Original Paper

Large Vision-Language Models for Pediatric Motor Tic Detection: Case-Control Diagnostic Study

Yujun Wu[1*], PhD; Yue Niu[1*], PhD; Zhao Xu[1], PhD; Jing Liu[1], PhD; Xianru Jiao[1], PhD; Xia Cui[2], PhD; Dan Sun[3], PhD; Zhixian Yang[1], MD, PHD

[1]Department of Pediatrics, Peking University People's Hospital, Beijing, China

[2]Department of Pediatrics, the Third Affiliated Hospital of Beijing University of Chinese Medicine, Beijing University of Chinese Medicine, Beijing, Beijing, China

[3]Department of Neurology, Tongji Medical College, Wuhan Children's Hospital, Huazhong University of Science and Technology, Hubei, China

[*]these authors contributed equally

Corresponding Author:
Zhixian Yang, MD, PHD
Department of Pediatrics
Peking University People's Hospital
11 Xizhimen South Street
Xicheng District
Beijing, 100034
China
Phone: 86 86 10 83573082
Fax: 86 86 10 66134261
Email: zhixian.yang@163.com

Abstract

Background: Clinical diagnosis of tic disorders (TDs) is often delayed by fluctuating symptoms and the subjective nature of assessments. While automated methods have been proposed to bridge this gap, traditional machine learning models frequently lack robustness in unconstrained environments and fail to capture the complex, multisite tics characteristic of severe TDs.

Objective: This study aimed to develop and evaluate a fine-tuned large vision-language model framework for the automated, interpretable detection of pediatric motor tics, addressing the clinical need for objective and efficient symptom assessment in unconstrained environments.

Methods: In this multicenter study, 168 pediatric participants (n=83, 49.4% with TDs and n=85, 50.6% healthy controls) underwent 12-hour continuous video-electroencephalography monitoring. We fine-tuned the Qwen3-VL-8B model using 3- to 5-minute video segments via low-rank adaptation. The framework used chain-of-thought prompting to generate both binary classifications and textual justifications. A strict patient-wise split ensured zero identity overlap between training and test sets. Diagnostic precision was evaluated using 95% CIs calculated with the Clopper-Pearson exact method.

Results: In total, 168 pediatric participants (age: mean 8.3, SD 2.3 years) were enrolled, providing 559 expert-validated video segments for analysis. On the independent test set (n=55), the framework achieved 89.1% (49/55; 95% CI 77.8%-95.9%) accuracy and an area under the curve of 0.950 (95% CI 0.892-1.000), significantly outperforming the landmark-based deep neural network (accuracy 74.1%; *P*=.004); recall reached 100% (25/25 positive episodes; 95% CI 86.3%-100.0%). Model-generated justifications showed substantial agreement with expert annotations (κ=0.784, 95% CI 0.625-0.943; *P*<.001).

Conclusions: These findings suggest that a vision-language model framework can capture clinically relevant tic features beyond conventional geometric descriptors. By combining high sensitivity (recall) with text-based diagnostic justifications aligned with clinical severity dimensions (eg, frequency and intensity), the approach may support automated video review and preliminary symptom quantification in controlled monitoring settings.

(JMIR AI 2026;5:e98782) doi: 10.2196/98782

KEYWORDS

tic disorders; automated tic detection; large vision-language model; LVLM; chain-of-thought reasoning; CoT; parameter-efficient fine-tuning; video-based monitoring; clinical interpretability; artificial intelligence; AI

XSL•FO
RenderX

Introduction

Tic disorders (TDs) are among the most prevalent neurodevelopmental conditions in children and adolescents, characterized by sudden, involuntary, and repetitive motor movements or vocalizations [1]. Typically emerging between ages 5 and 10 years, TDs affect boys 3 times more frequently than girls, with recent research confirming a persistent gender disparity [2]. These symptoms often follow a predictable cephalocaudal progression, initiating in the periorbital and perioral regions before extending to the neck, shoulders, and extremities [3]. With an estimated prevalence of 12% to 16% in school-aged populations, TDs are frequently complicated by comorbidities, such as attention-deficit/hyperactivity disorder and obsessive-compulsive disorder [4,5]. These overlapping conditions significantly impair academic performance and psychosocial development, making early, objective diagnosis a clinical necessity [6].

Despite the high prevalence of TDs, clinical diagnosis remains problematic due to the paroxysmal nature of symptoms and the "white-coat effect," where tics are often voluntarily suppressed during brief clinic visits [7]. This unpredictability is further compounded by the limitations of the Yale Global Tic Severity Scale (YGTSS)—the current gold standard—which relies heavily on retrospective parental recall and specialized expertise, making it impractical for high-volume clinical settings [8]. While tools such as the modified Rush Video Tic Rating Scale provide objective data, they necessitate tedious manual review that precludes large-scale screening [9]. Consequently, these challenges necessitate a transition toward "digital phenotyping"—the use of passive, continuous monitoring to capture real-world symptom dynamics [10]. Current automated detection primarily follows three technological paths: (1) sensor-based systems, which offer high objectivity but face challenges with pediatric compliance during long-term monitoring; (2) geometric landmark–tracking frameworks, which provide high precision but often necessitate specific recording angles (eg, selfie style) and are sensitive to occlusions [11,12]; and (3) global spatiotemporal models, which leverage raw pixels for end-to-end detection [13-15]. Such paradigm shifts have already shown transformative success in other movement disorders, such as Parkinson disease and epilepsy, where they have emerged as mature diagnostic adjuncts [16,17].

While automated detection has been explored to bridge this diagnostic gap in TDs [18], existing methodologies—primarily based on geometric landmarks or shallow convolutional neural networks—often exhibit limited robustness when transitioned to unconstrained clinical environments. This fragility primarily stems from the inherent limitations of geometric landmark tracking and frame-level optical extraction, which are highly sensitive to variations in ambient lighting, occlusions, and sudden movement [19-21]. Furthermore, because these models rely heavily on predefined facial anchor points, their spatial focus is frequently restricted to localized facial tics [22,23]. Such technical constraints preclude the identification of subtle "premonitory" muscle tension or complex, multisite tics involving the shoulders and trunk—hallmark features of severe TD [22,23]. To transcend these limitations and align TD monitoring with the advancements in generalist medical AI, this study aims to develop and evaluate a robust framework using a fine-tuned large vision-language model (LVLM), Qwen3-VL [24,25]. We hypothesized that by leveraging the superior spatiotemporal reasoning and clinical semantic understanding of foundation models, our framework could significantly outperform traditional architectures in unconstrained environments while providing textual diagnostic justifications aligned with clinical severity dimensions.

Methods

Multicenter Recruitment and Systematic Procedures

This multicenter study implemented a standardized, systematic recruitment and screening protocol between June 2023 and October 2025 at 3 tertiary clinical sites: Peking University People's Hospital, Wuhan Children's Hospital, and the Third Affiliated Hospital of Beijing University of Chinese Medicine. Candidates were consecutively recruited from high-volume pediatric neurology outpatient clinics. The systematic diagnostic procedure involved preliminary screening by attending physicians, followed by formal confirmation by 2 senior pediatric neurologists (each with more than 10 years of clinical experience), strictly based on the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition) criteria for TDs [19]. Clinical inclusion was restricted to patients presenting with active motor tics, defined as sudden, rapid, recurrent, and nonrhythmic motor movements involving localized or multiple muscle groups (eg, eye blinking, facial grimacing, or shoulder shrugging). To ensure a representative clinical spectrum, we enrolled 83 participants with TDs who exhibited a wide range of symptom severity, as quantified by the YGTSS (mean total tic score 22.4, SD 7.8). Concurrently, 85 age- and sex-matched healthy children were recruited as a negative control group during routine physical examinations. These controls underwent the same neurological screenings to exclude any prior history or familial patterns of neurodevelopmental or psychiatric disorders. The final sample size reflected the total number of eligible pediatric participants who completed standardized 12-hour video-electroencephalography (vEEG) monitoring during the multicenter recruitment period. No a priori sample size calculation was performed, as the study was designed as a convenience sample leveraging all available cases and controls that met the inclusion criteria.

