Wang *et al. Art Int Surg.* 2025;5:521-44
DOI: 10.20517/ais.2025.62

Artificial Intelligence Surgery

Original Article

Open Access

Check for updates

Assistive assessment of neurological dysfunction via eye movement patterns and ocular biometrics

Ziqi Wang[1], Jing Bi[2], Xiaomeng Zhao[3], Zhipeng Zheng[2], Jinglei Cui[2], Rong Cui[2], Junqi Zhang[2], Yuanchen Tang[4], Jiantao Liang[4], Kai Zhou[3], Jia Zhang[5]

[1]School of Software Technology, Zhejiang University, Ningbo 315100, Zhejiang, China.
[2]College of Computer Science, Beijing University of Technology, Beijing 100124, China.
[3]Beijing Neurorient Technology Co., Ltd., Beijing 100000, China.
[4]Department of Neurosurgery, Xuanwu Hospital, Capital Medical University, Beijing 100053, China.
[5]Department of Computer Science in Lyle School of Engineering, Southern Methodist University, Dallas, TX 75205, USA.

Correspondence to: Prof. Jing Bi, College of Computer Science, Beijing University of Technology, 100 Pingleyuan, Beijing, 100124, China. E-mail: bijing@bjut.edu.cn; Xiaomeng Zhao, Beijing Neurorient Technology Co., Ltd., Beijing 100000, China. E-mail: jasonzxm@neurorient.com

How to cite this article: Wang Z, Bi J, Zhao X, Zheng Z, Cui J, Cui R, Zhang J, Tang Y, Liang J, Zhou K, Zhang J. Assistive assessment of neurological dysfunction via eye movement patterns and ocular biometrics. *Art Int Surg.* 2025;5:521-44. https://dx.doi.org/10.20517/ais.2025.62

Received: 4 Jul 2025 **First Decision:** 31 Oct 2025 **Revised:** 19 Nov 2025 **Accepted:** 1 Dec 2025 **Published:** 15 Dec 2025

Academic Editor: Andrew Gumbs **Copy Editor:** Xing-Yue Zhang **Production Editor:** Xing-Yue Zhang

Abstract

Aim: Nerve dysfunction often manifests as abnormal eye behaviors, necessitating accurate and objective neurological assessment. Current deep learning-based facial analysis methods lack adaptability to inter-patient variability, making it difficult to capture subtle and rapid ocular dynamics such as incomplete eyelid closure or asymmetric eye movement. To address this, we propose a precise deep learning system for quantitative ocular state analysis, providing objective support for the evaluation of neurological dysfunction.

Methods: We propose the Ocular-enhanced Face Keypoints Net (OFKNet). It incorporates three key innovations: (1) a 40-point anatomically informed ocular landmark design enabling dense modeling of eyelid contours, canthus structure, and pupil dynamics; (2) a MobileNetV3-based region enhancement module that amplifies feature responses within clinically critical areas such as the internal canthus; (3) and an improved Path Aggregation Network combined with Squeeze-and-Excitation modules that enables adaptive multi-scale fusion and enhances sensitivity to subtle ocular deformations.

Results: Using clinically acquired data, OFKNet demonstrates substantial performance gains over state-of-the-art

baselines. It achieves a 65.3% reduction in normalized mean error on the 40-point dataset (0.029 *vs.* 0.084) and a 38.4% reduction on the 14-point dataset, with all improvements statistically significant ($P < 0.001$). Despite operating on high-resolution inputs, the system maintains real-time capability and provides stable frame-level landmark localization, enabling precise capture of dynamic ocular motion patterns.

Conclusion: OFKNet provides a reliable tool for real-time monitoring of eye movement patterns in patients with neurological disorders. By visualizing time-series graphs of bilateral eye openness, the system enables a more comprehensive understanding of ocular dynamics and supports timely clinical decision-making and treatment adjustment.

Keywords: Computer vision, nerve dysfunction, facial paralysis, convolutional neural networks, and eye movement tracking

INTRODUCTION

Nerve dysfunction comprises a group of disorders resulting from nerve damage, often leading to impairments in movement, sensation, or emotional regulation. Common neurological conditions involve difficulties in eyelid closure, which may cause serious complications, including vision impairment or even blindness due to exposure keratopathy. Analyzing eye movements provides crucial insights into these dysfunctions. However, diagnoses remain largely dependent on physicians' subjective evaluations, lacking standardized and objective assessment tools[1,2]. In facial paralysis, grading scales such as the House-Brackmann Grading System (H-BGS) and the Sunnybrook Facial Grading System (SF) are widely used[3,4]. However, they fail to capture the full complexity of facial nerve dysfunction and are still influenced by clinical experience. To address these limitations, Niu *et al.* propose a facial paralysis detection model based on facial action units and co-occurrence matrices[5], while Zhang *et al.* and Gao *et al.* develop deep learning-based systems to improve objectivity in static and dynamic evaluations[6,7]. Despite these advancements, current approaches require complex equipment, struggle to capture fast muscle movements, and remain sensitive to noise and individual variability, limiting their broader clinical application[8]. In recent years, deep learning has shown strong potential in medical image analysis. The Pseudo RGB-D (red, green, and blue-depth) Face Recognition framework[9] integrates depth-aware cues into facial feature extraction to improve discriminative representation, while the Simulated Multimodal Deep Facial Diagnosis model[10] demonstrates the feasibility of combining multiple imaging modalities for robust medical interpretation. These studies highlight the promise of deep learning in clinical diagnosis but also expose key limitations when applied to ocular analysis[8].

The assessment of ocular status plays a vital role in evaluating nerve dysfunction[9,11]. Clinically, patients frequently exhibit incomplete eyelid closure or asymmetric eyelid movement. These ocular signs offer important insights into the severity of nerve dysfunction and inform treatment planning. To enhance eye movement analysis, data visualization proves essential by converting complex ocular dynamics into interpretable formats, enabling physicians to detect trends and monitor disease progression[10,12]. For instance, line graphs showing the degree of eye opening in both eyes help identify asymmetry, while keyframes support detailed evaluation of eyelid motion patterns. Visualizing ocular data improves diagnostic accuracy and facilitates communication, giving patients a clearer understanding of their condition and promoting active involvement in treatment.

This work proposes the Ocular-enhanced Face Keypoints Net (OFKNet), a high-precision deep learning model explicitly designed for ocular state monitoring in patients with neurological dysfunction [Figure 1]. OFKNet enhances fine-grained landmark localization and quantifies dynamic eye movements through real-

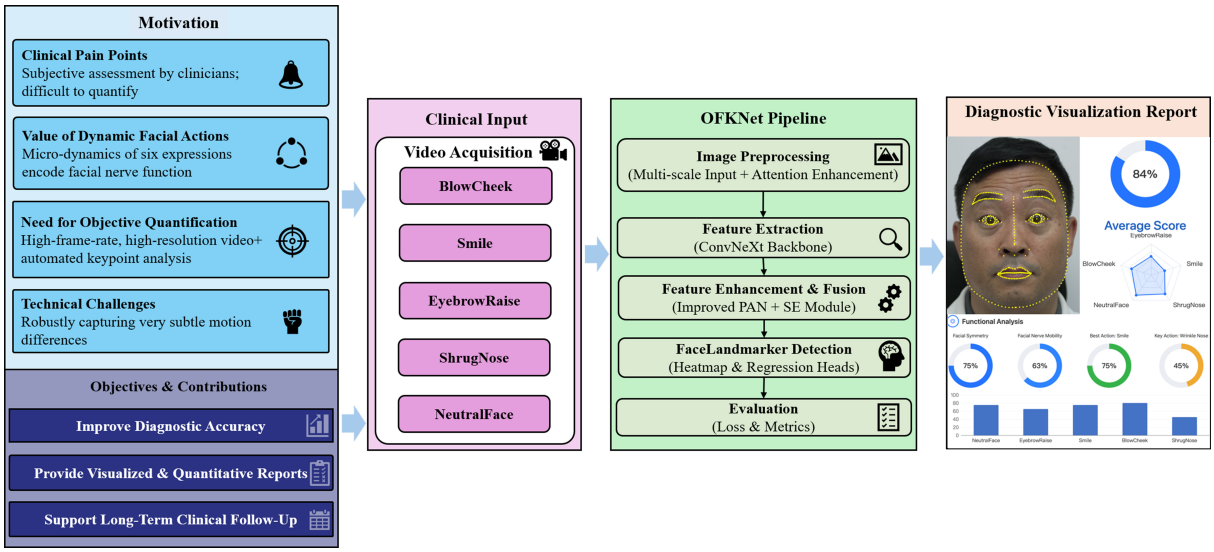

Figure 1. Motivation for this work.

time processing. It introduces three core innovations tailored to the challenges above: (1) a 40-point anatomically informed ocular landmark design that markedly increases spatial resolution around the eyelid and canthus; (2) a region enhancement module that strengthens feature representation in clinically critical periocular areas; and (3) a network that enables adaptive multi-scale fusion for robust tracking of subtle and fast ocular motion. Specifically, ConvNeXt is adopted as the backbone to preserve detailed spatial structure, a multi-scale input enhancement strategy improves adaptability across different facial resolutions[13], and a lightweight MobileNetV3-based region enhancement module refines feature extraction around the canthus[14]. An improved Path Aggregation Network (PAN) combined with Squeeze-and-Excitation (SE) modules ensures efficient multi-scale fusion and adaptive channel weighting, leading to accurate and stable keypoint detection under diverse clinical conditions.

The remainder of this paper is organized as follows. METHODS describes the overall architecture of OFKNet and its constituent modules. RESULTS presents the experimental setup, datasets, evaluation metrics, and reports the quantitative and visual comparison results. DISCUSSION discusses the clinical implications and potential extensions of the proposed framework, followed by conclusions in the end.

METHODS

We employ a high-speed, high-resolution camera to capture five facial movements of patients - neutral face (NeutralFace), eyebrow raising (EyebrowRaise), nose scrunching (ShrugNose), smiling (Smile), and cheek puffing (BlowCheek) - at 120 frames per second with 4K resolution. OFKNet is proposed to detect 40 specific key points in the ocular region based on the collected images. Specifically, the internal canthus of each eye is represented by a single key point, precisely defining its position. The right and left eye contours are represented by multiple key points distributed according to the eyes' morphological characteristics, outlining their boundaries. Additionally, the centers of the right and left pupils are included as critical single-point markers. Using these 40 ocular key points, OFKNet can capture dynamic changes during eye movements across several facial expressions. The overall framework of OFKNet is shown in Figure 2, which mainly comprises four modules.

