

Hierarchical Biomarker Modeling for Facial Paralysis Assessment and Rehabilitation Monitoring

Ziqi Wang, *Student Member, IEEE*, Jing Bi, *Senior Member, IEEE*, Zhipeng Zheng, Dong Liu, Jiahui Zhai, Junqi Zhang, Xiaomeng Zhao, Yu Shen, Po Hu, Jiantao Liang, and Ying Tang, *Fellow, IEEE*

Abstract—Facial paralysis assessment and rehabilitation monitoring require not only accurate analysis of facial biomarkers but also a clinically deployable workflow that spans standardized data acquisition, quantitative modeling, and interpretable reporting. We present a clinically implemented facial paralysis assessment system that integrates four modules: a Data Acquisition module for standardized 4K/120 fps video recording, a Key Point Extraction and Indicator Calculation module that derives 313 fine-grained indicators from 11 standardized facial actions, an Analysis module built on the Hierarchical Dynamic Attention Patient-Adaptive Network (HiDAPA), and a Result Visualization Report Generation module for interpretable clinical reporting. At the core of the system, HiDAPA models biomarker importance in a hierarchical and personalized manner. It organizes the 313 indicators into a three-level representation to capture both structured biomarker relationships and individual heterogeneity. Using Sunnybrook scores as a clinical reference, experiments on 200 subjects show that HiDAPA achieves a Pearson correlation of 0.980 and an Acc@10 of 90.5%. The learned importance patterns are clinically interpretable, highlighting the dominance of symmetry-related indicators and spontaneous blink asymmetry. The system enables an objective and interpretable assessment of facial paralysis and supports long-term monitoring of rehabilitation progress.

Index Terms—Facial paralysis assessment, rehabilitation monitoring, hierarchical biomarker modeling, clinical interpretability.

I. INTRODUCTION

Facial paralysis is a neurological disorder caused by dysfunction of the facial nerve (cranial nerve VII), leading to varying degrees of facial muscle weakness or paralysis. It may result from conditions such as Bell's palsy [1], trauma [2], tumors [3], or surgical complications [4], and can severely impair fundamental functions including eye closure, oral movement, speech, and feeding [5]. Population-level evidence reports an overall incidence of 30.5 cases per 100,000 adults, increasing from 17.6 to 47.4 cases per 100,000 over a 16-year period [6]. Accurate and reproducible assessment is therefore essential for initial severity evaluation and rehabilitation planning (Fig. 1).

In clinical practice, the House-Brackmann (HB) grading system and the Sunnybrook Facial Grading System are the most

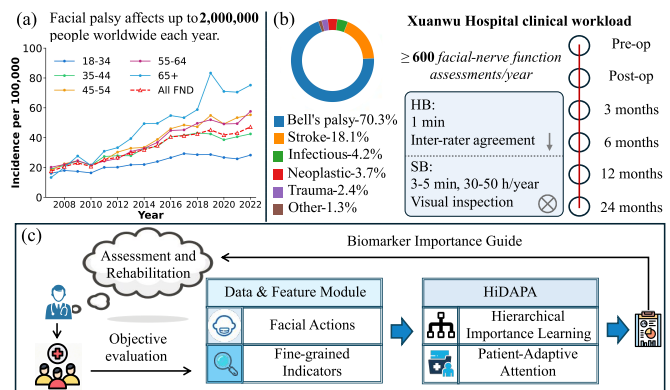


Figure 1: Motivation and framework overview. (a) The incidence of facial palsy has increased in a U.S. case-study cohort, indicating a growing clinical burden. (b) Subjective grading limits reliable long-term follow-up. (c) Our method enables objective, fine-grained assessment and rehabilitation monitoring.

widely adopted standardized assessment tools. While HB provides a coarse categorical classification of severity, Sunnybrook offers a more sensitive composite score by integrating resting symmetry, voluntary movement amplitude, and synkinesis. However, both systems rely heavily on subjective visual judgment, resulting in limited objectivity and inter-rater variability, which restricts their reproducibility and fine-grained analytical capacity. In practice, HB assessment can generally be completed within approximately one minute, whereas Sunnybrook scoring requires several minutes per evaluation [7]. Moreover, patients are commonly reassessed at 3, 6, and 12 months and may require continued follow-up for 1-2 years, making consistent cross-visit comparison difficult and substantially increasing cumulative clinical workload. More importantly, the lack of objective and interpretable biomarker-level analysis limits their ability to support personalized rehabilitation monitoring and data-driven [8] clinical decision-making.

Both clinician-based grading and automated analysis rely on informative facial biomarkers [9], but their diagnostic relevance is not uniform [10]. Different indicator types, facial actions, and fine-grained measurements reflect distinct functional impairments, and their importance varies across patients. Therefore, without explicit modeling of structured biomarker importance, global severity scores alone provide limited insight into individualized dysfunction patterns and insufficient support for personalized rehabilitation monitoring [11]. To address these limitations, we propose a clinically deployed end-to-end facial paralysis

Xiaomeng Zhao, Jiantao Liang, and Jing Bi are the corresponding authors. Ziqi Wang and Dong Liu contributed equally to this work.

Ziqi Wang, Jing Bi, Zhipeng Zheng, Jiahui Zhai, and Junqi Zhang are with the College of Computer Science, Beijing University of Technology.

Dong Liu and Jiantao Liang are with the Department of Neurosurgery, Xuanwu Hospital, Capital Medical University and International Neuroscience Institute (China-INI).

Xiaomeng Zhao, Yu Shen, and Po Hu are with the Beijing Neurorobot Technology Co., Ltd.

Ying Tang is with the Department of Electrical and Computer Engineering, Rowan University.

assessment system composed of four integrated modules: a Data Acquisition module for standardized 4K/120 fps video recording, a Key Point Extraction and Indicator Calculation module that derives 313 fine-grained indicators from 11 standardized facial actions, an Analysis module based on the Hierarchical Dynamic Attention Patient-Adaptive Network (HiDAPA), and a Result Visualization Report Generation module for interpretable clinical reporting. HiDAPA models the 313 indicators across seven clinically meaningful types: symmetry, angle, distance, area, duration, significance, and count. It progressively learns type-level importance, action-type interaction importance, and patient-specific fine-grained indicator importance. Patient-adaptive modeling is further enabled by dynamic graph-structure learning, which constructs individualized indicator-association graphs. The main contributions of this work are summarized as follows:

- 1) We develop a clinically deployed end-to-end facial paralysis assessment system integrating standardized data acquisition, key-point extraction and indicator calculation, HiDAPA-based analysis, and interpretable report generation. The system provides an objective and reproducible alternative to subjective visual assessment.
- 2) We propose HiDAPA, a hierarchical and patient-adaptive framework for fine-grained facial biomarker importance modeling. It models 313 indicators across seven clinically meaningful categories through type-level, action-type, and patient-specific importance learning.
- 3) Using Sunnybrook scores as a clinical reference, experiments on a real-world clinical dataset show that HiDAPA outperforms ten representative baselines, achieving a Pearson correlation of 0.980 and an Acc@10 of 90.5%. The learned importance distributions are clinically consistent and patient-specific.

The remainder of this paper is organized as follows. Section II reviews related work on facial paralysis assessment and biomarker-based modeling. Section III presents the overall workflow of the proposed system. Section IV details the design of HiDAPA. Section V presents the experiments and interpretability analyses. Finally, Section VI concludes the paper.

II. RELATED WORK

Objective facial paralysis assessment requires a workflow from standardized data acquisition to interpretable reporting. However, existing studies usually focus on individual components.

A. Indicator Extraction and Computational Assessment

Reliable facial paralysis assessment depends on standardized acquisition and facial action protocols, as variations in head pose, illumination, and frame rate can reduce biomarker reproducibility [12]. Although video-based analysis captures dynamic facial weakness, inconsistent acquisition settings may limit cross-session and cross-patient consistency [13]. Heinrich *et al.* [14] show that controlled acquisition supports reliable landmark-based symmetry analysis, while Feng *et al.* [15] demonstrate the value of dynamic facial motion for severity grading.

Early computational methods extract geometric asymmetry features from facial landmarks and apply machine learning for severity assessment. Rao *et al.* [16] model facial motion using

statistical distributions and likelihood-ratio tests. Ye *et al.* [17] enhance disease-related expression features through positional encoding and sparse attention, while Ge *et al.* [18] combine action-unit localization with local-global feature fusion. Recent deep learning methods further improve representation capacity. Gan *et al.* [19] combine local temporal asymmetry with global static features, Liu *et al.* [20] model regional and temporal information using hierarchical convolutional and recurrent networks, and Zhang *et al.* [21] integrate landmark symmetry with optical flow for HB scoring. However, existing methods lack an integrated acquisition workflow and focus on final score prediction without explaining indicator contributions or adapting biomarker importance to individual patients. Our system addresses these limitations through standardized acquisition, fine-grained indicator extraction, and patient-adaptive hierarchical modeling.