To ensure data integrity, a final quality control phase was implemented prior to analysis. This phase involved the systematic exclusion of participants with poor video quality, excessive visual occlusion (eg, by clothing or bedside equipment), or atypical motor stereotypies that did not meet the rigorous criteria for motor tics. The final cohort comprised 168 participants (n=83, 49.4% with TDs and n=85, 50.6% healthy controls). Several recordings with severe occlusion, poor visibility, or atypical motor stereotypies were excluded during the quality control phase. The exact number of excluded participants was not systematically recorded because exclusions occurred prior to construction of the analytic cohort; therefore, all analyses reflect the final cohort of 168 participants.

XSL•FO
RenderX

Ethical Considerations

The study protocol and multicenter data acquisition procedures were reviewed and approved by the Ethics Committee of Peking University People's Hospital, serving as the lead institution (approval 2023PHB245-001). The other participating centers formally acknowledged and adhered to the ethical standards and approval granted by the lead site. The study was conducted in strict accordance with the Declaration of Helsinki. Written informed consent was obtained from the legal guardians of all participants for participation in the study and for publication of any accompanying clinical images.

Systematic vEEG Monitoring and Ecologic Data Acquisition

To capture a more realistic representation of symptoms than brief clinical examinations, behavioral data were acquired during 12-hour continuous vEEG monitoring [26]. To maximize ecologic validity, recordings were conducted in a standardized naturalistic ward environment where participants were encouraged to engage in unstructured daily activities, including playing, reading, and eating, during both awake and sleep states.

The monitoring setup was strictly standardized across all clinical sites: a high-definition digital camera with infrared capabilities was fixed at a height of 2.0 m, positioned 2.5 m to 3.0 m from the bedside to ensure a full-body field of view even during nocturnal sleep. To maintain data consistency for multimodal analysis, the camera maintained a constant viewing angle. A rigorous data cleaning protocol was implemented during the acquisition phase: video segments were manually audited, and instances where the participant's face was fully occluded or turned away from the lens for more than 10 seconds were systematically discarded. This process ensured that the final dataset prioritized clear, expert-verifiable facial and motor manifestations.

Expert Annotation and Quality Control

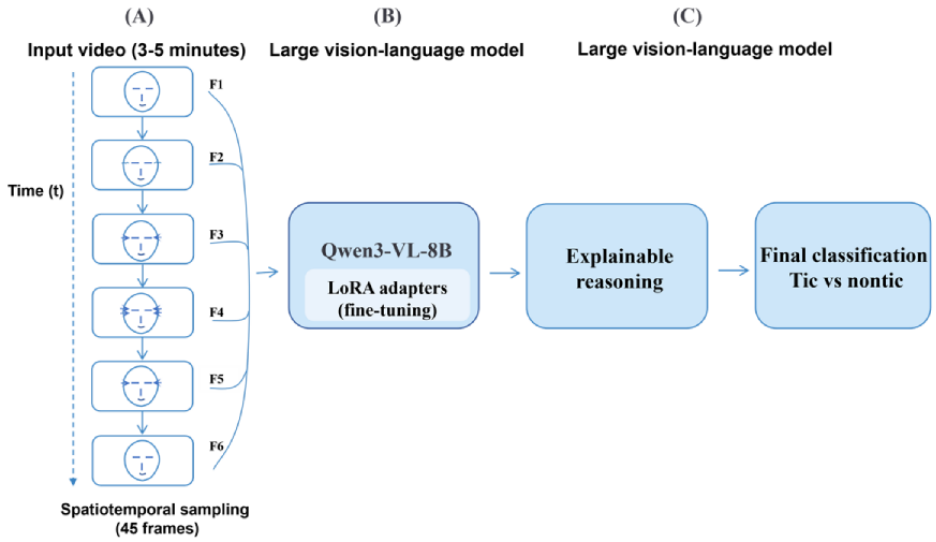
A total of 2 pediatric neurologists, blinded to clinical diagnoses, independently reviewed the 12-hour feeds to extract representative tic-positive and tic-negative segments. Each segment lasted 3 to 5 minutes. To handle the ambiguity of subtle movements, a third senior expert resolved any discrepancies. Interrater reliability was assessed, and a final consensus was reached. While specific phenotypes were tagged for internal validation, we collapsed these into a binary classification (tic vs nontic) to prioritize model robustness for general screening. After the predefined quality control procedures were applied, no missing data remained for any variable included in the analyses, and all retained video segments contained complete expert annotations.

Data Preprocessing and Spatiotemporal Sampling

To optimize computational efficiency, raw video recordings were standardized to a resolution of 1024×768 pixels. As depicted in Figure 1, we used a spatiotemporal sampling strategy to extract a sequence of 45 frames per segment. This sampling density was empirically determined to capture the complete kinetic progression of a motor tic across the temporal dimension (t).

The pool of expert-validated segments was partitioned into a training set and a test set using a stratified random sampling approach [27]. To ensure rigorous evaluation, we implemented a strict patient-wise split strategy. Under this protocol, all video clips derived from any single participant were assigned exclusively to either the training or the test cohort. By maintaining zero identity overlap, the framework is forced to focus on strictly pathological motion dynamics, ensuring the features are robust and generalizable.

Figure 1. Overall framework of the proposed Qwen3-VL system for automated pediatric motor tic detection. (A) Data acquisition and preprocessing. (B) Model architecture. (C) Diagnostic output and reasoning.



LVLM Architecture: Qwen3-VL-8B-Instruct

We used Qwen3-VL-8B-Instruct as the multimodal backbone to leverage its integrated vision transformer (ViT) and large language model (LLM) capabilities [28-30]. The model maps visual tokens from the video sequence to linguistic diagnostic instructions. We used a direct prompt requesting both a binary label (tic vs nontic) and a brief English rationale: "Please

XSL•FO
RenderX

analyze this video and determine whether a tic episode is present. Return (1) a label (tic/nontic) and (2) a short justification in English." This setup evaluates whether the model's pretrained semantic knowledge could distinguish purposeful actions from involuntary tics in unconstrained clinical monitoring videos.

Fine-Tuning With Low-Rank Adaptation

To adapt the 8B-parameter model while avoiding memory overflow errors on a single graphics processing unit, we used low-rank adaptation (LoRA), a parameter-efficient fine-tuning technique. These lightweight modules were integrated into the vision encoder, LLM layers, and multilayer perceptron (MLP) projections (as illustrated in Figure 1), with a rank of 16 and α=32.

