	max(P_u/Q_u,Q_u/P_u)	
f_{18}			Maximum value	max(V_l/A,V_r/A)	
f_{19}			Maximum value	max(P_l/W,Q_l/W)	
f_{20}			Maximum value	max(P_u/W,Q_u/W)	
f_{21}			Maximum value	max(W_l/W,W_r/W)	



Feature	Face Element	Distance	Description	Formula	Equation
f_{23}	Combined	**J,K** : Distance from the nose edge point to the middle mouth point. **T,U** : Distance from the third eyebrow point to the middle of the jaw point(left and right). **R,S** : Distance from the fourth eyebrow point to the middle of the jaw point(left and right). **C**: Distance from the middle mouth point to the middle of the jaw point. **X**: Distance from the middle nose tip to the outer middle point of the mouth.	Angle	$\lvert\angle(P22,P37)\rvert$	Same as above.
f_{24}			Maximum value	max(J/K, K/J)	
f_{25}			Maximum value	max(T/A, U/A)	
f_{26}			Maximum value	max(R/A, S/A)	
f_{27}			Ratio	C/A	
f_{28}			Ratio	X/A	

Fig. 7. The proposed 29 symmetric measurements in [4] for FPE include angles, inclinations, and distance ratios between facial landmarks.

algorithms under user interface guidance. The researchers set up 13 facial landmarks and selected 99 measurement values, such as triangle area and linear distance. Time difference vectors were extracted from each measurement value to quantify facial asymmetry. Similarly, Kim et al. [52] proposed an automatic diagnostic system based on a smartphone, which distinguishes patients with facial palsy from normal individuals by analyzing three types of facial movements (i.e., rest, smile, and raise eyebrows). The system calculates the displacement ratio of landmarks in the forehead and mouth areas to quantify the asymmetry index and uses SVM to classify healthy and facial palsy subjects. For computer platforms, Miller et al. [96] further developed an automated facial palsy evaluation tool called Auto-eFACE based on Emotrics [95], which can distinguish between normal faces, flaccid facial palsy, and severe facial palsy. In [77], the authors introduced a new evaluation tool called the automatic facial evaluation system (AFES), which uses 68 facial landmarks on the face to select and segment static and dynamic features. Specifically, the system extracts 32 static parameters that describe facial symmetry for static features, divides the face into six key areas, and calculates the optical flow difference between the left and right symmetric parts of these areas using the Horn-Schunck method. Finally, the system integrates static and dynamic feature data for comprehensive analysis and evaluates it using SVM to obtain a comprehensive facial function score.

B. Deep Learning Approaches for FPE

Deep learning, a rapidly developing subset of machine learning, has shown significant advantages over traditional

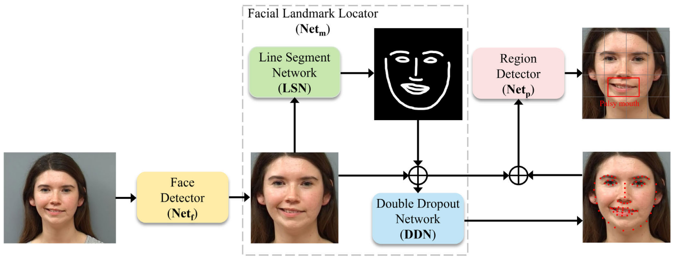


Fig. 8. The proposed Deep Hybrid Network (DHN) in [42] is composed of a face detector, a facial landmark locator, and a palsy region detector.

machine learning in evaluating facial palsy, bringing revolutionary advances in diagnosing and treating facial palsy. Traditional machine learning methods often use hand-crafted features that usually require design and selection, which present certain limitations. Deep learning, especially convolutional neural networks (CNNs), can deeply mine facial feature information and automatically learn unique feature representations of paralyzed patients from raw data, significantly improving diagnostic accuracy and providing more precise support for the treatment strategies of facial palsy. In facial palsy evaluation systems, commonly used deep learning models include VGGNet [118], GoogLeNet [119], ResNet [120], etc. In the existing literature, based on the different types of input data, deep learning-based FPE methods can be divided into two main categories: single-frame based approaches and multi-frame based approaches.

1) Single-frame Based Approaches: In the FPE systems based on single-frame images, the application of deep learning techniques is mainly reflected in two aspects: (a) Deep learning methods are used to extract facial features, which are used to assist in the in-depth evaluation of facial palsy; (b) By constructing end-to-end neural network models, the system can directly automate the evaluation of facial palsy.

For one thing, deep learning technology is extensively applied in detecting facial information, significantly improving the accuracy of facial information extraction and paving the way for subsequent relevant computations in FPE [59], [74], [97]. In this process, deep models such as deep convolutional neural networks (DCNNs) can automatically identify facial landmarks or extract facial features of the face or other key regions such as eyes, eyebrows, mouth, etc. These detected facial landmarks or extracted features can be combined with machine learning methods to evaluate facial palsy severity comprehensively. The study in [121] proposed an improved ResNet model to identify acupuncture points on the face, achieving quantitative analysis and grading of facial nerve damage by measuring the angle between the connecting lines of acupuncture points and the vertical lines of facial left-right segmentation. Li et al. [65] used Faster-RCNN as an object detection network to identify the nasolabial area in face images, followed by semantic segmentation using a global convolutional network (GCN) to extract nasolabial folds accurately. By analyzing the length, depth, and direction of these folds, the researchers establish a quantitative relationship between the asymmetry of nasolabial folds and the severity of facial palsy. Raj et al. [60] used a pre-trained CNN to extract deep features from images, which were then used to train a support vector regression (SVR) to automatically and objectively predict facial palsy grading indices. In another study [122], CNNs are used to detect facial landmarks accurately to quantify facial expression asymmetry.

For another, deep learning technology is also utilized to design end-to-end network models, which can automatically learn and extract features from input single-frame images and then directly output the evaluation results of facial palsy [92], [105]. Guo et al. [103] used a fine-tuned GoogLeNet model with transfer learning and data augmentation techniques to develop an end-to-end recognition method of the degree of unilateral peripheral facial palsy. Similarly, the authors in [55] combined GoogLeNet Inception V3 with DeepID to design a new network structure called Inception-DeepID-FNP (IDFNP) for FPE. This hybrid method implemented the evaluation of facial palsy by pre-training the IDFNP on ImageNet without the final classification layer and retraining it using the authors' dataset. Sajid et al. [56] proposed an improved CNN structure that automatically classifies facial palsy based on the HB scale. The method first applies preprocessed facial images and their mirrors to two independent CNNs, obtaining maximum pooling mappings of input facial images and their mirrors, and then inputs these maximum pooling mappings into the palsy grading structure composed of the pre-trained VGG16 network to finally classify facial palsy into 5 levels of facial asymmetry.

In addition, some works identify local facial palsy regions using target detection algorithms [123]. For instance, in [47], researchers proposed a hierarchical detection network (HDN) consisting of three components for facial palsy detection. The first component, based on YOLO-9000, is designed for detecting faces in images; the second component conducts landmark detection; the final component, based on the Darknet architecture, reduces the number of convolutional layers to optimize processing speed, effectively locating facial palsy regions. Furthermore, in a subsequent study [42], they further proposed a deep hierarchical network (DHN), building upon the HDN model, as shown in Fig. 8. The DHN network, also comprising three components, differs from [47] in that the second component connects facial landmarks according to facial structure, generating binary images of facial segments. This method enhances the accuracy of facial landmark and palsy area detection, capturing changes in facial palsy intensity over time and providing a new perspective for quantitative analysis of facial palsy syndrome.

Another interesting idea is to combine semantic segmentation networks to enhance diagnostic accuracy and efficiency in FPE [124]. For example, in the work of [41], a cascaded encoder network architecture was employed to automate the diagnosis of facial palsy through two main stages, as shown in Fig. 9. The first stage is a facial attribute semantic segmentation network, utilizing a fully convolutional network enhanced with multiscale attention modules to extract facial spatial information. The second stage is a facial palsy evaluation network, which predicts the HB level of facial palsy using a pre-trained VGG16 network structure based on the facial spatial information extracted in the first stage.