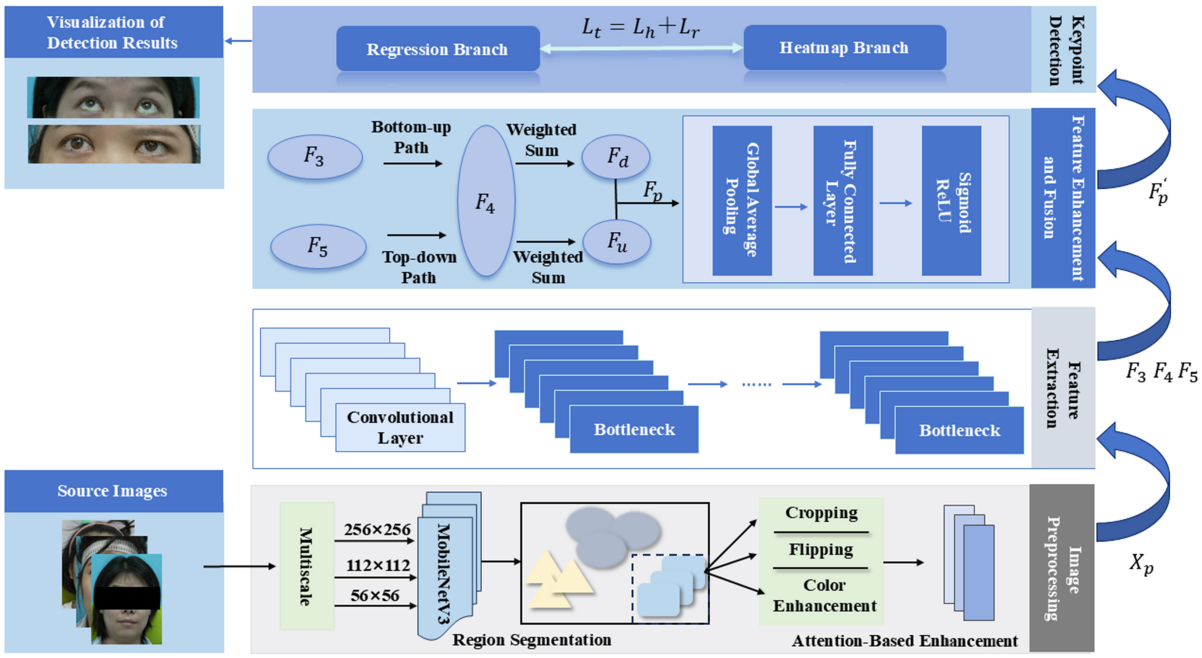

Figure 2. Framework of OFKNet. OFKNet: Ocular-enhanced Face Keypoints Net.

1. Image preprocessing module: It employs ConvNeXt[4] as the backbone network, removing fully connected and pooling layers to preserve the spatial structure. A multi-scale input strategy is introduced, adopting three resolutions (256 × 256, 112 × 112, and 56 × 56) of images to enhance the model's adaptability. MobileNetV3[5] is adopted for facial area segmentation, and a region-enhancement strategy is proposed to adjust the weights of each region.

2. Feature extraction module: It is based on ConvNeXt and extracts features at different levels from the preprocessed images. The bottleneck enhances feature representation while reducing the computational cost, and the down-sampling operation generates multi-scale feature maps.

3. Feature enhancement and fusion module: It receives these multi-scale feature maps and designs an improved PAN to fuse low-level and high-level features through top-down and bottom-up paths. SE further enhances the feature representation ability by weighting the channels and generating a new feature map.

4. Key point detection module: It conducts key point detection in a parallel manner using the heatmap and regression branches. The heatmap branch generates two-dimensional (2D) Gaussian heatmaps to represent the key point probabilities, and the regression branch obtains the key point coordinates. The total loss function combines the heatmap and regression losses to optimize the model and achieve accurate key point detection.

Image preprocessing module

During the image preprocessing stage, three different resolutions are generated for the input image, with one resolution randomly selected during training. Furthermore, MobileNetV3, a lightweight object detection model, is utilized to segment the input image and identify the facial region. Standard data augmentation techniques, including random cropping, scaling, horizontal flipping, small-angle rotations

(± 10°), and color jittering, are then applied to the detected facial region. To further refine the preprocessing stage, a region enhancement module based on an attention mechanism is designed to dynamically adjust the weights assigned to different regions in the image, thereby improving the model's focus on crucial areas.

The accuracy of canthus detection is crucial for the diagnosis, treatment, and prognostic assessment of facial nerve movement disorders. Therefore, an optimization procedure is designed for ocular regions to enhance ocular feature representation in OFKNet[6]. Specifically, the periocular region is finely divided into six functional subregions through a region-attention module: the upper eyelid, lower eyelid, internal canthus, lateral canthus, central fissure, and peri-pupil region. Dynamic weighting of these subregions is implemented based on the five facial movements. Specifically, during neutral expression and cheek puffing, the overall periorbital changes are minimal, resulting in relatively balanced weights across subregions. Eyebrow-raising elicits the highest weight in the upper eyelid region, while smiling significantly increases the weights of the lower eyelid and lateral canthus regions. Nose scrunching primarily involves elevated weights in the central fissure and lateral canthus areas, and cheek puffing emphasizes the lower eyelid and lateral canthus. Notably, the internal canthus region consistently maintains the highest or relatively high weighting across all five movements. This series of operations helps establish the internal canthus region as a critical area, and targeted cropping is applied to ensure that the extracted image preserves essential features surrounding the canthus. Subsequent adjustments are made based on the precise canthus location, focusing on capturing deformed features in this region. During color adjustment, contrast and color saturation in the canthus region are selectively enhanced to improve the distinguishability of its color distribution. Additionally, an image blur operation is applied while maintaining relative clarity in the canthus region. For other key regions beyond the canthus, conventional enhancement techniques, including cropping, flipping, and color adjustment, are employed, albeit with slightly lower intensity compared to the canthus region. In contrast, only mild enhancement operations are applied to background areas to ensure that feature optimization remains concentrated on the internal canthus region.

Feature extraction module

After preprocessing, the input image is denoted as x_p, which undergoes multiple convolutional operations to extract features at different levels. ConvNeXt is a convolution-based architecture that integrates a hierarchical modular design and efficient computational strategies, enabling superior feature extraction[4]. It has demonstrated outstanding performance in image classification and object detection tasks[4]. However, fully connected layers flatten the feature map into a vector, disrupting the spatial structure information. Since spatial structure is essential for accurately localizing key points in the facial region, fully connected layers are removed to preserve the original spatial layout of the feature map. Additionally, pooling layers are eliminated because they reduce the resolution of the feature map and lead to the loss of fine-grained details from the original image. In ConvNeXt[4], the convolution operation of each layer can be expressed by a general convolution operation. For the convolution layer l, its input and output are denoted as F_{l-1} and F_l, respectively. The kernel size is $k \times k$, the stride of the convolution kernel is s, the padding at the edges of the input feature map is p, and the number of output channels is denoted as C_l. Then, the output of convolution layer l is

$$F_l(i,j,c) = \mathrm{ReLU}\left(\sum_{m=0}^{k-1}\sum_{n=0}^{k-1}\sum_{d=0}^{C_{l-1}} w_l(m,n,d,c)\cdot F_{l-1}(i+m\cdot s-p, j+n\cdot s-p, d)\right) \quad (1)$$

where $w_l(m, n, d, c)$ is the weight of the convolution kernel l, represented as a four-dimensional tensor. m and n are position indices within the $k \times k$ convolution kernel, d is the channel index of the input feature map, c is the channel index of the output feature map, and (i, j) denotes the spatial position of the feature map. ReLU(·) denotes the activation function, i.e., the rectified linear unit.

When considering the convolution layer's computation, the spatial size of the feature map is changed for each convolution operation[15]. Let $H_l \times W_l$ denote the output feature map size of layer l, and its relation to the previous layer's output feature map size is calculated as

$$H_l = \left\lfloor \frac{H_{l-1} - k + 2p}{s} \right\rfloor + 1 \tag{2}$$

$$W_l = \left\lfloor \frac{W_{l-1} - k + 2p}{s} \right\rfloor + 1 \tag{3}$$

OFKNet includes a bottleneck module, which improves the expressive capacity of the network by expanding and compressing the number of channels, enhancing the representation of characteristics while reducing the computational cost. The input feature of the bottleneck module is denoted as x_i. After a 1×1 convolutional layer, the number of channels is expanded to get the intermediate feature x_m. Then, the feature x_n is obtained after a 3×3 convolutional layer. Finally, the number of channels is restored through a 1×1 convolutional layer to obtain the output feature x_o:

$$x_m = \mathrm{Conv}_{1\times1}(x_i, C_i \times t) \tag{4}$$

$$x_n = \mathrm{Conv}_{3\times3}(x_m, C_i \times t, s = 1) \tag{5}$$

$$x_o = \mathrm{Conv}_{1\times1}(x_n, C_i) \tag{6}$$

where Ci is the number of input channels of x_i, and t is the expansion factor. ConvNeXt performs downsampling operations to convert high-resolution feature maps into low-resolution ones. Let F_h denote the input high-resolution feature map and F_l denote the output low-resolution ones. We obtain:

$$F_l(i, j, c) = \mathrm{Conv}(F_h(i \times r, j \times r, c), k, s = r) \tag{7}$$

where r denotes the down-sampling factor. Based on this operation, the network extracts features at multiple scales, generating feature maps F_3, F_4, and F_5 at different resolutions, better capturing facial key points of varying sizes.

Feature enhancement and fusion module

Based on the multi-scale feature maps F_3, F_4, and F_5 extracted by the ConvNeXt, an improved PAN is employed to integrate features of different scales. PAN effectively merges low-level and high-level features through top-down and bottom-up paths. The high-level feature map F_5 is upsampled to the size of F_4 and weighted summed with F_4 to obtain an intermediate feature map F_u:

$$F_u = \psi(F_5) + F_4 \tag{8}$$

where $\Psi(\cdot)$ denotes the upsample process. Similarly, the low-level feature map F_3 is upsampled to the size of F_4 and weighted summed with F_4 to get the feature map F_d:

$$F_d = \psi(F_3) + F_4 \tag{9}$$

Then, the top-down and bottom-up feature maps F_u and F_d are fused to obtain the final multi-scale feature map F_p:

$$F_P = F_u + F_d \tag{10}$$

In this case, PAN facilitates information flow between feature maps[16]. It enables features at different scales to complement each other, ensuring that the network considers both local details and global structural information. SE is then adopted by using PAN's output to further enhance the feature representation ability[17]. SE applies global average pooling (ϕ) to obtain global features for each channel. It then computes channel importance z_c and channel weights s_c using fully connected layers, *i.e.*,

$$z_c = \phi\left(F_{P,c}\right) \tag{11}$$

$$s_c = \mathrm{ReLU}\left(W_2 \cdot \mathrm{ReLU}\left(W_1 \cdot z_c\right)\right) \tag{12}$$

where W_1 and W_2 are the weight matrices of the fully connected layers. Each channel's weight s_c is multiplied with the input feature to obtain the final feature map F'_P:

$$FP' = FP \cdot sc \tag{13}$$

The combination of SE and PAN enhances the representation ability of the new feature map F'_P, allowing it to better adapt to facial key points at different scales and complexities.