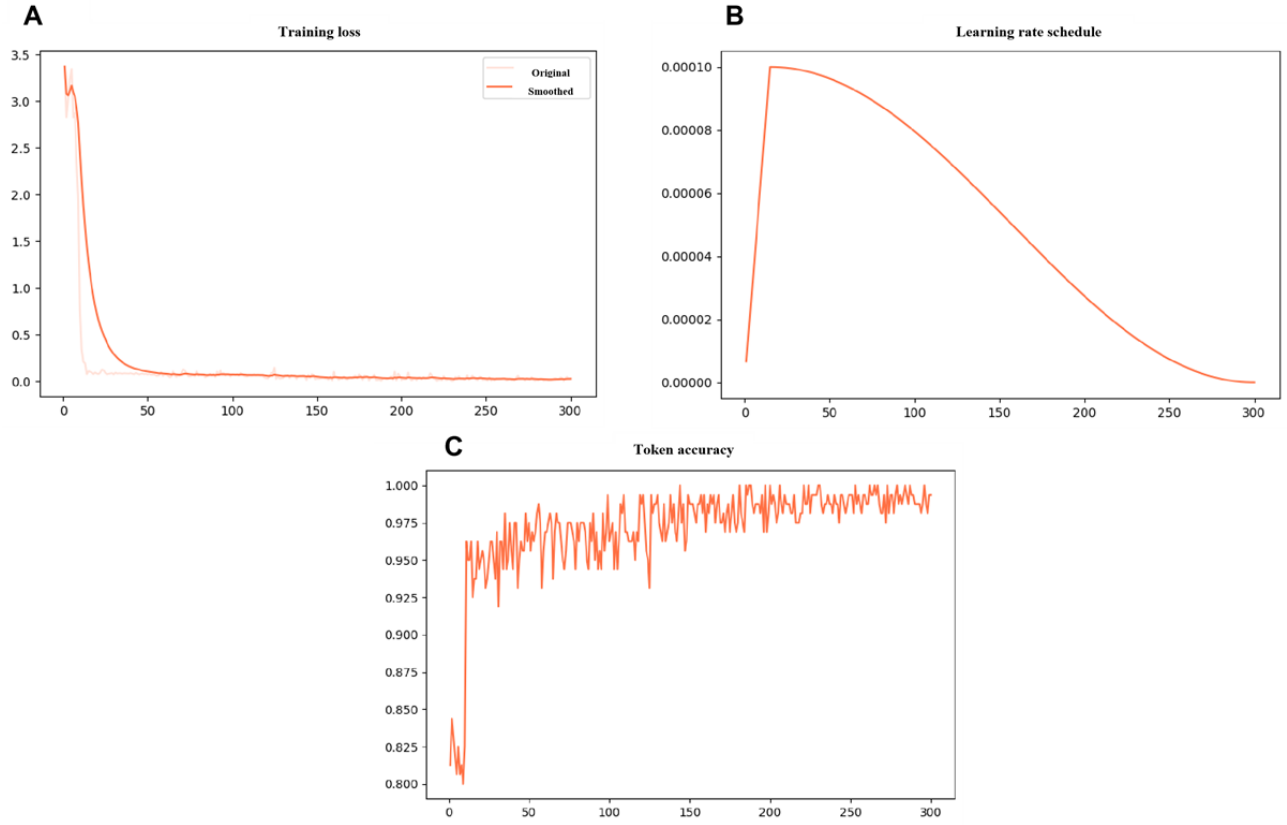
We used a cosine annealing scheduler with a learning rate of 1×10^{-4} and a 5% warmup over 12 epochs. The total training time on a cluster of 5 NVIDIA A100 GPUs (80 GB) was approximately 5.5 hours. We observed that increasing the batch size beyond 4 led to optimization instability; therefore, a small batch size combined with gradient accumulation was used.

Hyperparameter Optimization Strategy

The final hyperparameter configuration was determined through an iterative grid search over key training parameters. We systematically explored candidate learning rates ($1\times10^{-}$, $5\times10^{-}$, and 1×10^{-4}) and LoRA ranks ($r \in \{8, 16, 32\}$). All models were fine-tuned using the AdamW optimizer ($\beta1$=0.9, $\beta2$=0.999, and weight decay=0.01) via the DeepSpeed ZeRO-2 distributed strategy across 5 NVIDIA A100 (80 GB) GPUs, with a per-device batch size of 1 (effective total batch size=5). The cross-entropy loss function was used as the training objective. The optimal configuration—learning rate of 1×10^{-4} with LoRA rank r=16 and LoRA α=32—was selected based on the lowest validation loss and the most stable convergence behavior. Under this configuration, trainable modules included the LLM layers, ViT, and MLP projector. Training was conducted for 12 epochs using BFloat16 mixed precision, with a maximum sequence length of 40,000 tokens and a cosine annealing scheduler incorporating a 5% linear warmup ratio.

Figure 2 visualizes the training dynamics across approximately 300 total steps. Figure 2A (training loss) demonstrates rapid initial convergence—the loss drops sharply from approximately 3.4 to below 0.2 within the first 25 steps—followed by gradual refinement and stabilization near 0.05 by step 300, without obvious signs of training instability. Figure 2B (learning rate schedule) shows that the learning rate follows a cosine annealing schedule with a linear warmup, rising from near 0 to the peak value of 1×10^{4} within the first approximately 20 steps and then decaying smoothly toward 0 over the remaining steps. Figure 2C (token accuracy) shows a corresponding improvement in token-level prediction accuracy, rising from approximately 0.800 to approximately 0.950 within the first 50 steps and converging to a stable range of 0.950 to 0.995 thereafter. The absence of gradient explosion or accuracy degradation confirms the robustness of the selected hyperparameters.

Figure 2. Training dynamics and optimization stability of the Qwen3-VL framework. (A) Training loss: the loss curve shows rapid convergence within approximately 300 optimization steps. (B) Learning rate schedule: cosine annealing with linear warmup was used, consistent with the training configuration described in the Methods section. (C) Token accuracy: linguistic alignment accuracy gradually plateaued at approximately 98%.



XSL•FO
RenderX

Clinical Dimension Mapping via Chain-of-Thought Prompting

Qwen3-VL was specifically prompted to perform thin-slice analysis (3- to 5-minute segments) to extract three YGTSS-related metrics: (1) frequency (count of discrete events), (2) intensity (vigor of muscle contraction and duration), and (3) complexity (anatomical distribution). This allows the model to output a structured clinical justification alongside the binary label.

Comparative Baselines and Experimental Design

To rigorously evaluate the superiority of the proposed Qwen3-VL framework, we benchmarked it against representative architectures, spanning from traditional landmark-based deep learning to state-of-the-art self-supervised video foundation models.

Dual-Branch Facial Deep Neural Network (Traditional Cross-Validation Baseline)

Following the logic of established automated motor tic detection, we reproduced a dual-branch deep neural network (DNN) specifically tailored for facial movements [31].

Architecture

The model integrates a video branch (3D ResNet-18) for raw pixel sequences and a landmark branch (2D convolutional neural network) for temporal trajectory analysis.

Preprocessing

Frames were processed via RetinaFace to isolate the optimal facial bounding box across 30 uniformly sampled frames, resized to 64×64 pixels. Simultaneously, 5 facial landmarks (eyes, nose tip, and mouth corners) were extracted and normalized via *z* score to eliminate individual morphological variance.

Training

Features from both branches were fused via concatenation before a final MLP classifier. Optimization was performed using AdamW (learning rate=5×10^{-4} , weight decay=1×10^{-3}) with a cosine annealing scheduler (5-epoch linear warmup) and a batch size of 64.

VideoMAE-Large (Foundation Model Baselines)

To justify the transition from vision-only models to LVLMs, we implemented 2 fine-tuning strategies using VideoMAE-Large (pretrained on Kinetics-400 with a ViT-Large backbone) [32]. This model processes 16-frame clips to output global spatiotemporal representations.

Baseline A: Embed-Freeze (Linear Probing)

The entire VideoMAE backbone was frozen, with only the newly initialized MLP head being trained. This strategy evaluates the raw diagnostic power of pretrained video features, involving minimal trainable parameters (0.2% of the total).