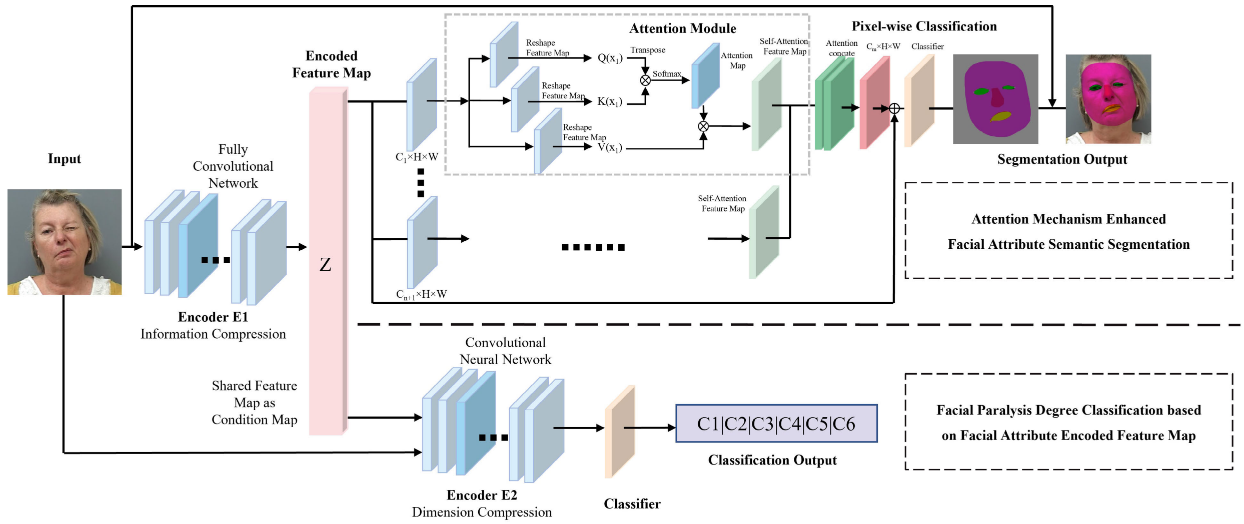


Fig. 9. The proposed cascaded encoder network architecture in [41]. The architecture comprises a facial attribute semantic segmentation network and a facial paralysis evaluation network. The two components share the same encoded feature map extracted by E1. This shared feature map contains compressed spatial information on facial attributes. Please note that E1 indicates information compression, while E2 refers to the spatial dimension reduction.

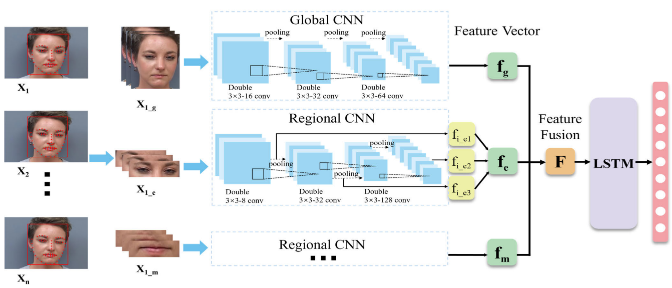


Fig. 10. The proposed parallel hierarchy convolutional neural network (PHCNN) framework in [17].

2) Multi-frame Based Approaches: The above methods based on deep learning only target static single-frame facial images by considering the facial asymmetry but ignore facial dynamic changes, which are crucial for in-depth analysis of facial palsy. Since facial palsy is dynamic, analysing still frames is less desirable for diagnosis and analysis. To accurately capture the dynamic characteristics of the face, researchers have started to adopt methods combining multi-frame video data to assess facial palsy (e.g., [17], [43], [44], [67], [68], [78]). The general idea of these studies is to input image sequences of the entire face or specific facial palsy evaluation areas into the designed network model. The model can obtain dynamic features of the whole face or the local regions through the network and then effectively integrate these features to achieve the final classification judgment. The multi-frame based methods improve the accuracy of facial palsy evaluation and provide new perspectives for research in related fields.

The most commonly used network architecture in multi-frame based approaches for FPE is the long short-term memory (LSTM). In [17], Liu et al. proposed a region-based parallel hierarchy convolutional neural network (PHCNN) combined with the LSTM network structure. The network structure of this method is shown in Fig. 10. The detected landmarks are first used to divide the eyebrows and mouth into two regions of interest (ROIs) for data preprocessing. Then, the global facial image and the two ROI images are sent to different sub-network branches. One sub-network is responsible for extracting the basic contours of facial organs and the asymmetry differences between the two sides of the face from the global facial image; the other two sub-networks focus on learning hierarchical features of the two ROIs, covering basic shapes and textures to higher-level abstract semantic information. Then, the obtained global and regional feature vectors are fused to form the static feature vector for each frame in the sequence. The static features of consecutive frames are concatenated into a feature sequence, which is input into the LSTM network to capture temporal dynamic changes for FPE. This method integrates global and local features extracted from facial palsy, capturing richer dynamic facial information and effectively reducing the impact of non-critical factors such as age, wrinkles, and changes in the shape and position of facial organs on feature learning. Furthermore, employing LSTM can capture temporal dynamic features in image sequences, significantly enhancing the accuracy of facial palsy diagnosis.

Also, using the LSTM network structure, Xu et al. [43] proposed a dual-path LSTM network based on a deep differential network (DP-LSTM-DDN), aiming to capture both global and local facial motion features simultaneously. The network structure of this method is shown in Fig. 11. In this study, the detected facial landmarks are first used to calculate the facial midline to achieve symmetry separation of the face and related areas. Then, the deep differential network (DDN) is used to extract static difference features between the symmetrical sides of the face or specific diagnostic regions. The DDN consists of two main parts: the first part is the bifurcated convolutional neural network (BCNN), which extracts features from image pairs, generates feature map pairs, and calculates depth difference information through the semi-global matching algorithm

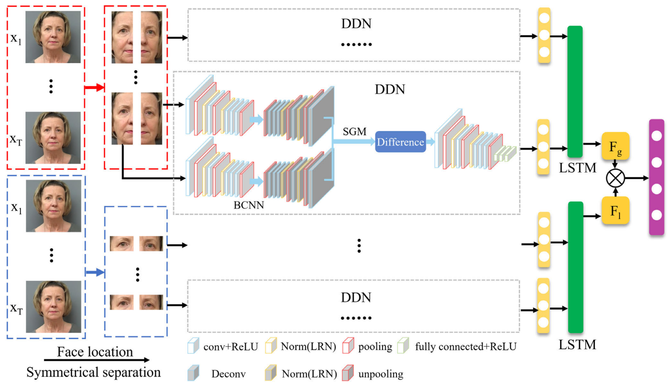


Fig. 11. The overall architecture of the proposed Adaptive Local-Global Relational Network (ALGRNet) for facial AU detection in [44].

(SGM). In the second part, these differential pieces of information are input into the CNN to obtain differential features from both sides of the face and related areas. These features are then sent to the LSTM network to learn their dynamic features. The DP-LSTM-DDN contains two LSTM paths, one focusing on learning global temporal features and the other on extracting local temporal features from areas involved in specific facial diagnostic actions. After feature extraction, global and local features are fused into a comprehensive feature vector for FPE.

Another solution is to use 3D CNN to replace 2D CNN, which can learn richer dynamic and spatiotemporal features in video data, thereby improving the recognition accuracy of FPE. In the study [67], the 3DPalsyNet model was proposed, specifically designed to recognize mouth movements and grade the severity of facial palsy according to the HB scale. 3DPalsyNet consists of two key stages: In the first stage, the input video data is preprocessed using an integrated deep model (IDM) for face detection and landmark localization, followed by cropping the images to retain the facial area and standardizing the number of frames in the video sequence to a fixed length. The second stage involves two 3D CNNs, one responsible for analyzing mouth movements and the other for assessing the degree of palsy. During the model training process, the researchers employed a joint supervision learning strategy using the softmax loss function and centre loss function, with the introduction of centre loss helping to improve the distribution between categories and the compactness within categories. Additionally, the researchers utilized transfer learning techniques, initially pretraining the model on the Kinetics dataset to complete action recognition tasks and transferring the model to the facial palsy dataset to fulfil the requirements for mouth movement analysis and facial palsy scoring.

Recently, studies have explored using action unit (AU) detection technology for facial palsy evaluation [125]. Each AU corresponds to a specific facial muscle activity, can be individually recognized, and shapes various facial expressions together. Similarly, the severity of facial palsy can also be estimated through features of facial muscle areas akin to the representation of facial expressions. In the study [44], a novel model named adaptive local-global relational network (ALGRNet) was designed, as shown in Fig. 12., for detecting AU and assessing the severity of facial palsy, which is the first time AU detection technology has been adopted in the field of facial palsy evaluation, providing a new perspective for the diagnosis and treatment of facial palsy. Specifically, the study uses a multi-branch network for AU detection, considering the information transfer between different branches and the diversity and individual differences in expression, and designed three key modules: an adaptive region learning module, a skip-BiLSTM module, and a feature Fusion&Refining module. The model extracts grid-based global features from the trunk network composed of multiple convolutional layers and extracts local AU features from the computational area based on the detected AU centres. It combines deep features of muscle areas closely related to facial palsy evaluation and global facial information to generate the final facial features for facial palsy grade classification.