B. Comprehensive Assessment Frameworks for Facial Paralysis

Facial paralysis assessment requires both overall severity scoring and quantitative analysis of facial motor function [22]. The Sunnybrook scale evaluates resting symmetry, voluntary movement, and synkinesis. Tedeschi *et al.* [23] note that discrete scores may not capture continuous rehabilitation changes, while Guan *et al.* [24] use dense facial landmarks for continuous assessment of eyelid and mouth function. However, existing frameworks mainly rely on predefined regions, fixed-weight indicators, and relatively flat feature organization, limiting their ability to represent hierarchical facial function and patient-specific heterogeneity. A unified system integrating standardized acquisition, indicator calculation, adaptive biomarker modeling, interpretable reporting, and rehabilitation monitoring remains lacking. This work integrates these components into an end-to-end clinical workflow.

III. SYSTEM OVERVIEW

The proposed system provides an end-to-end clinical assessment workflow that transforms subjective manual scoring into an objective, repeatable, and interpretable process. It consists of four core components (Fig. 2).

A. Data Acquisition module

It records subjects' facial motion videos under standardized conditions. The acquisition device (Fig. 2) consists of four components: a head and face alignment device, a high-speed 4K video capture system, a local AI computing center, and high-density LED lighting. Subjects perform 11 standardized facial actions (Table I), covering ocular, brow, nasal, oral, blinking, and resting facial motor functions mediated by the facial nerve. Static actions are maintained for 3 s, while dynamic actions are recorded for 10 s. The system records videos at 4K resolution and 120 fps to capture subtle facial movements [25], [26]. The alignment device reduces posture variation, and the LED lighting provides uniform illumination. Real-time preprocessing is performed locally at the AI computing center equipped with an NVIDIA RTX 3500 Ada GPU and 128 GB RAM, including frame extraction and quality screening, so that raw videos remain on site. During acquisition, an operator verifies that the subject faces the camera, maintains a level gaze, and remains aligned with the facial-positioning guides. Recordings with evident yaw, pitch, in-plane rotation, or head displacement are repositioned and reacquired.

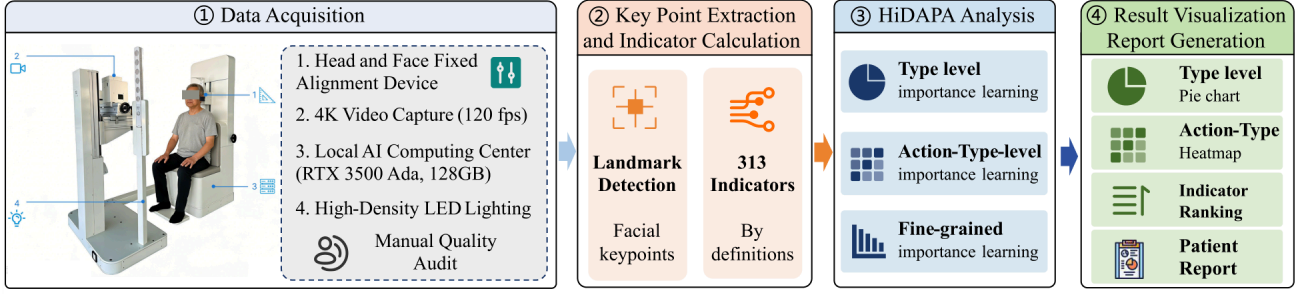


Figure 2: Overall framework of the proposed facial paralysis assessment system.

Table I: Facial action protocols for standardized data acquisition.

ID	Action Name	Action Description
A1	Cheek Puffing	The patient inflates both cheeks and holds the air.
A2	Forceful Eye Closure	The patient closes both eyes tightly with maximum effort.
A3	Gentle Eye Closure	The patient gently closes both eyes without applying force.
A4	Eyebrow Raising	The patient raises both eyebrows and wrinkles the forehead.
A5	Teeth Baring	The patient bares the teeth with the mouth stretched laterally.
A6	Nose Wrinkling	The patient wrinkles the nose by contracting the nasal muscles.
A7	Smile	The patient smiles naturally without showing the teeth.
A8	Spontaneous Eye Blink	The patient blinks naturally without deliberate control.
A9	Voluntary Eye Blink	The patient deliberately blinks three times on command.
A10	Lip Puckering	The patient protrudes and puckers the lips forward.
A11	Resting State	The patient maintains a relaxed and neutral facial expression.

B. Key Point Extraction and Indicator Calculation module

This module converts raw video frames into structured facial motion data for subsequent analysis. Facial landmarks are localized using a modular keypoint detection interface, which can be instantiated with either established open-source detectors [27] or our previously developed model [28]. Before indicator calculation, five registration landmarks (the bilateral pupil centers, nose tip, and bilateral mouth corners) are aligned with their corresponding positions in a predefined frontal-face template using a two-dimensional similarity transformation:

$$(\hat{s}, \hat{\mathbf{R}}, \hat{\mathbf{t}}) = \arg \min_{s, \mathbf{R}, \mathbf{t}} \sum_{i=1}^5 \|s\mathbf{R}\mathbf{p}_i + \mathbf{t} - \mathbf{q}_i\|_2^2, \quad (1)$$

subject to $\mathbf{R}^T \mathbf{R} = \mathbf{I}$, $\det(\mathbf{R}) = 1$, and $s > 0$, where $\mathbf{p}_i, \mathbf{q}_i \in \mathbb{R}^2$ denote the detected and template landmark coordinates, respectively, s is the isotropic scale factor, $\mathbf{R}$ is the in-plane rotation matrix, and $\mathbf{t}$ is the translation vector. All facial landmarks are then mapped as $\tilde{\mathbf{p}} = \hat{s}\hat{\mathbf{R}}\mathbf{p} + \hat{\mathbf{t}}$. Because the same global transformation is applied to all landmarks, it reduces differences in facial position, in-plane rotation, and imaging scale without introducing nonuniform scaling, shear, or side-specific deformation. Distance indicators are normalized by the interpupillary distance, and area indicators by its square. The vertical midline of the normalized frontal template, $x = x_c$, is used as the facial symmetry reference axis, where x_c denotes the horizontal coordinate of the template midline. Reflection across this axis is defined as $\mathcal{M}(x, y) = (2x_c - x, y)$. This

two-dimensional normalization does not fully correct perspective deformation caused by residual out-of-plane yaw or pitch. Such pose variations are minimized through standardized frontal acquisition and operator-guided quality control.

Based on the normalized landmark trajectories, the system computes 313 predefined facial motion indicators from 11 standardized facial actions. These indicators cover seven clinically meaningful types: symmetry, angle, distance, area, duration, significance, and count, as summarized in Table II. Fig. 3 illustrates the landmark configuration used for geometric analysis, while Table III presents representative indicator definitions for each facial action. The complete definitions of all 313 indicators are provided in the supplementary material.

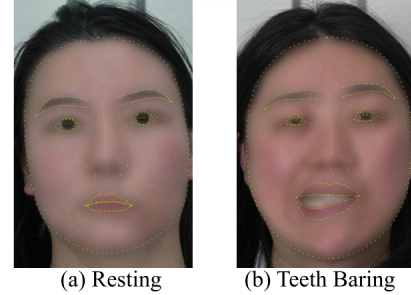


Figure 3: Representative facial landmark layouts from two patients: (a) resting state and (b) teeth baring. The detected landmarks are used to derive geometric and temporal indicators. The patient's face is blurred to ensure privacy.

Table II: Definitions of the Seven Indicator Types.

ID	Type	Description of different types of indicators
T1	Symmetry	Left-right asymmetry of facial landmarks.
T2	Angle	Angles of facial muscle movements and joint rotations.
T3	Distance	Distances between anatomical landmarks.
T4	Area	Areas of facial regions.
T5	Duration	Duration of each facial action and movement phases.
T6	Significance	Degree of motion amplitude and its deviation from baseline.
T7	Count	Frequency of discrete events.

C. HiDAPA Analysis module

The HiDAPA Analysis module receives the 313 structured indicators as input. To capture biomarker importance at different

Table III: Representative indicator definitions for each standardized facial action.

Facial action	Representative indicator	Definition
[A1] Cheek Puffing	Upper-Lip Shape Symmetry [T1]	Normalized bilateral upper-lip contour discrepancy: $E_{UL} = D(\mathcal{P}_L^{UL}, \mathcal{M}(\mathcal{P}_R^{UL}))/\ell$.
[A2] Forceful Eye Closure	Eye Closeness (R) [T6]	Right-eye closure significance: $C_R = 1 - A_R^{eye}/c_R^2$.
[A3] Gentle Eye Closure	Eye Closeness (L) [T6]	Left-eye closure significance: $C_L = 1 - A_L^{eye}/c_L^2$.
[A4] Eyebrow Raising	Left/Right Eyebrow-Eye Distance Ratio [T3]	Ratio of the bilateral eyebrow-eye distances: $r_{be} = d_L^{be}/d_R^{be}$.
[A5] Teeth Baring	Left/Right Upper-Lip Area Ratio [T4]	Ratio of the upper-lip areas separated by the facial midline: $r_{UL} = A_L^{UL}/A_R^{UL}$.
[A6] Nose Wrinkling	Vertical Tilt Angle of Face [T2]	Angle between the forehead-chin vector and facial midline: $\theta_{face} = \angle(\mathbf{p}_{chin} - \mathbf{p}_{forehead}, \mathbf{v}_{mid})$.
[A7] Smile	Bilateral Symmetry of Overall Facial Shape [T1]	Normalized bilateral facial-contour discrepancy: $E_{face} = D(\mathcal{P}_L^{face}, \mathcal{M}(\mathcal{P}_R^{face}))/\ell$.
[A8] Spontaneous Eye Blink	Blink Count [T7]	Number of complete five-phase blink cycles: $N_{blink} = \sum_{k=1}^K \mathbb{I}(\mathcal{S}_k = \tau_1 \rightarrow \tau_5)$.
[A9] Voluntary Eye Blink	Left/Right Blink-Minimum Duration Ratio [T5]	Ratio of the average bilateral closed-eye durations: $r_{min} = \overline{\Delta t}_{L, \tau_2:\tau_4} / \overline{\Delta t}_{R, \tau_2:\tau_4}$.
[A10] Lip Puckering	Mouth-Corner Horizontal-Level Difference [T3]	Normalized vertical difference between the bilateral mouth corners: $H_{mouth} = (\mathbf{p}_{mc,R} - \mathbf{p}_{mc,L})^\top \hat{\mathbf{v}}_{mid} /\ell$.
[A11] Resting State	Left Eye Area [T4]	Scale-normalized left palpebral-fissure area: $\tilde{A}_L^{eye} = A_L^{eye}/\ell^2$.