Key points detection module

Based on the processed feature map F'_P, key point detection is performed using both the heatmap and the regression branches. The heatmap branch generates more discriminative features that produce accurate 2D Gaussian heatmaps, making the probability distribution of key points locations more precise. The regression branch learns the mapping between key points' coordinates and features, leading to more accurate key point coordinate detection. The heatmap branch detects the 2D Gaussian heatmap for each key point through convolution operations. It is assumed that the input feature F_p is denoted as x, and the heatmap is obtained through two convolution operations:

$$y_1 = \mathrm{ReLU}(\mathrm{Conv}(x, 3 \times 3, 512)) \tag{14}$$

$$y_2 = \mathrm{ReLU}\left(\mathrm{Conv}\left(y_1, 3 \times 3, 256\right)\right) \tag{15}$$

$$h = \mathrm{Sigmoid}\left(\mathrm{Conv}\left(y_2, 1 \times 1\right)\right) \tag{16}$$

where y_1 is obtained by subjecting x to a single convolutional transformation and a ReLU activation, and y_2 is obtained by subjecting y_1 to another convolution and a ReLU activation. h is the heatmap for each key point, ranging from [0,1], indicating the probability distribution of the point in space. The regression branch directly detects the (x, y) coordinates of key points. When the input features are y_1 and y_2, the coordinates of the key points are detected through the convolutional layer as $\hat{x}_i$ and $\hat{y}_i$:

$$\hat{x}_i, \hat{y}_i = \mathrm{Conv}\left(y_2, 1 \times 1\right) \tag{17}$$

The total loss function L_t consists of heatmap loss and regression loss. The heatmap loss L_h is based on the mean squared error, ensuring that the detected heatmap is close to the true heatmap. The regression loss function L_r ensures accurate coordinate detection, with weights w_i distinguishing the importance of different

key points. They are obtained as:

$$L_h = \frac{1}{N}\sum_{i=1}^{N}(\widehat{y}_i - y_i)^2 \tag{18}$$

$$L_r = \sum_{i=1}^{M} w_i\,(\widehat{x}_i - x_i)^2 \tag{19}$$

$$L_t = L_h + L_r \tag{20}$$

where x_i and y_i are the true coordinates of the samples, N represents the number of samples used for calculating the heatmap loss, and M indicates the number of samples used for calculating the regression loss.

RESULTS

Experimental setting and evaluation metrics

All experiments are conducted using PyTorch on a single NVIDIA RTX 4090 GPU (24 GB memory). The network backbone adopts ReLU activations in all intermediate layers, while the final output layer employs a Sigmoid activation to generate the predicted outputs. The model is optimized with Adam for 60 epochs using a learning rate of 10^{-3}, batch size of 32, and a dropout rate of 0.25. Our dataset comprises 60 clinically diagnosed patients with facial nerve dysfunction, including 32 males and 28 females. All participants are assessed using the H-BGS and classified as grades II to V. Inclusion criteria require patients to have a confirmed diagnosis of facial nerve dysfunction without other neurological or facial disorders. Exclusion criteria include patients who failed to complete all facial movement tests. The clinical severity is categorized as mild, moderate, severe, and very severe, with each category reasonably represented in the dataset. We adopt a patient-level stratified splitting strategy to ensure that data from the same patient appear in only one subset. The distribution consisted of 42 patients for training, 9 for validation, and 9 for testing. We maintain proportional representation of different severity levels across all subsets during this division. The distribution of patients across severity grades is balanced among the training, validation, and test sets. Additionally, we ensure that different facial movement types are evenly distributed across all three subsets to maintain dataset representativeness and training fairness.

Each participant completes five facial movement tests under standardized conditions: NeutralFace, EyebrowRaise, ShrugNose, Smile, BlowCheek. Video capture uses a high-speed camera at 30 frames per second with 4K resolution. Each action is recorded for 5 s, generating approximately 63,000 video frames in total. Additionally, four eye movement tasks are performed: spontaneous eye blink (Spontaneous Eye Blink), voluntary eye blink (Voluntary Eye Blink), soft eye closure (Soft Eye Closure), and forced eye closure (Forced Eye Closure). For spontaneous and voluntary eye blinks, we increase the capture rate to 120 frames per second, with each action recorded for approximately 10 s.

From the collected video frames, we apply a keyframe extraction strategy. Specifically, for the annotation of 40-point eye landmarks, we select keyframes based on the following criteria: peak frames showing changes in eyelid opening and closing states (fully open, half-closed, and fully closed); transitional frames during action changes (from static to dynamic states, from closed to open); and frames where bilateral asymmetry is most pronounced. For the 14-point eye landmark annotations, the extraction criteria are relatively relaxed: in addition to the above keyframes, we include the beginning, middle, and end frames of each action sequence, as well as all frames showing significant changes in eyelid movement. We ensure that the selected keyframes possess high representativeness for subsequent analysis. Following keyframe selection, we conducted preprocessing on all videos. Initially, we perform quality screening to eliminate frames that

are unsuitable for precise annotation due to blurriness, occlusion, or abnormal lighting conditions. This process yielded 3,600 frames with fine annotations of 40-point eye landmarks (approximately 60 frames per patient, covering key moments across different actions) and 5,400 frames with annotations of 14-point eye landmarks (approximately 90 frames per patient). Subsequently, two annotators with professional medical knowledge independently label all images. To evaluate annotation consistency, the inter-annotator agreement shows an average Euclidean distance of less than 2 pixels, demonstrating high consistency and reliability of the annotation results.

The main objective of this work is to evaluate the robustness and accuracy of different key point detection models for eye-related feature points under pathological conditions. Two manually annotated datasets are selected to achieve this: one with 14 eye feature points and another with 40 eye feature points[14,18,19]. Five different facial landmark detection models are selected for comparative analysis, including MediaPipe Face Landmarker[20,21], InsightFace[22,23], Dlib68[15,21] and Dlib81[15,21]. By assessing their performance in detecting eye feature points, we aim to assess their applicability and precision in diagnosing facial nerve dysfunction. However, the models' ability to process eye feature points varies significantly, particularly for datasets with different numbers of points. Specifically, InsightFace, Dlib68, and Dlib81 cannot detect all 40 eye feature points. To ensure fairness and comparability, a dataset with 14 eye feature points was also used for comparison with models that cannot handle the full 40 points. Detailed descriptions of the 14- and 40-point eye feature sets are provided in Table 1.

For the 40-feature-point dataset, OFKNet and MediaPipe Face Landmarker are evaluated, as they can detect a more detailed set of eye feature points, including the complete eye contour, pupil position, and other key points. In contrast, the dataset with 14 eye feature points is used for a broader comparison across all models. In particular, Dlib68 and Dlib81 do not detect the pupil center. To quantitatively compare detection accuracy, three evaluation metrics are employed: mean Euclidean distance (MED), normalized mean error (NME), and eye opening degree (EO). MED calculates the deviation between the detected and manually annotated key points using Euclidean distance, quantifying the error for each key point. NME extends MED by considering the influence of inter-individual eye size differences. It normalizes the error by using the intercanthal distance, eliminating individual differences[16,23]. EO is an application-oriented metric that quantifies eye closure by analyzing changes in the eye fissure area, aiding in model sensitivity assessment for eye state changes in facial nerve dysfunction diagnosis[17-20,22,24-28]. Specifically, MED is defined as:

$$\mathrm{MED} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\left(x_i^{\mathrm{p}} - x_i^{\mathrm{g}}\right)^2 + \left(y_i^{\mathrm{p}} - y_i^{\mathrm{g}}\right)^2} \tag{21}$$

where N is the total number of key points. $\left(x_i^{\mathrm{p}}, y_i^{\mathrm{p}}\right)$ denote the predicted coordinates of the key point i at a given frame. x_i^{g}, and y_i^{g} represent the corresponding ground truth coordinates. NME is defined as:

$$\mathrm{NME} = \frac{d_{\mathrm{a}}}{\mathrm{dist}\left(P_{\mathrm{l}}, P_{\mathrm{r}}\right)} \tag{22}$$

where d_a is the average predicted deviation for the given image, P_{l} and P_{r} are the coordinates of the left and right internal canthus points, and $\mathrm{dist}(P_{\mathrm{l}}, P_{\mathrm{r}})$ is the Euclidean distance between the internal canthus points of both eyes. Finally, EO is computed as:

$$\mathrm{EO} = \frac{S\left(P_1, P_2, \ldots, P_{17}\right)}{\left|P_1 - P_{10}\right|^2} \tag{23}$$

Table 1. Description of eye feature point indices for two datasets

14 eye feature points		14 eye feature points	
Index	**Description**	**Index**	**Description**
0-5	Right eye contour	0	Right internal canthus
6-11	Left eye contour	1-18	Right eye contour
12	Right pupil center	19	Left internal canthus
13	Left pupil center	20-37	Left eye contour
-	-	38	Right pupil center
-	-	39	Left pupil center

where P_1 is the external canthus point, P_{10} is the internal canthus point, and $S(P_1, P_2, \ldots, P_{17})$ is the area of the eye fissure formed by these points.

Experiment results

One patient is selected to show eyebrow raise and voluntary blink actions as experimental data. Figure 3A and B compares the errors between the detected 40-point eye landmarks and the manually annotated points for OFKNet and MediaPipe Face Landmarker. It is shown that the results of the two models in detecting the eye contour and pupil location are different, with OFKNet exhibiting higher detection accuracy.

Figure 4A-E presents the detection results of different models using the 14-point eye feature dataset. OFKNet and MediaPipe Face Landmarker demonstrate higher accuracy, particularly in detecting the pupil center and the internal and external canthus points.

To complement the single-image overlays in Figures 3-4, we present dataset-level mean NME heatmaps for the eye landmarks in Figure 5A and B and Figure 6A-E. Each panel visualizes the canonical mean landmark positions after intercanthal alignment. The color and size of each marker encode the NME at that landmark (blue means low error, red means high error). The heatmaps show that OFKNet produces tightly localized errors with consistently low mean NME across both the 14-point and 40-point eye landmark sets, whereas competing methods exhibit higher errors at specific periocular landmarks, particularly near the pupil centers and the internal canthus.

For completeness, Tables 2 and 3 report the landmark-wise median NME under the 40-point and 14-point eye landmark protocols, respectively. These tables reveal that the largest discrepancies between OFKNet and the baseline models are concentrated at the pupil centers and around the internal canthus, rather than being uniformly distributed along the eyelid contours. Competing models exhibit notably higher errors at these locations, whereas OFKNet maintains consistently low NME. This observation is consistent with the fact that pupil centers are highly sensitive to specular highlights and low iris-sclera contrast, and that the internal canthus is frequently affected by eyelid occlusions.

Compared with existing facial landmark detection architectures, OFKNet is specifically designed for fine-grained analysis of ocular movements in neurological dysfunction assessment. MediaPipe Face Landmarker maintains high-resolution feature representations, but its parallel multi-branch design results in high computational overhead, limiting its practicality in clinical environments requiring lightweight and real-time inference. InsightFace is optimized for identity recognition and global facial embeddings and thus lacks the sensitivity to subtle ocular movement variations and local asymmetry, making it inadequate for capturing clinically relevant neuromuscular dysfunctions. To address these challenges, OFKNet focuses on high-precision ocular landmark representation by employing a denser, anatomically-informed annotation

Table 2. Landmark-wise median NME for 40-point eye landmarks

Index	Name	OFKNet	MediaPipe Face Landmarker
0	Right eye inner canthus	0.014625	0.069888
1	Right eye fissure pt1	0.025003	0.131626
2	Right eye fissure pt2	0.023567	0.143235
3	Right eye fissure pt3	0.019327	0.079514
4	Right eye fissure pt4	0.022034	0.034320
5	Right eye fissure pt5	0.021943	0.067446
6	Right eye fissure pt6	0.025659	0.033429
7	Right eye fissure pt7	0.029768	0.081202
8	Right eye fissure pt8	0.030149	0.102596
9	Right eye fissure pt9	0.032769	0.094153
10	Right eye fissure pt10	0.047507	0.071380
11	Right eye fissure pt11	0.029674	0.068432
12	Right eye fissure pt12	0.025813	0.066482
13	Right eye fissure pt13	0.023307	0.053581
14	Right eye fissure pt14	0.018082	0.062717
15	Right eye fissure pt15	0.020398	0.072940
16	Right eye fissure pt16	0.016673	0.045427
17	Right eye fissure pt17	0.019200	0.098480
18	Right eye fissure pt18	0.027434	0.125767
19	Left eye inner canthus	0.017090	0.060292
20	Left eye fissure pt1	0.025160	0.119178
21	Left eye fissure pt2	0.026532	0.125558
22	Left eye fissure pt3	0.024732	0.072863
23	Left eye fissure pt4	0.025440	0.049547
24	Left eye fissure pt5	0.021750	0.063389
25	Left eye fissure pt6	0.026195	0.048040
26	Left eye fissure pt7	0.031359	0.098658
27	Left eye fissure pt8	0.034447	0.125999
28	Left eye fissure pt9	0.039735	0.114241
29	Left eye fissure pt10	0.042385	0.084921
30	Left eye fissure pt11	0.036168	0.079271
31	Left eye fissure pt12	0.034386	0.072576
32	Left eye fissure pt13	0.027273	0.055829
33	Left eye fissure pt14	0.023680	0.070046
34	Left eye fissure pt15	0.020342	0.074843
35	Left eye fissure pt16	0.023130	0.053575
36	Left eye fissure pt17	0.021925	0.095407
37	Left eye fissure pt18	0.022810	0.114364
38	Right eye pupil center	0.015409	0.040883
39	Left eye pupil center	0.020024	0.048256

OFKNet: Ocular-enhanced Face Keypoints Net.

scheme with 40 key points in the periorbital region. This fine-grained key point layout enables more accurate characterization of eyelid contours, canthus morphology, and pupillary kinematics, significantly enhancing sensitivity to subtle neuromuscular impairments. Building on this design philosophy, OFKNet incorporates three architectural improvements to further enhance performance. To evaluate the effectiveness of each component in OFKNet, we conduct a series of ablation experiments by progressively

Table 3. Landmark-wise median NME for 14-point eye landmarks

Index	Name	OFKNet	MediaPipe Face Landmarker	InsightFace	Dlib68	Dlib81
0	Right eye inner canthus	0.027655	0.051430	0.351867	0.595532	0.612037
1	Right eye fissure pt1	0.027297	0.033783	0.301907	0.236468	0.248296
2	Right eye fissure pt2	0.023986	0.030429	0.417628	0.254163	0.245944
3	Right eye fissure pt3	0.054597	0.049162	0.356569	0.617627	0.623798
4	Right eye fissure pt4	0.031232	0.038899	0.363480	0.233890	0.240538
5	Right eye fissure pt5	0.029241	0.032880	0.288261	0.245039	0.248924
6	Left eye inner canthus	0.040101	0.024548	0.355001	0.048595	0.065375
7	Left eye fissure pt1	0.026622	0.106866	0.393546	0.023011	0.027597
8	Left eye fissure pt2	0.024279	0.046832	0.353652	0.035522	0.047803
9	Left eye fissure pt3	0.055815	0.062109	0.387172	0.039402	0.064830
10	Left eye fissure pt4	0.031324	0.044572	0.287926	0.035347	0.060404
11	Left eye fissure pt5	0.024939	0.031595	0.407323	0.026595	0.043264
12	Right eye pupil center	0.018353	0.034018	0.195228		
13	Left eye pupil center	0.023107	0.036842	0.213492		

NME: Normalized mean error; OFKNet: Ocular-enhanced Face Keypoints Net.

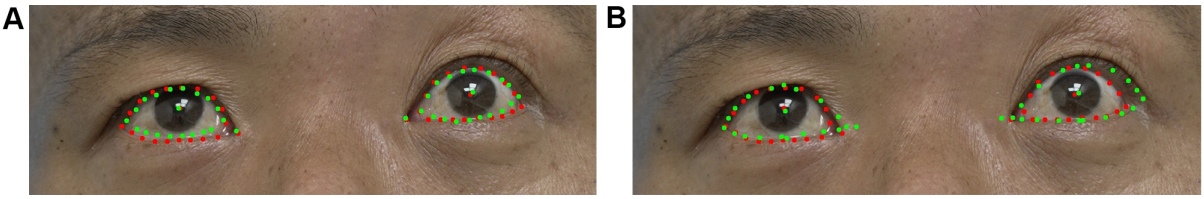

Figure 3. Error comparison for 40-point eye landmarks detection. Red points indicate the ground truth (manual annotation), while green points represent the model's predictions. (A) OFKNet; (B) MediaPipe Face Landmarker.

incorporating the proposed modules into the baseline framework. Specifically, ConvNeXt is adopted as the baseline feature extraction network, upon which we sequentially add the MobileNetV3-based local geometric enhancement module and the SE-PAN cross-scale adaptive feature fusion module. This results in three model variants: (a) ConvNeXt backbone; (b) ConvNeXt + MobileNetV3; and (c) ConvNeXt + MobileNetV3 + SE-PAN (i.e., the full OFKNet). The quantitative results are summarized in Table 4.

Table 4 shows that the ConvNeXt backbone exhibits relatively limited performance on the eye landmark detection task, particularly on the 40-point fine-level landmark set. This suggests that relying solely on high-level semantic features is insufficient for capturing the subtle geometric variations present in the eye region. Introducing the MobileNetV3 enhancement module yields a notable reduction in NME for both 40-point and 14-point settings. This improvement demonstrates that local geometric and contour-aware feature amplification is crucial for modeling the eyelid curvature, palpebral fissure shape, and canthus detail, thereby improving localization accuracy. Further incorporating the SE-PAN module leads to consistent performance gains. SE-PAN guides the network to adaptively fuse multi-scale features based on structural importance, enabling the model to dynamically focus on the most discriminative local regions under varying eye-opening states and individual anatomical differences. This confirms the effectiveness of the proposed cross-scale attention-guided fusion strategy. In summary, the ablation results indicate that the performance improvement of OFKNet does not stem merely from backbone selection or increased network capacity.

Table 4. Ablation study of OFKNet

	40-point eye landmark NME			14-point eye landmark NME		
	Median	**Mean**	**Std**	**Median**	**Mean**	**Std**
(a): ConvNeXt backbone	0.090576	0.094834	0.018902	0.060815	0.062735	0.016814
(b): (a) + MobileNetV3	0.052794	0.055437	0.012646	0.044178	0.045813	0.012136
(c): **(b) + SE-PAN**	**0.026285**	**0.028996**	**0.009228**	**0.033222**	**0.034307**	**0.00943**

OFKNet: Ocular-enhanced Face Keypoints Net; NME: normalized mean error; Std: standard deviation. Bold format indicates the whole part of the model.

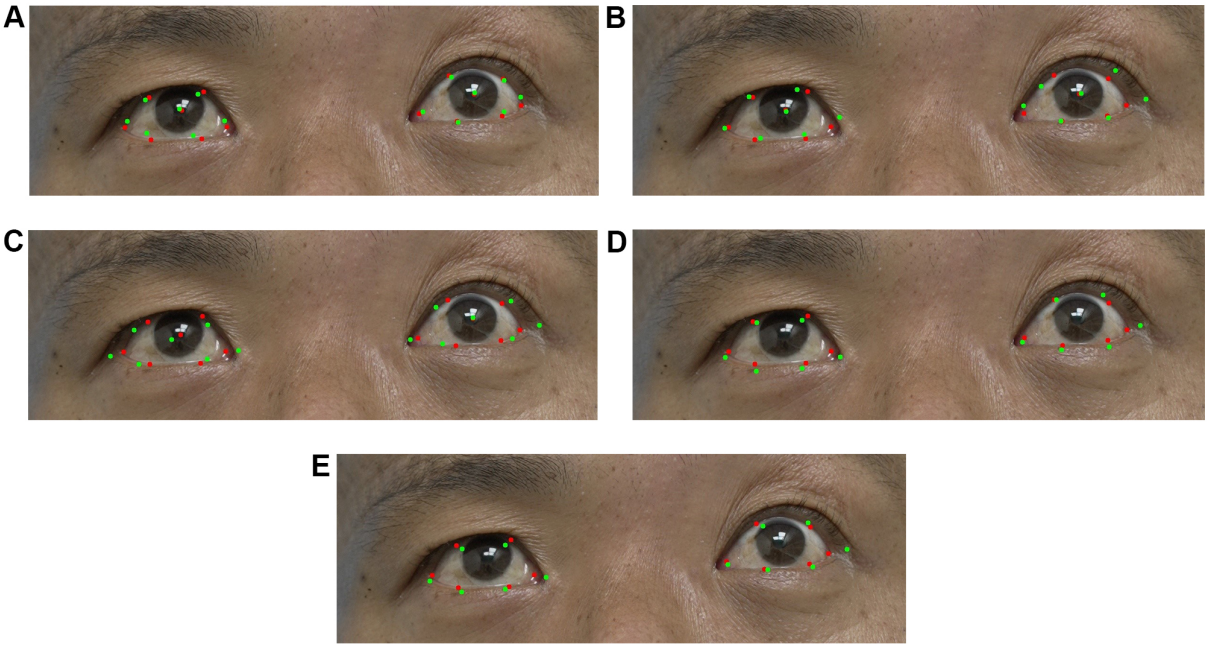

Figure 4. Error comparison for 14-point eye landmarks detection. Red points indicate the ground truth (manual annotation), while green points represent the model's predictions. (A) OFKNet; (B) MediaPipe Face Landmarker; (C) InsightFace; (D) Dlib68; (E) Dlib81. OFKNet: Ocular-enhanced Face Keypoints Net.

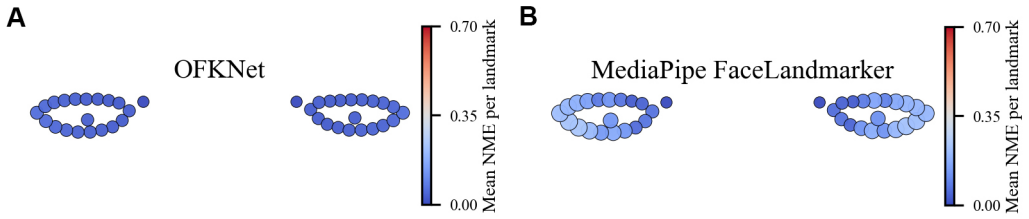

Figure 5. Mean error distribution of eye landmarks (40 points). (A) OFKNet; (B) MediaPipe Face Landmarker. OFKNet: Ocular-enhanced Face Keypoints Net.