3) System Applications: Numerous studies based on deep learning models have successfully developed a range of automatic scoring systems for evaluating facial palsy patient images [126]. In [127] and [128], the authors developed an automated scoring system designed for facial palsy patients, analyzing nine different facial expression images of patients and inputting them into a multi-layer neural network, which can automatically predict the HB level of facial palsy patients. Researchers have also developed a series of hardware-software integrated facial palsy detection systems, which provide a more comprehensive and convenient solution for diagnosing and evaluating facial palsy. Based on CNNs, the system in [129] combines Raspberry Pi with digital camera hardware to propose an automatic facial palsy detection system. Additionally, researchers have proposed an innovative, intelligent system [130] designed to assist facial palsy patients in automated physical therapy at home using 3D-printed headgear and a smartphone application. The system comprises two neural network models: the palsy prediction neural network (PPNN) for assessing the degree of palsy and the routine time suggestion neural network (RTSNN) for suggesting treatment times. This system aims to reduce the frequency of patients visiting clinics, thereby enhancing the convenience and efficiency of physical therapy.

C. Summary

A comparison of the existing AI-based FPE methods is summarized in Table III and Table IV. Recent studies commonly evaluate their algorithms and can achieve satisfactory performance on a specific database that is unpublic and inaccessible. The databases used in these studies are significantly different due to their inaccessibility, leading to incomparable results. Therefore, this survey summarizes performance comparisons of existing methods by describing the datasets they used. Some research has shown the impressive performance of the implications of machine learning in FPE; many hand-crafted features, including appearance features and geometric features, were used in a machine learning model to train and evaluate facial palsy severity. In many cases, geometric features based on facial landmarks, as the most mainstream, are used as the basis of many FPE methods based on machine learning. However, traditional methods for facial palsy evaluation based

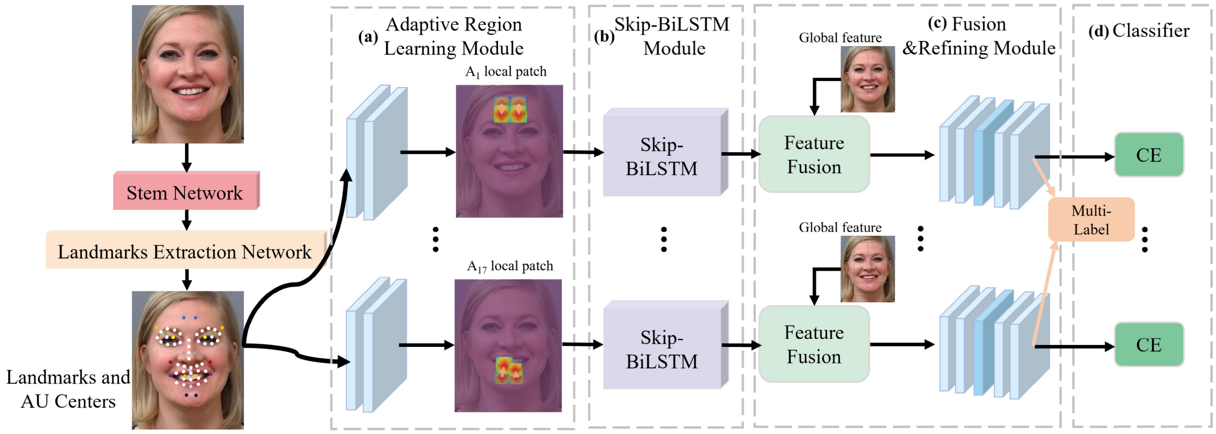


Fig. 12. The proposed dual-path LSTM with the deep differentiated network (DP-LSTM-DDN) in [43].

TABLE III
PERFORMANCE SUMMARY OF REPRESENTATIVE MACHINE LEARNING METHODS FOR FPE

References	Datasets	Data description	Evaluation Metrics	Preprocessing			Methods	Performance
[110]	U	Images from 75 patients and 10 healthy persons.	YGS	AdaBoost	-	-	LBP + Gabor + SVM	Average score error: 0.635
[112]	U	Images from 75 patients and 10 healthy persons.	YGS	-	-	-	CMF + SVM	Average score error: 0.6275
[109]	U	Images from 75 patients and 10 healthy persons.	T10P	-	-	-	LBP + Gabor / Wavelet Decomposition + SVM	Average recognition rate: 63.1%; Average score error: 0.49
[111]	U	Images from 75 patients and 10 healthy persons.	YGS	-	-	-	LO-MCGF + SVM	Accuracy rate of overall diagnosis: 81.2%
[40]	U	197 videos.	H-B	-	-	-	MLBP + SVM	Disagreement <=1(no difference, or one grade difference): 93.9%
[73]	U	Images from 62 patients.	H-B	-	ASM	-	LBP + SVM	Average recognition rate: 94%
[99]	U	156 videos from 23 patients.	H-B	-	-	✓	LBP + SVM	Accuracy: 85.7%
[117]	U	9 videos per person from 14 patients and 31 healthy subjects.	C (2 classes: healthy, unhealthy)	-	-	-	Landmark feature + SVM	Accuracy: 95.5%
[52]	U	Videos from 23 patients and 13 normal subjects.	H-B	-	iPAR-CLR	-	Landmark feature + SVM / LDA	Accuracy: 88.9%(SVM); Accuracy: 88.9%(LDA)
[115]	U	8 images of facial paralysis.	C (4 classes: normal, mild, moderate, severe)	-	SPSA	✓	Landmark feature + FFBPN	Accuracy: 94%
[58]	U	480 static images.	H-B	Haar face detector	SDM	✓	Landmark feature + SVM	Loose match (discrepancy = 1)rate: 90.01%
[91]	U	1,024 facial palsy images and 1081 healthy images.	C (2 classes: normal, patient)	OpenCV library	Emotrics	-	Landmark feature + SVM / Logistic Regression	Accuracy: 76.87%(SVM)
[116]	U	Images from 103 peripheral palsy patients, 40 central palsy patients, and 60 healthy people.	C (3 classes: healthy, peripheral palsy, central palsy)	-	-	-	Landmark feature + SVM / Gaussian naive Bayes	Accuracy: 85.1%(SVM); Accuracy: 80.7%(Gaussian naive Bayes)
[113]	U	Images from 75 patients and 10 healthy persons.	T10P	Adaboost	-	-	KLT + Landmark feature + SVM	Average score error: 0.31
[93]	U	33 facial videos.	H-B	-	Dlib	-	Landmark feature + SVM / RF	Accuracy: 72.2%(SVM); Accuracy: 88.9%(RF)
[53]	U	325 facial images from 50 patients and 15 healthy subjects.	H-B	-	LAC	-	Optimized Daugman's Algorithm + Landmark feature + hybrid classifier	Accuracy: 93.7%
[70]	U	440 facial images.	C (3 classes: healthy, central palsy, peripheral palsy)	-	ERT	-	Optimized Daugman's Algorithm + Landmark feature + hybrid classifier	Accuracy: 97.48%
[4]	MEEI and TNF	480 high-resolution images from the MEEI database and 261 videos from the TNF dataset.	C (2 classes: healthy, unhealthy)	Dlib	MEEshape predic-tor[75]	✓	Landmark feature + MLP	Accuracy: 97.22%
[62]	YFP and CK+	32 videos from the YFP database and 10,800 images from the CK+ database.	C (2 classes: healthy, unhealthy; 2 severity levels: slight, strong; 3 severity levels: healthy, slight, strong)	Dlib	MEEshape predic-tor[75]	✓	Landmark feature + SVM	Accuracy: 95.61%(2 classes); Accuracy: 95.05%(2 severity levels); Accuracy: 95.58%(3 severity levels)

Preprocessing: Face detection & Landmark detection & Data augmentation, U: unpublic datasets, YGS: The Yanagihara Grading System, T10P: Triage-10-Points Grading System, H-B: House-Brackmann Scale, LBP: Local Binary Pattern, Gabor: Gabor filter, SVM: Support Vector Machine, CMF: Concentric Modulation Filter, LO-MCGF: Limited-Orientation Modified Circular Gabor Filter, MLBP: Multiresolution Local Binary Pattern, JAFFE: Japanese Female Facial Expression database, ASM: Active Shape Model, iPAR-CLR: incremental Parallel Cascade of Linear Regression, LDA: Linear Discriminant Analysis, SPSA: Salient Point Selection Algorithm, FFBPN: Feed Forward Back Propagation Neural Network, SDM: Supervised Descent Method, RF: Random Forest, LAC: Localized Active Contour model, ERT: Ensemble of Regression Trees algorithm, MEEI: Massachusetts Eye and Ear Infirmary database, TNF: Toronto NeuroFace dataset, MLP: Multilayer Perceptron Classifier, YFP: The YouTube Facial Palsy database, CK+: The Extended Cohn-Kanade database

