Privacy-Preserving Video-Based Facial Palsy Recognition via Dynamic Facial Action Modeling

Yuance Chang, Bang Liu, Han Ding*, Cui Zhao, Fei Wang, Ge Wang, and Wei Xi

Xi'an Jiaotong University, Xi'an, China
*Corresponding author: dinghanxjtu@gmail.com

Abstract. Rapid stroke screening is crucial in remote and resource-limited settings where timely medical expertise may be unavailable. Facial palsy, a visible and early indicator of stroke, offers a practical and non-invasive basis for assessment, motivating the need for automated and reliable recognition methods. To address this need, we propose DP-Face, a video-based deep learning framework that captures dynamic facial asymmetry while preserving patient privacy. Rather than relying on raw RGB appearance, DP-Face integrates structural cues derived from facial landmarks and wrinkle patterns with temporal muscle dynamics modeled through facial Action Units. To promote reproducibility, we also construct Palsy-330, the largest publicly available video dataset to date for facial palsy classification, comprising 330 subjects. Evaluated under a patient-independent protocol, DP-Face outperforms existing baselines, demonstrating the promise of dynamic, privacy-preserving facial analysis for rapid stroke screening. The source code and dataset are available at https://github.com/ChangYuance/DP-Face.

Keywords: Stroke screening · Video-based facial palsy analysis · Computer vision.

1 Introduction

Facial palsy, characterized by asymmetry or drooping on one side of the face, is a key early indicator of stroke according to the widely adopted BEFAST (Balance, Eyes, Face, Arm, Speech, Time) criteria [18]. Detecting facial palsy therefore offers a rapid and non-invasive pre-screening cue for potential stroke, especially in environments where timely access to advanced imaging modalities (e.g., CT or MRI) or immediate neurological evaluation is unavailable. By focusing on this early and visually observable symptom, healthcare systems, or even individual users, can leverage videos captured by personal devices such as smartphones to enable rapid assessment in remote settings.

Recent advances in deep learning have provided promising tools for automated facial palsy analysis. Existing approaches can be broadly divided into two categories: single-frame methods and multi-frame (video-based) methods. Single-frame approaches analyze individual facial images, typically quantifying

facial asymmetry using detected landmarks [9, 11, 15], predefined regions of interest [12, 23, 10, 26], or end-to-end CNN models for direct palsy severity prediction [1, 21, 2]. While these methods achieve encouraging results, they inherently neglect the dynamic characteristics of facial movements, which are critical for accurately assessing facial palsy.

To address this limitation, multi-frame methods analyze sequences of facial images to capture temporal facial dynamics. These approaches extract spatiotemporal features using models such as LSTMs, 3D CNNs, or facial Action Unit (AU) representations [20, 6, 19]. While they better model movement patterns, most existing studies rely on raw facial appearance and are evaluated on non-public datasets due to privacy concerns. This lack of publicly accessible data restricts their reproducibility and wide adoption, highlighting the need for privacy-preserving approaches that maintain high detection performance.

To this end, we propose DP-Face, a video-based framework that jointly models facial **D**ynamics and **P**rivacy-aware representations for facial palsy recognition. DP-Face derives structural appearance cues from facial landmark connections and wrinkle patterns during expressions, while capturing temporal muscle movements through facial Action Units. Compared with single-frame methods, it explicitly leverages temporal dynamics; compared with prior multi-frame approaches, it reduces reliance on raw appearance features while effectively modeling facial palsy patterns.

This work makes three primary contributions: (1) we construct Palsy-330, the largest publicly available dataset (over 300 subjects) specifically designed for facial palsy classification; (2) we introduce privacy-preserving facial representations that retain diagnostic information while protecting sensitive identity cues; and (3) we propose a unified deep learning framework that integrates structural appearance and motion dynamics, demonstrating the potential of facial palsy detection to serve as a pre-screening cue for potential stroke.

2 Dataset

2.1 Overview of Palsy-330

Palsy-330 is, to our knowledge, the largest publicly available dataset for video-based facial palsy classification. It includes 330 participants divided into two diagnostic groups: positive (facial palsy present) and negative (facial palsy absent). Each participant performs a standardized clinical "show-teeth" smile task, a commonly used assessment protocol [5]. The data were collected by physicians and nurses from 10 hospitals in western China, and the dataset includes metadata such as age, gender, and anonymized hospital affiliation to facilitate future research. All data were collected under institutional ethics approval with informed consent from all participants. Videos were primarily captured from frontal views across diverse lighting conditions and real clinical environments, enhancing practical relevance. As summarized in Table 1, the dataset contains both male and female participants, with a higher proportion of males. The relatively balanced

Table 1. Statistics of the Palsy-330 dataset.