Baseline B: Embed-Partial (Partial Fine-Tuning)

To balance efficiency and task-specific adaptation, we froze the first 20 layers and unfroze the last 4 transformer layers and LayerNorm. A layer-wise learning rate was applied ($1\times10^{-}$ for the backbone and 5×10^{-4} for the head), with trainable parameters accounting for 16.7% of the total.

Optimization

Both VideoMAE variants used AdamW with a 5% warmup and an early stopping criterion (patience=10) to prevent overfitting on the pediatric dataset.

Evaluation Metrics

Model performance was evaluated using standard classification metrics, including accuracy, sensitivity (recall), specificity, and F_1-score. The area under the curve (AUC) was calculated to assess diagnostic robustness. Initial model metrics and baseline comparisons were calculated via the *scikit-learn* library in Python (Python Software Foundation) [33].

Validation of Model Justifications

To quantify the accuracy of the model's interpretability, we established a structured verification protocol. Beyond binary labels, experts generated blinded anatomical and kinematic tags (eg, specific muscle groups, rhythmicity, and duration) for each segment prior to model evaluation. For claims involving specific muscle groups, model-generated terms were compared against these predefined expert tags using a term-matching methodology. In instances of discrepancy between the two primary annotators, the final consensus established by the third senior expert served as the absolute ground truth. The model's reasoning was deemed "correct" only if it aligned with this final consensus, ensuring that justifications were benchmarked against the most authoritative clinical judgment.

Statistical Analysis

Statistical analyses were performed using SPSS (version 29.0; IBM Corp). For clinical validation, the agreement between LVLM justifications and the final consensus expert tags was quantified using the Cohen κ coefficient. This assessment focused on the alignment of specific clinical descriptors (eg, anatomical sites and contraction intensity) to ensure the model captured actual pathological features rather than nonspecific motion. To address the precision of our estimates, 95% CIs for all diagnostic metrics (accuracy, sensitivity, and specificity) were calculated using the Clopper-Pearson exact method in SPSS. Comparison of AUC values between models was performed using the DeLong test. Continuous variables were presented as mean (SD). All *P* values were 2-tailed, with $P<.05$ considered statistically significant.

Results

Demographic Profiles and Dataset Characterization

The clinical recruitment process yielded a final cohort of 168 pediatric participants, including 83 (49.4%) patients with TDs and 85 (50.6%) healthy controls. The cohort exhibited a mean age of 8.3 (SD 2.3; range 4-15) years and a characteristic 3:1 male-to-female ratio (Table 1). Analysis of the 12-hour vEEG recordings produced a dataset of 559 expert-validated video segments, comprising 255 (45.6%) tic-positive and 304 (54.4%) tic-negative instances. The annotation phase exhibited high diagnostic reliability, with an initial interrater agreement of

XSL•FO
RenderX

89.4%. Following a patient-wise split strategy, the data were partitioned into 504 (90.2%) training and 55 (9.8%) independent test clips with zero identity overlap.

Table 1. Summary of participant demographics, video segment characteristics, and experimental dataset partitioning.

Characteristics	Values
Participants (N=168)	
Diagnosis, n (%)	
TDs[a]	83 (49.4)
Healthy controls	85 (50.6)
Age (y), mean (SD; range)	8.3 (2.3; 4.0-15.0)
Sex, n (%)	
Male	126 (75)
Female	42 (25)
Male-to-female ratio	3:1
YGTSS[b] total tic score (TD group only), mean (SD)	22.4 (7.8)
Video segment (n=559)	
Total segments, n	559
Segment duration (min)	3.0-5.0
Tic-positive segments, n (%)	255 (45.6)
Tic-negative segments, n (%)	304 (54.4)
Initial interrater agreement[c] (%)	89.4
Experimental partitioning (n=559)[d,e]	
Training set (clips), n (%)	504 (90.2)
Test set (clips), n (%)	55 (9.8)

[a]TD: tic disorder.

[b]YGTSS: Yale Global Tic Severity Scale.

[c]Interrater agreement was calculated based on initial independent reviews by 2 blinded pediatric neurologists.

[d]The splitting strategy was a patient-wise split that ensured that video segments from the same participant were assigned exclusively to either the training or the test set, with 0% overlap in participant identity between the two datasets.

[e]There was no identity overlap between datasets.

Optimization Dynamics and Training Convergence

Fine-tuning of the 8B-parameter Qwen3-VL model was stable. As illustrated in Figure 2A, the training loss decreased rapidly before reaching a plateau at approximately 300 iterations. This behavior is consistent with the cosine annealing schedule described in the Methods section (Figure 2B). Token accuracy (Figure 2C) stabilized at 97.8%, suggesting effective alignment between visual features and diagnostic tokens. These dynamics indicate that the 16.7 million trainable LoRA parameters were sufficient to capture task-specific tic features while preserving the model's pretrained multimodal representation.

Quantitative Benchmarking and Performance Metrics

Qwen3-VL demonstrated superior overall performance compared with the baseline architectures, as shown in Table 2 and Figure 3. The framework attained an overall accuracy of 89.1% (95% CI 77.8%-95.9%). On the independent test set, recall reached 100% (25/25 positive episodes; 95% CI 86.3%-100.0%), whereas the landmark-based DNN missed 9 cases (recall=64%). Although the 100% recall should be interpreted cautiously given the limited test sample, the receiver operating characteristic analysis yielded a superior AUC of 0.950 (95% CI 0.892-1.000) for Qwen3-VL, which was significantly higher than that for the traditional DNN baseline (AUC=0.840; *P*=.004, DeLong test).

XSL•FO
RenderX

Table 2. Comparison of performance metrics across different model architectures for automated tic detection[a].

Model	Accuracy, % (95% CI)	Precision, % (95% CI)	Recall, % (95% CI)	Specificity, % (95% CI)	F_1-score	AUC[b] (95% CI)
DNN[c]	74.1 (60.3-85.0)	76.2 (52.8-91.8)	64.0 (42.5-82.0)	82.8 (64.2-94.2)	0.696	0.840 (0.73-0.95)
Embed-freeze[d]	75.9 (62.4-86.5)	77.3 (54.6-92.2)	68.0 (46.5-85.1)	82.8 (64.2-94.2)	0.723	0.829 (0.71-0.94)
Embed-partial[e]	79.6 (66.5-89.4)	81.8 (59.7-94.8)	72.0 (50.6-87.9)	86.2 (68.3-96.1)	0.766	0.863 (0.76-0.96)
Qwen3-VL	*89.1 (77.8-95.9)* [f]	*80.7 (62.5-92.5)*	*100.0 (86.3-100)*	*80.0 (61.4-92.3)*	*0.893*	*0.950 (0.89-1.00)*

[a]All metrics were evaluated on an independent test set (n=55) using a patient-wise split strategy. The 100% recall reflects the performance on the current independent test set (n=25 positive episodes) and should be validated in larger cohorts. 95% CIs for proportion-based metrics (accuracy, precision, and specificity [recall]) were calculated using the Clopper-Pearson exact method; area under the curve CIs were derived using the DeLong method.

[b]AUC: area under the curve.

