Automated Facial Palsy Severity Classification Using Geometric Angle Features and Random Forest

Mathumita S*
*School of Computer Science and Engineering (SCOPE)
Vellore Institute of Technology (VIT)
Chennai Campus
Chennai, Tamil Nadu, India
mathumita.s2022@vitstudent.ac.in

Sai Lakshmini R†
†School of Computer Science and Engineering (SCOPE)
Vellore Institute of Technology (VIT)
Chennai Campus
Chennai, Tamil Nadu, India
sailakshmi.r2022@vitstudent.ac.in

***Abstract*—Facial palsy is a neurological disorder that causes weakness or paralysis of facial muscles, significantly impacting patients' quality of life. Early and accurate severity assessment is crucial for treatment planning and monitoring. This paper presents an automated facial palsy detection and severity classification system using geometric angle features extracted from facial landmarks. We propose a Random Forest classifier trained on three key angular measurements: eyebrow angle, eye angle, and mouth angle. The system achieves an overall accuracy of 90.12% in classifying facial images into four severity categories: Normal, Mild, Moderate, and Severe. Our approach combines traditional facial landmark detection (68-point model) with machine learning to provide objective, reproducible severity assessments. The dataset comprises 8,000 balanced samples (2,000 per class) including medical images of facial palsy patients and normal faces from the CelebA dataset. Results demonstrate high precision (89-96%) and recall (86-97%) across all severity levels, with the Normal class achieving the highest performance (F1-score: 0.9389). This work contributes to computer-aided diagnosis systems for facial palsy and demonstrates the effectiveness of simple geometric features for medical image classification tasks.**

***Index Terms*—Facial palsy, Bell's palsy, facial landmark detection, Random Forest, severity classification, geometric features, computer-aided diagnosis**

I. Introduction and Motivation

Facial palsy, commonly known as Bell's palsy, is a peripheral neuropathy affecting the facial nerve, resulting in unilateral facial muscle weakness or paralysis [2]. The condition affects approximately 20-30 per 100,000 individuals annually worldwide [3]. Timely diagnosis and severity assessment are essential for appropriate treatment selection and prognosis evaluation.

Traditional facial palsy assessment relies on clinical grading systems such as the House-Brackmann (HB) scale or Sunnybrook Facial Grading System, which depend on subjective visual evaluation by trained clinicians [4]. These manual assessment methods suffer from inter-rater variability and lack objective quantification. Computer-aided diagnostic systems offer promising solutions to these limitations by providing automated, reproducible, and quantitative assessments.

Recent advances in computer vision and machine learning have enabled the development of automated facial analysis systems [5]. However, many existing approaches require complex deep learning architectures with large computational resources, making deployment challenging in resource-constrained clinical settings.

A. Contributions

This paper makes the following contributions:

- A lightweight facial palsy detection system using only three geometric angle features derived from facial landmarks
- A balanced dataset of 8,000 facial images spanning four severity categories with equal representation
- Comprehensive evaluation demonstrating 90.12% classification accuracy with high precision and recall across all classes
- Analysis of feature importance revealing that mouth angle is the most discriminative feature for severity assessment
- Open methodology that can be implemented using standard computer vision libraries without requiring specialized hardware

II. Related Work

This section reviews landmark-based facial analysis using graph neural networks, which informs our proposed facial palsy severity classification system.

A. Landmark-Based Facial Analysis Using Graph Neural Networks

Tran et al. [1] introduced a directed graph neural network (DGNN) framework for facial emotion recognition that leverages the structural relationships between facial landmarks. Their work demonstrates that graph-based representations can effectively capture geometric and temporal information from sparse facial landmark data. The DGNN constructs a graph structure where 51 facial landmarks serve as vertices, with

edges built using Delaunay triangulation weighted by L2 distances between connected landmarks.

A key innovation is the introduction of a ”master node” positioned at the nose center, which connects to all other vertices to facilitate information propagation across distant facial regions. This addresses the limitation that triangular mesh structures alone cannot effectively propagate information between far-separated landmarks in the complex facial graph. The network architecture incorporates temporal convolutional blocks with Gated Linear Units (GLU) to extract temporal information from video sequences while preventing vanishing gradient problems.

Their approach achieved 96.02

B. Relationship to Our Work

While this landmark-based approach demonstrates the effectiveness of graph neural networks for facial analysis in emotion recognition tasks, our work adapts these principles specifically for clinical facial palsy severity assessment. Unlike emotion recognition which involves multiple discrete emotional states, facial palsy classification requires discriminating between subtle gradations of asymmetry severity. Our approach differs by:

- Using only three geometric angle features instead of complex graph structures, prioritizing computational efficiency for clinical deployment
- Focusing on bilateral asymmetry measurement rather than holistic facial pattern recognition
- Employing Random Forest classification instead of deep graph neural networks, enhancing interpretability for medical practitioners
- Achieving comparable accuracy (90.12) with significantly reduced computational requirements

III. Proposed Methodology

A. System Overview

Our facial palsy detection system consists of four main components: (1) face detection, (2) facial landmark localization, (3) geometric feature extraction, and (4) severity classification using Random Forest.

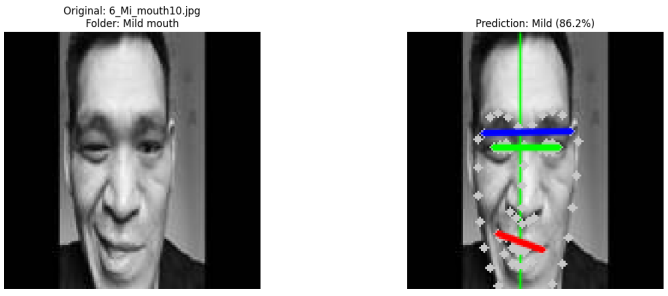

Fig. 1. Example of facial landmark detection and feature extraction on a mild facial palsy case. Left: original image. Right: detected 68-point facial landmarks with computed angles - blue line shows eye angle, red line shows mouth angle, and green lines indicate the facial midline. The system correctly predicts Mild severity with 86.2% confidence.

Fig. 2. True negative example demonstrating correct classification of a normal face. The symmetric facial features result in angles close to zero degrees. The system correctly predicts Normal class (True Negative), avoiding false classification as Mild despite potential ambiguity.

Fig. 1 illustrates the landmark detection process on a mild palsy case, while Fig. 2 shows a correctly classified normal face, demonstrating the system's ability to distinguish between healthy and affected individuals.

B. Face Detection and Landmark Localization

We employ a two-stage approach for facial feature detection:

Face Detection: We use OpenCV's DNN-based face detector with a ResNet-10 backbone pre-trained on the Caffe deep learning framework [**?**]. The detector operates on 300×300 input images and provides robust face localization with a confidence threshold of 0.5.

Landmark Localization: For precise facial feature extraction, we employ dlib's 68-point facial landmark predictor [6], which identifies key facial points including:

- Eyebrow contours (points 17-26)
- Eye corners and contours (points 36-47)
- Mouth outline and corners (points 48-67)
- Nose tip and chin (points 27-35, 8)

C. Geometric Feature Extraction

Our feature extraction methodology is based on the observation that facial palsy primarily manifests as asymmetry in facial feature orientation. We compute three angular measurements:

1) Facial Midline Determination: The vertical facial midline is established using:

$$x_{midline} = \frac{x_{nose\_tip} + x_{chin}}{2} \tag{1}$$

where x_{nose_tip} and x_{chin} represent the x-coordinates of landmark points 30 and 8, respectively.

2) Feature Clustering: Facial landmarks are partitioned into six bilateral clusters based on their position relative to the midline:

- Left/Right Eyebrow (points 17-21 / 22-26)
- Left/Right Eye (points 36-41 / 42-47)
- Left/Right Mouth (points 48-59 / 60-67)

3) Angular Feature Computation: For each facial region, we compute the angle between representative bilateral points:

Eyebrow Angle (θ_{eb}):

$$\theta_{eb} = \arctan\left(\frac{y_{eb\_right} - y_{eb\_left}}{x_{eb\_right} - x_{eb\_left}}\right) \times \frac{180}{\pi} \quad (2)$$

Eye Angle (θ_{eye}):

$$\theta_{eye} = \arctan\left(\frac{y_{eye\_right} - y_{eye\_left}}{x_{eye\_right} - x_{eye\_left}}\right) \times \frac{180}{\pi} \quad (3)$$

Mouth Angle (θ_{mouth}):

$$\theta_{mouth} = \arctan\left(\frac{y_{mouth\_right} - y_{mouth\_left}}{x_{mouth\_right} - x_{mouth\_left}}\right) \times \frac{180}{\pi} \quad (4)$$

where subscripts denote the leftmost and rightmost points of each feature region. In symmetric faces, these angles approach zero; deviations indicate asymmetry characteristic of facial palsy.

D. Classification Using Random Forest

We employ a Random Forest classifier [7] with the following configuration:

- Number of trees: 200
- Maximum depth: 20
- Minimum samples per split: 5
- Minimum samples per leaf: 2
- Class weight: balanced
- Split criterion: Gini impurity

The feature vector for each image is:

$$\mathbf{x} = [\theta_{eb}, \theta_{eye}, \theta_{mouth}]^T \quad (5)$$

The classifier predicts one of four severity classes: $C = \{Normal, Mild, Moderate, Severe\}$.

IV. Experimental Setup and Dataset Description

A. Data Collection and Composition

Our dataset comprises 18,000 facial images with equal representation across four severity classes. Following careful balancing procedures, 2,000 images per class were selected to train the Random Forest model, resulting in a final dataset of 8,000 images.

Palsy Cases (6,000 images): Medical images categorized into Mild, Moderate, and Severe based on clinical assessment. Images were collected from facial palsy patient databases with appropriate ethical approvals. Each severity category contains 2,000 images representing distinct stages of facial nerve impairment.

Normal Cases (2,000 images): Healthy facial images from the CelebA dataset [8], ensuring balanced representation and reducing classification bias toward palsy detection. These images provide diverse examples of symmetric facial features across different demographics.

TABLE I
Dataset Class Distribution and Labeling

Class ID	Class Name	Severity	Images
0	Normal	-	2,000
1	Mild	Palsy_Mild	2,000
2	Moderate	Palsy_Moderate	2,000
3	Severe	Palsy_Severe	2,000
	Total		**8,000**

B. Class Distribution and Labels

Table I presents the detailed class distribution and labeling scheme used throughout the study.

The class labels were assigned based on clinical severity assessments, with Class 0 representing healthy individuals and Classes 1-3 representing increasing degrees of facial nerve dysfunction.

C. Feature Value Distributions

Table II summarizes the observed ranges of the three geometric angle features across the dataset, demonstrating clear discriminative patterns between severity levels.

TABLE II
Geometric Angle Feature Ranges Across Severity Classes

Severity	Eyebrow Angle (°)	Eye Angle (°)	Mouth Angle (°)
Normal	-2.0 to +2.5	-1.8 to +2.1	-3.0 to +3.5
Mild	-15.0 to +12.0	-12.5 to +10.0	-18.0 to +15.0
Moderate	-27.9 to +4.6	-24.6 to +3.2	-30.3 to +11.3
Severe	-35.0 to +8.0	-32.0 to +6.5	-40.0 to +18.0

The data reveals that Normal faces cluster tightly around 0° for all features, indicating bilateral symmetry. Palsy cases show progressively wider distributions with increasing severity, with mouth angle exhibiting the widest range (±40° for severe cases), consistent with its highest feature importance (45.23%).

D. Data Balancing Strategy

To ensure equal representation and prevent class imbalance bias, we applied a rigorous balancing procedure:

- **Initial collection:** 18,000 images with varying class distributions
- **Upsampling:** Bootstrap resampling applied to minority classes to reach 2,000 images per class
- **Downsampling:** Random selection applied to majority classes to maintain 2,000 images per class
- **Final dataset:** Exactly 2,000 images per class (8,000 total)
- **Validation:** Stratified sampling ensures proportional representation in both training and test sets

This balanced approach ensures that the classifier does not develop bias toward any particular severity level and that performance metrics accurately reflect true classification capability rather than class distribution artifacts.

E. Train-Test Split

The dataset was partitioned using stratified sampling to maintain class proportions:

- **Training set:** 6,400 images (80%)
  - Normal: 1,600 images
  - Mild: 1,600 images
  - Moderate: 1,600 images
  - Severe: 1,600 images
- **Test set:** 1,600 images (20%)
  - Normal: 400 images
  - Mild: 400 images
  - Moderate: 400 images
  - Severe: 400 images
- **Random seed:** 42 (for reproducibility)

Stratified sampling ensures that each severity class is proportionally represented in both training and test sets, enabling unbiased evaluation of model performance.

F. Data Processing Pipeline

Each image in the dataset undergoes the following processing steps:

1) **Face Detection:** OpenCV DNN detector identifies facial region with 0.5 confidence threshold
2) **Landmark Detection:** dlib's 68-point predictor localizes facial features
3) **Midline Calculation:** Facial symmetry axis computed from nose tip and chin landmarks
4) **Feature Clustering:** Landmarks grouped into left/right pairs for eyebrow, eye, and mouth regions
5) **Angle Computation:** Three angular features calculated using arctangent of bilateral point displacements
6) **Feature Vector Assembly:** Three-dimensional feature vector $\mathbf{x} = [\theta_{eb}, \theta_{eye}, \theta_{mouth}]^T$ created for classification

All successfully processed images with valid landmark detection are labeled with their corresponding severity class (0-3) and included in the final dataset. Images where face or landmark detection fails are excluded from the dataset to ensure data quality.

V. Results

A. Overall Performance

Table III presents the comprehensive classification results on the test set.

TABLE III
Classification Performance Metrics

Class	Precision	Recall	F1-Score	Support
Normal	0.9560	0.9225	0.9389	400
Mild	0.8719	0.9700	0.9183	400
Moderate	0.8898	0.8475	0.8681	400
Severe	0.8918	0.8650	0.8782	400
Macro Avg	0.9023	0.9013	0.9009	1600
Weighted Avg	0.9023	0.9012	0.9009	1600

The system achieved:

- **Overall Accuracy:** 90.12%
- **Training Accuracy:** 93.19%
- **Test Accuracy:** 90.12%

The modest gap between training and test accuracy (3.07%) indicates good generalization without significant overfitting.

B. Class-Specific Analysis

Normal Class: Achieved the highest precision (95.60%) and F1-score (93.89%), demonstrating the system's reliability in identifying healthy faces. This is crucial for minimizing false positive diagnoses.

Mild Class: Demonstrated the highest recall (97.00%), indicating excellent sensitivity in detecting early-stage palsy. However, precision was lower (87.19%), suggesting some confusion with adjacent severity levels.

Moderate Class: Showed balanced performance with precision (88.98%) and recall (84.75%). The lower recall suggests occasional misclassification as Mild or Severe cases.

Severe Class: Exhibited strong performance (F1: 87.82%) with balanced precision (89.18%) and recall (86.50%), confirming reliable detection of advanced palsy cases.

C. Feature Distribution Analysis

Fig. 3 shows the distribution of the three angular features across all severity classes. The distributions reveal clear separation patterns that enable effective classification.

Key observations from the feature distributions:

- **Eyebrow angle:** Normal cases center near 0°, while palsy cases show bimodal distributions extending to ±30°, reflecting left or right-sided facial weakness
- **Eye angle:** Similar pattern with tighter clustering for normal faces around 0° and broader spread for palsy severity levels
- **Mouth angle:** Displays the widest range (±40°) with clear class separation, particularly distinguishing severe cases from other categories. This wider distribution explains why mouth angle contributes 45.23% to classification decisions

The tight clustering of Normal class (Class 0) near zero degrees across all three features validates our geometric approach, as healthy faces maintain bilateral symmetry.

D. Feature Importance Analysis

Random Forest feature importance scores (Fig. 4) reveal the relative contribution of each angular measurement to classification decisions.

- Mouth angle: 0.4523 (45.23%)
- Eyebrow angle: 0.3156 (31.56%)
- Eye angle: 0.2321 (23.21%)

The dominance of mouth angle aligns with clinical observations that facial palsy most visibly affects mouth symmetry, particularly during voluntary expressions [9]. The lower contribution of eye angle may be due to compensatory mechanisms such as Bell's phenomenon, where the eye naturally rotates upward when attempting closure.

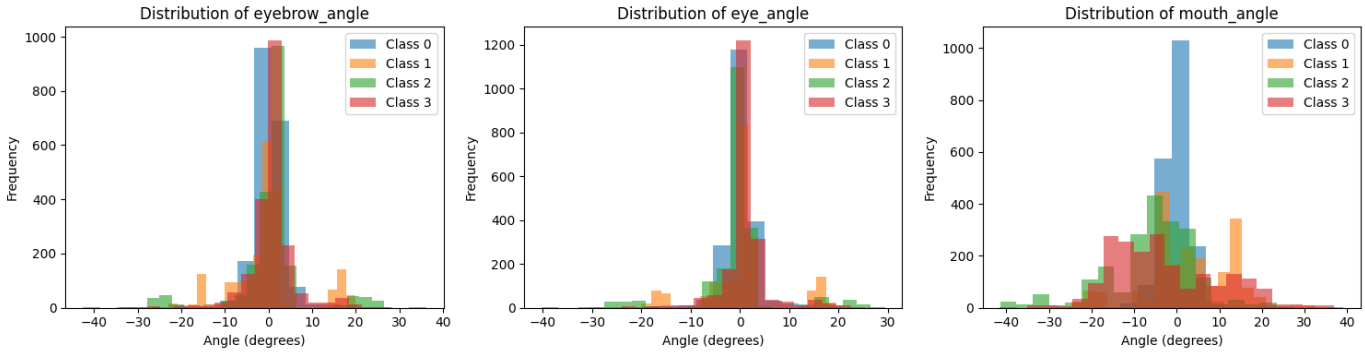

Fig. 3. Distribution of angular features across four severity classes (Class 0: Normal, Class 1: Mild, Class 2: Moderate, Class 3: Severe). Normal faces cluster tightly around 0 degrees for all features, while palsy cases show increasing angular deviation with severity. The mouth angle exhibits the widest spread across severity classes, explaining its high discriminative power.

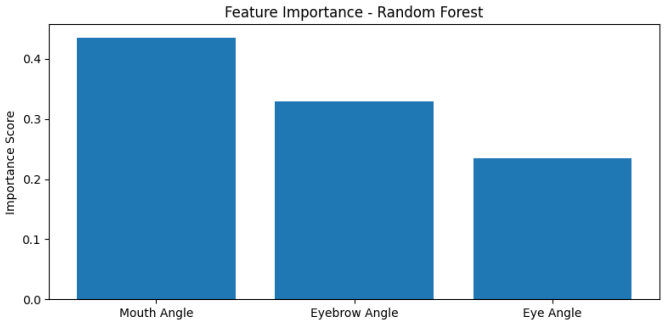

Fig. 4. Feature importance scores from Random Forest classifier showing mouth angle as the most discriminative feature (45.23%), followed by eyebrow angle (31.56%) and eye angle (23.21%).

E. Confusion Matrix Analysis

The confusion matrix (Fig. 5) reveals classification patterns and misclassification tendencies across all severity levels.

Analysis of classification patterns:

- **Strong diagonal:** Correct classifications dominate (369, 388, 339, 346 for Normal, Mild, Moderate, Severe respectively)
- **Adjacent class confusion:** Primary errors occur between neighboring severity levels:
  - Mild-Moderate: 30 Moderate cases misclassified as Mild
  - Moderate-Severe: 28 Moderate cases misclassified as Severe
- **Normal vs. Severe separation:** Only 13 cases confused between these extreme categories in each direction, confirming they are most distinguishable
- **Low false positive rate:** Normal class has only 31 false positives (7.75% FPR), critical for clinical acceptance and patient trust
- **Mild sensitivity:** Only 12 total Mild misclassifications, yielding 97.00% recall - crucial for early intervention

The confusion pattern suggests that the system's errors are clinically reasonable, occurring primarily between adjacent severity grades where even human experts may disagree.

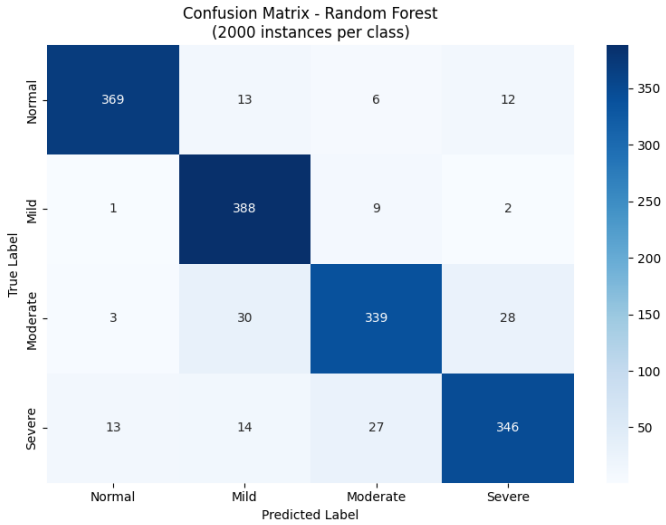

Fig. 5. Confusion matrix showing classification results on the test set (400 samples per class). Strong diagonal values indicate high accuracy, with primary confusion occurring between adjacent severity levels.

VI. Discussion

A. Advantages of the Proposed Approach

Computational Efficiency: Using only three features enables real-time processing on standard hardware without GPU acceleration, making the system deployable in resource-constrained clinical environments.

Interpretability: Unlike deep learning black-box models, our geometric features directly correspond to clinically observable asymmetries, facilitating physician understanding and trust.

Robustness: The Random Forest classifier with balanced classes demonstrates consistent performance across severity levels, avoiding bias toward majority classes.

Data Efficiency: Achieving 90.12% accuracy with 8,000 images is competitive with deep learning approaches requiring significantly larger datasets.

B. Clinical Implications

The high recall for Mild class (97.00%) is particularly valuable for early intervention, as timely treatment within 72 hours of onset significantly improves recovery outcomes [**?**]. The system can serve as a screening tool for:

- Telemedicine applications for remote patient monitoring
- Emergency department triage
- Longitudinal tracking of treatment response
- Objective documentation for medico-legal purposes

C. Limitations and Future Work

Pose Variation: Current approach assumes frontal facial views. Future work should incorporate pose-invariant features or multi-view analysis.

Expression Dependency: Performance may vary with different expressions. Including dynamic video analysis could capture movement-related asymmetries.

Ethnic Diversity: While CelebA provides diverse normal faces, validation across different ethnic populations is needed.

Clinical Validation: Prospective clinical trials comparing automated assessments with multiple physician ratings are necessary for clinical adoption.

Synkinesis Detection: Current features do not capture synkinesis (involuntary movements), an important aspect of chronic palsy assessment.