Attributes	All	Positive	Negative
Number of participants	330	210	120
Number of samples	1518	975	543
Samples of each participant (mean ± std)	4.64 ± 0.48	4.66 ± 0.47	4.61 ± 0.49
Sex (F/M)	85 / 245	53 / 157	32 / 88
Age (mean ± std)	62 ± 11	63 ± 12	60 ± 11

Fig. 1. Images from the Palsy-330 dataset: (a) positive and (b) negative samples.

class distribution and multiple samples per subject provide sufficient variability for robust model training and evaluation. Representative examples are shown in Fig. 1.

2.2 Data Collection and Preprocessing

Most videos were recorded using smartphones, with a smaller portion acquired using professional cameras. During recording, patients performed the standardized smiling task. Video resolutions ranged from 480×640 to 2160×3840 pixels, most commonly 720×1280 or 1080×1920, with frame rates between 24 and 60 fps (over 90% at approximately 30 fps). All data and annotations were uploaded via a dedicated mobile application, with labels provided by attending clinicians and verified by a senior chief physician.

For analysis, we standardized these videos through preprocessing. Faces were detected in each frame, cropped, and resized to 512×512 pixels. All videos were then downsampled to 20 fps to address frame rate variations. Given that a smile typically lasts about 4 seconds, we uniformly sampled 16 consecutive frames and grouped them into one sequence to capture the complete expression. These processed sequences were used as model inputs.

3 Methodology

This section presents DP-Face, our framework for video-based facial palsy recognition Fig. 2). The pipeline consists of three components: privacy-preserving face representation generation, facial Action Unit extraction, and a multi-frame fusion network.

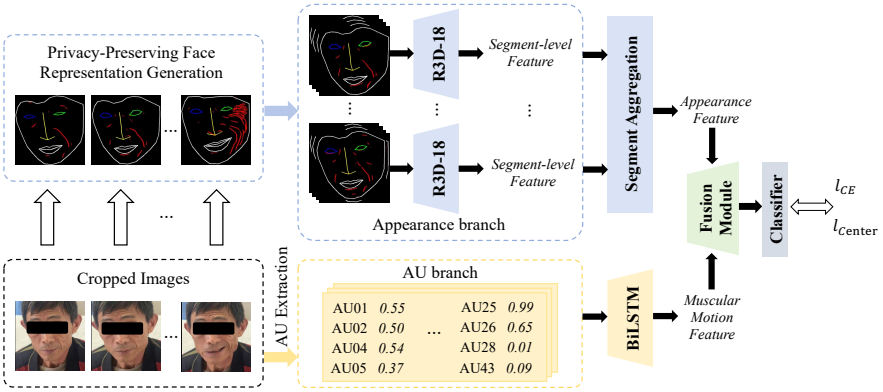

Fig. 2. An overview of the proposed DP-Face framework.

3.1 Privacy-Preserving Face Representation Generation

Facial palsy primarily manifests as asymmetric facial movements during expressions.

To capture these structural changes while avoiding identity leakage, we first detect 68 facial landmarks from cropped images using STAR [27], a pretrained landmark detector that employs Self-adapTive Ambiguity Reduction loss to produce robust features even under severe asymmetry. The landmarks are then connected to form facial contours, providing geometry-based appearance cues without exposing raw facial textures.

However, landmarks alone cannot fully reflect wrinkle asymmetry, which is clinically important—particularly in regions such as the nasolabial folds. To address this limitation, we further extract facial wrinkles using a dedicated detection method [13] and overlay them onto the landmark maps. The resulting representation is termed the PP image (Privacy-Preserving image). Edge constraints confine the wrinkles within facial regions. As illustrated in Fig. 3, the face contour is rendered in white, eyes in blue and green, nose in yellow, mouth in white, and wrinkles in red. Notably, facial palsy subjects exhibit observable asymmetry in nasolabial folds and mouth shapes during smiling, demonstrating that this representation preserves key diagnostic cues while removing identifiable appearance information. Moreover, it suppresses external variations from background clutter and lighting changes that may otherwise hinder palsy recognition.