Note: L and R denote the left and right facial sides, and [T1]–[T7] correspond to the indicator types in Table II. τ_1 – τ_5 are the five phases of a complete blink. $\mathcal{P}_s^\Omega$ denotes the landmark set of region Ω on side s , $\mathcal{M}(\cdot)$ denotes reflection about the facial midline, and $D(\cdot, \cdot)$ is the mean landmark distance after similarity alignment; smaller E_{UL} and E_{face} indicate greater symmetry. Polygon area is calculated as $A(\mathcal{P}) = \frac{1}{2} \left| \sum_{i=1}^N (x_i y_{i+1} - x_{i+1} y_i) \right|$, with cyclic indexing. c_s is the inner-to-outer canthus distance, ℓ is the reference facial length, and d_s^{be} is the eyebrow-eye distance. $\mathbf{p}_{forehead}$, $\mathbf{p}_{chin}$, and $\mathbf{p}_{mc,s}$ denote the forehead, chin, and mouth-corner landmarks. $\hat{\mathbf{v}}_{mid} = \mathbf{v}_{mid}/\|\mathbf{v}_{mid}\|_2$. $\mathcal{S}_k$ is the phase sequence of the k -th blink candidate, K is the number of candidates, and $\mathbb{I}(\cdot)$ is the indicator function.

granularities, it organizes the indicators into a three-level representation and progressively learns type-level importance, action-type interaction importance, and patient-specific fine-grained indicator importance. The resulting personalized importance distribution supports interpretable reporting and clinical assessment.

D. Result Visualization Report Generation module

This module converts HiDAPA's hierarchical importance distributions into an interpretable clinical report. It includes a type-level importance pie chart (Fig. 5), an action-type heatmap (Fig. 6), and a ranked list of the top-10 patient-specific indicators (Fig. 8). The complete 313 indicator values and importance scores are also retained for review and longitudinal follow-up.

IV. HiDAPA ANALYSIS MODULE

HiDAPA learns patient-specific hierarchical importance patterns of facial motion indicators¹ to identify clinically meaningful biomarkers. As shown in Fig. 4, it follows a three-stage design:

- 1) Data preprocessing. Indicators are organized into a three-level representation, and clinical attributes are encoded as patient-specific conditional information.
- 2) Hierarchical importance learning. Type-level and action-type importance signals are learned to capture global diagnostic relevance and structured medical semantics.
- 3) Fine-grained indicator importance learning. Guided by the upper-level importance signals, dynamic graphs and adaptive attention are used to model patient-specific indicator relationships and infer fine-grained biomarker importance.

¹Implementation available at: <https://github.com/ZhipengZheng607/HiDAPA>

A. Data Preprocessing

Given facial motion data from N subjects, our objective is to learn the relative contribution of each indicator to facial paralysis assessment. To capture medical semantics at different granularities, the indicators are organized into a three-layer representation. The first layer summarizes each indicator type using statistical descriptors, forming type features $\mathbf{T} \in \mathbb{R}^{7 \times 5}$. The five dimensions correspond to the mean, standard deviation, skewness, kurtosis, and anomaly rate, providing a global overview of category-level characteristics. The second layer models action-type interactions via an action-type feature matrix $\mathbf{A} \in \mathbb{R}^{11 \times 7}$, where each entry

$$\mathbf{A}[a, t] = \frac{1}{|\mathcal{I}_{a,t}|} \sum_{j \in \mathcal{I}_{a,t}} x_j \quad (2)$$

represents the average standardized response of indicators belonging to action a and type t . $\mathcal{I}_{a,t}$ is the set of indicators belonging to action a and type t , and x_j is the indicator value standardized by Z-score. The third layer retains the standardized fine-grained indicator features $\mathbf{X} \in \mathbb{R}^{313 \times 1}$ for subsequent patient-specific modeling. To enable patient adaptation, clinical attributes are encoded into a conditional vector $\mathbf{c} = [\text{HB}/5, \text{P}, \text{L}, \text{R}]$, where HB denotes the House-Brackmann grade [29], P denotes a binary patient-status indicator: P=1 for patients with facial paralysis and P=0 for healthy controls. It is used only during training as an auxiliary contextual cue to help the model distinguish facial-paralysis-related representations from normal facial-motion patterns, and L/R denote the affected side (left/right). These attributes are projected into a high-dimensional patient embedding $\mathbf{e}_p \in \mathbb{R}^{128}$, which conditions all subsequent modules.

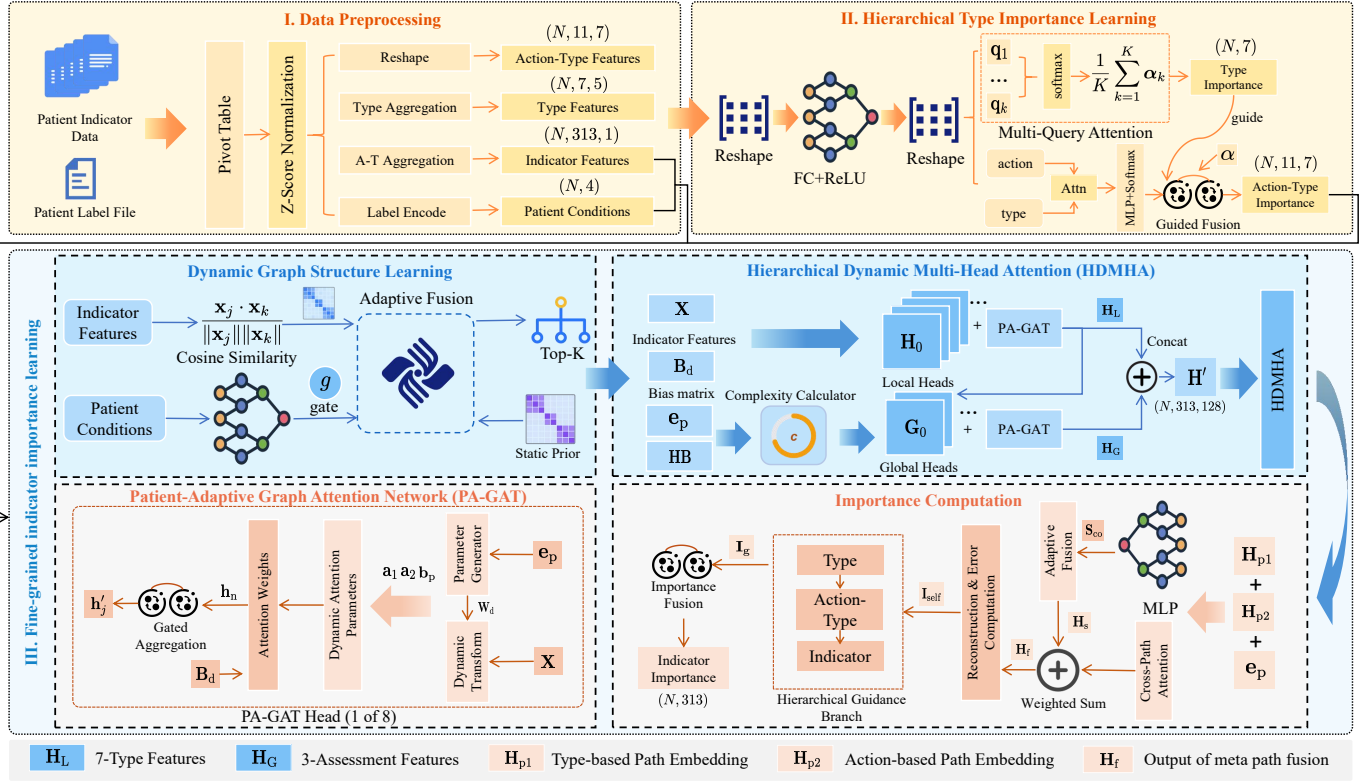


Figure 4: Overview of the proposed HiDAPA for hierarchical and patient-adaptive facial biomarker importance modeling.

B. Hierarchical Type Importance Learning

1) *Type-Level Importance Learning*: Different indicator types contribute unequally to facial paralysis assessment. To capture their global diagnostic relevance, the type features $\mathbf{T}$ are projected into a latent space and processed by a multi-head attention module with K learnable global queries. Each query produces a type-level attention vector $\alpha_k \in \mathbb{R}^7$, and the final type-level importance vector is obtained by averaging the attention weights across all queries:

$$\mathbf{I}_t = \frac{1}{K} \sum_{k=1}^K \alpha_k \in \mathbb{R}^7. \quad (3)$$

2) *Action-Type Interaction Importance Learning*: The diagnostic importance of each indicator type varies across facial actions. To model this action-dependent relevance, action-type features derived from $\mathbf{A} \in \mathbb{R}^{11 \times 7}$ and type-level representations are projected into a shared latent space to obtain self-learned action-type importance scores $\mathbf{I}_{at}$. To maintain consistency with global type-level semantics, a hierarchical guidance mechanism is introduced:

$$\mathbf{I}'_{at} = \alpha \cdot \mathbf{I}_{at} + (1 - \alpha) \cdot \mathbf{I}'_t, \quad (4)$$

where $\mathbf{I}'_t$ denotes the broadcasted $\mathbf{I}_t$, and α balances action-specific adaptation with global semantic consistency.

C. Fine-grained Indicator Importance Learning

1) *Dynamic Graph Structure Learning*: Existing methods often use a fixed adjacency structure shared by all patients, implicitly assuming uniform indicator relationships across individuals.

This assumption is limited for facial paralysis, where lesion location, severity, and compensatory patterns may lead to different indicator coupling patterns. We therefore construct a patient-specific graph by combining data-driven indicator relevance with medical prior knowledge.

Specifically, pairwise indicator similarity is computed to form $\mathbf{S}$. At the input layer, the normalized inner product of standardized scalar indicators provides a directional-consistency cue. In subsequent layers, cosine similarity is recomputed from learned multidimensional indicator embeddings to capture continuous subject-specific similarities. In parallel, a static prior graph $\mathbf{A}_{static}$ is constructed to encode category-level medical prior knowledge. Specifically, $\mathbf{A}_{static}[j, k] = 1$ if indicators j and k belong to the same clinical indicator type, and 0 otherwise. A patient-conditioned gate is then used to balance the data-driven similarity and the static prior:

$$g = \sigma \left(\mathbf{e}_p \mathbf{w}_g + \frac{\mathbf{HB}}{5} \right), \quad (5)$$

where $\sigma(\cdot)$ denotes the sigmoid function, $\mathbf{w}_g$ is a learnable parameter, and $g \in [0, 1]$ controls the relative contribution of the two graph components. The fused adjacency matrix is obtained as

$$\tilde{\mathbf{A}} = g \cdot \mathbf{S} + (1 - g) \cdot \mathbf{A}_{static}. \quad (6)$$

To further reflect pathological asymmetry, connections involving indicators associated with the affected side are strengthened by an enhancement factor. Top- K sparsification is then applied to $\tilde{\mathbf{A}}$ to retain the strongest connections for each node and suppress weak or noisy similarities. The resulting sparse

matrix is symmetrized and augmented with self-connections to ensure structural consistency. Finally, temperature scaling is used to adjust edge-weight sharpness, producing the dynamic graph bias matrix $\mathbf{B}_d$ for subsequent attention computation. Thus, the constructed graph is adapted to each patient according to clinical severity and affected-side information.

2) *Patient-Adaptive Graph Attention Network (PA-GAT)*: Conventional graph attention networks share transformation and attention parameters across all samples, limiting their ability to model patient-specific indicator interactions. To address this limitation, we design PA-GAT, where key attention parameters are conditioned on the patient embedding $\mathbf{e}_p$. Specifically, the transformation matrix $\mathbf{W}_d$ is dynamically generated from $\mathbf{e}_p$, and the feature of node j , denoted as $\mathbf{x}_j$, is projected as $\mathbf{h}_j = \mathbf{W}_d \mathbf{x}_j$.

The attention parameters $\mathbf{a}_1$, $\mathbf{a}_2$, and bias b_p are also generated from $\mathbf{e}_p$. The attention logit between nodes j and k is computed as

$$e_{jk} = \text{LeakyReLU}((\mathbf{a}_1)^\top \mathbf{h}_j + (\mathbf{a}_2)^\top \mathbf{h}_k + b_p). \quad (7)$$

$\mathbf{B}_d$ is then incorporated into the attention normalization:

$$\alpha_{jk} = \frac{\exp(e_{jk} + \mathbf{B}_d[j, k])}{\sum_{k' \in \mathcal{N}_d(j)} \exp(e_{jk'} + \mathbf{B}_d[j, k'])}, \quad (8)$$

where $\mathcal{N}_d(j)$ denotes the neighbor set of node j in the dynamic graph. In this way, both neighborhood structure and attention strength are modulated by patient-specific information.

Finally, a gating vector $\mathbf{g}_1$ is generated from $\mathbf{e}_p$ to balance self-information and neighborhood aggregation. The final node representation is obtained by

$$\mathbf{h}'_j = \mathbf{g}_1 \odot \mathbf{h}_n + (1 - \mathbf{g}_1) \odot \mathbf{h}_j, \quad (9)$$

where $\odot$ denotes element-wise multiplication, and $\mathbf{h}_n = \sum_{k \in \mathcal{N}_d(j)} \alpha_{jk} \mathbf{h}_k$ is the aggregated neighborhood representation. This gate allows PA-GAT to adaptively control the contribution of neighboring indicators according to patient-specific conditions.

3) *Hierarchical Dynamic Multi-Head Attention (HDMHA)*: Patients exhibit different levels of pathological complexity, requiring representations at both local and global granularities. Based on PA-GAT, we design an HDMHA mechanism that combines local heads and global heads. Local heads focus on type-specific indicator relationships and are constrained by a type-specific mask $\mathbf{B}_m$, while global heads capture cross-type functional patterns. The output of a single local head and a single global head are computed as

$$\mathbf{H}_l = \text{PA-GAT}(\mathbf{X}, \mathbf{B}_d + \mathbf{B}_m, \mathbf{e}_p), \quad (10)$$

$$\mathbf{H}_g = \text{PA-GAT}(\text{Dense}_{\text{sym}}(\mathbf{F}_s), \mathbf{B}_d, \mathbf{e}_p), \quad (11)$$

where $\mathbf{F}_s$ denotes the concatenated outputs of selected local heads, and $\text{Dense}_{\text{sym}}$ is a fully connected layer that projects them into a unified dimension. A patient complexity score derived from $\mathbf{e}_p$ and HB grade determines the number of active local and global heads through linear interpolation. Fewer heads are activated for mild cases, while more heads are used for severe cases to capture more diverse abnormal patterns. The outputs of active local heads and active global heads are uniformly averaged to obtain the group-level representations $\mathbf{H}_L$ and $\mathbf{H}_G$, respectively. They are concatenated as $\mathbf{H}' = [\mathbf{H}_L \parallel \mathbf{H}_G]$ and fed into subsequent meta-path processing.

4) *Importance Computation*: To capture complementary indicator relationships, we construct two meta-paths for fine-grained modeling. The type-based meta-path connects indicators belonging to the same clinical type to model intra-type dependencies, while the action-based meta-path connects indicators involved in the same facial action to capture action-specific coordination. Both paths use the same input features $\mathbf{X}$ but different adjacency matrices. After being processed by HDMHA, the type-based and action-based embeddings are obtained as $\mathbf{H}_{p1}, \mathbf{H}_{p2} \in \mathbb{R}^{313 \times 512}$, respectively.

Patient-adaptive path fusion. The two meta-path embeddings are integrated through a patient-conditioned fusion strategy. A compatibility score s_{co} is estimated from the pooled path representations and $\mathbf{e}_p$ using an MLP. The fused representation is then computed using s_{co} and the patient-dependent path weights w_1 and w_2 :

$$\mathbf{H}_s = s_{co}(\mathbf{H}_{p1} + \mathbf{H}_{p2}) + (1 - s_{co})(w_1 \mathbf{H}_{p1} + w_2 \mathbf{H}_{p2}), \quad (12)$$

where w_1 and w_2 are generated from the pooled path representations and $\mathbf{e}_p$. To further enhance cross-path interaction, the cross-path enhanced embeddings are defined as

$$\begin{aligned} \tilde{\mathbf{H}}_{p1} &= \mathbf{H}_{p1} + \beta \text{CrossAttn}(\mathbf{H}_{p1}, \mathbf{H}_{p2}, \mathbf{e}_p), \\ \tilde{\mathbf{H}}_{p2} &= \mathbf{H}_{p2} + \beta \text{CrossAttn}(\mathbf{H}_{p2}, \mathbf{H}_{p1}, \mathbf{e}_p). \end{aligned} \quad (13)$$

The enhancement term is then computed as:

$$\mathbf{H}_e = s_{co}(\tilde{\mathbf{H}}_{p1} + \tilde{\mathbf{H}}_{p2}) + (1 - s_{co})(w_1 \tilde{\mathbf{H}}_{p1} + w_2 \tilde{\mathbf{H}}_{p2}), \quad (14)$$

where β is the scaling factor for cross-path attention. The final indicator embedding is obtained by combining the fusion and enhancement terms: $\mathbf{H}_f = \theta \mathbf{H}_s + (1 - \theta) \mathbf{H}_e$.

Self-supervised importance estimation. Given the final embedding $\mathbf{H}_f$, the original indicators are reconstructed as $\hat{\mathbf{X}} = \text{Dense}(\mathbf{H}_f)$. The importance of indicator j is quantified by its reconstruction difficulty:

$$e_j = \|\mathbf{X}[j] - \hat{\mathbf{X}}[j]\|^2 \cdot w_s[j] \cdot w_{HB}, \quad (15)$$

where $w_s[j]$ and w_{HB} are predefined weights. Specifically, $w_{HB} = 0.5 + \text{HB}/6$ for patients and 0.5 for healthy controls. $w_s[j]$ assigns a boost factor of 1.5 to indicators on the affected side and 1.0 otherwise, followed by row-wise normalization and clipping to $[0.5, 2.0]$. Indicators with larger reconstruction difficulty are considered more informative because they reflect less redundant or more patient-specific functional variations. To reduce the influence of global error magnitude across severity levels, reconstruction errors are normalized at the patient level:

$$\mathbf{I}_{\text{self}}[j] = \frac{e_j - e_{\min}}{e_{\max} - e_{\min} + \epsilon}, \quad (16)$$

where $e_{\min}$ and $e_{\max}$ are the minimum and maximum reconstruction errors among all indicators of the current patient, and ϵ is a small constant to prevent division by zero.

Hierarchical guidance integration. The action-type importance $\mathbf{I}'_{at}$ learned at (4) is projected onto each indicator j according to its action a_j and type t_j , yielding a guidance signal $\mathbf{I}_g[j]$. The final importance is obtained by linearly combining the self-learned importance and hierarchical guidance:

$$\mathbf{I}_{\text{final}}[j] = \eta \mathbf{I}_{\text{self}}[j] + (1 - \eta) \mathbf{I}_g[j], \quad (17)$$

where η controls the balance between reconstruction-based importance and upper-level medical guidance. This integration promotes consistency between fine-grained indicator importance and hierarchical medical semantics.

D. Loss function design

The overall training objective is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_r + \lambda_1 \mathcal{L}_{\text{at}} + \lambda_2 \mathcal{L}_{\text{type}} + \lambda_3 \mathcal{L}_c + \mathcal{L}_{\text{L2}}, \quad (18)$$

where λ_1, λ_2 , and λ_3 are trade-off coefficients. The reconstruction loss $\mathcal{L}_r$ is weighted by affected-side and HB-grade information to emphasize clinically relevant facial motion deviations. The hierarchical prior terms $\mathcal{L}_{\text{at}}$ and $\mathcal{L}_{\text{type}}$ regularize the learned action-type and type-level importance distributions, encouraging consistency with medical priors and reducing overfitting to noisy indicator interactions. To improve separability between patients and healthy controls in the representation space, we introduce a comparative embedding loss:

$$\mathcal{L}_c = \frac{1}{N_p} \sum_{i:y_i=1} \max(0, 1 - \bar{r}^{(i)}) + \frac{1}{N_n} \sum_{i:y_i=0} \bar{r}^{(i)}, \quad (19)$$

where $\bar{r}^{(i)} = \frac{1}{M} \sum_{j=1}^M \|\mathbf{h}_j^{(i)}\|_2^2$ denotes the mean squared norm of the indicator embeddings for subject i , M is the number of indicators, N_p and N_n denote the numbers of patients and healthy subjects, respectively, $y_i \in \{0, 1\}$ is the sample label, and $\mathbf{h}_j^{(i)}$ is the embedding of the j th indicator for subject i . This loss encourages patient representations to deviate from the normal baseline while keeping healthy representations compact. The L2 regularization term $\mathcal{L}_{\text{L2}}$ is used to reduce model overfitting.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Setting and Evaluation Metrics

1) *Dataset*: The dataset is derived from a five-year standardized clinical data collection conducted at the Department of Neurosurgery, Xuanwu Hospital, Capital Medical University². 200 subjects are included, comprising 143 patients with facial paralysis and 57 healthy controls. The HB grades of the patients range from II to VI [30], covering different severity levels. Two experienced surgeons independently assign Sunnybrook Facial Nerve Scores [31] to provide clinical reference annotations, achieving an intraclass correlation coefficient (ICC) of 0.91, which indicates excellent inter-rater reliability. The dataset is split at the subject level into training, validation, and test sets at a ratio of 6:2:2. The three sets are fully subject-disjoint. All examinations and longitudinal visits from the same patient are assigned to the same set.

2) *Parameter Setting*: HiDAPA uses a two-layer graph attention network with hidden dimension 64 and ELU activation. The HDMHA module includes up to 7 local heads and 3 global heads, with the number of active heads adaptively adjusted from 4 to 10 depending on the complexity of the patient. Specifically, the number of local heads ranges from 3 to 7, and the number of global heads ranges from 1 to 3. The meta-path attention

dimension is set to 128. In the type-level importance learning module, the number of global queries K is set to 4. In dynamic graph learning, the temperature parameter is 0.5, the Top- K sparsification parameter is 20, and the affected-side enhancement factor is 1.5. The model is optimized using Adam with an initial learning rate of 0.001 and an L2 regularization coefficient of 1×10^{-4} . For the loss function, λ_1, λ_2 , and λ_3 are set to 0.01, 0.5, and 0.5, respectively. The normalization constant ϵ is set to 10^{-6} . The action-type hierarchical guidance coefficient α is set to 0.8, the cross-path attention weight β is set to 0.2, the embedding fusion coefficient θ is set to 0.7, and the final importance fusion coefficient η is set to 0.85.

3) *Clinical Consistency Evaluation*: To evaluate whether the learned importance is consistent with established clinical assessment criteria, we derive an indicator-based severity index and compare it with the Sunnybrook Facial Nerve Score. We first compute a type-wise aggregated loss:

$$L_t[t] = \mathbf{I}_t[t] \sum_{j \in \mathcal{T}_t} \mathbf{I}_{\text{final}}[j] e_j, \quad (20)$$

where $L_t[t]$ denotes the weighted loss of indicator type t , $\mathcal{T}_t$ is the index set of indicators belonging to type t . The overall severity index is then computed as $L_{\text{score}} = \sum_{t=1}^7 L_t[t]$. To compare the learned indicator-based severity index with the Sunnybrook reference scale, L_{score} is aligned to the Sunnybrook score range through a monotonic inverse calibration using training-set statistics only. This step serves as an evaluation interface to existing clinical criteria and does not participate in model training.

4) *Comparison Methods*: To comprehensively evaluate HiDAPA, we compare it with ten representative baselines, including feature selection methods (Concrete-AE [32], SCARF [33], SubTab [34]), autoencoder-based methods (IP-CAE [35], SDAE [36]), and graph neural network models (GATv2 [37], HGT [38]). We further include three facial-paralysis-specific methods: MLST-Net [15], UPFP-SG [19], and PHCNN [20]. These methods are adapted to the common 313-indicator representation while preserving their core modeling principles. MLST-Net models decoupled action and type-level dependencies, UPFP-SG organizes static and dynamic indicators according to facial-nerve branches, and PHCNN models global and regional action sequences through hierarchical and temporal feature learning. To assess the clinical validity of the learned importance distribution, we compare the calibrated indicator-based severity scores with the Sunnybrook reference scores using the evaluation procedure described in Section V-A3. Higher agreement indicates that the learned importance patterns are more consistent with established clinical assessment criteria. Five metrics are reported: Pearson correlation coefficient (r) for linear agreement, Spearman correlation (ρ) for ranking consistency, mean absolute error (MAE) and root mean squared error (RMSE) for score deviation, and ± 10 -point accuracy (Acc@10) for clinical tolerance. All results are reported as the mean $\pm$ sample standard deviation over 10 independent runs using the same subject-disjoint split.

B. Comparative Study

Table IV summarizes the comparative results. HiDAPA achieves the best performance across all metrics, with a Pearson

²The study was approved by the Ethics Committee of Xuanwu Hospital, Capital Medical University, Beijing, China (Linyan Review 2021[221], Project Number: KD2021221).

correlation of 0.980, MAE of 4.90, and Acc@10 of 90.5%. Feature selection and autoencoder-based methods show limited performance, suggesting that treating the 313 indicators as independent features or learning representations without explicit structure is insufficient for modeling facial paralysis-related biomarker interactions. Graph neural networks improve the results by incorporating indicator relationships, with GATv2 and HGT achieving Pearson correlations above 0.94.

Table IV: Comparison of experimental results of various methods.

Method	Pearson $\uparrow$	Spearman $\uparrow$	MAE $\downarrow$	RMSE $\downarrow$	Acc@10 $\uparrow$
Concrete-AE [32]	0.869 $\pm$.021	0.865 $\pm$.019	11.07 $\pm$ 1.24	14.79 $\pm$ 2.18	58.7% $\pm$ 5.8%
SCARF [33]	0.796 $\pm$.031	0.818 $\pm$.028	13.62 $\pm$ 1.57	18.17 $\pm$ 2.93	55.6% $\pm$ 6.4%
SubTab [34]	0.772 $\pm$.034	0.839 $\pm$.025	13.68 $\pm$ 1.63	19.18 $\pm$ 3.05	54.0% $\pm$ 7.1%
IP-CAE [35]	0.860 $\pm$.024	0.852 $\pm$.022	11.44 $\pm$ 1.31	15.22 $\pm$ 2.34	57.1% $\pm$ 6.1%
SDAE [36]	0.799 $\pm$.029	0.835 $\pm$.026	13.15 $\pm$ 1.48	17.75 $\pm$ 2.76	52.4% $\pm$ 6.7%
GATv2 [37]	0.949 $\pm$.009	0.941 $\pm$.008	5.83 $\pm$ 0.61	9.14 $\pm$ 1.12	85.7% $\pm$ 3.9%
HGT [38]	0.953 $\pm$.011	0.937 $\pm$.010	6.65 $\pm$ 0.74	9.00 $\pm$ 1.08	77.8% $\pm$ 4.3%
MLST-Net [15]	0.961 $\pm$.004	0.932 $\pm$.006	6.67 $\pm$ 0.24	8.38 $\pm$ 0.35	82.2% $\pm$ 1.3%
UPFP-SG [19]	0.964 $\pm$.002	0.927 $\pm$.002	6.64 $\pm$ 0.09	8.15 $\pm$ 0.23	82.5% $\pm$ 1.3%
PHCNN [20]	0.944 $\pm$.005	0.928 $\pm$.006	7.95 $\pm$ 0.33	10.37 $\pm$ 0.59	75.1% $\pm$ 1.3%
HiDAPA	0.980$\pm$.002	0.958$\pm$.003	4.90$\pm$0.23	6.18$\pm$0.35	90.5%$\pm$1.6%

Note: MLST-Net, UPFP-SG, and PHCNN are adapted to the common 313-indicator representation while preserving their core modeling strategies.

The facial-paralysis-specific baselines also achieve competitive results, with UPFP-SG reaching a Pearson correlation of 0.964 and an Acc@10 of 82.5%. Compared with the strongest task-specific results, HiDAPA improves Pearson and Spearman correlations by 0.016 and 0.026, reduces MAE and RMSE by 1.74 and 1.97, and increases Acc@10 by 8.0 percentage points. Paired two-sided Wilcoxon signed-rank tests over the 10 matched runs, followed by Holm correction, confirm that HiDAPA significantly outperforms all methods in terms of Acc@10 ($p_{\text{Holm}} < 0.05$). The fixed or predefined structures of the comparison methods limit their ability to capture patient-specific heterogeneity. In contrast, HiDAPA combines patient-specific graph construction with hierarchical indicator modeling, leading to consistent improvements in both correlation and error-based metrics. These results support the clinical consistency and validity of the learned indicator importance distribution.

C. Ablation Study

Table V summarizes the ablation results. Removing the patient-adaptive module causes the largest degradation, indicating the importance of patient-conditioned feature transformation and attention learning. Removing the dynamic graph module or hierarchical attention also leads to clear performance drops, showing that patient-specific graph construction and multi-level dependency modeling both contribute to indicator importance learning. The removal of prior regularization and label enhancement causes relatively smaller but consistent degradation, suggesting their auxiliary roles in improving robustness.

To examine whether the model relies excessively on HB information, we separately remove the HB-guided reconstruction

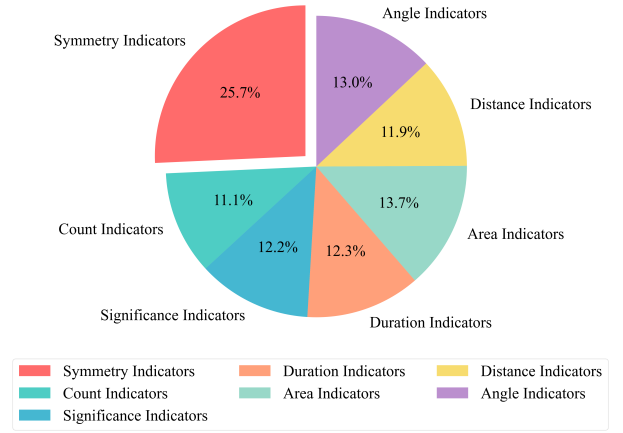


Figure 5: Type-level importance distribution.

weighting and active-head allocation, and remove HB from the conditioning vector. Both variants show consistent but limited degradation compared with the full model. In particular, the model without HB conditioning retains a Pearson correlation of 0.979, an MAE of 5.17, and an Acc@10 of 88.49%. These results indicate that HB serves as an auxiliary clinical prior, while the learned facial-indicator representations remain the primary basis for Sunnybrook estimation. We also include the Heterogeneous Graph Attention Network (HAN), which hierarchically aggregates meta-path-based neighbors through node-level and semantic-level attention, as a graph-based reference model for biomarker relation modeling [39]. Compared with HAN, the complete HiDAPA substantially improves all metrics. These results indicate that the proposed components provide complementary contributions to personalized and hierarchical biomarker modeling.

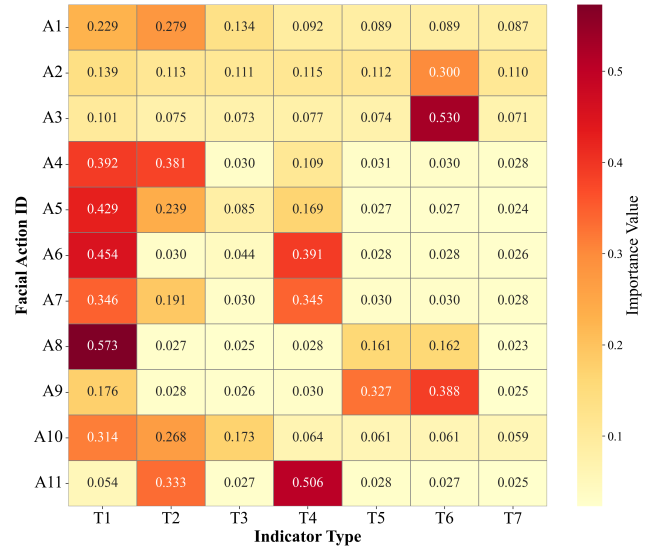


Figure 6: Action-type interaction importance.

D. Repeatability and Longitudinal Responsiveness

Large-scale repeated recordings under an unchanged clinical condition are not available in the current retrospective cohort

Table V: Comparison of results from model ablation experiments.

Method	Description	Pearson r	Spearman	MAE	RMSE	Acc@10
w/o Patient Adaptive	Remove Patient Adaptive	0.936 $\pm$ 0.014	0.922 $\pm$ 0.011	7.74 $\pm$ 0.87	10.43 $\pm$ 1.76	75.56% $\pm$ 4.17%
w/o Dynamic Graph	Remove Dynamic Graph	0.973 $\pm$ 0.006	0.944 $\pm$ 0.006	5.73 $\pm$ 0.49	7.29 $\pm$ 0.97	84.76% $\pm$ 3.66%
w/o Hierarchical Attention	Remove Hierarchical Attention	0.970 $\pm$ 0.011	0.942 $\pm$ 0.011	5.92 $\pm$ 0.79	7.58 $\pm$ 1.57	84.78% $\pm$ 5.21%
w/o Prior	Remove Prior	0.976 $\pm$ 0.003	0.945 $\pm$ 0.009	5.57 $\pm$ 0.27	6.99 $\pm$ 0.53	84.4% $\pm$ 2.34%
w/o Label Enhancement	Remove Label Enhancement	0.978 $\pm$ 0.003	0.942 $\pm$ 0.011	5.27 $\pm$ 0.35	6.51 $\pm$ 0.54	86.4% $\pm$ 3.82%
w/o HB-Guided Weighting	Remove w_{HB} and HB-guided active heads	0.978 $\pm$ 0.002	0.942 $\pm$ 0.008	5.41 $\pm$ 0.24	6.58 $\pm$ 0.36	87.30% $\pm$ 2.90%
w/o HB Conditioning	Use $\mathbf{c} = [P, L, R]$	0.979 $\pm$ 0.001	0.937 $\pm$ 0.005	5.17 $\pm$ 0.16	6.28 $\pm$ 0.19	88.49% $\pm$ 1.52%
HAN [39]	Heterogeneous Graph Attention Network	0.919 $\pm$ 0.008	0.925 $\pm$ 0.007	9.46 $\pm$ 0.42	12.09 $\pm$ 0.53	65.1% $\pm$ 2.26%
HiDAPA	Full Model	0.980$\pm$0.002	0.958$\pm$0.003	4.90$\pm$0.23	6.18$\pm$0.35	90.5%$\pm$1.60%

Spontaneous Eye Blink-Ratio of Left/Right Blink Minima Duration (Min) Score: 0.137	Spontaneous Eye Blink-Ratio of Left/Right Blink Minima Duration (Avg) Score: 0.123	Spontaneous Eye Blink-Ratio of Left/Right Blink Closing Duration (Max) Score: 0.122	Pout-Bilateral Symmetry of the Overall Facial Shape Score: 0.120
Spontaneous Eye Blink-Ratio of Left/Right Blink Duration (Avg) Score: 0.141	Show Teeth-Shape Symmetry of Eyes Score: 0.123	Spontaneous Eye Blink-Ratio of Left/Right Blink Closing Duration (Avg) Score: 0.122	Show Teeth-Ratio of Lower Lip Left and Right Side Areas Score: 0.121
Spontaneous Eye Blink-Ratio of Left/Right Blink Opening Duration (Avg) Score: 0.144	Pout-Ratio of Lower Lip Left and Right Side Areas Score: 0.124	Show Teeth-Rotation Symmetry of Lower Lip with Left and Right side Score: 0.122	Closing Eyes Gently-Eye Closeness Score: 0.122
Spontaneous Eye Blink-Ratio of Left/Right Blink Closing Duration (Min) Score: 0.147	Spontaneous Eye Blink Action Symmetry Score: 0.137	Spontaneous Eye Blink-Ratio of Left/Right Blink Opening Duration (Min) Score: 0.137	Spontaneous Eye Blink-Ratio of Left/Right Blink Duration (Min) Score: 0.136

Figure 7: Top-20 indicators ranked by learned importance.

because most patients proceed to surgery or rehabilitation after assessment. We therefore present two complementary case-level analyses in Table VI.

For the clinically stable patient, the reference score remains unchanged at 92 across two independent assessments conducted seven days apart, while the predicted scores are 91.35 and 91.28. The importance distributions across the seven indicator types remain highly consistent, with a Pearson correlation of 0.99 and a cosine similarity of 0.99. Moreover, 9 of the top-10 fine-grained indicators are retained between the two assessments. These results provide preliminary evidence of intra-system repeatability under a stable clinical condition. For the second patient, the reference score increases from 33 to 59 after surgery, while the predicted score increases from 35 to 63. The seven-type importance vectors remain highly consistent, whereas the Top-10 fine-grained indicator change rate increases to 30%. This suggests that the system preserves a stable high-level explanatory structure while reflecting longitudinal changes through its fine-grained indicators. The two cases suggest that the system produces consistent outputs when the clinical condition remains stable and reflects longitudinal changes when the patient's condition improves.

In addition, to examine how landmark-detection errors propagate through the indicator calculation and severity prediction pipeline, independent zero-mean Gaussian noise is added to the

horizontal and vertical coordinates of each detected landmark. The noise standard deviation is set to 5%, 10%, or 20% of the frame-wise interpupillary distance. At each perturbation level, the similarity normalization and calculation of all 313 indicators are repeated, followed by model training and evaluation using 10 independent random seeds. As shown in Table VII, the difference between the original and perturbed indicator vectors increases progressively with the perturbation level. However, the indicator representations remain highly correlated under 5% and 10% perturbations, and the final prediction performance remains relatively stable. Under the stronger 20% perturbation, the scoring error increases and Acc@10 decreases, indicating that substantial landmark displacement propagates to the downstream prediction while not completely disrupting severity ranking.

E. Interpretability Analysis of Indicator Importance

1) *Type-level importance*: Fig. 5 shows the global importance distribution across the seven indicator types. Symmetry indicators contribute the largest proportion (25.7%), followed by area (13.7%) and angle indicators (13.0%). This distribution is consistent with clinical assessment, in which bilateral symmetry and movement amplitude are central indicators of facial dysfunction.

2) *Action-type interaction importance*: Fig. 6 presents the learned interactions between 11 facial actions and seven indicator types. Different actions emphasize distinct functional dimensions: eye-related actions mainly highlight symmetry and significance, whereas the resting state places greater importance on spatial features. These structured patterns show that HiDAPA assigns importance according to the functional characteristics of each action rather than uniformly across indicators.

3) *Indicator-level importance and case study*: Fig. 7 presents the top-20 indicators ranked by global importance. Spontaneous-blink indicators occupy several leading positions, particularly those measuring bilateral differences in closing, opening, and total blink duration. Indicators from voluntary eye and mouth actions provide complementary information on spatial asymmetry and movement amplitude. Fig. 8 further shows the top-5 indicators for four representative patients. Although common high-importance indicators appear across patients, their rankings differ substantially, reflecting patient-specific ocular and oral impairment patterns. These personalized importance distributions provide interpretable evidence beyond the overall severity

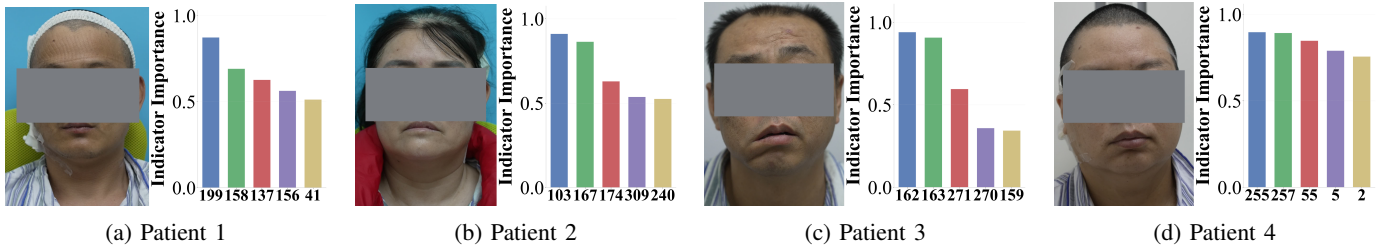


Figure 8: Patient-specific top-5 indicator importance distributions for four representative patients, illustrating individualized biomarker priorities for clinical interpretation. The corresponding numbers for indicators are shown in the supplementary material.

Table VI: Case-level evidence of intra-system repeatability and longitudinal responsiveness.

Case	Assessment 1	Assessment 2	Score comparison	Indicator comparison
Stable condition 7-day interval			Reference score: 92 → 92 Predicted score: 91.35 → 91.28 Absolute prediction error: 0.65 → 0.72 Between-assessment difference: 0.07	Seven-type importance vector: Pearson $r = 0.99$ Cosine similarity = 0.99 Mean absolute difference = 0.0030 Action-level Top-1 consistency: 11/11 Action-level Top-3 overlap: 27/33 Top-10 retained indicators: 9/10 Top-10 indicator change rate: 10%
Clinical improvement Pre / post-surgery			Reference score: 33 → 59 Clinical change: +26 Predicted score: 35 → 63 Predicted change: +28 Absolute prediction error: 2 → 4 Difference in change: 2 points	Seven-type importance vector: Pearson $r = 0.9992$ Cosine similarity = 0.9995 Mean absolute difference = 0.0053 Action-level Top-1 consistency: 11/11 Action-level Top-3 overlap: 30/33 Top-10 retained indicators: 7/10 Top-10 indicator change rate: 30%

Note: Pearson correlation, cosine similarity, and mean absolute difference are calculated between the importance vectors across the seven indicator types. Action-level Top-1 consistency and Top-3 overlap compare the indicator types selected for the 11 facial actions. The Top-10 change rate denotes the proportion of fine-grained indicators that are not retained between the two assessments. The stable case shows highly consistent predictions and indicator patterns, whereas the clinically improving case shows a corresponding score increase and greater changes in the fine-grained indicator ranking.

Table VII: Sensitivity to landmark perturbation at the indicator and score levels.

Noise	Indicator-level consistency				Score-level performance				
	Pearson	Cosine	MAE	RMSE	Pearson	Spearman	MAE	RMSE	Acc@10
5%	0.9984 ± 0.0007	0.9984 ± 0.0007	0.0399	0.0498	0.977 ± 0.002	0.936 ± 0.007	5.33 ± 0.16	6.60 ± 0.29	87.46% ± 2.42%
10%	0.9937 ± 0.0029	0.9938 ± 0.0028	0.0794	0.0999	0.978 ± 0.003	0.940 ± 0.010	5.33 ± 0.29	6.54 ± 0.54	87.10% ± 2.99%
20%	0.9756 ± 0.0114	0.9758 ± 0.0113	0.1595	0.1991	0.974 ± 0.003	0.940 ± 0.008	5.65 ± 0.26	7.16 ± 0.53	85.71% ± 3.26%

score and can support individualized assessment and longitudinal rehabilitation monitoring.

VI. CONCLUSION

We develop an end-to-end facial paralysis assessment system that integrates data acquisition, facial landmark extraction, HiDAPA-based analysis, and visual report generation. HiDAPA hierarchically models 313 facial biomarkers through multi-level representation learning, dynamic graph construction, and patient-conditioned attention. Experimental results demonstrate improved Sunnybrook-referenced clinical consistency over representative baselines. The system further converts model outputs

into interpretable biomarker reports, enabling clinicians to identify dominant functional deficits beyond a single global score.

Future work will include additional repeated assessments to further evaluate system repeatability. We will conduct multicenter studies with larger and more diverse patient cohorts, including clinically confirmed HB-I and mild facial paralysis cases, to assess generalizability across the full severity spectrum. We will also incorporate explicit head-pose estimation and investigate 3D facial acquisition to improve robustness to out-of-plane pose and depth-related deformation. Finally, the identified biomarkers will be further evaluated to support the development of standardized, biomarker-guided assessment and rehabilitation protocols.

REFERENCES

- [1] M. Zhang, Y. Sun and X. Sun, "Development of an Adaptive Serious Game System for Facial Paralysis Rehabilitation: A Facial Movement Recognition Pilot Study," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 34, pp. 237–250, Dec. 2025.
- [2] J. J. Nash, D. R. Friedland, K. J. Boorsma, and J. S. Rhee, "Management and Outcomes of Facial Paralysis from Intratemporal Blunt Trauma: A Systematic Review," *The Laryngoscope*, vol. 120, no. 7, pp. 1397–1404, 2010.
- [3] E. Chung, D. Matic, K. Fung, S. D. MacNeil, A. C. Nichols, R. Kiwan, K. Tay, and J. Yoo, "Bell's Palsy Misdiagnosis: Characteristics of Occult Tumors Causing Facial Paralysis," *Journal of Otolaryngology-Head Neck Surgery*, vol. 51, no. 1, p. 39, 2022.
- [4] J. Finsterer, "Management of peripheral facial nerve palsy," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 265, no. 7, pp. 743–752, Jul. 2008.
- [5] T. C. ten Harkel, S. Vinayahalingam, K. J. A. O. Ingels, S. J. Berge, T. J. J. Maal and C. M. Speksnijder, "Reliability and Agreement of 3D Anthropometric Measurements in Facial Palsy Patients Using a Low-Cost 4D Imaging System," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 8, pp. 1817–1824, Aug. 2020.
- [6] E. X. Wei, A. Green, J. P. Pepper, and S. R. Akkina, "Increasing incidence of facial nerve disorders in the United States from 2007 to 2022," *The Laryngoscope*, vol. 135, no. 6, pp. 2008–2013, 2025.
- [7] E. Mengi, *et al.*, "Comparison of the reliability of the House-Brackmann, facial nerve grading system 2.0, and Sunnybrook facial grading system for the evaluation of patients with peripheral facial paralysis," *The Journal of International Advanced Otolaryngology*, vol. 20, no. 1, pp. 14–18, Jan. 2024.
- [8] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," *Nat. Med.*, vol. 28, pp. 31–38, Jan. 2022.
- [9] F. Nerrise, N. Schütz, Q. Zhao *et al.*, "A framework of digital biomarkers for neurodegenerative diseases," *Nat. Rev. Bioeng.*, Apr. 2026.
- [10] Y. Xia, C. Nduka, R. Yap Kannan, E. Pescarini, J. Enrique Berner and H. Yu, "AFLFP: A Database With Annotated Facial Landmarks for Facial Palsy," *IEEE Transactions on Computational Social Systems*, vol. 10, no. 4, pp. 1975–1985, Aug. 2023.
- [11] N. Ikezawa, T. Okamoto, Y. Yoshida, S. Kurihara, N. Takahashi, T. A. Nakada, and H. Haneishi, "Toward an application of automatic evaluation system for central facial palsy using two simple evaluation indices in emergency medicine," *Sci. Rep.*, vol. 14, no. 1, p. 3429, Feb. 2024.
- [12] Y. Zhang, W. Gao, H. Yu, J. Dong, and Y. Xia, "Artificial intelligence-based facial palsy evaluation: a survey," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 3116–3134, 2024.
- [13] Y. Zhuang, M. M. McDonald, C. M. Aldridge, M. A. Hassan, O. Uribe, D. Arteaga, A. M. Southerland and G. K. Rohde, "Video-Based Facial Weakness Analysis," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 9, pp. 2698–2705, Sept. 2021.
- [14] A. Heinrich, G. F. Volk, C. Dobel, and O. Guntinas-Lichius, "Deep learning-based automatic facial symmetry scoring in peripheral facial palsy," *Sci. Rep.*, vol. 15, Aug. 2025.
- [15] Z. Feng, T. Zhou and T. Han, "MLST-Net: Multi-Task Learning Based Spatial-Temporal Disentanglement Scheme for Video Facial Paralysis Severity Grading," *IEEE Journal of Biomedical and Health Informatics*, vol. 29, no. 8, pp. 5675–5686, Aug. 2025.
- [16] A. A. Rao, J. J. Greene and T. P. Coleman, "Machine Learning Methods to Track Dynamic Facial Function in Facial Palsy," *IEEE Transactions on Biomedical Engineering*, vol. 72, no. 11, pp. 3359–3373, Nov. 2025.
- [17] J. Ye *et al.*, "CmdVIT: A Voluntary Facial Expression Recognition Model for Complex Mental Disorders," *IEEE Transactions on Image Processing*, vol. 34, pp. 3013–3024, 2025.
- [18] X. Ge, J. M. Jose, P. Wang, A. Iyer, X. Liu and H. Han, "ALGRNet: Multi-Relational Adaptive Facial Action Unit Modelling for Face Representation and Relevant Recognitions," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 5, no. 4, pp. 566–578, Oct. 2023.
- [19] W. Gan *et al.*, "UPFP-SG: A New Benchmark for Unilateral Peripheral Facial Paralysis Severity Grading," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 33, pp. 3744–3754, 2025.
- [20] X. Liu, Y. Xia, H. Yu, J. Dong, M. Jian and T. D. Pham, "Region Based Parallel Hierarchy Convolutional Neural Network for Automatic Facial Nerve Paralysis Evaluation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 10, pp. 2325–2332, Oct. 2020.
- [21] Y. Zhang *et al.*, "The Feasibility of An Automatic Facial Evaluation System Providing Objective and Reliable Results for Facial Palsy," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1680–1686, 2023.
- [22] A. Gaber, M. F. Taher, M. A. Wahed, N. M. Shalaby, and S. Gaber, "Comprehensive assessment of facial paralysis based on facial animation units," *PLOS ONE*, vol. 17, no. 12, pp. 1–18, Dec. 2022.
- [23] R. Tedeschi, D. Donati, and F. Giorgi, "Beyond the smile: a systematic review of diagnostic tools for peripheral facial paralysis," *Acta Neurologica Belgica*, vol. 124, no. 6, pp. 1805–1822, Aug. 2024.
- [24] X. Guan, H. Wang, D. Zhang *et al.*, "Continuous and componentized facial palsy measurement alignment and clinical interpretable model," *npj Digital Medicine*, vol. 8, p. 694, Nov. 2025.
- [25] D. Ouyang, B. He, A. Ghorbani *et al.*, "Video-based AI for beat-to-beat assessment of cardiac function," *Nature*, vol. 580, pp. 252–256, Mar. 2020.
- [26] A. Haque, A. Milstein and L. Fei-Fei, "Illuminating the dark spaces of healthcare with ambient intelligence," *Nature*, vol. 585, pp. 193–202, Sept. 2020.
- [27] J. Sumitra and N. Kulkarni, "A comprehensive survey and evaluation of mediapipe face mesh for human emotion recognition," *2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. IEEE, 2024.
- [28] Z. Wang, J. Bi, X. Zhao, Z. Zheng, J. Cui, R. Cui, J. Zhang, Y. Tang, J. Liang, K. Zhou and J. Zhang, "Assistive Assessment of Neurological Dysfunction via Eye Movement Patterns and Ocular Biometrics," *Artificial Intelligence Surgery*, vol. 5, pp. 521–544, Dec. 2025.
- [29] S. D. Reitzen, J. S. Babb, and A. K. Lalwani, "Significance and reliability of the House-Brackmann grading system for regional facial nerve function," *Otolaryngol. Head Neck Surg.*, vol. 140, no. 2, pp. 154–158, Feb. 2009.
- [30] A. Williams, N. Eapen, A. Kochar, B. Lawton, J. Hort, A. West, S. George, R. Berkowitz, K. J. Lee, S. R. Dalziel, S. Hearps, and F. E. Babl, "Agreement between House-Brackmann and Sunnybrook facial nerve grading systems in Bell's palsy in children: Secondary analysis of a randomized, placebo-controlled multicenter trial," *J. Child Neurol.*, vol. 38, no. 1–2, pp. 44–51, Feb. 2023.
- [31] J. Lou, H. Yu and F. -Y. Wang, "A Review on Automated Facial Nerve Function Assessment From Visual Face Capture," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, pp. 488–497, Feb. 2020.
- [32] M. F. Baln, A. Abid, and J. Zou, "Concrete Autoencoders: Differentiable Feature Selection and Reconstruction," *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, pp. 444–453, 2019.
- [33] D. Bahri, H. Jiang, Y. Tay and D. Metzler, "SCARF: Self-Supervised Contrastive Learning Using Random Feature Corruption," *Proceedings of the 10th International Conference on Learning Representations*, Apr. 2022.
- [34] T. Uçar, E. Hajiramezanali, and L. Edwards, "SubTab: Subsetting Features of Tabular Data for Self-Supervised Representation Learning," *Advances in Neural Information Processing Systems*, vol. 34, pp. 18853–18865, 2021.
- [35] A. Nilsson, K. Wijk, S. B. C. Gutha, E. Engleson, A. Hotti, C. Saccardi, O. Kviman, J. Lagergren, R. Vinuesa and H. Azizpour, "Indirectly Parameterized Concrete Autoencoders," *Proceedings of the 41st International Conference on Machine Learning*, May 2024.
- [36] W. Hassanieh and A. Chehade, "Selective Deep Autoencoder for Unsupervised Feature Selection," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 12322–12330, Mar. 2024.
- [37] S. Brody, U. Alon and E. Yahav, "How Attentive are Graph Attention Networks?" *Proceedings of the 10th International Conference on Learning Representations*, Jan. 2022.
- [38] Z. Hu, Y. Dong, K. Wang, and Y. Sun, "Heterogeneous Graph Transformer," *Proceedings of The Web Conference 2020*, pp. 2704–2710, 2020.
- [39] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous Graph Attention Network," *The World Wide Web Conference*, 2019, pp. 2022–2032.