TABLE IV
PERFORMANCE SUMMARY OF REPRESENTATIVE DEEP LEARNING METHODS FOR FPE

References	Datasets	Data description	Evaluation Metrics	Preprocessing			Methods	Performance
[103]	U	720 UPFP images.	H-B	-	-	✓	DCNN	Accuracy: 91.25%
[56]	U	2,000 face images.	H-B	-	-	✓	CNN	Average Classification Accuracy: 92.60%
[55]	U	1,049 clinical images.	C (7 classes: normal, left mild dysfunction, left moderate dysfunction, left severe dysfunction, right mild dysfunction, right moderate dysfunction, right severe dysfunction)	-	-	✓	Inception-DeepID-FNP CNN	Accuracy: 97.5%
[41]	U	AFLW dataset and 12,000 facial paralysis images.	H-B	-	-	✓	CNN	Accuracy: 95.6%
[60]	U	Image series of 52 multi-ethnical patients.	eFace	Dlib	-	-	VGG-16 + SVR	Average grading error: 11%
[122]	U	Images from 128 patients and 2 healthy persons.	H-B, SGS	-	CNN	-	Landmark feature	Correlation with SGS: r = 0.905; Correlation with H-B: r = 0.783
[97]	YFP and CK+	31 videos from the YFP database and 1,211 normal samples from the CK+ database.	C (6 classes: SlightPalsy Eyes, StrongPalsy Eyes, SlightPalsy Mouth, StrongPalsy Mouth, Normal Mouth, Normal Eyes)	MTCNN	-	-	EfficientNet + RC-SSELM^{vc}	Accuracy: 85.5%
[59]	YFP and Caltech Face Database	1,105 palsy face images from the YFP database and 450 face images from the Caltech Face database.	C (2 classes: normal, palsy)	Improved Viola–Jones face detection method	-	✓	SqueezeNet + ECOC-SVM	Accuracy: 99.34%
[127]	U	Images from 51 patients and 10 healthy persons.	H-B	-	-	-	CNN	Accuracy: 98%
[128]	U	9 images per person from 86 patients.	H-B	-	-	✓	Neural network	Accuracy: 100%
[105]	U	1,024 facial paralysis images and 23,700 normal face images.	C (2 classes: normal, stroke)	-	-	✓	CNN	Accuracy: 96.63%
[104]	U	Images from 116 UPFP patients and 9 healthy persons.	SGS	-	-	✓	CNN	Correlation with SGS: r = 0.87
[92]	YFP, CK+, and Tufts Face Database	3,958 images.	C (2 classes: healthy, unhealthy)	-	Dlib	✓	MobileNetV2	Accuracy: 98.93%
[101]	YFP, CK+ and MEEI	32 videos from the YFP database, 350 images from the CK+ database, and videos of 60 subjects from the MEEI database.	C (6 classes: normal, near normal, mild, moderate, severe, complete)	Dlib	-	✓	FP-VGGFace	Accuracy: 99.3%
[129]	U	19,000 normal images and 1600 FP images.	C (3 classes: right palsy, left palsy, normal)	Dlib	Haar Cascade	-	CNN	Accuracy: 98%
[126]	U	570 images.	C (3 classes: right palsy, left palsy, normal)	Dlib	-	-	CNN	Accuracy: 98%
[17]	YFP and CK+	5,088 sequences from the YFP database and 332 image sequences from the CK+ database.	H-B	AdaBoost	IntraFace	✓	PHCNN + LSTM	Accuracy: 94.81%
[43]	U	3003 facial videos.	C (4 classes: normal, mildly ill, moderate ill, critically ill)	Faster RCNN	AAM	-	DP-LSTM-DDN	Accuracy: 73.47%
[68]	U	Facial videos of 75 facial paralysis patients.	C (4 classes: normal, mildly ill, moderate ill, critically ill)	Faster RCNN	-	-	Triple-stream LSTM	Average accuracy: 86.37%
[78]	U	236 measurements of 127 patients.	H-B	-	AAM	✓	Triple-Path CNN	Accuracy: 89%
[67]	U	397 samples.	H-B	IDM	-	✓	3DPalsyNet	Accuracy: 86%
[44]	U	89 videos of facial palsy patients.	C (4 classes: Normal, Low, Medium, High)	-	-	✓	ALGRNet	Average F1-frame score: 75.4%

UPFP: Unilateral Peripheral Facial Paralysis, DCNN: Deep Convolutional Neural Network, AFLW: Annotated Facial Landmarks in the Wild dataset, SVR: Support Vector Regression, SGS: Sunnybrook Grading System, MTCNN: Multi-Task Cascaded Convolutional Network, RC-SSELM^{VC}: Regularized Correntropy Criterion based on Semi-Supervised Extreme Learning Machine, ECOC-SVM: Error-Correcting Output Codes based Support Vector Machine, PHCNN: Parallel Hierarchy Convolutional Neural Network, LSTM: Long Short-Term Memory network, DP-LSTM-DDN: Dual-Path LSTM with Deep Differentiated Network, Faster-RCNN: Faster-Regions with Convolutional Neural Network, AAM: Active Appearance Model, IDM: Integrated Deep Model, ALGRNet: Adaptive Local-Global Relational Network

on hand-crafted features and machine learning still have specific limitations, such as being easily affected by the changes in lighting conditions and the accuracy of facial landmark detection.

Recently, multiple FPE methods based on deep learning have demonstrated higher accuracy and reliability in facial palsy evaluation, particularly when dealing with large-scale datasets. However, clinical evaluation of facial palsy usually depends on static facial asymmetry at maximum movement and dynamic changes during motion. Most FPE methods based on deep learning focus solely on static asymmetry. Besides, research in the medical field also shows that features of facial palsy normally occur in particular areas [131], [132]. Therefore, some researchers have started to study dynamic changes and the asymmetry of specific facial areas in deep learning-based FPE methods.

In summary, various relevant research and rehabilitation bodies have committed to researching more objective and universal AI-based FPE systems. These systems can reduce the influence of subjective factors and facilitate comparing clinical outcomes. Moreover, AI-based FPE is important in reducing healthcare costs. It is not meant to replace the assessment from clinicians completely but to offer a rapid automated estimation of the patient's data. The AI-based FPE can compensate for the deficiency and inherent problems of traditional manual evaluation manners and integrate human intelligence and machine efficiency into one unit [133], which has proven its feasibility [77].

V. Research Challenges

Automated evaluation of facial palsy using AI techniques such as machine learning or deep learning offers a promising solution to the limitations of current traditional assessment methods, which are time-consuming, labour-intensive, and subject to clinicians' bias. Over the past several years, facial palsy evaluation based on AI has been extensively researched and developed. However, the use of AI in facial paralysis is a relatively new concept. Current methods are still far from satisfying clinical requirements, and some limitations still need to be overcome to develop a clinically usable tool. This section briefly presents the research challenges for AI-based FPE, concentrating on the following aspects.

A. Insufficient and Unavailable Databases

AI-based FPE is a data-driven task that relies on large-scale and accurate labelled datasets. However, current datasets related to facial palsy are not only small in scale, but it is also challenging to collect a large annotated facial palsy database, requiring specific expertise and a time-consuming annotation process. In general, the facial palsy database should have facial images of varying age ranges, genders, and ethnicities, not just of severity but also of other facial attributes. However, existing databases are far from meeting these requirements, which may impact related research on the cross-group FPE. Also, imbalanced class distribution is a common issue in the database, frequently occurring in naturalistic settings. For example, collecting a normal face is simple, but collecting a face with severe facial palsy can be difficult. AI algorithms can perpetuate and amplify these imbalances, which may result in reduced accuracy across different categories. Another common issue relates to privacy and ethical concerns. Some facial palsy databases (e.g., [40], [56], [58], [71], [117], [134]) are not publicly available. Currently, there is no widely accepted and publicly available benchmark database.

B. Inconsistent and Coarse Annotations

Another current limitation of AI-based FPE is the accuracy and reliability of data annotations. A common approach is to manually annotate the data with FPE grade scales under the guidance of clinicians. The quality of annotations mainly depends on the experience of the annotator, which may lead to inconsistency across different annotators or databases. In addition, most existing works classify the severity of facial images of the patient into a few discrete levels. While categorical grade scales like House-Brackmann [20] have been widely used in FPE, they simply classify the severity and cannot describe symptoms in more detail. The continuous-valued scales may be a better choice for this problem. For example, the dimensional FPE grade scales, namely, Sunnybrook [21] and eFACE [22], are proposed to describe a detailed range of facial palsy severity and continuously encode small visible appearance changes in the severity of facial palsy. However, these dimensional FPE grade scales need to grade each facial region by clinicians, which is more time-consuming and expensive to collect the data.