3.2 Action Unit Extraction

Beyond structural asymmetry, facial palsy also affects underlying muscle activation patterns. Facial Action Units (AUs) provide a fine-grained and anatomically grounded description of facial muscle movements, making them well suited for palsy analysis. Traditionally, AU coding required labor-intensive manual annotation following the Facial Action Coding System (FACS), which decomposes facial behavior into muscle-specific action units defined by Ekman and Friesen [4].

To enable automatic and scalable analysis, we extract 20 AUs per frame using Py-Feat [3] in an offline pre-processing step, where each AU is represented as a continuous value in $[0, 1]$ (Fig. 2). These signals capture subtle muscle activations associated with asymmetric expressions. For example, AU10 (Upper Lip Raiser) reflects upper lip elevation, while AU12 (Lip Corner Puller) characterizes smiling behavior.

3.3 Multi-Frame Palsy Recognition

Given the privacy-preserving face sequences and AU features, we design a fusion model to learn holistic representations. For a sample with t frames, the appearance input has size $t \times 3 \times h \times w$, and the AU sequence has size $t \times 20$.

Appearance Branch. Facial palsy cues evolve over time, so modeling temporal dynamics is essential. Inspired by video facial expression recognition [22], we adopt a segment-based strategy. The t frames are divided into n segments of k frames each. A pretrained ResNet3D-18 [8] extracts segment-level appearance features $\{F_1, \ldots, F_n\}$, which are then aggregated by a BiLSTM followed by a Self-Attention module to capture long-range temporal dependencies.

AU Branch. In parallel, AU sequences are processed by a separate BiLSTM to model temporal muscle activation patterns complementary to appearance features.

Fusion. We fuse the two modalities using FiLM [17], which conditionally modulates appearance features with AU information. Given appearance features x, FiLM generates scaling γ and shifting θ from AU features via fully connected layers:

$$y = \gamma \cdot x + \theta, \tag{1}$$

enabling AU-guided refinement of appearance representations.

Training Objective. The fused features are fed into a fully connected (FC) classifier for palsy recognition. To improve feature discriminability, we combine Cross-Entropy loss with Center Loss [24], which encourages intra-class compactness. The overall loss is:

$$\mathcal{L} = \alpha \, \mathcal{L}_{\text{CE}} + \beta \, \mathcal{L}_{\text{center}}, \tag{2}$$

where α and β balance the two terms.

4 Experiments and Results

4.1 Experimental Setups

Dataset Partitioning. Previous studies [10, 9, 16, 15] often rely on random splits, which may cause models to memorize subject-specific traits rather than disease-related patterns. To mitigate this issue, we adopt five-fold cross-validation at the participant level, ensuring that all samples from the same participant remain within the same fold. In each round, three folds are used for training, one

(a) Positive examples (b) Negative examples

Fig. 3. Privacy-preserving face representations. Top row: cropped faces; bottom row: connected landmarks with wrinkles (*i.e.*, PP images).

Table 2. Performance comparison of various methods (%). Metrics include Accuracy (Acc.), Precision (Prec.), Sensitivity (Sens.), Specificity (Spec.), F1-score, and Area Under Curve (AUC). Methods denoted with (*) utilize single-frame inputs, while (†) indicates sequence-based models.

Method	**Inputs**	**Acc.**	**Prec.**	**Sens.**	**Spec.**	**F1**	**AUC**
FP-VGGFace* [1]	Cropped image	69.9	77.9	74.4	64.2	75.3	72.4
AFPD-RI* [15]	Hand-crafted features	64.8	71.8	72.7	50.3	72.2	65.5
3DPalsyNet† [20]	Cropped image	73.2	79.9	79.9	63.2	78.9	76.5
AUSTIN† [25]	Cropped image+Audio	66.2	74.5	74.2	54.4	73.4	66.5
BiLSTM-AU†	AUs	71.7	75.6	**83.3**	50.4	78.9	72.4
DP-Face† (Ours)	PP image+AUs	**75.1**	**84.6**	77.0	**68.9**	**80.3**	**78.7**

for validation, and one for testing. The folds are rotated so that each fold serves as the test set exactly once.

Evaluation Metrics. Model performance is evaluated using common classification metrics, including *Accuracy*, *Precision*, *Sensitivity* (*Recall*), *Specificity*, *F1-score*, and *AUC*. Among them, *Specificity* measures the proportion of true negatives that are correctly identified, which is particularly important for avoiding false alarms in clinical screening.