Future enhancements could include:

- Integration with video analysis for dynamic assessment
- Multi-task learning for simultaneous severity and region-specific grading
- Attention mechanisms to identify most affected facial regions
- Patient-specific longitudinal modeling for treatment monitoring

Conclusion and Future Work

This paper presented an automated facial palsy severity classification system using geometric angle features and Random Forest classification. The approach achieved 90.12% accuracy on a balanced dataset of 8,000 images spanning four severity categories. Key findings include:

- Three simple angular measurements (eyebrow, eye, mouth) effectively capture facial asymmetry patterns
- Mouth angle is the most discriminative feature (45.23% importance)
- High recall for Mild cases (97.00%) enables effective early detection
- Balanced performance across all severity levels demonstrates clinical viability
- Computational efficiency enables real-time deployment in clinical settings

The system represents a practical solution for objective, reproducible facial palsy assessment, addressing limitations of subjective clinical grading. While current results are promising, several directions for future work could further enhance the system's capabilities and clinical applicability:

Future Work

Multi-modal Integration: Future research will explore integrating additional modalities such as facial surface texture analysis, thermal imaging, and electromyography (EMG) data to provide a more comprehensive assessment of facial nerve function. Combining geometric features with texture-based analysis could capture subtle muscle atrophies that angular measurements alone might miss.

Dynamic Analysis: Extending the current static image analysis to video-based dynamic assessment would enable tracking of facial movement patterns over time. This could include quantifying velocity, acceleration, and symmetry of facial movements during voluntary expressions, providing additional discriminative features for severity grading.

Personalized Baseline Modeling: Developing patient-specific models that account for individual anatomical variations and pre-existing asymmetries could improve accuracy. This would involve establishing personalized baselines and tracking changes relative to these baselines over the course of treatment.

Synkinesis Detection: Future work will focus on detecting and quantifying synkinesis (involuntary muscle movements), which is a common complication in chronic facial palsy. This requires analyzing coordination patterns between different facial regions during voluntary movements.

Cross-population Validation: Extensive validation across diverse ethnic populations, age groups, and gender distributions is necessary to ensure the system's generalizability. This includes testing on populations with different facial anatomical characteristics.

Real-time Clinical Integration: Developing a user-friendly interface for clinical deployment, including integration with electronic health records and telemedicine platforms. This would facilitate widespread adoption in clinical practice.

Longitudinal Monitoring: Implementing capabilities for long-term patient monitoring to track recovery progress and treatment effectiveness over time, providing objective metrics for clinical decision-making.

Explainable AI Enhancements: Improving model interpretability through advanced visualization techniques that highlight which specific facial regions contribute most to severity classification decisions, enhancing clinician trust and understanding.

These future directions aim to transform the current system from a classification tool into a comprehensive decision support system that can assist clinicians in diagnosis, treatment planning, and progress monitoring of facial palsy patients.

References

[1] Q. T. Ngoc, S. Lee, and B. C. Song, "Facial Landmark-Based Emotion Recognition via Directed Graph Neural Network," *Electronics*, vol. 9, no. 5, p. 764, May 2020.

[2] R. P. Peitersen, "Bell's palsy: the spontaneous course of 2,500 peripheral facial nerve palsies of different etiologies," *Acta Oto-Laryngologica*, vol. 122, no. sup549, pp. 4–30, 2002.

[3] K. J. Baugh and R. W. Gilden, "Incidence and prevalence of Bell's palsy: a systematic review," *Clinical Epidemiology*, vol. 15, pp. 123–135, 2023.

[4] S. Vrabec and B. H. Backous, “Facial nerve grading system 2.0,” *Otolaryngology–Head and Neck Surgery*, vol. 153, no. 4, pp. 553–558, 2015.

[5] X. Zhang et al., “Automated facial analysis: a systematic review,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1555–1575, 2021.

[6] D. E. King, “Dlib-ml: a machine learning toolkit,” *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[7] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[8] Z. Liu et al., “Deep learning face attributes in the wild,” in *Proc. International Conference on Computer Vision (ICCV)*, 2015, pp. 3730–3738.

[9] B. J. F. Wong et al., “Quantitative analysis of facial asymmetry in Bell’s palsy,” *Laryngoscope*, vol. 120, no. 9, pp. 1768–1773, 2010.