C. Low Reliability and Explainability

Although the main research focus in FPE has shifted to analysis with AI technologies, existing methods cannot be effectively and widely used in clinical practice. One of the challenges to prevent this is the lack of reliability and sufficient clinical validation. The AI algorithms, especially deep learning approaches, need large data with annotations to ensure performance. However, in FPE, most existing studies are based on databases with a small amount of facial image data. Thus, they have limited performance and are not reliable enough for practical applications. Moreover, most studies just applied an existing classical network without designing a particular network structure for the task of FPE [56], [103]. Most importantly, researchers often consider deep learning models as black boxes with low explainability. In clinical practice, users need to understand the mechanisms of AI-based FPE to trust them in practical applications.

D. Limited Application and Real-Time Performance

AI-based FPE can automate and accelerate the workflow by rapidly doing standardized assessments of facial palsy based on patient data. Usually, a clinician needs to examine dozens of patients per day. The lack of clinicians in some areas leads to an increased workload for clinicians, which makes them more prone to human error. Automatic FPE systems are urgently needed for clinical diagnosis. Currently, no FPE system applications can be widely used in clinical practice. Additionally, health applications on handheld devices, such as weight-loss support and heart rate monitoring, have become an integral part of the clinical workflow over the past decade [128], [135]. The FPE can be deployed onto portable devices with a camera, like smartphones, and is easily available to anyone. Therefore, an ideal FPE system should have high real-time performance. However, most researchers usually choose and design deeper network structures to achieve the higher performance of FPE. These networks often have more parameters, calculations, and memory requirements, which may have limited real-time performance in applications.

VI. Future Directions

In addition to the research challenges reviewed above, we further introduce a few future directions related to AI-based FPE.

A. Databases Construction and Synthesis

Since the performance of AI-based FPE mainly depends on the training data, collecting more facial data from clinical settings is needed. For the annotation quality of data, the inter-rater annotation reliability should be reported for each database to filter out noisy annotations. Building automatic labelling tools [95], [96] refined by experts is an alternative that can efficiently provide consistent annotations in case of disagreements. In addition, the recent success of technologies of artificial intelligence generated content (AIGC) has made it possible to generate various specific facial images. Without manually collecting and labelling large datasets, synthesizing the images of facial palsy patients based on generative models, such as generative adversarial networks (GANs) [56], [89], [136], is a novel direction and a more feasible alternative for obtaining the data. Overall, constructing a large, public, and high-quality database with annotations for FPE that meets AI technology is an important and long-term task in this field, which is beneficial to develop and push an AI-based FPE system to clinical use.

B. Utilizing Other Training Manners and Modal Data

One solution to overcome the low reliability problems because of limited data is transfer learning, which uses or fine-tunes pre-trained models from large databases to transfer the knowledge to the FPE task instead of training a model from scratch [101]. Other alternatives are semi-supervised learning [97], or designing a cost-sensitive loss function [17]. Moreover, facial palsy severity can be encoded from different modal data, although collecting multiple types of data on patients is challenging as it involves specific devices and costs. The authors in [76] and [137] proposed to predict facial palsy severity based on eye information instead of facial images. In [138], the authors proposed the FPE method based on infrared thermal images which can extract the features of temperature distribution on specific facial areas, and [72] proposed an approach to provide a quantitative evaluation of facial palsy by analysing the blood flow images of relevant facial regions. In addition, EMG information can promote FPE methods based on facial images by distinguishing some micro-expressions [139]. The information obtained from 3D face reconstruction or 3D facial landmarks also provides the opportunity for FPE [140], [141], [142], [143], [144]. The fusion of other modalities is becoming a promising research direction. Future research could investigate whether diverse multi-modal data could result in better reliable and generalized outcomes.

C. Incorporation of Prior Knowledge

To enhance the explainability of the deep learning network, a novel solution is to introduce or incorporate the prior knowledge, which makes the shift from the data-driven model to the knowledge-driven model. The facial representations via landmark detection allow the extraction of clinically interpretable measurement outcomes about symptoms of facial palsy [3], [145]. These representations are valuable prior knowledge, which applies to different people since facial landmarks are some predefined locations unaffected by skin colour, ethnicity, or gender. Another novel viewpoint argues that FPE is closely related to the problem of facial expression analysis since they both study facial muscle movements and changes [146]. Some works have also demonstrated a correlation between facial expression analysis and FPE. During attempted smiles in facial palsy patients, researchers found they may have some negative emotions like contempt [45]. The FPE methods can be studied with prior knowledge of facial expression analysis. For instance, it is well-known that facial action units (AUs) rarely appear in isolation during spontaneous facial behaviour, and some AUs cannot co-occur. Therefore, a possible solution is to rely on the prior knowledge of AU temporal co-occurrence and consistency, which adds additional constraints to complement the AI model for improving the performance to some extent. Some interesting associated works extended the AU detection models of facial expressions to the assessment of facial paralysis grades [44], [147], [148], [149]. Overall, the utilization of prior knowledge for interpreting models has been a hot research topic in the AI field. However, it still has received limited attention in the field of FPE. Therefore, how to incorporate the knowledge to help solve the problem of explainability of FPE is an important future research direction.

D. Network Optimization and Medical Large Models

AI-based systems have the potential to rapidly triage patients with varying levels of palsy severity according to a facial nerve grading scale, which may reduce the workloads of clinicians, prevent subjective bias, and increase time-efficient disease evaluation [77]. In addition, FPE can be further integrated with other mobile health applications to collect various information on patients, such as heart rate, sleep quality, and blood pressure. The clinically relevant data from these mobile health applications may give a comprehensive evaluation report of the patient's situation, which is beneficial for more refined therapy decisions and evaluating their effectiveness. With the mobile application, people can complete the initial self-diagnosis at any time and thus get real-time feedback, which is highly efficient and cost-effective for patients who have difficulties visiting therapists or are in rehabilitation training. The application also helps people establish a correct perception of their facial function and determine whether they can be cured at home or need medical support, reducing the waste of medical resources. As a result, optimising the network framework to design a lighter network for real-time and easy-to-use FPE mobile applications is a future direction.

Recently, the emergence of large models in natural language processing has driven progress in various tasks, attracting a growing research interest. Although in the early research stages, some researchers have begun to extend large models to the medical field, resulting in medical large models that can help clinicians and patients [150]. For example, a recent study proposed a medical large model for dental diagnosis [151]. However, there are currently no medical large models for FPE, and training large models is still very challenging. Therefore, how to costly and effectively train and utilize large medical models can be a promising direction in FPE.

VII. Conclusion

Facial palsy evaluation with high performance is beneficial to facial functional treatment and rehabilitation, which plays an important role in the field of healthcare. This paper reviews the state-of-the-art development of AI-based facial palsy evaluation, from databases to approaches. Furthermore, this paper also provides a detailed discussion of current research challenges and potential future directions. Overall, although AI-based facial palsy evaluation has been developed over the past years, it is still in its infant stage, and many problems remain yet to be solved. We hope that this survey will provide helpful guidelines to researchers conducting research in the future.

References

[1] J. Finsterer, "Management of peripheral facial nerve palsy," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 265, no. 7, pp. 743–752, Jul. 2008.

[2] J. A. McCaul, L. Cascarini, D. Godden, D. Coombes, P. A. Brennan, and C. J. Kerawala, "Evidence based management of Bell's palsy," *Brit. J. Oral Maxillofacial Surg.*, vol. 52, no. 5, pp. 387–391, 2014.

[3] 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.

[4] G. S. Parra-Dominguez, R. E. Sanchez-Yanez, and C. H. Garcia-Capulin, "Facial paralysis detection on images using key point analysis," *Appl. Sci.*, vol. 11, no. 5, p. 2435, Mar. 2021.

[5] A. Bandini et al., "A new dataset for facial motion analysis in individuals with neurological disorders," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1111–1119, Apr. 2021.

[6] L. Fu, C. Bundy, and S. A. Sadiq, "Psychological distress in people with disfigurement from facial palsy," *Eye*, vol. 25, no. 10, pp. 1322–1326, Oct. 2011.

[7] D. T. Walker, M. J. Hallam, S. N. Mhurchadha, P. McCabe, and C. Nduka, "The psychosocial impact of facial palsy: Our experience in one hundred and twenty six patients," *Clin. Otolaryngol.*, vol. 37, no. 6, pp. 474–477, Dec. 2012.

[8] 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," *Cirugía y Cirujanos*, vol. 89, no. 6, pp. 718–727, Nov. 2023.

[9] Y. Xia, H. Yu, X. Wang, M. Jian, and F.-Y. Wang, "Relation-aware facial expression recognition," *IEEE Trans. Cognit. Develop. Syst.*, vol. 14, no. 3, pp. 1143–1154, Sep. 2022.

[10] 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, 2012.

[11] 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.

[12] J. Tavares-Brito, M. M. van Veen, J. R. Dusseldorp, F. Bahmad, and T. A. Hadlock, "Facial palsy-specific quality of life in 920 patients: Correlation with clinician-graded severity and predicting factors," *Laryngoscope*, vol. 129, no. 1, pp. 100–104, Jan. 2019.