Implementation Details. All models are implemented in PyTorch. Each sample contains 16 frames with a spatial resolution of 112×112, and the batch size is set to 32. Random data augmentation, including horizontal flipping and color jittering, is applied to increase data diversity. Label smoothing with a ratio of 0.1 is also used. The training objective combines Cross-Entropy loss and Center Loss, with α and β set to 0.4 and 0.6, respectively. Models are trained for 200 epochs using the AdamW optimizer with an initial learning rate of 5×10^{-4} and a weight decay of 0.05.

4.2 Overall Performance

To evaluate the effectiveness of our approach, we compare it against several representative methods, including both single-frame and sequence-based models. 1) FP-VGGFace [1]: a single-frame model obtained by fine-tuning the pre-trained VGGFace model [14]. 2) AFPD-RI [15]: a single-frame method that quantifies

facial asymmetry using facial landmarks to construct hand-crafted geometric features (*e.g.*, selected inter-landmark distances and angles), followed by a simple 3-layer MLP classifier. 3) 3DPalsyNet [20]: a sequence-based model built on a pre-trained 3D ResNet [8] for mouth motion analysis and palsy grading. 4) AUSTIN [25]: a sequence-based multimodal model that leverages cropped facial images and audio to capture temporal dynamics. 5) BiLSTM-AU: our implemented BiLSTM baseline that directly uses AU sequences for palsy recognition.

Table 2 summarizes the quantitative results. Single-frame methods that ignore temporal dynamics perform relatively poorly, with FP-VGGFace and AFPD-RI achieving accuracies of 69.9% and 64.8%, respectively. In contrast, sequence-aware approaches such as 3DPalsyNet and BiLSTM-AU show improved performance (71.7% and 73.2%), highlighting the importance of temporal modeling in facial palsy analysis. AUSTIN yields lower performance, likely because smiling produces minimal meaningful audio; consequently, the speech signals used as auxiliary input introduce noise rather than useful cues. Our method achieves the best overall performance, reaching 75.1% accuracy, 80.3% F1-score, and 78.7% AUC. Notably, Palsy-330 is, to our knowledge, the first large-scale video dataset with over 300 subjects for facial palsy analysis, providing a representative benchmark for sequence-based modeling.

To further evaluate the privacy-preserving property of PP images, we conducted a person re-identification (ReID) experiment. We extract features from RGB and PP images using separate ResNet-50 networks and train them with a triplet loss. For comparison, we also run the same experiment with RGB images only. The RGB-only setting achieved 100% Rank-1 accuracy, whereas the PP→RGB retrieval setting achieved only 5.11% Rank-1 accuracy on a 100-subject gallery (chance level: 1%). This substantial degradation indicates that PP features substantially suppress identity information while preserving diagnostic utility, highlighting that our design removes identity-sensitive appearances while retaining task-relevant geometric dynamics.

To evaluate generalization, we also test FP-VGGFace, 3DPalsyNet, BiLSTM-AU, and our method on the publicly available MEEI dataset [7]. The MEEI dataset is a clinically curated benchmark containing standardized facial palsy photos and video recordings collected at Massachusetts Eye and Ear Infirmary. However, compared with Palsy-330, MEEI is substantially smaller and highly imbalanced (10 normal vs. 50 palsy subjects). As shown in Table 3, despite the limited scale and class imbalance of MEEI, our method achieves competitive performance, demonstrating strong cross-dataset generalization. Specifically, DP-Face achieves the highest Accuracy (87.3%) and Recall (94.5%), while its lower Specificity (52.5%) is due to the severe class imbalance (palsy:normal = 5:1). For stroke pre-screening, we prioritize recall to minimize missed diagnoses, making this trade-off clinically acceptable. These results further validate that the proposed privacy-preserving dynamic representation learns disease-relevant patterns rather than dataset-specific biases, while the larger and more diverse Palsy-330 better supports stable model development and benchmarking.

Table 3. Evaluation on the MEEI dataset (%).

Method	Inputs	Acc.	Prec.	Sens.	Spec.	F1	AUC
FP-VGGFace* [1]	Cropped image	74.4	**93.3**	75.5	**70.0**	81.2	**73.9**
3DPalsyNet† [20]	Cropped image	84.0	90.6	90.5	52.0	90.4	66.9
BiLSTM-AU†	AUs	85.0	88.9	**94.9**	36.0	91.3	62.8
DP-Face† (Ours)	PP image + AUs	**87.3**	90.9	94.5	52.5	**92.5**	72.3

4.3 Ablation Study

We conduct an ablation study (Table 4) to examine the effects of input types, frame numbers, and fusion strategies.