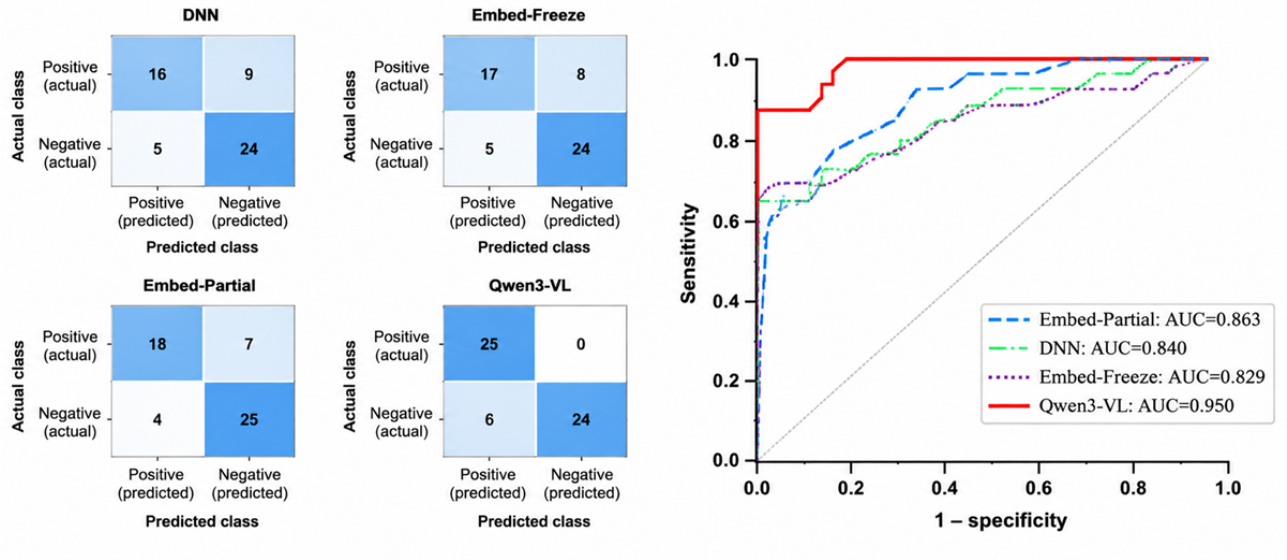
[c]DNN: deep neural network (landmark-based baseline).

[d]Embed-freeze: VideoMAE-large with frozen backbone.

[e]Embed-partial: VideoMAE-large with partially unfrozen layers.

[f]Values in italics represent the best performance across each metric.

Figure 3. Quantitative evaluation of diagnostic performance across 4 model variants. Left: Confusion matrices: comparison of classification results for deep neural network (DNN), Embed-Freeze, Embed-Partial, and the proposed Qwen3-VL. Note the 0 false-negative rate achieved by Qwen3-VL, indicating its superior sensitivity for clinical screening. Right: Receiver operating characteristic curves: comparative analysis of discriminative robustness. The Qwen3-VL model (red curve) achieves the highest area under the curve (AUC=0.950), significantly outperforming traditional landmark-based and vision-only architectures.



Observation of Qualitative Reasoning and Interpretability

The model generated clinically plausible textual justifications (Figure 4). For positive segments, the chain-of-thought (CoT) output described asymmetrical periocular contractions with durations exceeding 500 ms (Figure 4A). For negative instances, it described symmetrical eyelid closure and physiological blink timing (~250 ms), qualitatively differentiating involuntary tics from normal blinking (Figure 4B). As shown in Table 3, the model's CoT outputs successfully identified clinical markers, such as prolonged contraction (>500 ms) and multisite involvement, providing a qualitative basis for severity assessment.

XSL•FO
RenderX

Figure 4. Qualitative analysis of explainable diagnostic reasoning in motor tic identification. (A) Positive case showing recognition of tic-related motor features. (B) Negative case showing identification of physiological blinking features.

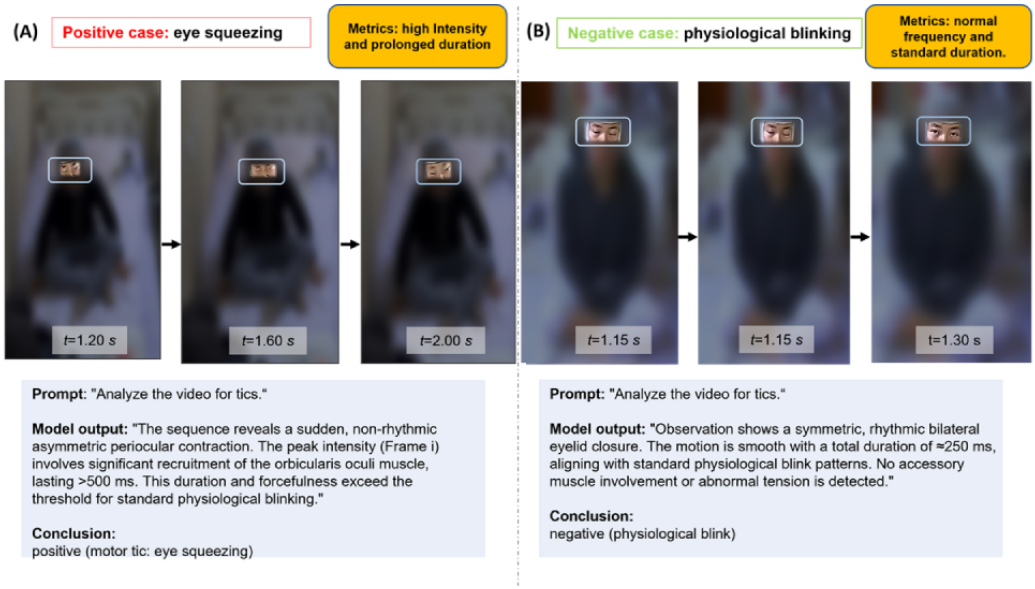
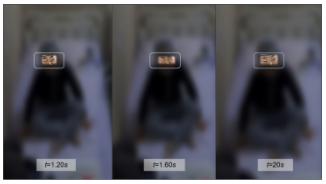
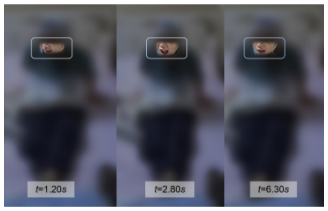
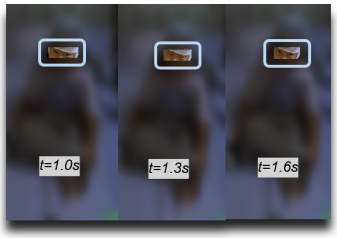


Table 3. Representative case analysis: mapping LVLM[a] reasoning to YGTSS[b] dimensions.

Case type	Visual sequence (frame 1...n)	Model CoT[c] reasoning	YGTSS dimension mapping
Simple motor tic (case A)		"Detected a sudden, paroxysmal eye-squeezing episode. The muscle contraction is forceful and asymmetric, lasting ~800ms, which significantly exceeds the normal physiological blink threshold (~250 ms)."	Intensity: high (forceful contraction); duration: prolonged (>500 ms)
Complex motor tic (case B)		"Analysis reveals a coordinated sequence: an involuntary paroxysmal mouth opening accompanied by intense mid-facial muscle twitching (grimacing). Multiple muscle groups (perioral and zygomaticus) are involved in a nonrhythmic pattern."	Complexity: multisite facial involvement (oral+mid-face); form: complex motor tic sequence
Physiological blink (case C)		"The movement exhibits symmetric eyelid closure with a rapid duration of ~250ms. No accessory facial muscle involvement or head movement is observed. Pattern matches normal physiological frequency."	Frequency: normal range; differential: successfully excluded from pathological tics
False positive (case D: allergy-driven blinking)		"Identified recurrent, high-frequency, forceful bilateral eye blinking. The motion features elevated kinematics and rhythmic intensity, misinterpreting ocular irritation as a motor tic sequence."	Intensity: high (kinematic overuse); mismatch with YGTSS: misclassified nontic physiological irritation (allergic conjunctivitis) as a severe motor tic, highlighting limitations in context and etiology reasoning

[a]LVLM: large vision-language model.

[b]YGTSS: Yale Global Tic Severity Scale.