[13] 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.

[14] 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," *Otology Neurotol.*, vol. 25, no. 6, pp. 1014–1019, Nov. 2004.

[15] D. Jayatilake, T. Isezaki, Y. Teramoto, K. Eguchi, and K. Suzuki, "Robot assisted physiotherapy to support rehabilitation of facial paralysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 3, pp. 644–653, May 2014.

[16] J. T. Heaton, C. J. Knox, J. S. Malo, J. B. Kobler, and T. A. Hadlock, "A system for delivering mechanical stimulation and robot-assisted therapy to the rat whisker pad during facial nerve regeneration," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 6, pp. 928–937, Nov. 2013.

[17] 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.

[18] J. Thevenot, M. B. López, and A. Hadid, "A survey on computer vision for assistive medical diagnosis from faces," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1497–1511, Sep. 2018.

[19] L. Liu, G. Cheng, J. Dong, S. Wang, and H. Qu, "Evaluation of facial paralysis degree based on regions," in *Proc. 3rd Int. Conf. Knowl. Discov. Data Mining*, Jan. 2010, pp. 514–517.

[20] J. W. House, "Facial nerve grading systems," *Laryngoscope*, vol. 93, no. 8, pp. 1056–1069, 1983.

[21] B. Ross, G. Fradet, and J. Nedzelski, "Development of a sensitive clinical facial grading system," *Otolaryngol. Head Neck Surg.*, vol. 114, no. 3, pp. 380–386, Mar. 1996.

[22] C. A. Banks, P. K. Bhama, J. Park, C. R. Hadlock, and T. A. Hadlock, "Clinician-graded electronic facial paralysis assessment: The eFACE," *Plastic Reconstructive Surg.*, vol. 136, no. 2, pp. 223e–230e, Aug. 2015.

[23] R. Niziol, F. P. Henry, J. I. Leckenby, and A. O. Grobbelaar, "Is there an ideal outcome scoring system for facial reanimation surgery? A review of current methods and suggestions for future publications," *J. Plastic, Reconstructive Aesthetic Surg.*, vol. 68, no. 4, pp. 447–456, Apr. 2015.

[24] A. S. Ahuja, "The impact of artificial intelligence in medicine on the future role of the physician," *PeerJ*, vol. 7, p. e7702, Oct. 2019.

[25] R. A. Miller, "Medical diagnostic decision support systems-past, present, and future: A threaded bibliography and brief commentary," *J. Amer. Med. Inform. Assoc.*, vol. 1, no. 1, pp. 8–27, Jan. 1994.

[26] P. Szolovits, R. S. Patil, and W. B. Schwartz, "Artificial intelligence in medical diagnosis," *Ann. Internal Med.*, vol. 108, no. 1, pp. 80–87, 1988.

[27] F. T. de Dombal, D. J. Leaper, J. R. Staniland, A. P. McCann, and J. C. Horrocks, "Computer-aided diagnosis of acute abdominal pain," *BMJ*, vol. 2, no. 5804, pp. 9–13, Apr. 1972.

[28] M. Kundu, M. Nasipuri, and D. K. Basu, "Knowledge-based ECG interpretation: A critical review," *Pattern Recognit.*, vol. 33, no. 3, pp. 351–373, Mar. 2000.

[29] R. A. Miller, M. A. McNeil, S. M. Challinor, F. E. Masarie Jr., and J. D. Myers, "The internist-1/quick medical reference project-status report," *Western J. Med.*, vol. 145, no. 6, p. 816, 1986.

[30] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: Management, analysis and future prospects," *J. Big Data*, vol. 6, no. 1, pp. 1–25, Dec. 2019.

[31] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.

[32] I. H. Sarker, A. S. M. Kayes, and P. Watters, "Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage," *J. Big Data*, vol. 6, no. 1, pp. 1–28, Dec. 2019.

[33] I. H. Sarker and A. S. M. Kayes, "ABC-RuleMiner: User behavioral rule-based machine learning method for context-aware intelligent services," *J. Netw. Comput. Appl.*, vol. 168, Oct. 2020, Art. no. 102762.

[34] C. Díez-Sanmartín and A. S. Cabezuelo, "Application of artificial intelligence techniques to predict survival in kidney transplantation: A review," *J. Clin. Med.*, vol. 9, no. 2, p. 572, Feb. 2020.

[35] T. B. Murdoch and A. S. Detsky, "The inevitable application of big data to health care," *JAMA*, vol. 309, no. 13, p. 1351, Apr. 2013.

[36] S. E. Dilsizian and E. L. Siegel, "Artificial intelligence in medicine and cardiac imaging: Harnessing big data and advanced computing to provide personalized medical diagnosis and treatment," *Current Cardiol. Rep.*, vol. 16, no. 1, pp. 1–8, Jan. 2014.

[37] E. Kolker, V. Özdemir, and E. Kolker, "How healthcare can refocus on its super-customers (patients, n = 1) and customers (Doctors and Nurses) by leveraging lessons from Amazon, Uber, and Watson," *OMICS, J. Integrative Biol.*, vol. 20, no. 6, pp. 329–333, Jun. 2016.

[38] F. Jiang et al., "Artificial intelligence in healthcare: Past, present and future," *Stroke Vascular Neurol.*, vol. 2, no. 4, pp. 230–243, 2017.

[39] J. Dong et al., "An approach for quantitative evaluation of the degree of facial paralysis based on salient point detection," in *Proc. Int. Symp. Intell. Inf. Technol. Appl. Workshops*, Dec. 2008, pp. 483–486.

[40] 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.

[41] 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.

[42] G. 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.

[43] P. Xu et al., "Automatic evaluation of facial nerve paralysis by dual-path LSTM with deep differentiated network," *Neurocomputing*, vol. 388, pp. 70–77, May 2020.

[44] X. Ge, J. M. Jose, P. Wang, A. Iyer, X. Liu, and H. Han, "Automatic facial paralysis estimation with facial action units," 2022, *arXiv:2203.01800*.

[45] J. J. Greene et al., "The spectrum of facial palsy: The MEEI facial palsy photo and video standard set," *Laryngoscope*, vol. 130, no. 1, pp. 32–37, Jan. 2020.

[46] 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.*, vol. 10, no. 4, pp. 1975–1985, Aug. 2023.

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

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

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

[50] 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.

[51] 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.

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

[53] J. Barbosa et al., "Efficient quantitative assessment of facial paralysis using iris segmentation and active contour-based key points detection with hybrid classifier," *BMC Med. Imag.*, vol. 16, no. 1, pp. 1–18, Dec. 2016.

[54] A. Moayedikia, K.-L. Ong, Y. L. Boo, W. G. Yeoh, and R. Jensen, "Feature selection for high dimensional imbalanced class data using harmony search," *Eng. Appl. Artif. Intell.*, vol. 57, pp. 38–49, Jan. 2017.

[55] A. Song, Z. Wu, X. Ding, Q. Hu, and X. Di, "Neurologist standard classification of facial nerve paralysis with deep neural networks," *Future Internet*, vol. 10, no. 11, p. 111, Nov. 2018.

[56] M. Sajid, T. Shafique, M. J. A. Baig, I. Riaz, S. Amin, and S. Manzoor, "Automatic grading of palsy using asymmetrical facial features: A study complemented by new solutions," *Symmetry*, vol. 10, no. 7, p. 242, Jun. 2018.

[57] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Dec. 2001, pp. I-511–I-518.

[58] Z. Guo et al., "An unobtrusive computerized assessment framework for unilateral peripheral facial paralysis," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 3, pp. 835–841, May 2018.

[59] O. O. Abayomi-Alli, R. Damaševič ius, 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.

[60] A. Raj, O. Mothes, S. Sickert, G. F. Volk, O. Guntinas-Lichius, and J. Denzler, "Automatic and objective facial palsy grading index prediction using deep feature regression," in *Proc. 24th Annu. Conf. Med. Image Understand. Anal. (MIUA)*, Oxford, U.K. Cham, Switzerland: Springer, 2020, pp. 253–266.

[61] Z. Guo, W. Li, J. Dai, J. Xiang, and G. Dan, "Facial imaging and landmark detection technique for objective assessment of unilateral peripheral facial paralysis," *Enterprise Inf. Syst.*, vol. 16, nos. 10–11, pp. 1556–1572, Oct. 2022.

[62] 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.

[63] G. S. Parra-Dominguez, R. E. Sanchez-Yanez, and C. H. Garcia-Capulin, "Towards facial gesture recognition in photographs of patients with facial palsy," *Healthcare*, vol. 10, no. 4, p. 659, Mar. 2022.

[64] D. E. King, "Dlib-ML: A machine learning toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, Jan. 2009.

[65] H. Li, K. Wu, H. Cheng, C. Gu, and X. Guan, "Nasolabial folds extraction based on neural network for the quantitative analysis of facial paralysis," in *Proc. 2nd Int. Conf. Imag., Signal Process. Commun. (ICISPC)*, Jul. 2018, pp. 54–58.

[66] G. J. Hsu and M.-H. Chang, "Deep hybrid network for automatic quantitative analysis of facial paralysis," in *Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Nov. 2018, pp. 1–7.

[67] G. Storey, R. Jiang, S. Keogh, A. Bouridane, and C.-T. Li, "3DPalsyNet: A facial palsy grading and motion recognition framework using fully 3D convolutional neural networks," *IEEE Access*, vol. 7, pp. 121655–121664, 2019.

[68] C. Yang et al., "Automatic degree evaluation of facial nerve paralysis based on triple-stream long short term memory," in *Proc. 3rd Int. Symp. Image Comput. Digit. Med.*, Aug. 2019, pp. 7–11.

[69] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, Mar. 2021.

[70] J. Barbosa, W.-K. Seo, and J. Kang, "ParaFaceTest: An ensemble of regression tree-based facial features extraction for efficient facial paralysis classification," *BMC Med. Imag.*, vol. 19, no. 1, pp. 1–14, Dec. 2019.

[71] T. Wang, J. Dong, X. Sun, S. Zhang, and S. Wang, "Automatic recognition of facial movement for paralyzed face," *Bio-Med. Mater. Eng.*, vol. 24, no. 6, pp. 2751–2760, 2014.

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

[73] T. Wang, S. Zhang, J. Dong, L. Liu, and H. Yu, "Automatic evaluation of the degree of facial nerve paralysis," *Multimedia Tools Appl.*, vol. 75, no. 19, pp. 11893–11908, Oct. 2016.

[74] H. Yoshihara, M. Seo, T. H. Ngo, N. Matsushiro, and Y.-W. Chen, "Automatic feature point detection using deep convolutional networks for quantitative evaluation of facial paralysis," in *Proc. 9th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2016, pp. 811–814.

[75] 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.

[76] 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.

[77] Y. Zhang et al., "The feasibility of an automatical facial evaluation system providing objective and reliable results for facial palsy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1680–1686, 2023.

[78] M. Shayestegan, J. Kohout, K. Štícha, and J. Mareš, "Advanced analysis of 3D Kinect data: Supervised classification of facial nerve function via parallel convolutional neural networks," *Appl. Sci.*, vol. 12, no. 12, p. 5902, Jun. 2022.

[79] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[80] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, Aug. 2001.

[81] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 532–539.

[82] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Incremental face alignment in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1859–1866.

[83] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1867–1874.

[84] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2930–2940, Dec. 2013.

[85] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *Proc. 12th Eur. Conf. Comput. Vis.*, Florence, Italy. Cham, Switzerland: Springer, 2012, pp. 679–692.

[86] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2879–2886.

[87] 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*, Dec. 2013, pp. 397–403.

[88] Z. Zhang, W. Dai, Z. Xie, W. Wang, and W. Wang, "A hybrid evaluation system for facial paralysis rehabilitation based on machine learning and doctor experience," in *Proc. IEEE 17th Int. Conf. Ind. Informat. (INDIN)*, vol. 1, Jul. 2019, pp. 1243–1248.

[89] T. Sakai, M. Seo, N. Matsushiro, and Y.-W. Chen, "Simulation of facial palsy using conditional generative adversarial networks and face shape normalization," in *Proc. IEEE 10th Global Conf. Consum. Electron. (GCCE)*, Oct. 2021, pp. 793–797.

[90] D. Bose, K. Somandepalli, T. Tai, C. Voelker, S. Narayanan, and A. Kochhar, "Automatic analysis of asymmetry in facial paralysis patients using landmark-based measures," *Facial Plastic Surg. Aesthetic Med.*, vol. 24, no. 6, pp. 491–493, Dec. 2022.

[91] A. Arora, A. Sinha, K. Bhansali, R. Goel, I. Sharma, and A. Jayal, "SVM and logistic regression for facial palsy detection utilizing facial landmark features," in *Proc. 14th Int. Conf. Contemp. Comput.*, Aug. 2022, pp. 43–48.

[92] Z. Baig and D. van der Haar, "Facial paralysis recognition using face mesh-based learning," in *Proc. 12th Int. Conf. Pattern Recognit. Appl. Methods*, 2023, pp. 881–888.

[93] Y. Liu, Z. Xu, L. Ding, J. Jia, and X. Wu, "Automatic assessment of facial paralysis based on facial landmarks," in *Proc. IEEE 2nd Int. Conf. Pattern Recognit. Mach. Learn. (PRML)*, Jul. 2021, pp. 162–167.

[94] F. De la Torre, W.-S. Chu, X. Xiong, F. Vicente, X. Ding, and J. Cohn, "IntraFace," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, vol. 1, May 2015, pp. 1–8.

[95] 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 Surgery*, vol. 20, no. 4, pp. 335–337, Jul. 2018.

[96] M. Q. Miller, T. A. Hadlock, E. Fortier, and D. L. Guarin, "The auto-eFACE: Machine learning–enhanced program yields automated facial palsy assessment tool," *Plastic Reconstructive Surgery*, vol. 147, no. 2, pp. 467–474, 2021.

[97] X. Tan, J. Yang, and J. Cao, "Facial nerve paralysis assessment based on regularized correntropy criterion SSELMvc and cascade CNN," in *Proc. 55th Asilomar Conf. Signals, Syst., Comput.*, Oct. 2021, pp. 1043–1047.

[98] 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)*, Oct. 2017, pp. 1021–1030.

[99] P. Li et al., "A two-stage method for assessing facial paralysis severity by fusing multiple classifiers," in *Proc. 14th Chin. Conf. Biometric Recognit. (CCBR)*, Zhuzhou, China. Cham, Switzerland: Springer, 2019, pp. 231–239.

[100] Y. Wu and Q. Ji, "Facial landmark detection: A literature survey," *Int. J. Comput. Vis.*, vol. 127, no. 2, pp. 115–142, Feb. 2019.

[101] W. Ali, M. Imran, M. U. Yaseen, K. Aurangzeb, N. Ashraf, and S. Aslam, "A transfer learning approach for facial paralysis severity detection," *IEEE Access*, vol. 11, pp. 127492–127508, 2023.

[102] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn–Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2010, pp. 94–101.

[103] Z. Guo et al., "Deep assessment process: Objective assessment process for unilateral peripheral facial paralysis via deep convolutional neural network," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 135–138.

[104] T. C. T. Harkel, G. de Jong, H. A. M. Marres, K. J. A. O. Ingels, C. M. Speksnijder, and T. J. J. Maal, "Automatic grading of patients with a unilateral facial paralysis based on the sunnybrook facial grading system—A deep learning study based on a convolutional neural network," *Amer. J. Otolaryngol.*, vol. 44, no. 3, May 2023, Art. no. 103810.

[105] S. M. Hossain, Z. Jamal, A. A. Noshin, and M. M. Khan, "Comparative study of deep learning algorithms for the detection of facial paralysis," in *Proc. IEEE 13th Annu. Inf. Technol. Electron. Mobile Commun. Conf. (IEMCON)*, May 2022, pp. 368–377.

[106] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.

[107] B. F. O'Reilly, J. J. Soraghan, S. McGrenary, and S. He, "Objective method of assessing and presenting the house-brackmann and regional grades of facial palsy by production of a facogram," *Clin. Otolaryngol.*, vol. 36, no. 4, p. 369, Aug. 2011.

[108] N. Perveen, C. Mohan, and Y. Chen, "Quantitative analysis of facial paralysis using GMM and dynamic kernels," in *Proc. 15th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2020, pp. 173–184.

[109] T. H. Ngo, M. Seo, N. Matsushiro, and Y.-W. Chen, "Evaluation of facial paralysis based on spatial features of filtered images," *Int. J. Biosci., Biochem. Bioinf.*, vol. 6, no. 1, pp. 1–8, 2016.

[110] T. H. Ngo, M. Seo, Y.-W. Chen, and N. Matsushiro, "Quantitative assessment of facial paralysis using local binary patterns and Gabor filters," in *Proc. 5th Symp. Inf. Commun. Technol.*, 2014, pp. 155–161.

[111] T. H. Ngo, M. Seo, N. Matsushiro, W. Xiong, and Y.-W. Chen, "Quantitative analysis of facial paralysis based on limited-orientation modified circular Gabor filters," in *Proc. Int. Conf. Pattern Recognit.*, 2016, pp. 349–354.

[112] T. H. Ngo, M. Seo, N. Matsushiro, and Y. Chen, "Quantitative analysis of facial paralysis based on filters of concentric modulation," in *Proc. 12th Int. Conf. Fuzzy Syst. Knowl. Discovery (FSKD)*, Aug. 2015, pp. 1758–1763.

[113] T. H. Ngo, Y.-W. Chen, N. Matsushiro, and M. Seo, "Quantitative assessment of facial paralysis based on spatiotemporal features," *IEICE Trans. Inf. Syst.*, vol. 99, no. 1, pp. 187–196, 2016.

[114] T. Nishida, Y.-W. Chen, N. Matsushiro, and K. Chihara, "An image based quantitative evaluation method for facial paralysis," in *Proc. 2nd Int. Conf. Softw. Eng. Data Mining*, Jun. 2010, pp. 706–709.

[115] K. Anguraj and S. Padma, "Evaluation and severity classification of facial paralysis using salient point selection algorithm," *Int. J. Comput. Appl.*, vol. 123, no. 7, pp. 23–29, Aug. 2015.

[116] C. V. Vletter, H. L. Burger, H. Alers, N. Sourlos, and Z. Al-Ars, "Towards an automatic diagnosis of peripheral and central palsy using machine learning on facial features," 2022, *arXiv:2201.11852*.

[117] O. Azoulay, Y. Ater, L. Gersi, Y. Glassner, O. Bryt, and D. Halperin, "Mobile application for diagnosis of facial palsy," in *Proc. 2nd Int. Conf. Mobile Inf. Technol. Med*, 2014, pp. 1–4.

[118] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[119] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[120] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[121] K. Huang, Y. Li, and F. Zhu, "Quantitative analysis of facial paralysis based on TCM acupuncture point identification," in *Proc. IEEE 3rd Int. Conf. Softw. Eng. Artif. Intell. (SEAI)*, Jun. 2023, pp. 163–167.

[122] S. A. Lee, J. Kim, J. M. Lee, Y.-J. Hong, I.-J. Kim, and J. D. Lee, "Automatic facial recognition system assisted-facial asymmetry scale using facial landmarks," *Otol. Neurotol.*, vol. 41, no. 8, pp. 1140–1148, 2020.

[123] H. Shi et al., "Intelligent bell facial paralysis assessment: A facial recognition model using improved SSD network," *Sci. Rep.*, vol. 14, no. 1, p. 12763, Jun. 2024.

[124] S. Umirzakova, S. Ahmad, S. Mardieva, S. Muksimova, and T. K. Whangbo, "Deep learning-driven diagnosis: A multi-task approach for segmenting stroke and Bell's palsy," *Pattern Recognit.*, vol. 144, Dec. 2023, Art. no. 109866.

[125] H. Niu, J. Liu, X. Sun, X. Zhao, and Y. Liu, "Facial paralysis symptom detection based on facial action unit," *IEEE Access*, vol. 12, pp. 52400–52413, 2024.

[126] A. S. Amsalam, A. Al-Naji, A. Yahya Daeef, and J. Chahl, "Computer vision system for facial palsy detection," *J. Techn.*, vol. 5, no. 1, pp. 44–51, Mar. 2023.

[127] L. Knoedler et al., "A ready-to-use grading tool for facial palsy examiners—Automated grading system in facial palsy patients made easy," *J. Personalized Med.*, vol. 12, no. 10, p. 1739, Oct. 2022.

[128] L. Knoedler et al., "Towards a reliable and rapid automated grading system in facial palsy patients: Facial palsy surgery meets computer science," *J. Clin. Med.*, vol. 11, no. 17, p. 4998, Aug. 2022.

[129] A. S. Amsalam, A. Al-Naji, A. Y. Daeef, and J. Chahl, "Automatic facial palsy, age and gender detection using a raspberry pi," *BioMed-Informatics*, vol. 3, no. 2, pp. 455–466, Jun. 2023.

[130] A. M. Ridha, W. Shehieb, P. Yacoub, K. Al-Balawneh, and K. Arshad, "Smart prediction system for facial paralysis," in *Proc. 7th Int. Conf. Electr. Electron. Eng. (ICEEE)*, Apr. 2020, pp. 321–327.

[131] W. S. W. Samsudin, R. Samad, M. Z. Ahmad, and K. Sundaraj, "Forehead lesion score for facial nerve paralysis evaluation," in *Proc. IEEE Int. Conf. Autom. Control Intell. Syst. (I2CACIS)*, Jun. 2019, pp. 102–107.

[132] M. P. Lafer and O. Tm, "Management of long-standing flaccid facial palsy: Static approaches to the brow, midface, and lower lip," *Otolaryngologic Clinics North Amer.*, vol. 51, no. 6, pp. 1141–1150, 2018.

[133] M. N. Islam, T. T. Inan, S. Rafi, S. S. Akter, I. H. Sarker, and A. K. M. N. Islam, "A systematic review on the use of AI and ML for fighting the COVID-19 pandemic," *IEEE Trans. Artif. Intell.*, vol. 1, no. 3, pp. 258–270, Dec. 2020.

[134] Y. Kihara, G. Duan, T. Nishida, N. Matsushiro, and Y.-W. Chen, "A dynamic facial expression database for quantitative analysis of facial paralysis," in *Proc. 6th Int. Conf. Comput. Sci. Converg. Inf. Technol. (ICCIT)*, Nov. 2011, pp. 949–952.

[135] M. N. Islam, I. Islam, K. Md. Munim, and A. K. M. N. Islam, "A review on the mobile applications developed for COVID-19: An exploratory analysis," *IEEE Access*, vol. 8, pp. 145601–145610, 2020.

[136] S. Yaotome, M. Seo, N. Matsushiro, and Y.-W. Chen, "Simulation of facial palsy using conditional generative adversarial networks," in *Proc. IEEE 8th Global Conf. Consum. Electron. (GCCE)*, Oct. 2019, pp. 579–582.

[137] S. Ali Ansari, K. Rao Jerripothula, P. Nagpal, and A. Mittal, "Eye-focused detection of Bell's palsy in videos," 2022, *arXiv:2201.11479*.

[138] 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.

[139] P. Zhu, H. Wang, L. Zhang, and X. Jiang, "Deep learning-based surface nerve electromyography data of E-Health electroacupuncture in treatment of peripheral facial paralysis," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–12, May 2022.

[140] D.-P. Nguyen, T.-N. Nguyen, S. Dakpé, M.-C. H. B. Tho, and T.-T. Dao, "Fast 3D face reconstruction from a single image using different deep learning approaches for facial palsy patients," *Bioengineering*, vol. 9, no. 11, p. 619, Oct. 2022.

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

[142] J. Kim et al., "Numerical approach to facial palsy using a novel registration method with 3D facial landmark," *Sensors*, vol. 22, no. 17, p. 6636, Sep. 2022.

[143] M. A. Alagha, A. Ayoub, S. Morley, and X. Ju, "Objective grading facial paralysis severity using a dynamic 3D stereo photogrammetry imaging system," *Opt. Lasers Eng.*, vol. 150, Mar. 2022, Art. no. 106876.

[144] D. P. Tanwar, "Evaluation of 3D facial paralysis using fuzzy logic," *Int. J. Eng. Technol.*, vol. 7, no. 4, p. 2325, Sep. 2018.

[145] L. Modersohn and J. Denzler, "Facial paresis index prediction by exploiting active appearance models for compact discriminative features," in *Proc. 11th Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2016, pp. 271–278.

[146] N. Perveen, C. K. Mohan, and Y. W. Chen, "Expression modeling using dynamic kernels for quantitative assessment of facial paralysis," in *Proc. Int. Joint Conf. Comput. Vis., Imag. Comput. Graph.* Cham, Switzerland: Springer, 2020, pp. 372–393.

[147] A. Gaber, M. F. Taher, M. A. Wahed, N. M. Shalaby, and S. Gaber, "Classification of facial paralysis based on machine learning techniques," *Biomed. Eng. OnLine*, vol. 21, no. 1, p. 65, Sep. 2022.

[148] A. Gaber, M. F. Taher, M. Abdel Wahed, N. M. Shalaby, and S. Gaber, "Comprehensive assessment of facial paralysis based on facial animation units," *PLoS ONE*, vol. 17, no. 12, Dec. 2022, Art. no. e0277297.

[149] G. B. Dell'Olio and M. Sra, "FaraPy: An augmented reality feedback system for facial paralysis using action unit intensity estimation," in *Proc. 34th Annu. ACM Symp. User Interface Softw. Technol.*, 2021, pp. 1027–1038.

[150] H. Zhou et al., "A survey of large language models in medicine: Progress, application, and challenge," 2023, *arXiv:2311.05112*.

[151] H. Huang et al., "ChatGPT for shaping the future of dentistry: The potential of multi-modal large language model," *Int. J. Oral Sci.*, vol. 15, no. 1, p. 29, Jul. 2023.