Input Types. We first compare cropped images, PP images without wrinkles, and full PP images. PP images without wrinkles achieve performance comparable to cropped images. Incorporating wrinkle information further improves results, allowing PP images to consistently outperform raw cropped images, reaching 74.8% accuracy and 79.3% F1, compared with 72.6% and 78.2%, respectively. This indicates that subtle facial wrinkles provide complementary diagnostic cues. While without data augmentation (DA), the performance slightly degrades, indicating that DA contributes to improved generalization.

Number of Frames. We then study the impact of temporal length. Increasing the number of frames generally benefits performance, as richer temporal dynamics are captured. Using 16 frames yields the best overall accuracy (74.8%), while shorter sequences (8 or 4 frames) lead to slight degradation, demonstrating the importance of sufficient temporal context.

Fusion Methods. Finally, we compare several fusion strategies for combining PP images and AU features. Our FiLM-based method achieves the best results, reaching 75.1% accuracy and 80.3% F1, outperforming attention-, addition-, and concatenation-based fusion. This confirms the effectiveness of conditional modulation for integrating visual and AU information.

Table 4. Ablation study on various inputs, frames numbers and fusion strategies (%).

Fusion	Frames	Inputs	Acc.	Prec.	Sens.	Spec.	F1	AUC
~	16	Cropped image	72.6	80.7	76.6	68.3	78.2	75.0
~	16	PP image w/o Wrinkles	72.3	78.8	73.6	**70.0**	75.9	77.2
~	16	PP image w/o DA	74.2	79.9	78.5	67.3	79.1	77.5
~	16	PP image	74.8	80.5	78.6	68.1	79.3	77.9
~	8	PP image	74.6	79.7	**79.1**	65.7	79.3	75.9
~	4	PP image	72.8	78.4	77.1	63.9	77.6	75.9
attention	16	PP image + AUs	72.1	79.2	76.5	64.0	77.4	76.1
add	16	PP image + AUs	73.1	78.5	78.9	61.2	78.5	73.9
concat	16	PP image + AUs	72.2	80.3	75.5	67.6	77.5	76.9
Ours (FiLM)	16	PP image + AUs	**75.1**	**84.6**	77.0	68.9	**80.3**	**78.7**

5 Conclusion

In this work, we presented DP-Face, a privacy-preserving video-based framework for facial palsy recognition that emphasizes dynamic facial action modeling rather than identity-dependent appearance cues. By jointly leveraging geometry-aware PP images and temporally modeled Action Units, the proposed method effectively captures clinically meaningful facial asymmetry. To support standardized evaluation, we also released Palsy-330, currently the largest publicly available video dataset for facial palsy classification with subject-independent protocols. Comprehensive experiments and ablation studies verify the advantages of wrinkle-enhanced structural representation and FiLM-based multimodal fusion. These findings highlight the potential of privacy-aware dynamic facial analysis for scalable neurological screening in real-world settings. Future work will explore broader clinical validation and finer-grained severity assessment to further advance privacy-preserving medical AI.

Acknowledgments. This work was supported by Noncommunicable Chronic Diseases-National Science and Technology Major Project 2024ZD0527800, the NSFC Grant No. 62372365, 62572390, 62572383. All authors are supported by the State Key Laboratory of Human-Machine Hybrid Augmented Intelligence, Xi'an Jiaotong University.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ali, W., Imran, M., Yaseen, M.U., Aurangzeb, K., Ashraf, N., Aslam, S.: A transfer learning approach for facial paralysis severity detection. IEEE Access **11**, 127492–127508 (2023)
2. Baig, Z.M., Van Der Haar, D.: Facial paralysis recognition using face mesh-based learning. In: Proceedings of ICPRAM (2023)
3. Cheong, J.H., Jolly, E., Xie, T., Byrne, S., Kenney, M., Chang, L.J.: Py-feat: Python facial expression analysis toolbox. Affective Science **4**(4), 781–796 (2023)
4. Ekman, P., Friesen, W.V.: Facial action coding system. Environmental Psychology & Nonverbal Behavior (1978)
5. Eroglu, S., Ozsoy, O., Yildirim, Y., Ozsoy, U., Uysal, H.: Comparison of 3d surface and landmark-based analysis methods: The reliability and efficiency in determining asymmetry after facial palsy. Journal of Plastic, Reconstructive & Aesthetic Surgery **104**, 469–478 (2025)
6. Ge, X., Jose, J.M., Wang, P., Iyer, A., Liu, X., Han, H.: Automatic facial paralysis estimation with facial action units. arXiv preprint arXiv:2203.01800 (2022)
7. Greene, J.J., Guarin, D.L., Tavares, J., Fortier, E., Robinson, M., Dusseldorp, J., Quatela, O., Jowett, N., Hadlock, T.: The spectrum of facial palsy: The meei facial palsy photo and video standard set. The Laryngoscope **130**(1), 32–37 (2020)
8. Hara, K., Kataoka, H., Satoh, Y.: Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In: Proceedings of IEEE/CVF CVPR (2018)
9. Huang, K., Li, Y., Zhu, F.: Quantitative analysis of facial paralysis based on tcm acupuncture point identification. In: Proceedings of IEEE SEAI (2023)