[c]CoT: chain-of-thought.

XSL•FO
RenderX

Statistical Significance and Reliability

The DeLong test indicated that the improvement in AUC was statistically significant (P=.004) relative to the DNN [33]. The McNemar test revealed a significant reduction in false negatives (P=.02). The consistency between model reasoning and expert annotations attained a Cohen κ of 0.784 (95% CI 0.625-0.943; P<.001), indicating substantial agreement [34].

Discussion

Principal Findings

The main goal of this study was to develop a reliable, objective tool for detecting tics in children that overcomes the limitations of brief clinical examinations [7,8]. Our findings demonstrate that the fine-tuned Qwen3-VL framework achieved an overall accuracy of 89.1% and a superior AUC of 0.950. A major advantage of this work is its foundation in a multicenter study involving 3 large hospitals (Peking University People's Hospital, Wuhan Children's Hospital, and the Third Affiliated Hospital of Beijing University of Chinese Medicine), which helped avoid single-center bias. Using 12-hour continuous vEEG monitoring for our 83 patients (all recruited from specialized clinics and diagnosed by senior experts), we captured the full range of naturalistic tic expressions. This rigorous data collection ensured that the model learned from high-quality, real-world clinical data. Notably, the framework achieved 100% recall on the independent test set. While this high sensitivity likely reflects the focused nature of our current test set (25 positive episodes) and requires further validation, it underscores the model's potential as a highly sensitive screening adjunct. The consistent results achieved through a patient-wise split strategy further suggest that the model identifies actual pathological features rather than memorizing specific facial identities.

Detailed Discussion

Interpretations, Comparisons, and Implications: From Geometric Landmarks to Semantic Mapping

Traditional automated detection often relies on geometric descriptors (eg, measuring intereyelid distance) [18,22]. As noted during our evaluation, our LVLM framework does not explicitly measure these physical metrics but instead learns a mapping from high-dimensional spatiotemporal features to clinical semantics. While this "black box" nature is a characteristic of foundation models, the use of CoT prompting transforms internal features into human-readable justifications. By linking specific motion features (eg, asymmetry and rhythmicity) to clinical dimensions, the model provides a form of semantic interpretability that aligns with expert observation, bridging the gap between raw pixel analysis and clinical diagnosis.

The transition toward foundation models is already reshaping generalist medical AI. In broad generalist medical AI tasks, these models have demonstrated promising zero-shot generalization and multimodal reasoning, often achieving strong or competitive performance across diverse clinical scenarios. Recent studies have validated this potential across varied domains, ranging from concept-based medical image classification to specialized cervical cytology screenings [14,15]. Shifting from geometric landmarks to LVLMs represents a paradigm shift in how we analyze pathological movements. Older methods, such as those from Loewenstern et al [22], rely on tracking specific facial coordinates. While useful for reducing data size, these methods often miss fine details, such as skin texture changes or subtle muscle tension defining a mild tic. Our Qwen3-VL framework addresses these limitations by analyzing raw pixels through end-to-end alignment. By leveraging a pretrained semantic space, the model distinguishes between voluntary expressions and involuntary tics based on clinical features rather than moving dots, enhancing robustness against variations in face shape or lighting. Our comparisons were intended to test the value of adopting an LVLM-based framework under an identical patient-wise split and evaluation protocol rather than to rank LVLM backbones. We therefore benchmarked against a landmark-based facial DNN and 2 vision-only VideoMAE baselines; Qwen3-VL is reported here as a feasible instantiation of the LVLM framework in our setup, without claiming it is optimal among all LVLM alternatives.

The high performance is also attributed to the length of our analyzed video segments. Many automated systems use ultrashort clips, such as the 1-second windows noted by Brügge et al [18], which are often insufficient to distinguish tics from normal blinks. By analyzing 3- to 5-minute segments, Qwen3-VL accesses richer temporal context, capturing rhythm, recurrence, and associated muscle involvement more reliably than brief snapshots. Furthermore, while the performance of the models by Brügge et al [18] often dropped when tic intensity was low, our framework's focus on broader spatiotemporal dynamics maintained high diagnostic precision across varying intensities.

Clinical Interpretability and Safety

A primary challenge for AI in medicine is understanding the "why" behind a decision. Common tools such as saliency maps only show pixel focus without providing clinical explanations [35]. Our framework moves beyond "black box" detection by generating structured semantic reports through CoT prompting. For instance, the model can highlight asymmetrical eye contractions or prolonged durations (>500 ms; Figure 4). This "thin-slice" clinical observation approach provides objective frequency counts and intensity cues, potentially reducing the bias associated with retrospective parental recall in standard YGTSS assessments [36]. The substantial agreement (κ=0.784) between model reasoning and expert annotations is encouraging for its future clinical utility. Our post hoc audit of error cases revealed that the CoT justifications act as a "safety switch" for clinical reliability. For instance, in a negative segment where the model produced a false positive, it misinterpreted an allergy-driven blink as a motor tic, specifically citing "forceful, asymmetric contraction." While the model's diagnostic inference was incorrect, its reasoning was transparent and grounded in the observed kinematic intensity. This ensures that its errors are interpretable rather than arbitrary, allowing clinicians to quickly identify the misjudgment as a lack of behavioral context (eg, ocular irritation) rather than a "hallucinated" medical claim. If the model had provided a plausible rationale for a movement that did not exist, the system would be truly unreliable; instead,

XSL•FO
RenderX

our findings suggest that the LVLM remains anchored to physical pixels even when its clinical conclusion is flawed.

Unlike the smartphone-based selfie video methods of Loewenstern et al [22] that require patient cooperation, our data were collected in a naturalistic ward environment. While the relevance to home settings is a potential application rather than a demonstrated result, the framework's tolerance to routine monitoring conditions suggests that passive, less-controlled deployment settings may be feasible after external validation, allowing for longitudinal symptom tracking without highly scripted recordings.

Practical Utility and Future Directions

Beyond diagnostic accuracy, the practical utility of this LVLM framework lies in its potential as an objective screening adjunct for continuous monitoring. In high-volume pediatric clinics, experts could use the model to prescreen long-duration vEEG feeds, focusing their review on "high probability" segments identified by the AI. This "human-in-the-loop" approach addresses the white-coat effect by capturing tics in naturalistic settings. Looking forward, the integration of audio modalities for phonic tic detection and the deployment of quantized models for edge devices could further enhance the system's utility. By evolving toward a comprehensive multimodal monitoring tool, this framework may ultimately support clinicians in objective symptom quantification and facilitate more personalized treatment planning for children with TDs.

Limitations

Several limitations in this study deserve mention. From a statistical standpoint, although the observed 100% recall is encouraging, the relatively modest size of our independent test set (n=25) warrants caution; these preliminary findings require rigorous validation in larger, multicenter cohorts to ensure statistical power. Clinically, the current system should be viewed as a screening adjunct rather than a stand-alone diagnostic tool, as it lacks audio integration for phonic tic detection and cannot capture the "interference" dimension of the YGTSS—a metric that fundamentally requires observing social disruption over longer durations than our 3- to 5-minute clips allow. Technically, the computational overhead of the 8B-parameter backbone remains a barrier to edge-device deployment. Transitioning this framework to home-based monitoring will likely rely on 4-bit quantization or cloud-based inference to balance diagnostic precision with local processing constraints. Finally, expanding this model to adult populations and moving beyond binary classification to identify specific tic phenotypes, such as eye squeezing and shoulder shrugging, will be essential for providing the granular data clinicians need for tailored treatment planning. We did not benchmark multiple LVLM backbones under identical fine-tuning and prompting settings; cross-backbone comparisons are left for future work. Additionally, while our framework provides semantic explainability through CoT prompting, integrating visual explainability methods (such as spatiotemporal attention heatmap visualizations) remains an important avenue for future work to further bridge model transparency with clinical interpretation.