Figure 7A shows the cumulative error distribution curve for each model in the 14-eye feature points detection task, while Figure 7B presents the corresponding results for the 40 eye feature points. It is shown that the iteration curves for OFKNet and MediaPipe Face Landmarker are relatively smooth with low error, with OFKNet being the best. In contrast, comparative models show larger errors and a faster increase in error. Table 5 summarizes the per-image NME statistics for all models under the 40-point and 14-point eye landmark settings. For each baseline model, we perform a two-sided paired comparison against OFKNet on

Table 5. NME for eye landmark detection across different models and datasets

Model	40-point eye landmark NME				14-point eye landmark NME			
	Median	Mean	Std	*P*-value	Median	Mean	Std	*P*-value
OFKNet	0.026285	0.028996	0.009228	< 0.001	0.033222	0.034307	0.00943	< 0.001
MediaPipe Face Landmarker	0.080603	0.083615	0.01862	< 0.001	0.045109	0.049395	0.015408	< 0.001
InsightFace	-	-	-	-	0.335803	0.354827	0.081584	< 0.001
Dlib68	-	-	-	-	0.208735	0.203527	0.022459	< 0.001
Dlib81	-	-	-	-	0.218109	0.216568	0.01873	< 0.001

NME: Normalized mean error; Std: standard deviation.

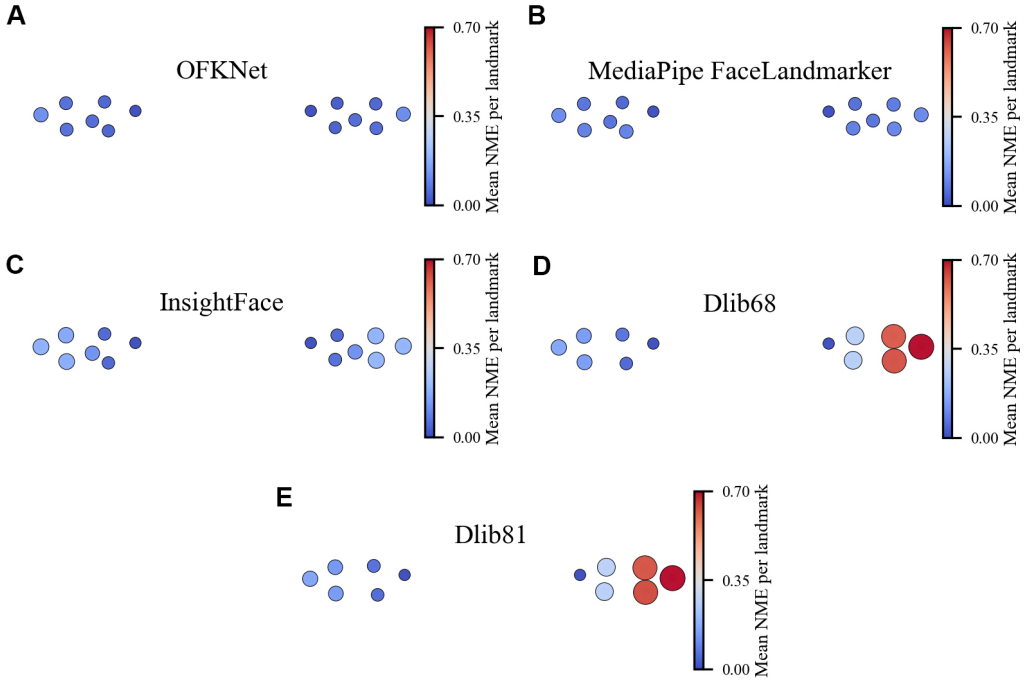

Figure 6. Mean Error Distribution of Eye Landmarks (14 points). (A) OFKNet; (B) MediaPipe Face Landmarker; (C) InsightFace; (D) Dlib68; (E) Dlib81. OFKNet: Ocular-enhanced Face Keypoints Net.

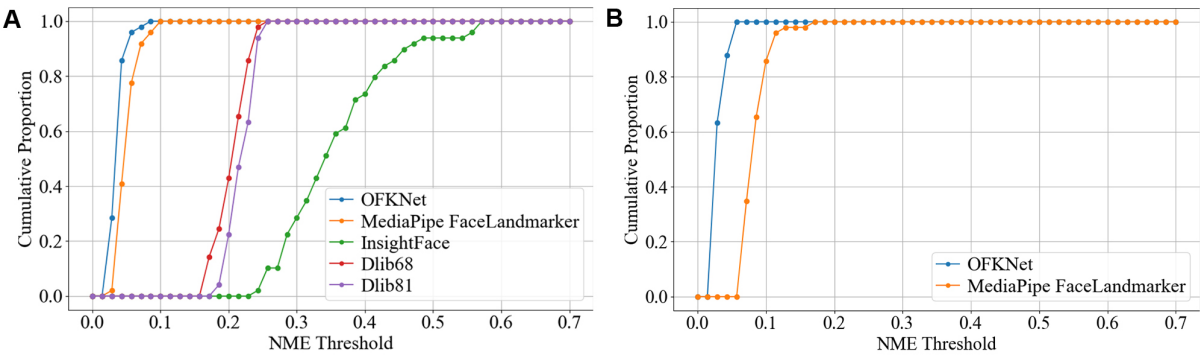

Figure 7. Cumulative error distribution curve for 14 and 40-point eye landmarks. (A) 14 eye feature points; (B) 40 eye feature points.

the common image subset. Two-sided paired Wilcoxon signed-rank tests between OFKNet and each

baseline model on the common image subset using per-image NME as paired samples. Holm correction is applied to control the family-wise error rate across multiple baseline comparisons. The resulting p-values are reported in the rightmost columns. For both protocols, OFKNet achieves the lowest median NME, and all pairwise comparisons against OFKNet are statistically significant ($P < 0.001$), confirming that the improvements are not due to random variation. Beyond accuracy, we also benchmark the efficiency of all models in terms of model size (MB) and inference time (ms). The results are summarized in Table 6.

The EO is a key indicator of the severity of facial nerve dysfunction, providing an objective measure of eyelid closure for each eye. Figure 8(A–E) illustrates the EO degree over time for representative voluntary blinking sequences, comparing OFKNet with all four baseline models. It is shown that OFKNet produces a more precise and nuanced eye-opening degree curve than other methods. In all cases, the OFKNet curve closely tracks the subtle changes in eyelid position throughout the blink cycle, including minor partial closures and re-openings, resulting in a smooth, detailed waveform for the eye-opening degree. In contrast, the compared methods yield coarser or less responsive curves. Their outputs tend to miss or flatten minor variations in eye-opening, showing only significant dips corresponding to complete blinks but not finer oscillations during slight or incomplete blinks. These visual comparisons highlight the superiority of OFKNet in capturing minor variations in eye-opening degrees during voluntary blinking.

In addition, offering a data-driven representation of the eye's opening and closing states enables a quantitative assessment of eye-opening ability throughout the blinking cycle. This measurement is applicable across different blinking stages and facilitates bilateral comparisons between the left and right eyes, thereby characterizing neural motor function asymmetries. The dynamic variations in EO degree over time offer a clear visual representation of facial nerve impairment and recovery, making it particularly valuable in the clinical assessment of facial nerve dysfunction. In this work, we dynamically measure the EO degrees of both eyes in patients exhibiting significant facial nerve dysfunction symptoms. To enhance interpretability, we generate time-series graphs to illustrate temporal changes in EO degree. Furthermore, keyframes extracted from patient data highlight the accuracy and reliability of the measurements, reinforcing their potential utility in neurological evaluation and treatment monitoring. Therefore, we conduct measurements of four characteristic eye movements in patients with facial nerve dysfunction: Spontaneous Eye Blink, Voluntary Eye Blink, Soft Eye Closure and Forced Eye Closure. These movements effectively simulate the majority of eye activities performed by individuals with facial paralysis in daily life, enabling a comprehensive analysis of EO degree variations in both eyes. A detailed introduction to each movement is provided below.

• Spontaneous Eye Blink refers to the involuntary blinking of the eyes that occurs without external stimuli or conscious control. In this work, patients are instructed to relax their facial muscles through verbal cues, allowing them to blink naturally for over 10 s while maintaining a relaxed state.

• Voluntary Eye Blink refers to a consciously controlled and deliberately initiated blink. In this study, patients are instructed to keep their heads still, maintain direct eye contact with the camera, and blink three times consecutively.

• Soft Eye Closure is a natural, unconscious action, and neurological abnormalities often first manifest in such subtle movements. By assessing the ability to close the eyes softly, we can preliminarily evaluate whether the orbicularis oculi muscle is functioning properly.

Table 6. Model efficiency and complexity comparison

Model	Model size (MB)	Inference time (ms)
OFKNet	158.56	713.38
MediaPipe Face Landmarker	3.58	9.87
InsightFace	20.94	61.68
Dlib68	95.08	1,597.16
Dlib81	18.83	1,601.14

OFKNet: Ocular-enhanced Face Keypoints Net.

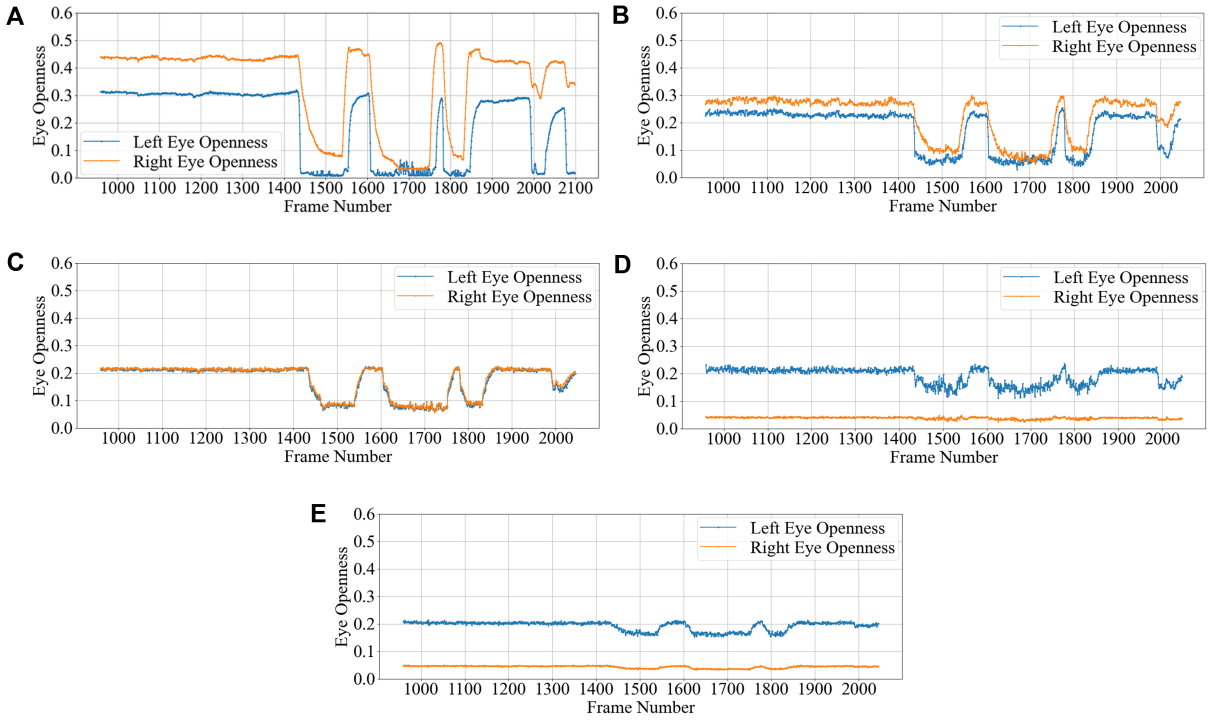

Figure 8. Comparison of eye opening degree curves for different models during voluntary blinking. (A) OFKNet; (B) MediaPipe Face Landmarker; (C) InsightFace; (D) Dlib68; (E) Dlib81. OFKNet: Ocular-enhanced Face Keypoints Net.

• Forced Eye Closure is a conscious action that serves as an assessment of the strength of the eyelid muscles. In patients with nerve damage or muscle weakness, insufficient muscle tension may be observed during forceful eye closure, resulting in incomplete eyelid closure or difficulty closing the eyes. By having the patient perform a forceful eye closure, we can evaluate the intensity of the action and analyze the speed, symmetry, and smoothness of the closure through video analysis.

The line charts derived from the EO degree data clearly illustrate the temporal variation in eye openness under different eye movement tasks. In each chart, the x-axis denotes the video frames, while the y-axis represents the EO degree. The blue and red lines correspond to the EO degrees of the left and right eyes, respectively. These plots reflect the patient's ability to control facial muscles and serve as a basis for the quantitative assessment of facial nerve dysfunction. From these line charts, one can intuitively observe dynamic changes in EO degrees across time during various standard eye movements. Notably, at specific moments, the difference between the EO degrees of the two eyes reaches a maximum, indicating a marked asymmetry in binocular motor function. The most pronounced asymmetry consistently occurs during the

peak amplitude phase of eyelid closure and exhibits a reproducible pattern across different patients and tasks. These phases, serving as spatiotemporal windows where features of muscle weakness or paralysis are most evident, hold important diagnostic value. Identifying and selecting these time points as key frames facilitates accurate characterization of the condition's severity and provides reliable reference data for clinical evaluation. To further validate this observation, we extracted key frames from three patients diagnosed with facial nerve dysfunction across four movement types and visualized their EO degree time series. Figures 9-20 show that in each case, one eye is fully closed while the contralateral eye remains partially open. This is consistent with the typical clinical manifestations of facial nerve dysfunction.

This phenomenon carries two important academic implications. First, it may extend beyond the scope of traditional clinical observation; our method is capable of effectively capturing and documenting it, thereby addressing a gap in current diagnostic and therapeutic approaches within specific observational dimensions. Second, even if this phenomenon has been qualitatively recognized in clinical practice, conventional methods often lack the precision and reproducibility required for quantitative analysis, which constrains the development and application of targeted treatment strategies. Our approach overcomes these limitations by enabling objective and repeatable assessment, thereby offering novel support for the refinement and scientific advancement of clinical diagnosis and intervention.

DISCUSSION

From a practical standpoint, OFKNet supports real-time processing on standard clinical hardware, making it readily deployable within routine examinations. Beyond feasibility, several implications of our findings deserve emphasis. Current landmark models often struggle with ocular subtleties because the periocular region contains small, rapidly deforming structures and clinically important areas such as the internal canthus are prone to occlusion and low contrast. OFKNet's fine-grained 40-point design and region-enhanced representation directly address these challenges by allocating greater sensitivity to these difficult zones. Fine-grained EO tracking may also integrate naturally into clinical workflows. EO trajectories offer objective measures of blink completeness and interocular asymmetry typically judged subjectively. These quantitative indicators could supplement existing grading scales such as House-Brackmann and Sunnybrook by providing more continuous, reproducible assessments of eyelid function. Finally, OFKNet's region-specific modeling aligns with clinical heuristics, as neurologists commonly rely on internal-canthus behavior to detect early or subtle muscle weakness.

Limitations

Our analysis focuses on the ocular region. While OFKNet can detect keypoints across the entire face, we have not yet fully examined how features from other facial areas, such as the mouth corners or forehead, might contribute to neurological assessment. Given that facial nerve dysfunction affects multiple muscle groups, a more comprehensive analysis integrating these regions could yield richer diagnostic insights. In addition, testing the system's performance with standard medical imaging equipment would help determine its practical deployment potential. Finding ways to reduce hardware requirements without sacrificing accuracy represents an important direction for future development.

We demonstrate OFKNet's effectiveness in facial paralysis assessment, but its utility for other neurological conditions (Parkinson's disease and Alzheimer's disease) remains to be established. Different neurological disorders may manifest distinct ocular movement patterns, requiring algorithm adaptations and dedicated clinical validation. Additionally, longitudinal studies tracking patients over extended periods would clarify whether this approach can effectively monitor disease progression and treatment response.

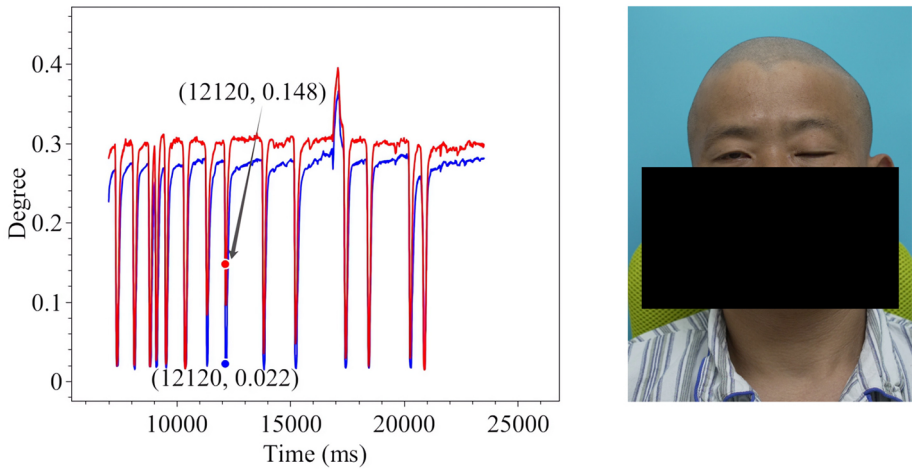

Figure 9. Spontaneous Eye Blink in Patient 1.

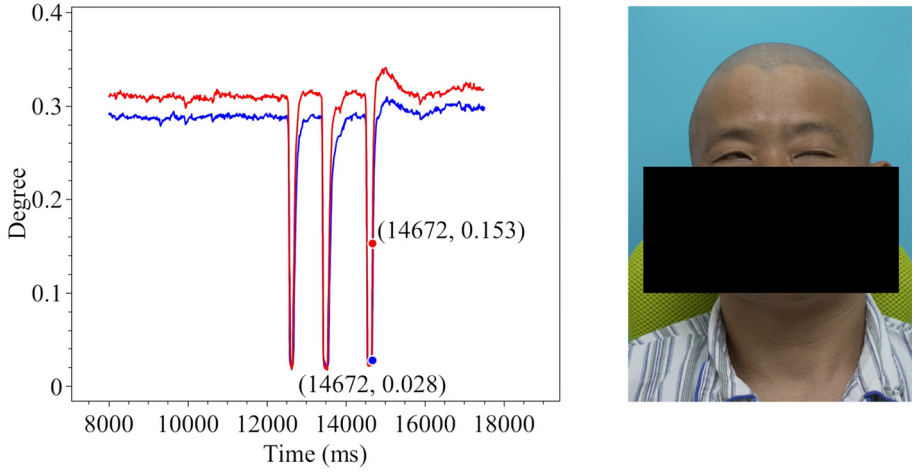

Figure 10. Voluntary Eye Blink in Patient 1.

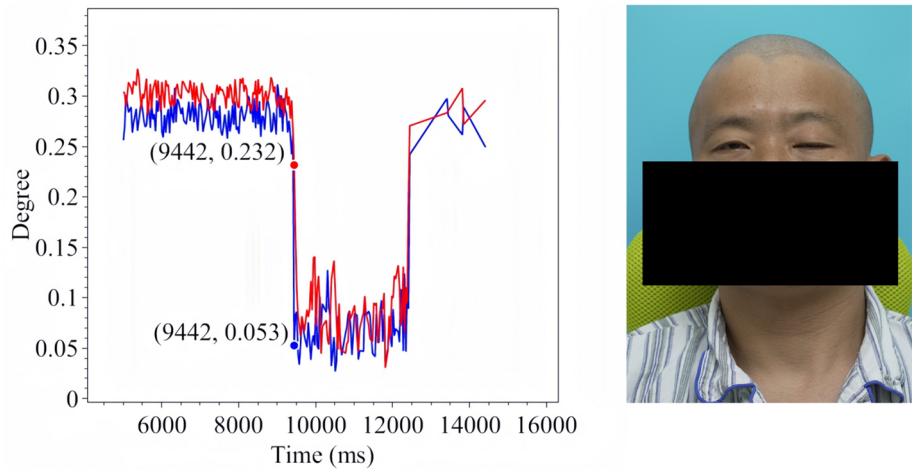

Figure 11. Soft Eye Closure in Patient 1.

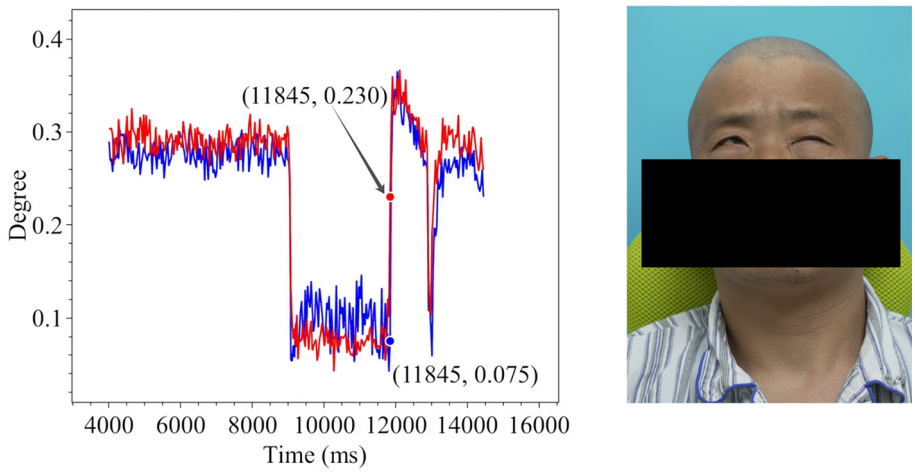

Figure 12. Forced Eye Closure in Patient 1.