10. Jison Hsu, G.S., Huang, W.F., Kang, J.H.: Hierarchical network for facial palsy detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 580–586 (2018)
11. Lee, S.A., Kim, J., Lee, J.M., Hong, Y.J., Kim, I.J., Lee, J.D.: Automatic facial recognition system assisted-facial asymmetry scale using facial landmarks. Otology & Neurotology **41**(8), 1140–1148 (2020)
12. Li, H., Wu, K., Cheng, H., Gu, C., Guan, X.: Nasolabial folds extraction based on neural network for the quantitative analysis of facial paralysis. In: Proceedings of IEEE ICISPC (2018)
13. Moon, J., Chung, H., Jang, I.: Facial wrinkle segmentation for cosmetic dermatology: Pretraining with texture map-based weak supervision. In: Proceedings of Springer ICPR (2024)
14. Parkhi, O., Vedaldi, A., Zisserman, A.: Deep face recognition. In: Proceedings of BMVC (2015)
15. Parra-Dominguez, G.S., Garcia-Capulin, C.H., Sanchez-Yanez, R.E.: Automatic facial palsy diagnosis as a classification problem using regional information extracted from a photograph. Diagnostics **12**(7), 1528 (2022)
16. Parra-Dominguez, G.S., Sanchez-Yanez, R.E., Garcia-Capulin, C.H.: Facial paralysis detection on images using key point analysis. Applied Sciences **11**(5), 2435 (2021)
17. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI (2018)
18. Pickham, D., Valdez, A., Demeestere, J., Lemmens, R., Diaz, L., Hopper, S., de la Cuesta, K., Rackover, F., Miller, K., Lansberg, M.G.: Prognostic value of befast vs. fast to identify stroke in a prehospital setting. Prehospital Emergency Care **23**(2), 195–200 (2019)
19. Shayestegan, M., Kohout, J., Štícha, K., Mareš, J.: Advanced analysis of 3d kinect data: supervised classification of facial nerve function via parallel convolutional neural networks. Applied Sciences **12**(12), 5902 (2022)
20. Storey, G., Jiang, R., Keogh, S., Bouridane, A., Li, C.T.: 3dpalsynet: A facial palsy grading and motion recognition framework using fully 3d convolutional neural networks. IEEE access **7**, 121655–121664 (2019)
21. Ten Harkel, T.C., de Jong, G., Marres, H.A., Ingels, K.J., Speksnijder, C.M., Maal, T.J.: 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. American Journal of Otolaryngology **44**(3), 103810 (2023)
22. Wang, H., Li, B., Wu, S., Shen, S., Liu, F., Ding, S., Zhou, A.: Rethinking the learning paradigm for dynamic facial expression recognition. In: Proceedings of IEEE/CVF CVPR (2023)
23. Wang, T., Zhang, S., Liu, L., Wu, G., Dong, J.: Automatic facial paralysis evaluation augmented by a cascaded encoder network structure. IEEE Access **7**, 135621–135631 (2019)
24. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: Proceedings of Springer ECCV (2016)
25. Yang, S., Cai, T., Ni, H., Ma, W., Xue, Y., Wong, K., Volpi, J., Wang, J.Z., Huang, S.X., Wong, S.T.: Enhancing ai-assisted stroke emergency triage with adaptive uncertainty estimation. In: Proceedings of Springer MICCAI (2025)
26. Zhang, Y., Gao, W., Yu, H., Dong, J., Xia, Y.: Artificial intelligence-based facial palsy evaluation: a survey. IEEE Transactions on Neural Systems and Rehabilitation Engineering (2024)

27. Zhou, Z., Li, H., Liu, H., Wang, N., Yu, G., Ji, R.: Star loss: Reducing semantic ambiguity in facial landmark detection. In: Proceedings of IEEE/CVF CVPR (2023)