Conclusions

In conclusion, we developed and evaluated a framework for automated pediatric tic detection using a fine-tuned Qwen3-VL LVLM. By leveraging 3- to 5-minute video segments and semantic-level reasoning, the system achieved an accuracy of 89.1% and a recall of 100% on the independent test set. Beyond high diagnostic accuracy, the clinical value of this framework lies in its high sensitivity and interpretability, as textual justifications bridge the gap between AI outputs and professional clinical review. Furthermore, these findings underscore the transformative potential of generalist medical AI in movement disorder neurology [16,17], supporting automated video review and symptom quantification in controlled settings. While focused on visual cues, this framework provides a baseline for integrating additional modalities (eg, audio) to eventually enhance longitudinal symptom monitoring in ecologic settings. Nevertheless, further external validation is required before claims regarding home deployment or routine longitudinal use can be made.

Acknowledgments

The authors thank the patients and their families for participating. The authors confirm that GPT-4 (OpenAI) was used for language polishing and grammatical editing during the preparation of this manuscript. The tool was used solely to improve the linguistic quality of the text. All AI-generated suggestions were critically reviewed and validated by the authors, who take full responsibility for the accuracy and integrity of the final content.

Funding

This work was supported by the National Natural Science Foundation of China (grants 82171436 and 82571644), the General Program of Beijing Municipal Natural Science Foundation (7252146), the North Nature Focus Program (L254033), the Peking University People's Hospital Talent Introduction Start-up Fund (2023-T-02), the Peking University People's Hospital R&D Fund Unveiling Project (RDGS2023-10), the Peking University People's Hospital Intramural Development Fund (RDEB2024-11), and the Research and Development Fund of Peking University People's Hospital (RDX 2025-06).

Data Availability

The data that support the findings of this study are available on reasonable request from the corresponding author.

The fine-tuning and training pipeline in this study was implemented using the open-source MS-Swift framework [37]. However, the specific data processing scripts and trained model weights cannot be publicly shared due to institutional data governance and

XSL•FO
RenderX

security policies for pediatric clinical video data and because parts of the pipeline are integrated with internal hospital systems for secure data access. Additional methodological details and hyperparameter configurations are provided in the Methods section and may be further clarified by the corresponding author upon reasonable request.

Authors' Contributions

ZY conceptualized and designed the study, coordinated the research, and revised the manuscript; YW and YN co-designed the study, drafted the initial manuscript, and revised the manuscript; ZX, JL, XJ, XC, and DS collected and summarized data and revised the manuscript. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Pringsheim T, Okun MS, Müller-Vahl K, Martino D, Jankovic J, Cavanna AE, et al. Practice guideline recommendations summary: treatment of tics in people with Tourette syndrome and chronic tic disorders. Neurology. May 07, 2019;92(19):896-906. [FREE Full text] [doi: 10.1212/WNL.0000000000007466] [Medline: 31061208]
2. Garcia-Delgar B, Servera M, Coffey BJ, Lázaro L, Openneer T, Benaroya-Milshtein N, et al. Tic disorders in children and adolescents: does the clinical presentation differ in males and females? A report by the EMTICS group. Eur Child Adolesc Psychiatry. Oct 2022;31(10):1539-1548. [doi: 10.1007/s00787-021-01751-4] [Medline: 33944988]
3. Cohen SC, Leckman JF, Bloch MH. Clinical assessment of Tourette syndrome and tic disorders. Neurosci Biobehav Rev. Jul 2013;37(6):997-1007. [FREE Full text] [doi: 10.1016/j.neubiorev.2012.11.013] [Medline: 23206664]
4. Szejko N, Robinson S, Hartmann A, Ganos C, Debes NM, Skov L, et al. European clinical guidelines for Tourette syndrome and other tic disorders-version 2.0. Part I: assessment. Eur Child Adolesc Psychiatry. Mar 2022;31(3):383-402. [FREE Full text] [doi: 10.1007/s00787-021-01842-2] [Medline: 34661764]
5. Banaschewski T, Neale BM, Rothenberger A, Roessner V. Comorbidity of tic disorders and ADHD: conceptual and methodological considerations. Eur Child Adolesc Psychiatry. Jun 2007;16 Suppl 1:5-14. [doi: 10.1007/s00787-007-1002-8] [Medline: 17665278]
6. Leckman JF. Tic disorders. BMJ. Jan 05, 2012;344:d7659. [doi: 10.1136/bmj.d7659] [Medline: 22223834]
7. Nilles C, Martino D, Fletcher J, Sarna JR, Pringsheim T. Challenging current diagnostic and classification criteria for primary tic disorders. Eur Child Adolesc Psychiatry. Dec 2025;34(12):3883-3890. [doi: 10.1007/s00787-025-02732-7] [Medline: 40531361]
8. Wen F, Gu Y, Yan J, Liu J, Wang F, Yu L, et al. Revisiting the structure of the Yale Global Tic Severity Scale (YGTSS) in a sample of Chinese children with tic disorders. BMC Psychiatry. Aug 09, 2021;21(1):394. [FREE Full text] [doi: 10.1186/s12888-021-03399-5] [Medline: 34372795]
9. Cernera S, Pramanik L, Boogaart Z, Cagle JN, Gomez J, Moore K, et al. The Human Tic Detector: an automatic approach to tic characterization using wearable sensors. Clin Neurophysiol. Feb 2022;134:102-110. [doi: 10.1016/j.clinph.2021.10.017] [Medline: 34952803]
10. Hau C, Xia W, Ryan S, Firth J, Linardon J, Torous J. Smartphone monitoring and digital phenotyping apps for schizophrenia: a review of the academic literature. Schizophr Res. Jul 2025;281:237-248. [doi: 10.1016/j.schres.2025.05.019] [Medline: 40413837]
11. Qiang J, Hong W, Sun Y, Lyu X, Pan Z, Wu D, et al. Evaluation of an AI facial recognition system for Turner syndrome screening and facial complexity: a prospective cohort. Int J Med Inform. Nov 2025;203:105976. [doi: 10.1016/j.ijmedinf.2025.105976] [Medline: 40479775]
12. Danecki D, Sage A, Miodońska Z, Zowada S, Lipowicz A, Myśliwiec A, et al. Objective dynamic assessment of facial movement asymmetry in children using a marker-based video method. J Clin Med. Feb 28, 2026;15(5):1870. [FREE Full text] [doi: 10.3390/jcm15051870] [Medline: 41827286]
13. Zhang J, Feng B, Cheng Y, Zhou F, Lu X, Zhu X, et al. Artificial intelligence-assisted quality assessment of mid-trimester ultrasound examinations using large vision-language models. BMC Pregnancy Childbirth. Apr 22, 2026;26(1):614. [FREE Full text] [doi: 10.1186/s12884-026-09073-6] [Medline: 42021182]
14. Hong Q, Liu S, Wu L, Lu Q, Yang P, Chen D, et al. Evaluating the performance of large language and visual-language models in cervical cytology screening. NPJ Precis Oncol. May 23, 2025;9(1):153. [FREE Full text] [doi: 10.1038/s41698-025-00916-7] [Medline: 40410424]
15. Patrício C, Rio-Torto I, Cardoso JS, Teixeira LF, Neves JC. CBVLM: training-free explainable concept-based large vision language models for medical image classification. Comput Biol Med. Nov 2025;198(Pt A):111145. [FREE Full text] [doi: 10.1016/j.compbiomed.2025.111145] [Medline: 41075479]
16. Lucas A, Revell A, Davis KA. Artificial intelligence in epilepsy - applications and pathways to the clinic. Nat Rev Neurol. Jun 2024;20(6):319-336. [doi: 10.1038/s41582-024-00965-9] [Medline: 38720105]