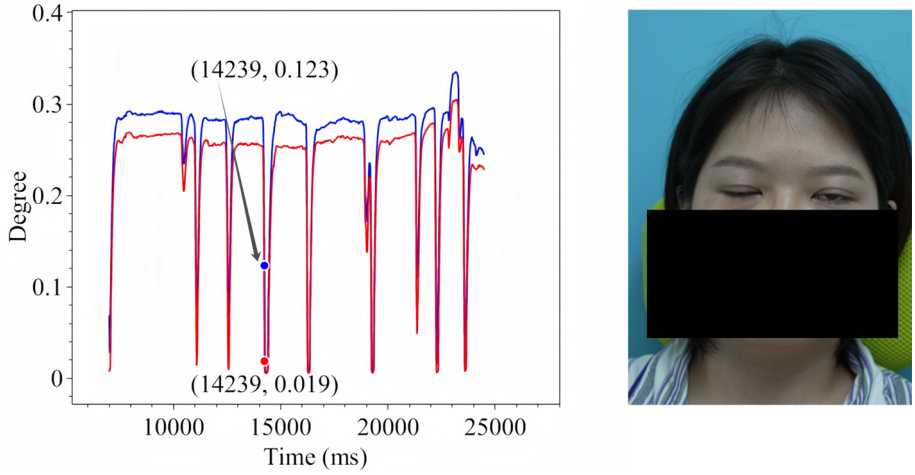

Figure 13. Spontaneous Eye Blink in Patient 2.

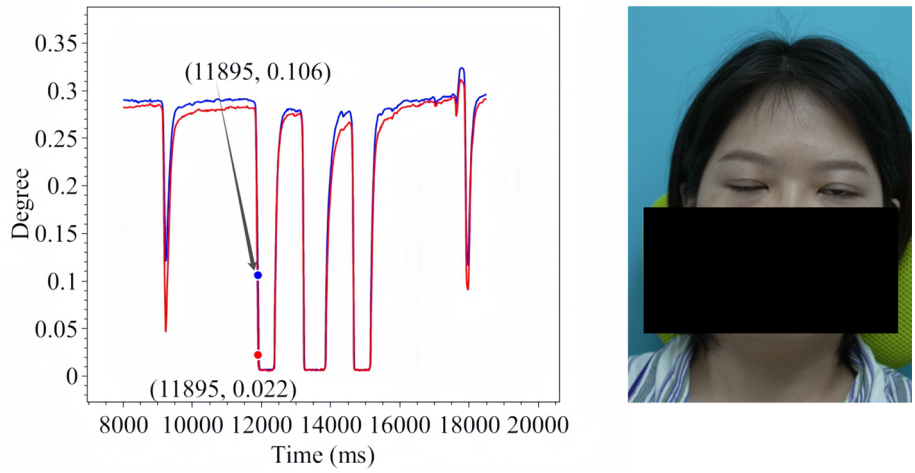

Figure 14. Voluntary Eye Blink in Patient 2.

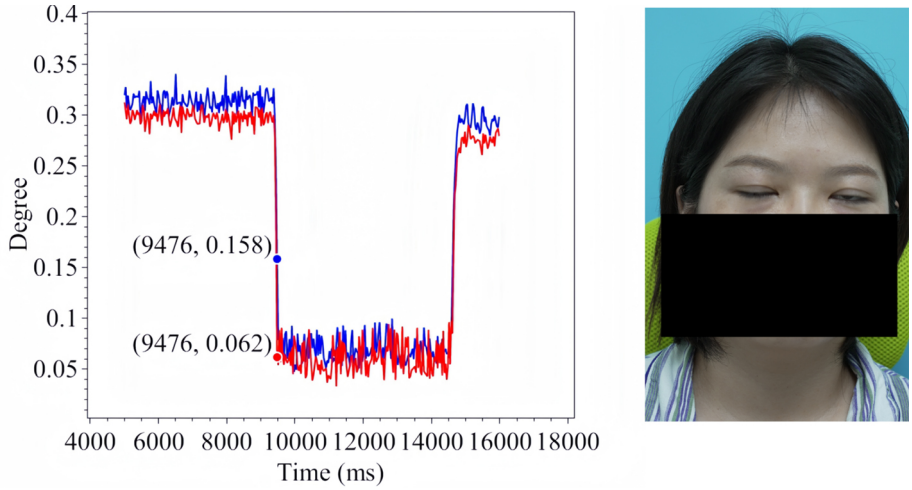

Figure 15. Soft Eye Closure in Patient 2.

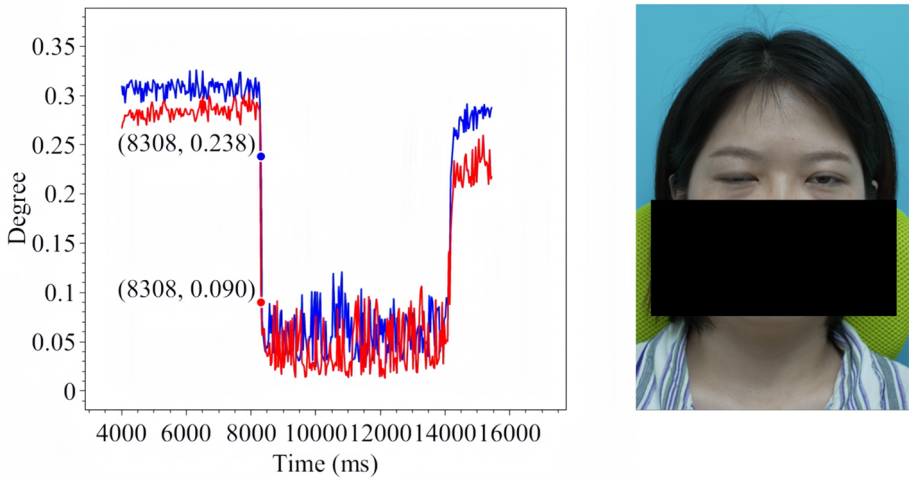

Figure 16. Forced Eye Closure in Patient 2.

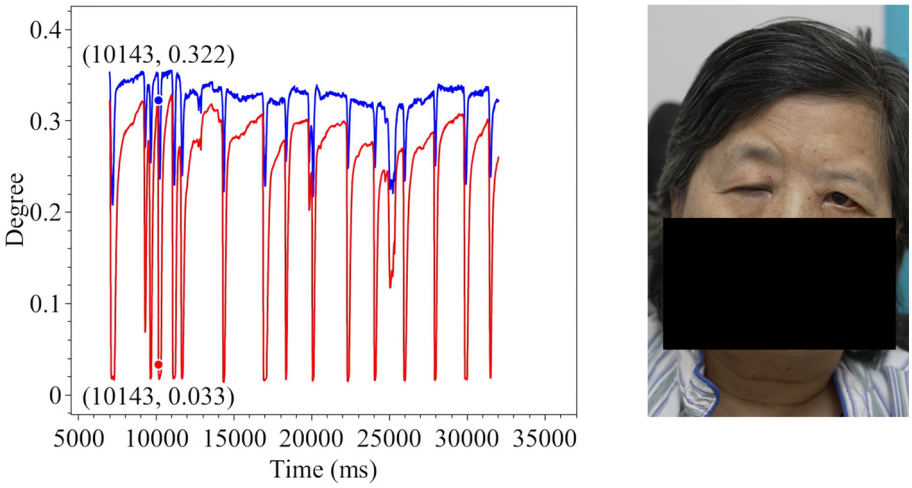

Figure 17. Spontaneous Eye Blink in Patient 3.

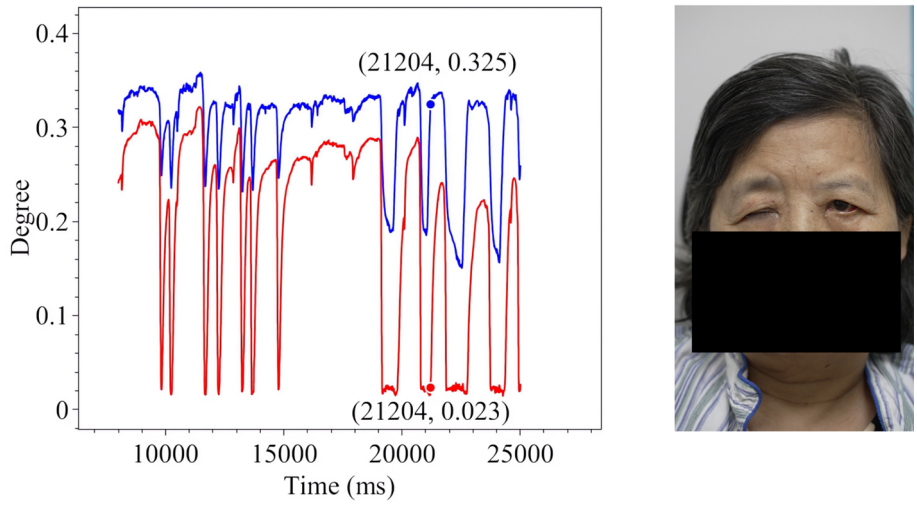

Figure 18. Voluntary Eye Blink in Patient 3.

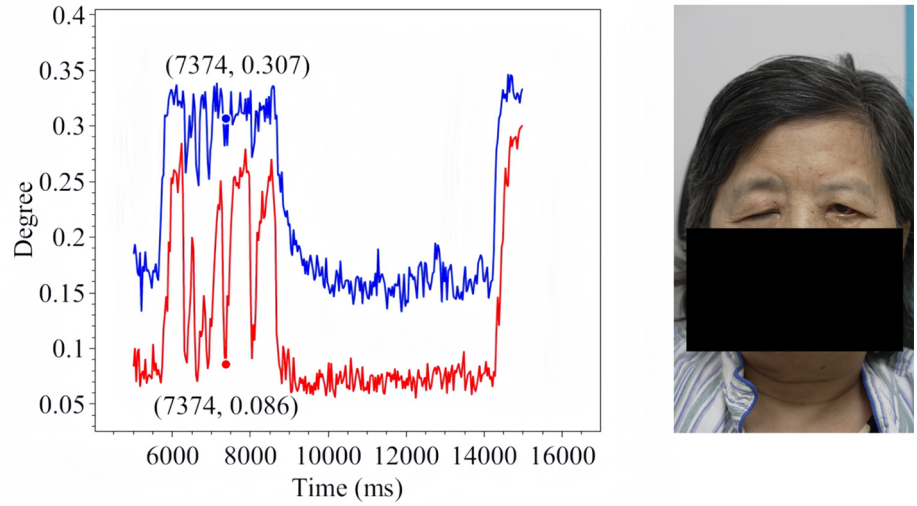

Figure 19. Soft Eye Closure in Patient 3.

In addition, our study has several limitations arising from the controlled nature of data collection. All recordings are captured under standardized indoor clinical conditions with stable lighting, fixed patient positioning, and high-resolution imaging equipment. As a result, the current evaluation does not fully reflect factors such as non-uniform illumination. Furthermore, the patient cohort consisted primarily of East Asian individuals, and periocular morphology varies substantially across ethnic groups. Therefore, cross-ethnic validation is required to establish generalizability. Finally, because the dataset is derived from continuous video sequences, the frame-level redundancy inherent in temporal sampling may introduce subtle correlations that increase the risk of model overfitting, even though patient-level data splitting is applied to prevent identity leakage.

In conclusion, this work presents OFKNet, a fine-grained ocular landmark detection framework that substantially improves the accuracy of eye-related keypoint localization during dynamic eyelid motion. Beyond technical performance, OFKNet provides quantitative eye-opening trajectories and visualized asymmetry indices that can support clinical decision-making. These measurements help clinicians evaluate

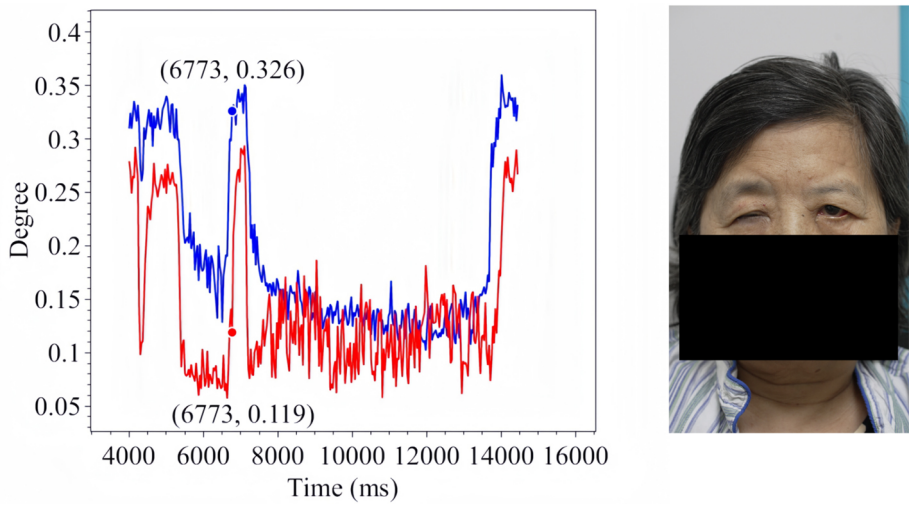

Figure 20. Forced Eye Closure in Patient 3.

blink completeness, eyelid closure fatigue, and interocular imbalance, which are important factors for treatment planning, prognostic assessment, and follow-up monitoring. OFKNet operates in real time on standard clinical computing hardware, indicating that it can be deployed within existing examination rooms without major infrastructure changes. This supports practical integration into routine neurological assessments and facial nerve clinics.

Future work will focus on several translational directions. The next step is prospective multicenter validation to ensure robustness across diverse imaging devices, ethnic groups, and recording environments, which is necessary for broader clinical adoption and regulatory evaluation. Quantitative blink dynamics obtained from OFKNet may also be incorporated into current grading systems to complement House-Brackmann or Sunnybrook scales with objective and continuous measurements. In addition, expanding analysis beyond the ocular region and integrating multimodal signals such as surface electromyography, electroencephalography, or professional eye-tracking will help establish a more comprehensive and data-driven framework for neurological function assessment.

DECLARATIONS

Acknowledgments

Authors' contributions

Conceptualization, investigation, methodology, software, validation, visualization, writing: Wang Z, Bi J
Conceptualization, data curation, review and editing: Zhao X, Zheng Z, Cui J, Cui R
Conceptualization, writing-review: Zhang J, Tang Y
Conceptualization, writing-review, supervision: Liang J
Conceptualization, data curation, writing-review, supervision: Zhou K, Zhang J
All authors participated equally in the research design, data analysis, and manuscript preparation.

Availability of data and materials

Materials will be made available upon request.

Financial support and sponsorship

This work was supported by the Beijing Natural Science Foundation under Grants L233005 and 4232049, and by the National Natural Science Foundation of China under Grants 62173013 and 62473014.

Conflicts of interest

Zhao X and Zhou K are affiliated with Beijing Neurorient Technology Co., Ltd. The authors declare that the company had no influence on the study design, data collection, analysis, interpretation, or decision to publish, and that no other competing interests exist.

Ethical approval and consent to participate

The study was approved by the Ethics Committee of Xuanwu Hospital, Capital Medical University, Beijing, China (Linyan Review 2021[221], Project Number: KD2021221). Written informed consent was obtained from all participants to ensure compliance with ethical standards for research involving human subjects.

Consent for publication

All participants whose images are shown in this study provided written informed consent for publication.

Copyright

REFERENCES

1. Skaramagkas V, Pentari A, Kefalopoulou Z, Tsiknakis M. Multi-modal deep learning diagnosis of Parkinson's disease-a systematic review. *IEEE Trans Neural Syst Rehabil Eng.* 2023;31:2399-423. DOI PubMed
2. Lou J, Yu H, Wang FY. A review on automated facial nerve function assessment from visual face capture. *IEEE Trans Neural Syst Rehabil Eng.* 2020;28:488-97. DOI PubMed
3. Liu X, Jin Y, Li X, Wu M, Guo Y. Facial paralysis evaluation based on improved residual network. In: 2023 2nd International Conference on Advanced Sensing, Intelligent Manufacturing (ASIM); 2023 May 12-14; Changsha City, China. New York: IEEE; 2023. pp. 36-40. DOI
4. Liu X, Xia Y, Yu H, Dong J, Jian M, Pham TD. Region based parallel hierarchy convolutional neural network for automatic facial nerve paralysis evaluation. *IEEE Trans Neural Syst Rehabil Eng.* 2020;28:2325-32. DOI PubMed
5. Ge X, Jose JM, Wang P, Iyer A, Liu X, Han H. ALGRNet: multi-relational adaptive facial action unit modelling for face representation and relevant recognitions. *IEEE Trans Biom Behav Identity Sci.* 2023;5:566-78. DOI
6. Zhang Y, Ding L, Xu Z, et al. The feasibility of an automatical facial evaluation system providing objective and reliable results for facial palsy. *IEEE Trans Neural Syst Rehabil Eng.* 2023;31:1680-6. DOI PubMed
7. Zhang Y, Gao W, Yu H, Dong J, Xia Y. Artificial intelligence-based facial palsy evaluation: a survey. *IEEE Trans Neural Syst Rehabil Eng.* 2024;32:3116-34. DOI PubMed
8. Xia Y, Nduka C, Yap Kannan R, Pescarini E, Enrique Berner J, Yu H. AFLFP: a database with annotated facial landmarks for facial palsy. *IEEE Trans Comput Soc Syst.* 2023;10:1975-85. DOI
9. Jin B, Cruz L, Gonçalves N. Pseudo RGB-D face recognition. *IEEE Sensors J.* 2022;22:21780-94. DOI
10. Jin B, Gonçalves N, Cruz L, Medvedev I, Yu Y, Wang J. Simulated multimodal deep facial diagnosis. *Expert Syst Appl.* 2024;252:123881. DOI
11. Sachs NA, Chang EL, Vyas N, Sorensen BN, Weiland JD. Electrical stimulation of the paralyzed orbicularis oculi in rabbit. *IEEE Trans Neural Syst Rehabil Eng.* 2007;15:67-75. DOI PubMed
12. Linhares CDG, Lima DM, Ponciano JR, et al. ClinicalPath: a visualization tool to improve the evaluation of electronic health records in clinical decision-making. *IEEE Trans Vis Comput Graph.* 2023;29:4031-46. DOI PubMed
13. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 19-24; New Orleans, LA, USA. New York: IEEE; 2022. pp. 11966-76. Available from: https://openaccess.thecvf.com/content/CVPR2022/html/Liu_A_ConvNet_for_the_2020s_CVPR_2022_paper.html [accessed 9 December 2025].
14. Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 2019 Oct 27-Nov 2; Seoul, South Korea. New York: IEEE; 2019. pp. 1314-24. Available from: https://openaccess.thecvf.com/content_ICCV_2019/html/Howard_Searching_for_MobileNetV3_ICCV_2019_paper.html [accessed 9 December 2025].
15. Paszke A, Gross S, Massa F, et al. Pytorch: an imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R, Editors. Advances in Neural Information Processing Systems 32. NeurIPS 2019; 2019 Dec 8-14; Vancouver, Canada. NeurIPS; 2019. Available from: https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html [accessed 9 December 2025].
16. Liu S, Qi L, Qin H, Shi J, Jia J. Path aggregation network for instance segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018 Jun 18-22; Salt Lake City, UT, USA. New York: IEEE; 2018. pp. 8759-68. Available from: https://openaccess.thecvf.com/content_cvpr_2018/html/Liu_Path_Aggregation_Network_CVPR_2018_paper.html

[accessed 9 December 2025].

17. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018 Jun 18-22; Salt Lake City, UT, USA. New York: IEEE; 2018. pp. 7132-41. Available from: https://openaccess.thecvf.com/content_cvpr_2018/html/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.html [accessed 9 December 2025].
18. Moncaliano MC, Ding P, Goshe JM, Genther DJ, Ciolek PJ, Byrne PJ. Clinical features, evaluation, and management of ophthalmic complications of facial paralysis: a review. *J Plast Reconstr Aesthet Surg.* 2023;87:361-8. DOI PubMed
19. Ding L, Martinez AM. Features versus context: an approach for precise and detailed detection and delineation of faces and facial features. *IEEE Trans Pattern Anal Mach Intell.* 2010;32:2022-38. DOI PubMed PMC
20. Lugaresi C, Tang J, Nash H, et al. MediaPipe: a framework for perceiving and processing reality. In: Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR); 2019 Jun 16-20; Long Beach, CA, USA. New York: IEEE; 2019. Available from: https://static1.squarespace.com/static/5c3f69e1cc8fedbc039ea739/t/5e130ff310a69061a71cbd7c/1578307584840/NewTitle_May1_MediaPipe_CVPR_CV4ARVR_Workshop_2019.pdf [accessed 9 December 2025].
21. King DE. Dlib-ml: a machine learning toolkit. *J Mach Learn Res* 2009;10:1755-8. Available from: https://www.jmlr.org/papers/volume10/king09a/king09a.pdf [accessed 9 December 2025].
22. Deng J, Guo J, An X, Zhu Z, Zafeiriou S. Masked face recognition challenge: the insightface track report. In: IEEE/CVF International Conference on Computer Vision (ICCV) Workshops; 2021 Oct 11-17; Virtual. New York: IEEE; 2021. pp. 1437-44. Available from: https://openaccess.thecvf.com/content/ICCV2021W/MFR/html/Deng_Masked_Face_Recognition_Challenge_The_InsightFace_Track_Report_ICCVW_2021_paper.html [accessed 9 December 2025].
23. Wagner K, Doroslovacki M. Proportionate-type normalized least mean square algorithms with gain allocation motivated by mean-square-error minimization for white input. *IEEE Trans Signal Process.* 2011;59:2410-5. DOI
24. Mandal B, Li L, Wang GS, Lin J. Towards detection of bus driver fatigue based on robust visual analysis of eye state. *IEEE Trans Intell Transport Syst.* 2017;18:545-57. DOI
25. Chen S, Epps J. Efficient and robust pupil size and blink estimation from near-field video sequences for human-machine interaction. *IEEE Trans Cybern.* 2014;44:2356-67. DOI PubMed
26. Aloudat M, Faezipour M, El-Sayed A. Automated vision-based high intraocular pressure detection using frontal eye images. *IEEE J Transl Eng Health Med.* 2019;7:3800113. DOI PubMed PMC
27. Liu S, Liu X, Yan D, et al. Alterations in patients with first-episode depression in the eyes-open and eyes-closed conditions: a resting-state EEG study. *IEEE Trans Neural Syst Rehabil Eng.* 2022;30:1019-29. DOI PubMed
28. Chiranjeevi P, Gopalakrishnan V, Moogi P. Neutral face classification using personalized appearance models for fast and robust emotion detection. *IEEE Trans Image Process.* 2015;24:2701-11. DOI PubMed