XSL•FO
RenderX

17. Reddy A, Reddy RP, Roghani AK, Garcia RI, Khemka S, Pattoor V, et al. Artificial intelligence in Parkinson's disease: early detection and diagnostic advancements. Ageing Res Rev. Aug 2024;99:102410. [doi: 10.1016/j.arr.2024.102410] [Medline: 38972602]
18. Brügge NS, Sallandt GM, Schappert R, Li F, Siekmann A, Grzegorzek M, et al. Automated motor tic detection: a machine learning approach. Mov Disord. Jul 2023;38(7):1327-1335. [doi: 10.1002/mds.29439] [Medline: 37166278]
19. Farkhod A, Abdusalomov AB, Mukhiddinov M, Cho YI. Development of real-time landmark-based emotion recognition CNN for masked faces. Sensors (Basel). Nov 11, 2022;22(22):8704. [FREE Full text] [doi: 10.3390/s22228704] [Medline: 36433303]
20. Friji R, Chaieb F, Drira H, Kurtek S. Geometric deep neural network using rigid and non-rigid transformations for landmark-based human behavior analysis. IEEE Trans Pattern Anal Mach Intell. Nov 2023;45(11):13314-13327. [doi: 10.1109/TPAMI.2023.3291663] [Medline: 37399164]
21. Wang Z, Ma H, Zhai J. Low-rank adaptation for edge AI. Sci Rep. Sep 26, 2025;15(1):33109. [FREE Full text] [doi: 10.1038/s41598-025-16794-9] [Medline: 41006599]
22. Loewenstern Y, Benaroya-Milshtein N, Belelovsky K, Bar-Gad I. Automatic identification of facial tics using selfie-video. IEEE J Biomed Health Inform. Jan 2025;29(1):409-419. [doi: 10.1109/JBHI.2024.3488285] [Medline: 39475732]
23. Schappert R, Verrel J, Brügge NS, Li F, Paulus T, Becker L, et al. Automated video-based approach for the diagnosis of Tourette syndrome. Mov Disord Clin Pract. Sep 2024;11(9):1136-1140. [FREE Full text] [doi: 10.1002/mdc3.14158] [Medline: 38973244]
24. Bernabei M, Preatoni E, Mendez M, Piccini L, Porta M, Andreoni G. A novel automatic method for monitoring Tourette motor tics through a wearable device. Mov Disord. Sep 15, 2010;25(12):1967-1972. [FREE Full text] [doi: 10.1002/mds.23188] [Medline: 20669298]
25. Gdoura A, Degünther M, Lorenz B, Effland A. Combining CNNs and Markov-like models for facial landmark detection with spatial consistency estimates. J Imaging. May 22, 2023;9(5):104. [doi: 10.3390/jimaging9050104] [Medline: 37233323]
26. Xiang W, Zhu G, Hou Y, Mei Z, Wan L, Zhang L, et al. Development and validation of an interpretable machine learning model for predicting tic disorders and severity in children based on electroencephalogram data. IEEE Trans Neural Syst Rehabil Eng. 2025;33:2416-2427. [doi: 10.1109/TNSRE.2025.3579763] [Medline: 40522788]
27. Guo C, Zhang L. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Trans Image Process. Jan 2010;19(1):185-198. [doi: 10.1109/TIP.2009.2030969] [Medline: 19709976]
28. Yao T, Li Y, Pan Y, Mei T. HIRI-ViT: scaling vision transformer with high resolution inputs. IEEE Trans Pattern Anal Mach Intell. Sep 2024;46(9):6431-6442. [doi: 10.1109/TPAMI.2024.3379457] [Medline: 38502628]
29. Ozcan A, Coudert FX, Rogge SM, Heydenrych G, Fan D, Sarikas AP, et al. Artificial intelligence paradigms for next-generation metal-organic framework research. J Am Chem Soc. Jul 09, 2025;147(27):23367-23380. [FREE Full text] [doi: 10.1021/jacs.5c08214] [Medline: 40551706]
30. Nguyen HV, Park H, Yoo N, Yang J. Resource-efficient fine-tuning of large vision-language models for multimodal perception in autonomous excavators. Front Artif Intell. Nov 18, 2025;8:1681277. [FREE Full text] [doi: 10.3389/frai.2025.1681277] [Medline: 41341806]
31. Baptista A, Barp A, Chakraborti T, Harbron C, MacArthur BD, Banerji CR. Deep learning as Ricci flow. Sci Rep. Oct 08, 2024;14(1):23383. [FREE Full text] [doi: 10.1038/s41598-024-74045-9] [Medline: 39379488]
32. Li Y, Li Y, Naqvi SM. Unified VideoMAE framework for detection of multi-disorder ADHD and depression. IEEE J Biomed Health Inform. Dec 01, 2025:1-13. [doi: 10.1109/JBHI.2025.3638973] [Medline: 41325113]
33. Zou L, Choi YH, Guizzetti L, Shu D, Zou J, Zou G. Extending the DeLong algorithm for comparing areas under correlated receiver operating characteristic curves with missing data. Stat Med. Sep 20, 2024;43(21):4148-4162. [doi: 10.1002/sim.10172] [Medline: 39013403]
34. Martín Andrés A, Álvarez Hernández M. Hubert's multi-rater kappa revisited. Br J Math Stat Psychol. Feb 2020;73(1):1-22. [doi: 10.1111/bmsp.12167] [Medline: 31056757]
35. Wong CY, Antaki F, Woodward-Court P, Ong AY, Keane PA. The role of saliency maps in enhancing ophthalmologists' trust in artificial intelligence models. Asia Pac J Ophthalmol (Phila). Jul 2024;13(4):100087. [FREE Full text] [doi: 10.1016/j.apjo.2024.100087] [Medline: 39069106]
36. Hebenstreit K, Praas R, Kiesewetter LP, Samwald M. A comparison of chain-of-thought reasoning strategies across datasets and models. PeerJ Comput Sci. Apr 30, 2024;10:e1999. [FREE Full text] [doi: 10.7717/peerj-cs.1999] [Medline: 38855241]
37. modelscope/ms-swift. GitHub. URL: https://github.com/modelscope/ms-swift [accessed 2026-10-12]

Abbreviations

AUC: area under the curve
CoT: chain-of-thought
DNN: deep neural network
LLM: large language model
LoRA: low-rank adaptation

XSL•FO
RenderX

LVLM: large vision-language model
MLP: multilayer perceptron
TD: tic disorder
ViT: vision transformer
YGTSS: Yale Global Tic Severity Scale

Edited by I Steenstra; submitted 19.Apr.2026; peer-reviewed by R Martínez-García-Peña, C Wu; comments to author 20.May.2026; revised version received 01.Oct.2026; accepted 06.Oct.2026; published 06.Nov.2026

Please cite as:
Wu Y, Niu Y, Xu Z, Liu J, Jiao X, Cui X, Sun D, Yang Z
Large Vision-Language Models for Pediatric Motor Tic Detection: Case-Control Diagnostic Study
JMIR AI 2026;5:e98782
URL: https://ai.jmir.org/2026/1/e98782
doi: 10.2196/98782
PMID:

©Yujun Wu, Yue Niu, Zhao Xu, Jing Liu, Xianru Jiao, Xia Cui, Dan Sun, Zhixian Yang. Originally published in JMIR AI (https://ai.jmir.org), 06.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on https://www.ai.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX