Development and Validation of an Interpretable Machine Learning Model for Predicting Tic Disorders and Severity in Children Based on Electroencephalogram Data

Wanting Xiang, Gang Zhu, Yichong Hou, Zhandong Mei, Lin Wan, Li Zhang, Guang Yang, and Jian Zu

***Abstract*— Accurate diagnosis of Tic disorders (TD) and its severity based on electroencephalogram (EEG) data were of great clinical importance. This study analyzed EEG data from 90 children with TD and 88 healthy controls (HC). A two-stage progressive diagnosis framework based on EEG data and machine learning methods was developed. To achieve individualized prediction and reduce the feature dimension, we proposed a novel individual-based feature-weighted integration method in machine learning, as well as a new SHAP-driven feature selection and weighting (SFSW) strategy to improve the prediction accuracy. Based on 13 weighted features, Logistic Regression model achieved an average accuracy of 94.2% (95% CI, 90.6%-97.9%) in diagnosing TD, with a sensitivity of 92.4% (95% CI, 85.3%-99.5%) and a specificity of 96.1% (95% CI, 92.9%-99.2%). The Decision Tree model attained an average accuracy of 81.5% (95% CI, 68.6%-94.5%) in predicting severity, with a sensitivity of 81.5% (95% CI, 68.6%-94.5%) and a specificity of 89.9% (95% CI, 82.1%-97.6%). In the hold-out set validation, the method demonstrated accuracy rates of 95.7% in diagnosing TD and 83.3% in predicting severity. Interpretability analysis revealed that the top three main features affecting TD diagnosis were the mean frequency (MNF) of P3 channel β band, age and MNF of C3 channel γ band. This work offered a more efficient approach to individualized diagnosis of TD and had substantial practical value for clinical auxiliary diagnosis and intervention.**

***Index Terms*— Tic disorders, severity, electroencephalogram, machine learning, interpretability.**

I. Introduction

TIC disorders (TD) was a complex neurodevelopmental disorder characterized by involuntary, sudden, rapid, repetitive, and non-rhythmic motor and/or vocal tics [1]. Epidemiological studies had shown that the incidence of TD in the population was increasing. Foreign studies had reported that the prevalence rate in the group of children aged 5-18 years was approximately 0.4–3.8% [2]. The impact of TD on the physical and mental health of children was multidimensional, not only affecting their learning, socialization, and daily life functions, but also easily causing psychological problems such as low self-esteem and depression, which seriously affected the quality of life [3].

In terms of clinical diagnosis, the current diagnosis of TD relied mainly on clinical observation and scale assessment [4]. Although TD exhibited some significant clinical features, it may pose diagnostic difficulties when tic manifestations overlapped with autonomous behaviors or psychiatric symptoms [5]. Functional brain research had provided unique insights into exploring the neural mechanisms of TD, physiological signaling changes in the brains of patients, and proposing classification models [6], [7]. Greene et al. [8] extracted resting-state functional connectivity (RSFC) as features and used a multivariate method to identify and classify Tourette syndrome (TS). Liao et al. [9] applied voxel-mirrored homotopic connectivity (VMHC) to extract features and used multivariate pattern analysis (MVPA) to diagnose TS. Wen et al. [10] constructed a brain network for each subject based on fusion of similar network, combined with a

Received 16 April 2025; accepted 10 June 2025. Date of publication 16 June 2025; date of current version 20 June 2025. This work was supported in part by Shaanxi Fundamental Science Research Project for Mathematics and Physics under Grant 22JSY039, in part by the National Natural Science Foundation of China under Grant 12471470, in part by the General Project of National Key Research and Development Program of China under Grant 2023YFC2706405, and in part by Beijing Natural Science Foundation under Grant 7222187. *(Wanting Xiang and Gang Zhu contributed equally to this work.) (Corresponding authors: Guang Yang; Jian Zu.)*

Wanting Xiang, Yichong Hou, Zhandong Mei, and Jian Zu are with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: jianzu@xjtu.edu.cn).

Gang Zhu and Lin Wan are with the Department of Pediatrics, The First Medical Center, Chinese PLA General Hospital, Beijing 100853, China, also with the Senior Department of Pediatrics, The Seventh Medical Center, Chinese PLA General Hospital, Beijing 100700, China, and also with the Medical School, Chinese People's Liberation Army, Beijing 100853, China.

Guang Yang is with the Department of Pediatrics, The First Medical Center, Chinese PLA General Hospital, Beijing 100853, China, also with the Senior Department of Pediatrics, The Seventh Medical Center, Chinese PLA General Hospital, Beijing 100700, China, also with the Medical School, Chinese People's Liberation Army, Beijing 100853, China, and also with The Second School of Clinical Medicine, Southern Medical University, Guangzhou 510280, China (e-mail: yangg301@126.com).

Li Zhang is with Chinese Institute for Brain Research, Beijing 102206, China (e-mail: zhangli@cibr.ac.cn).

Digital Object Identifier 10.1109/TNSRE.2025.3579763

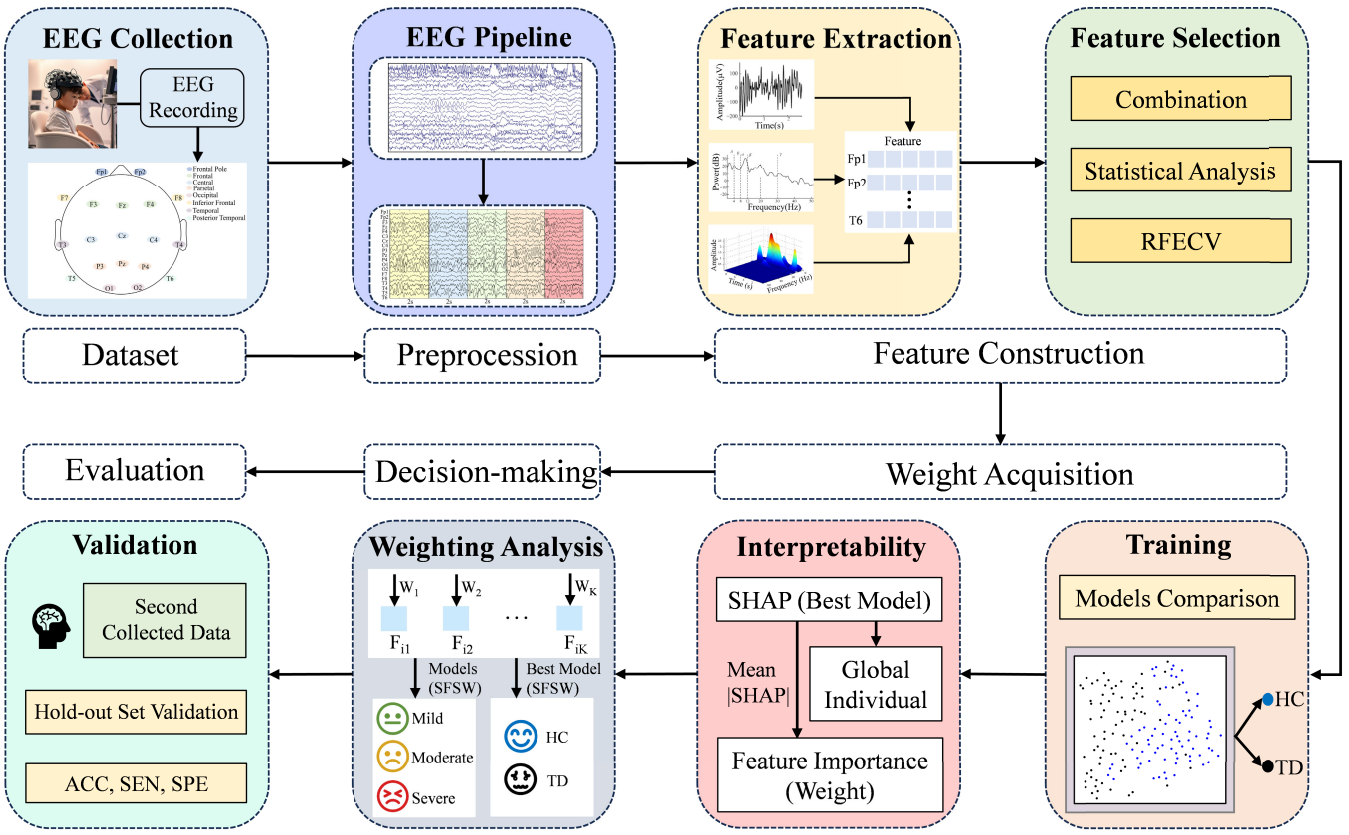


Fig. 1. Overview of the proposed framework for predicting the presence and severity of TD in children.

multi-threshold strategy, and diagnosed TS based on SVM. Electroencephalogram (EEG), as one of the non-invasive neuroimaging technologies with the highest temporal resolution at present [11], objectively reflected the dynamic changes in the physiological state of the brain and could be continuously applied during childhood development. EEG was very suitable for the study of brain function in children with TD. Duan et al. [7] extracted spatial features, such as the clustering coefficient and global efficiency, and used the spatial pattern of the network (SPN) to diagnose TS. Wang et al. [11] constructed a deep learning-based diagnostic model for pediatric TD. EEG analysis based on machine learning (ML) had made great progress in the research of intelligent diagnosis and prediction of neuropsychiatric diseases, such as Alzheimer's disease [12], epilepsy [13], and attention deficit hyperactivity disorder (ADHD) [14]. The combination of signal processing methods, feature selection and ML was considered as the key factor to achieve optimal classification accuracy [12]. Durongbhan et al. [12] successfully diagnosed Alzheimer's disease by extracting frequency and time-frequency features from EEG, with a classification accuracy of up to 99% during the short (4s) eyes-open EEG signal segments. Hernández et al. [13] utilized the information of time domain, frequency domain, and time-frequency domain to identify epilepsy, extracted 52 features from EEG, and the accuracy of all models was around 90%. However, to the best of our knowledge, few studies had been conducted to diagnose TD and its severity based on EEG data, and interpretability was unclear.

Therefore, this study proposed a novel ML framework for two-stage diagnosis of children with TD using EEG data, with the overall framework illustrated in Fig. 1. We evaluated the performance of the framework on an independently collected data, comparing it against recent state-of-the-art methods, including Residual Neural Network (ResNet) [11], Convolutional Neural Network (CNN) [15], and Deep Neural Network (DNN) [16]. Additionally, we focused on the interpretability of the model and explored the potential for application in clinical practice. By identifying the key neurophysiological features associated with TD, this study was expected to provide clinicians with a better understanding and assist them in more accurately identifying and explaining the pathogenesis of TD children.

The main contributions of this study were: (1) To achieve individualized prediction and overcome the difficulty of inconsistent numbers of EEG segments per individual, we proposed a novel individual-based feature-weighted integration method in machine learning to improve the prediction accuracy. (2) In order to reduce the feature dimension and strengthen the importance of selected features, based on the mean absolute SHAP value, we designed a SHAP-driven feature selection method and weighting strategy to further improve the performance of model prediction. (3) We constructed a two-stage progressive diagnosis framework for TD and severity, which

TABLE I
DESCRIPTIVE STATISTICAL ANALYSIS OF BASIC VARIABLES OF TD AND HC CHILDREN

Samples	Variables	Numbers	Gender (Male)	Gender (Female)	Age (Mean±SD)	EEG segments
Total samples	TD	90	76	14	9.09±1.84	11838
	HC	88	61	27	7.95±1.65	11222
	Mild	30	22	8	8.66±1.67	3885
	Moderate	40	35	5	9.02±1.92	5233
	Severe	17	16	1	10.03±1.68	2276
Training set (First collected)	TD	78	64	14	8.96±1.90	100479
	HC	77	55	22	7.64±1.47	100226
	Mild	26	18	8	8.46±1.57	3486
	Moderate	36	31	5	8.95±2.00	4780
	Severe	13	12	1	9.99±1.93	1769
Hold-out set (Second collected)	TD	12	12	0	9.89±1.14	1359
	HC	11	6	5	10.10±1.16	996
	Mild	4	4	0	9.94±1.97	399
	Moderate	4	4	0	9.61±0.78	453
	Severe	4	4	0	10.14±0.32	507

effectively improved the identification accuracy of TD severity and achieved individualized accurate prediction.

II. MATERIALS

A. Dataset

In this study, EEG data from 178 children aged 4-12 years collected in the Department of Pediatrics of the First Medical Center of the General Hospital of the People's Liberation Army (PLA) from February 2022 to September 2023 were selected. 90 children were diagnosed with TD by two chief or deputy pediatric chief physicians based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) in a double-masked manner. They were accompanied by completing EEG examinations and Yale Global Tic Severity Scale (YGTSS) assessments. And they did not have severe organic brain dysfunction, concomitant epilepsy and cerebral palsy, and severe chronic diseases. Their mean age was 9.09 ± 1.84 years. The healthy controls (HC) group was recruited from 88 healthy volunteers with normal mental and intellectual abilities, normally attending kindergarten or elementary school, ensuring that there was no history of neuropsychiatric diseases such as neurological diseases, mental retardation, emotional disorders, ADHD, and pervasive developmental disorders, with a mean age of 7.95 ± 1.65 years. The guardians provided their written informed consent and the study was approved by the Ethics Committee of PLA General Hospital (S2021-488-01).

The EEG data acquisition instrument was recorded by the U32 EEG system of Neuracle Technology (Changzhou) Co., Ltd., with a sampling frequency of 1000 Hz. Electrode arrangement was performed according to the international 10-20 system, with a total of 19 electrodes (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Cz, Fz, Pz). For the analysis of brain regions, this study focused on eight regions: frontal pole, frontal, central, parietal, occipital, inferior frontal, temporal, and posterior temporal. Each subject recorded a resting-state EEG for about 5 minutes during the eyes-open and eyes-closed periods. TD severity was assessed using the YGTSS [17], noting =<24 as mild, 25-39 as moderate, and > = 40 as severe. TD (n = 90) and HC samples (n = 88) were included. TD children were asked to complete the YGTSS (n = 87, with 3 samples missing due to chance), and physicians assisted patients throughout the assessment. The descriptive statistical results were presented in Table I.

B. Preprocessing

This study focused on processing and analyzing EEG data during the eyes-closed period, as this state was accompanied by minimal external stimuli. The EEGLAB toolbox was used to preprocess the raw EEG signals. Initially, channel selection and location were performed to ensure the recorded signals effectively represented neural activity in specific brain regions. Data were then resampled to 256 Hz to facilitate faster analysis. Next, all channels were re-referenced by averaging to reduce global noise and enhance the clarity of local signals. A bandpass filter (0.5-70Hz) was applied to eliminate high-frequency and low-frequency noise, while a notch filter at 50 Hz was used to remove interference from power frequencies. The pop_rejchan function in the Harvard EEG Automatic Processing Pipeline (HAPPE) was then employed to detect problematic channels using amplitude and spectral thresholds. Faulty channels were corrected using the spherical spline interpolation method [18]. Independent Component Analysis (ICA) was applied to isolate and remove physiological artifacts, such as those caused by blinking, horizontal eye movements, muscle activity, and heartbeats, resulting in cleaner EEG signals. Data from 5 minutes eyes-closed period were then extracted and divided into 2s segments. (If the eyes-closed period lasted less than 5 minutes, all available data were used.) Any segment with significant non-physiological fluctuations was manually reviewed and removed. Finally, 11,838 segments were collected from children with TD, and 11,222 segments were gathered from HC.

C. Data Division

The data used in this study were collected in two batches from the same research center using the same EEG acquisition

device. We used the first collected data (TD, n = 78) and (HC, n = 77) as the training set, and the second collected data (TD, n = 12) and (HC, n = 11) as the hold-out set [19], [20].

Based on the YGTSS results, TD samples (n = 75) were categorized into 3 levels of severity: mild (n = 26), moderate (n = 36) and severe (n = 13). During the model training, an imbalance in sample distribution between categories was observed. To alleviate the problem, the Borderline-SMOTE algorithm was employed for data enhancement [21]. The Borderline-SMOTE algorithm was an advanced oversampling technique based on SMOTE, which synthesized minority class samples near the decision boundary to avoid overfitting. After applying the algorithm, balanced samples for mild (n = 36), moderate (n = 36), and severe (n = 36) were obtained, as the training set. TD samples (n = 12) in the hold-out set were divided into mild (n = 4), moderate (n = 4), and severe (n = 4).

III. Methods

A. Feature Extraction

This study extracted features from time domain, frequency domain, and time-frequency domain. We extracted 14 features of time domain, including mean [22], standard deviation [22], normalized first-order difference [22], energy [22], minimums and maximums (MMAX) [23], kurtosis [24], skewness [24], Katz fractal dimension [25], Higuchi fractal dimension [25], Non-Stationary Index (NSI) [22], zero-crossing rate [23], Hjorth parameter [22] (activity, mobility, complexity). The power spectral density (PSD) was used to convert the EEG signals into a correspondence between frequency and power, and the features of frequency domain were extracted from the PSD. In order to solve the problem of excessive variance in the periodogram, the Welch method was used for analysis [26]. The decomposition of the signal into several frequency bands was temporary and lacked a direct biological basis, with slight variations in the band definitions across different authors [27]. The following definitions were used in this study: δ (0.5-4Hz), θ (4)-8Hz), α (8-12Hz), β (13-30Hz), γ (31-70Hz). Within the range of 0.5-70 Hz, the following features were calculated for each frequency band: mean PSD [25], standard deviation of PSD [25], mean frequency (MNF) [28], mean frequency band power [27], power spectrum entropy (PSE) [29], differential entropy (DE) [30], and spectral entropy (SE) [25]. Additionally, continuous wavelet transform (CWT) was used to extract the features of time-frequency domain from EEG, including energy [31], absolute mean [31], standard deviation [31], recursive energy efficiency [32], and wavelet time frequency entropy [33].

B. Feature Selection

This study adopted a feature selection method based on statistical analysis. Initially, features that showed significant differences and low correlation between the TD and HC were selected. Then, Recursive Feature Elimination with Cross-Validation (RFECV) was applied to further refine the selection of features from specific channels. The selected features were incorporated into the models for training, significantly reducing the dimension of features.

1) Demographic Analysis: Comparisons were made between the gender and age of children with TD and HC. The chi-square test was used to compare gender, while the Mann-Whitney U test was applied to age after checking for unequal variances using the Kolmogorov-Smirnov test. Adjustment was made for variance instability when necessary. Statistical analysis was two-sided, with $p < 0.05$ considered significant, and was performed using SPSS version 26.0.

2) Statistical Analysis: Statistical analyses of EEG features from children with TD and HC were performed in three categories: time domain, frequency domain, and time-frequency domain, using MATLAB. First, inter-group differences were analyzed. The Kolmogorov-Smirnov test was applied to assess the consistency of variances. After confirming unequal variances, the Mann-Whitney U test was used to identify features with the most significant differences between the two groups, with $p < 0.01$. Then, Mutual Information (MI) was calculated for the selected features, and highly correlated features were removed to ensure the independence and validity of the final feature set.

3) Recursive Feature Elimination With Cross-Validation: A systematic feature selection strategy was employed here, specifically RFECV using random forest as the estimator. The RFECV algorithm ranked the features according to their importance and iteratively removed those that had minimal impact on the classification performance. Redundant and weakly contributing features were discarded, while independent features that significantly improved model performance were retained [34].

C. Individual-Based Feature-Weighted Integration Module

Fig. 2 in part A illustrated the overall framework of individual-based feature-weighted integration module. This study selected 5 minutes of EEG data (256 Hz) for each subject and divided it into 150 non-overlapping 2s segments. To ensure data quality, both automated artifact detection and manual inspection were employed to remove segments with substantial interference, such as noise, artifacts, or breakpoints. Suppose the set of valid segments for the subject i was:

$$X_i = \left\{ x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(M_i)} \right\}, x_i^{(m)} \in \mathbb{R}^{C \times 512} \tag{1}$$

C represented the number of electrode channels, 512 represented the number of time points in each segment, and $M_i \leq 150$ represented the number of valid segments for the subject i to be retained. Subsequently, we performed feature extraction on each retained segment $x_i^{(m)}$, so the feature matrix of each segment was recorded as:

$$D_i^{(m)} \in \mathbb{R}^{C \times Q} \tag{2}$$

Q represented the number of extracted features. To ensure consistent input dimensions across subjects and to minimize the influence of transient neural fluctuations on the modeling process, the features from all valid segments of the subject i were uniformly integrated (equal weights assigned to each segment) across the segment number to obtain a consistent

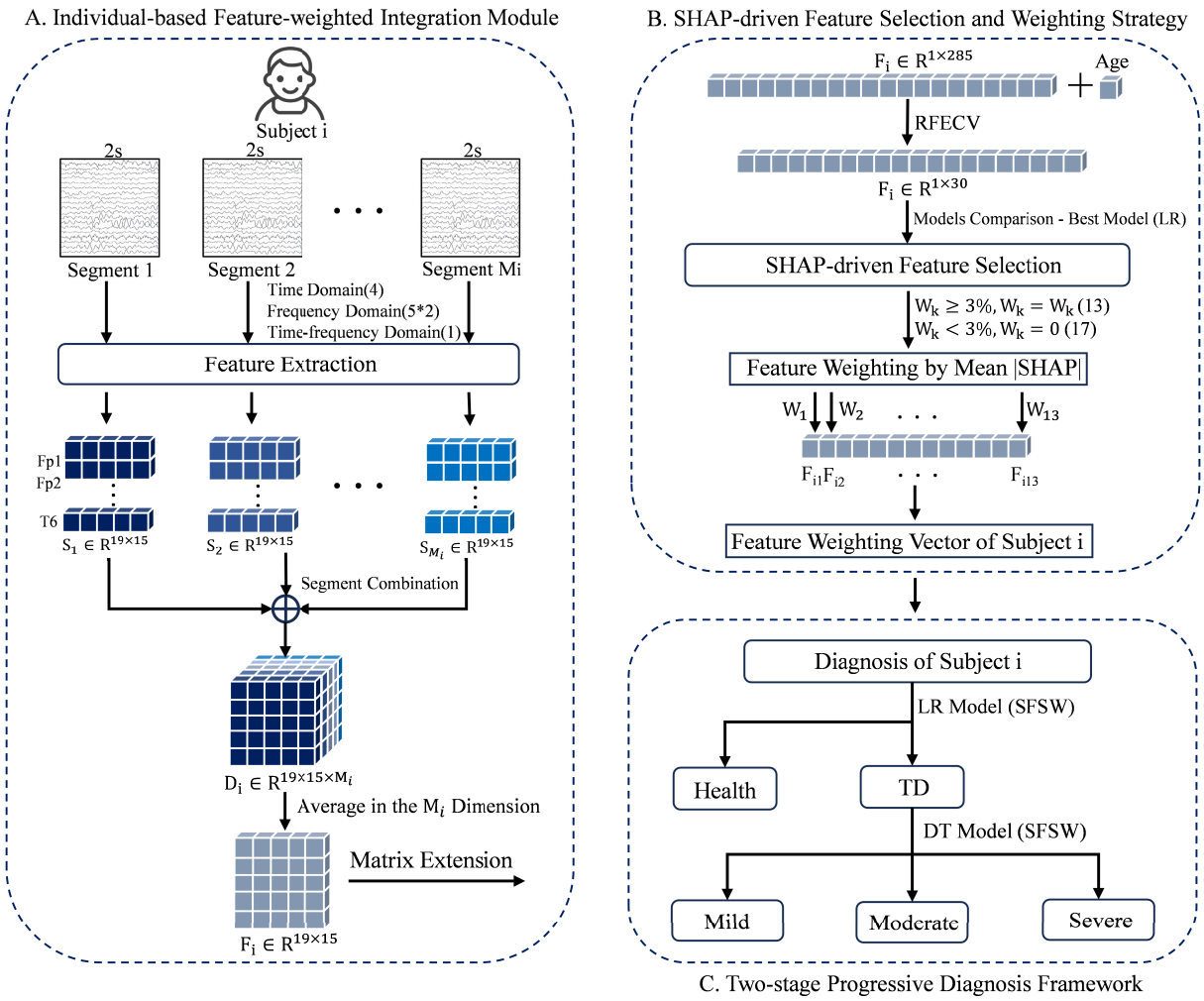


Fig. 2. Overview of the proposed two-stage progressive diagnosis framework based on feature-weighted integration module and SFSW Strategy. A. Individual-based feature-weighted integration. Constructed a unified feature representation. B. SFSW Strategy. Optimized feature importance. C. Two-stage Progressive Diagnosis Framework. Performed TD diagnosis based on the reweighted features.

individual-level feature representation:

$$F_i = \sum_{m=1}^{M_i} \omega_i D_i^{(m)} \in \mathbb{R}^{C\times Q}, \omega_i = \frac{1}{M_i} \tag{3}$$

Finally, the integration module output a feature matrix $F_i \in \mathbb{R}^{C\times Q}$ with a consistent structure across all subjects. This standardized representation served as the input for subsequent steps, including feature selection, feature weighting, and classification modeling.

This study introduced an individual-based feature-weighted integration module to address the challenge of data inconsistency caused by varying numbers of EEG valid segments across different subjects. Most current EEG classification studies treated each segment as an independent sample [15], [16], which helped maximize data usage. However, this strategy could not capture the relationship between segments within an individual, which limited the accurate expression of the overall cognitive state and had difficulties in individual prediction tasks. Additionally, in the face of inconsistent numbers of valid segments for different individuals, current methods lacked effective strategies to handle the input heterogeneity. To address these issues, the proposed integration module enabled a smooth shift from segment-level to individual-level modeling, thereby enhancing the robustness and generalizability of the model in real applications. Although we adopted a uniform weighting scheme, the proposed module was designed with flexibility to accommodate future incorporation of segment-specific weights, making it a potentially feature-weighted framework.

D. SHAP-Driven Feature Selection and Weighting (SFSW) Strategy

To improve the interpretability and prediction performance of the model, we proposed a novel strategy termed SHAP-driven Feature Selection and Weighting (SFSW), which integrated SHAP-based interpretability with adaptive feature contribution optimization. Fig. 2 in part B illustrated the overall process of SFSW strategy. Algorithm 1 summarized the detailed steps.

First, a feature matrix with consistent dimensions was constructed for each subject using the individual-based

Algorithm 1 Two-Stage TD Diagnosis Algorithm With SFSW Strategy

Input: EEG dataset $X \in \mathbb{R}^{N\times C\times T}$ (N: samples, C: channels, T: time series), labels $y \in \{0, 1\}$ (HC, TD), $z \in \{1, 2, 3\}$ (Mild, Moderate, Severe)

Output: Predictions $\hat{y}$, $\hat{z}$ for hold-out data

1: **Feature extraction:**
for i = 1 to N **do**
Obtain pre-processed EEG data X_i with M_i segments
for m = 1 to M_i **do**
Extract features $D_i^{(m)} \in \mathbb{R}^{C\times Q}$ (Q: features)
end for
Combine features from all M_i segments for subject i, $D_i \in \mathbb{R}^{C\times Q\times M_i}$
Aggregate features across segments, $F_i \in \mathbb{R}^{C\times Q}$
Flatten F_i into a vector, $F_i \in \mathbb{R}^{1\times(C\times Q)}$
Add demographic features, $F_i \in \mathbb{R}^{1\times(C\times Q+L)}$ (L: demographic features)
end for
2: Apply RFECV to obtain $F \in \mathbb{R}^{N\times K}$ (K: selected features), where $K < (C \times Q + L)$
3: Put into ML models to choose the best model h
Compute feature weights using SHAP analysis, W_k=mean|SHAP(F_k)|, for k=1, . . . , K
if $W_k \geq 3\%$**then** W_k=W_k**else** W_k=0 **end if**
for i = 1 to N **do** F_i^{w}=$F_i \odot W$ **end for**
4: **Train**:
Train the parameter θ of model $h_\theta()$ until convergence with $F_{\text{train}}^{\mathrm{w}}$
Choose the best model g, train ζ of model $g_\zeta()$ until convergence with $F_{\text{train}}^{\mathrm{w}}$
5: **Predict**:
$\hat{y}$=$h_\theta(F_{\text{hold-out}}^{\mathrm{w}})$
if $\hat{y}$=1**then** $\hat{z}$=$g_\zeta(F_{\text{hold-out}}^{\mathrm{w}})$ **end if**
return $\hat{y}$, $\hat{z}$

feature-weighted integration module. Then feature selection was performed to identify the final input feature set $F \in \mathbb{R}^{N\times K}$ (N represented the number of samples, K represented the number of features) for model training. Subsequently, based on the optimal model obtained from the preliminary training, we used the KernelSHAP algorithm to calculate the SHAP value of each input feature to quantify the contribution of the feature in model prediction. To assess the overall importance of each feature at the global level, the mean absolute SHAP value was calculated across all samples to obtain the overall contribution of each feature [35].

$$W_k = \frac{1}{N}\sum_{i=1}^{N} |\mathrm{SHAP}(F_{ik})|, \quad k = 1, 2, \ldots, K \tag{4}$$

To further eliminate redundant features that contributed less to the model, we set a fixed threshold (3% in this study) and only retained features whose contribution exceeded the threshold. The specific weight update rule was as follows:

$$W_k = \begin{cases} W_k, & \text{if } W_k \geq 3\% \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

We used W as the weighting coefficient and multiplied it by the original feature value to obtain a weighted feature vector, which was then used for final model training.

$$F_i^{\mathrm{w}} = F_i \odot W \tag{6}$$

The SFSW strategy proposed in this study aimed to improve classification accuracy while reducing the number of features and identifying the most important features for distinguishing TD. The original feature selection method based on SHAP value used a basic binary rule (retain or discard) [35]. While this could improve model efficiency, it ignored the varying degrees of importance that different features may have, and how these differences influenced model decisions. This method may fall short in both optimizing performance and enhancing interpretability. To address these limitations, the SFSW strategy introduced a feature weighting mechanism guided by interpretability. By applying a threshold and constructing weights, this enhanced the influence of key features while effectively removing those with minimal impact, achieving fine-grained control at the feature level. This strategy provided a more accurate and explanatory optimization path for feature selection.

E. Two-Stage Progressive Diagnosis Framework

The diagnostic task was structured into two stages. The first stage involved determining whether a child was TD or HC, while the second stage assessed the severity of TD child (mild, moderate, or severe). The specific framework was shown in part C of Fig. 2. Both stages employed five widely used ML models: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), eXtreme Gradient Boosting (XGBoost) and K-Nearest Neighbors (KNN). All models were implemented in a consistent computational environment (Python 3.7), and the input features were standardized prior to model training. A uniform data division strategy and consistent evaluation metrics were applied across all models to ensure fair comparison. All models were trained and evaluated using five-fold cross-validation. In each fold, the dataset was divided into training and validation sets, with the class distribution kept consistent across both to mitigate potential bias from class imbalance. Parameter optimization was performed using Grid Search within a predefined search space, and the parameter combination with the best validation performance was selected. Subsequently, using the selected parameters, the model was retrained on the entire training set to produce the final trained model for diagnosis. Model performance was assessed using multiple evaluation metrics, including F-score, accuracy (ACC), sensitivity (SEN), specificity (SPE), precision (PRE), and Area Under the Receiver Operating Characteristic (AUROC). Additionally, the average training time was recorded to assess computational efficiency. All performance metrics were reported as the mean values along with their 95% Confidence Intervals (95% CI), providing a basis for comparing the stability and generalizability of different models.

In the first stage, the constructed individual-level feature matrix (30 features) was used to train models. After obtaining the optimal model, feature contributions were quantified

using the KernelSHAP algorithm. The SFSW strategy was then applied to weight the features. The weighted features were reintroduced into the optimal model to perform TD classification task. In the second stage, the weighted features obtained in the first stage were directly used to further assess the severity in the diagnosed children. The two-stage progressive diagnosis framework design closely aligned with the typical clinical process, that was, initially identifying the presence of TD, followed by assessing the severity, which improved the clinical applicability of the model. Compared to directly building a four-classification model (HC, mild, moderate, severe), we adopted the proposed strategy based on two considerations. First, the four-classification task involved significant class imbalance, which could lead to insufficient learning of the model for a few categories and reduced overall model accuracy. Second, a four-classification model needed to learn complex boundaries between multiple categories simultaneously, which was easy to introduce noise and was not conducive to the recognition of key biomarkers. Our two-stage strategy allowed the model to first identify the core features related to TD, and then further capture differential expression in severity levels. This design aligned more closely with the physiological mechanism of disease progression.

F. Hold-Out Set Validation

To further assess the generalization capability of the proposed model, we re-collected EEG data from 23 children (TD, n = 12 and HC, n = 11) as a hold-out set validation. The EEG feature extraction procedure was identical to that used during the training phase to ensure consistency. Data from the hold-out set were input into the final model to predict whether a child had TD and to further assess the severity for the child diagnosed with TD.

G. Model Interpretability

For the model interpretability, the SHAP framework was employed to interpret the model from both global and individual perspectives. Based on the SHAP value obtained, the influence of features on the EEG channel in diagnosing TD was identified and quantified. A positive SHAP value indicated that an increase in the feature value was associated with a higher likelihood of the target variable, while a negative value suggested the opposite relationship. The mean absolute SHAP value of each feature was used to determine the importance [35] and served as the weight for subsequent analysis. At the individual level, we used the SHAP waterfall plot to visualize the prediction process of the model for a specific child. The base value $E[f(x)]$ represented the average predicted probability of TD for all data without feature information. The $f(x)$ represented the probability that the model predicted the individual was TD. When $f(x)$ exceeded $E[f(x)]$, it was predicted that the child has TD, otherwise it was not. The SHAP waterfall plot explained the contribution of each input feature to the prediction possibility $f(x)$ for TD (positive push was red, negative push was blue).Additionally, a detailed stratified analysis of the age was conducted using the SHAP value to explore the specific impact of age on the diagnosis of TD at different stages of life.

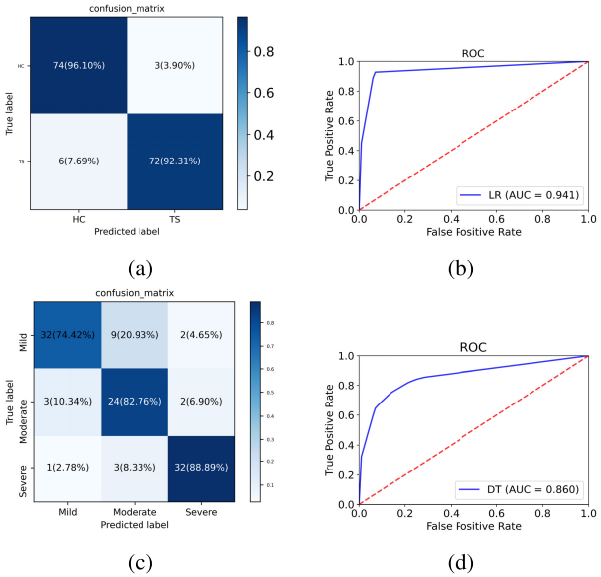


Fig. 3. Prediction results of the optimal model in the validation set. (a) Confusion matrix of TD diagnosis. (b) ROC curve of TD diagnosis. (c) Confusion matrix of severity. (d) ROC curve of severity.

IV. Results

A. Feature Selection Result

Gender distribution did not differ significantly between TD and HC groups (p = 0.117), despite the TD group having a higher male proportion (82.1%) consistent with clinical reports [36]. This balance was achieved through intentional gender matching to minimize EEG-related confounding, as gender may influence brain oscillatory patterns [37]. Age was included as a covariate due to significant difference (p<0.001). Each subject had 19 electrodes, from which 54 features were extracted. After statistical testing, the following variables were selected: MMAX, skewness, kurtosis, complexity, MNF, PSE, and wavelet time frequency entropy, with age selected as a covariate. The RFECV algorithm was applied to further refine the feature set. The results showed that the model achieved optimal classification performance when 30 features were retained. Therefore, these 30 features were selected for distinguishing children with TD from HC, as presented in Table II.

We put the features into the ML models to diagnose TD. The best LR model was selected and SHAP analysis was performed. The 13 key features (12 EEG features and age) were selected based on the mean average SHAP value, each contributing more than 3% to overall importance. These features were retained for subsequent model training. The weight was shown in Table II.

B. Prediction Result Based on the Proposed Framework

In the TD diagnosis task, the results of the five ML models were shown in Table III. LR model performed the best in the five validation sets. Subsequently, feature weighting was performed using the SFSW strategy, and the weighted features were reintroduced into the LR model for retraining. The results indicated that the average ACC improved to 94.2%, with a SEN of 92.4% and a SPE increasing to 96.1%. The confusion matrix and the ROC curve were presented in Figs. 3(a) and 3(b). On the hold-out set validation, the model achieved

TABLE II
SELECTED FEATURES USING RFECV AND THEIR WEIGHTS BASED ON SHAP ANALYSIS (29 EEG FEATURES AND AGE)

Channel	Feature	P-value	Mean |SHAP|	Weight
Fp1	skewness	0.02	1.02%	0
Fz	skewness	0.00	2.93%	0
F4	**PSE_γ**, **MMAX**, complexity	0.00, 0.00, 0.00	5.68%, 3.39%, 1.14%	**5.68%**, **3.39%**, 0
C3	**MNF_γ**, **complexity**	0.00, 0.00	7.43%, 3.53%	**7.43%**, **3.53%**
Cz	**complexity**	0.00	5.27%	**5.27%**
C4	**complexity**	0.00	5.49%	**5.49%**
P3	**MNF_β**, complexity	0.32, 0.00	10.87%, 1.84%	**10.87%**, 0
Pz	**complexity**	0.00	3.31%	**3.31%**
P4	complexity, skewness	0.00, 0.00	2.43%, 0.79%	0, 0
O1	**MMAX**, PSE_γ, skewness	0.00, 0.00, 0.03	5.55%, 1.84%, 1.27%	**5.55%**, 0, 0
O2	PSE_γ, MMAX, skewness, kurtosis	0.00, 0.00, 0.00, 0.00	2.57%, 2.17%, 1.45%, 1.41%	0, 0, 0, 0
F7	**kurtosis**, **skewness**, MNF_γ	0.00, 0.01, 0.00	6.90%, 3.72%, 0.73%	**6.90%**, **3.72%**, 0
F8	PSE_γ	0.00	1.18%	0
T3	**complexity**, PSE_γ, skewness	0.00, 0.00, 0.00	4.38%, 1.13%, 0.95%	**4.38%**, 0, 0
T5	complexity	0.02	1.18%	0
-	**age**	0.00	8.45%	**8.45%**

TABLE III
MEAN AND 95% CI OF THE PREDICTION RESULTS OF DIFFERENT MODELS ABOUT TD DIAGNOSIS ON THE VALIDATION SET (FIVE-FOLD CROSS-VALIDATION)

Model	F2-score	ACC	SEN	SPE	PRE	AUROC
LR (SFSW)	**0.930 (0.872, 0.988)**	**0.942 (0.906, 0.979)**	**0.924 (0.853, 0.995)**	**0.961 (0.929, 0.992)**	**0.962 (0.931, 0.992)**	**0.943 (0.906, 0.979)**
LR	0.925 (0.872, 0.979)	0.930 (0.893, 0.966)	0.924 (0.853, 0.995)	0.936 (0.869, 1.003)	0.942 (0.885, 0.999)	0.930 (0.894, 0.966)
LR (Original SHAP [35])	0.893 (0.803, 0.983)	0.910 (0.860, 0.960)	0.887 (0.775, 0.999)	0.934 (0.893, 0.976)	0.935 (0.896, 0.975)	0.910 (0.861, 0.960)
XGBoost	0.895 (0.841, 0.948)	0.890 (0.822, 0.959)	0.897 (0.846, 0.947)	0.884 (0.782, 0.987)	0.893 (0.800, 0.985)	0.890 (0.822, 0.959)
RF	0.896 (0.859, 0.933)	0.891 (0.842, 0.941)	0.898 (0.867, 0.928)	0.885 (0.814, 0.956)	0.890 (0.823, 0.958)	0.891 (0.842, 0.941)
KNN	0.820 (0.786, 0.853)	0.871 (0.854, 0.889)	0.795 (0.750, 0.840)	0.949 (0.903, 0.995)	0.944 (0.896, 0.992)	0.872 (0.854, 0.890)
DT	0.767 (0.744, 0.790)	0.762 (0.726, 0.797)	0.768 (0.732, 0.804)	0.754 (0.662, 0.847)	0.769 (0.696, 0.842)	0.761 (0.725, 0.797)
ResNet [11]	0.787 (0.668, 0.906)	0.807 (0.732, 0.882)	0.782 (0.643, 0.922)	0.833 (0.753, 0.914)	0.829 (0.768, 0.890)	0.808 (0.734, 0.881)
CNN [15]	0.816 (0.732, 0.900)	0.833 (0.776, 0.890)	0.809 (0.716, 0.902)	0.857 (0.835, 0.880)	0.849 (0.810, 0.888)	0.833 (0.776, 0.890)
DNN [16]	0.883 (0.824, 0.942)	0.884 (0.857, 0.910)	0.885 (0.804, 0.966)	0.883 (0.822, 0.945)	0.890 (0.844, 0.936)	0.884 (0.859, 0.909)

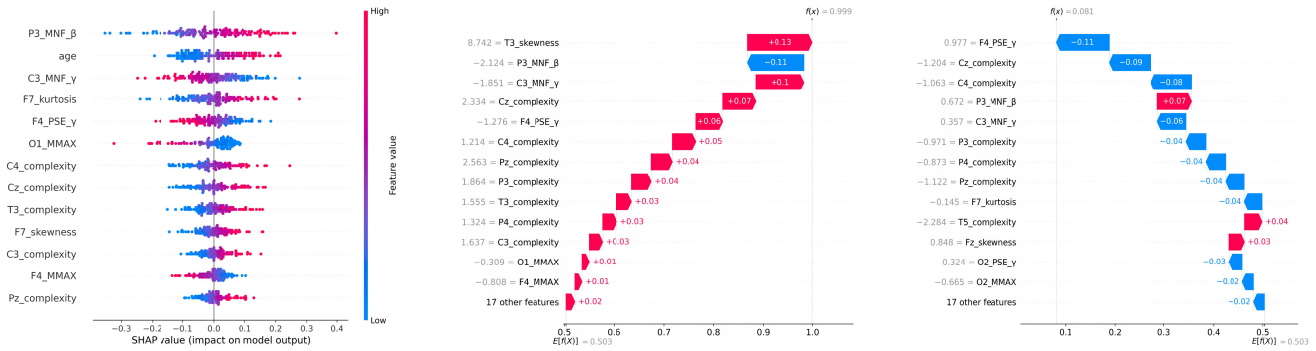
TABLE IV
MEAN AND 95% CI OF THE PREDICTION RESULTS OF DIFFERENT MODELS ABOUT SEVERITY ON THE VALIDATION SET (FIVE-FOLD CROSS-VALIDATION)

Model	F1-score	ACC	SEN	SPE	PRE	AUROC
DT with SFSW	**0.809 (0.670, 0.949)**	**0.815 (0.686, 0.945)**	**0.815 (0.686, 0.945)**	**0.899 (0.821, 0.976)**	**0.833 (0.693, 0.973)**	**0.862 (0.765, 0.958)**
DT (No SFSW)	0.584 (0.451, 0.718)	0.601 (0.493, 0.709)	0.601 (0.493, 0.709)	0.752 (0.654, 0.849)	0.591 (0.422, 0.760)	0.701 (0.620, 0.782)
DT (Original SHAP [35])	0.489 (0.298, 0.680)	0.501 (0.315, 0.688)	0.501 (0.315, 0.688)	0.678 (0.523, 0.833)	0.533 (0.297, 0.770)	0.626 (0.486, 0.766)
XGBoost	0.666 (0.597, 0.735)	0.675 (0.599, 0.751)	0.675 (0.599, 0.751)	0.820 (0.772, 0.868)	0.710 (0.598, 0.822)	0.802 (0.749, 0.856)
RF	0.693 (0.571, 0.815)	0.705 (0.591, 0.818)	0.705 (0.591, 0.818)	0.832 (0.759, 0.905)	0.726 (0.601, 0.850)	0.819 (0.710, 0.929)
KNN	0.625 (0.486, 0.764)	0.644 (0.516, 0.772)	0.644 (0.516, 0.772)	0.787 (0.689, 0.884)	0.675 (0.520, 0.830)	0.778 (0.706, 0.850)
LR (RBF kernel)	0.638 (0.501, 0.775)	0.649 (0.513, 0.784)	0.649 (0.513, 0.784)	0.790 (0.696, 0.884)	0.660 (0.519, 0.801)	0.798 (0.698, 0.897)
ResNet [11]	0.526 (0.390, 0.662)	0.582 (0.453, 0.712)	0.545 (0.380, 0.711)	0.719 (0.606, 0.833)	0.396 (0.153, 0.640)	0.784 (0.670, 0.897)
CNN [15]	0.492 (0.367, 0.617)	0.555 (0.443, 0.667)	0.487 (0.294, 0.680)	0.649 (0.453, 0.845)	0.449 (0.235, 0.663)	0.717 (0.657, 0.776)
DNN [16]	0.507 (0.369, 0.646)	0.572 (0.446, 0.699)	0.543 (0.404, 0.683)	0.653 (0.468, 0.838)	0.454 (0.240, 0.669)	0.687 (0.573, 0.801)

an ACC of 95.7%, with a SEN of 100% and a SPE of 90.9%. In the task of identifying the severity of TD, the classification performance of each ML model was shown in Table IV. DT model demonstrated the highest performance, achieving an average ACC of 81.5%, a SEN of 81.5%, and a SPE of 89.9%. Further details, including the confusion matrix and the ROC curve, were provided in Fig. 3(c) and Fig. 3(d). The DT model was selected as the optimal severity prediction model. The result of the hold-out set validation confirmed the robustness of the model, with an ACC of 83.3%, a SEN of 83.3%, and a SPE of 91.1%.

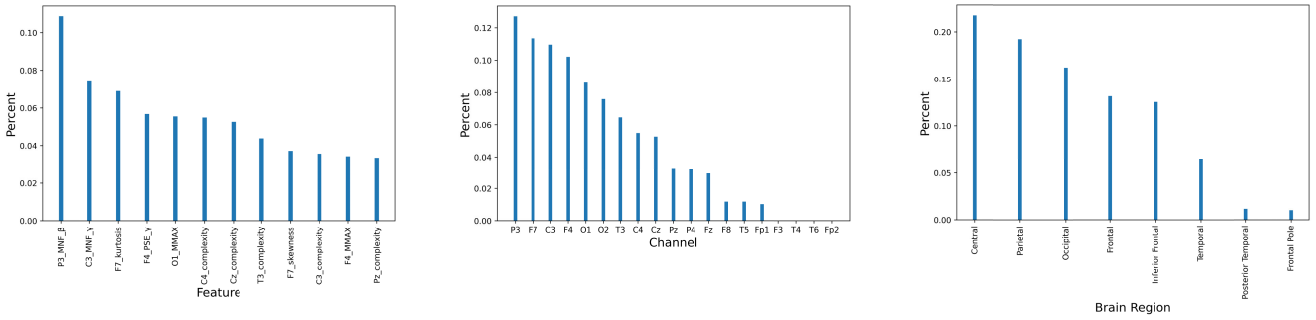
C. Interpretability Analysis Result

After identifying the LR model as the best classifier for TD, global interpretation at the feature level and local interpretation at the individual level were presented in Fig. 4. From the global perspective in Fig. 4(a), certain EEG features were strongly associated with TD. Specifically, a higher MNF in the P3 channel β band, a lower MNF in the C3 channel γ band, a higher kurtosis in the F7 channel, and a lower γ band PSE in the F4 channel made a diagnosis of TD more likely. The features of the P3, F7, C3 and F4 channels played a critical role in determining the presence of TD. Changes in the central, parietal, and occipital brain regions were particularly significant for diagnosing TD in children. In Fig. 4(b), the model predicted a 99.9% probability that the child had TD, aligning with the clinical diagnosis. The skewness of the T3 channel promoted the increase of this prediction probability. In Fig. 4(c), the predicted probability of TD for the child shown was only 8.1%, leading to a classification as a healthy subject. The PSE of the γ band in



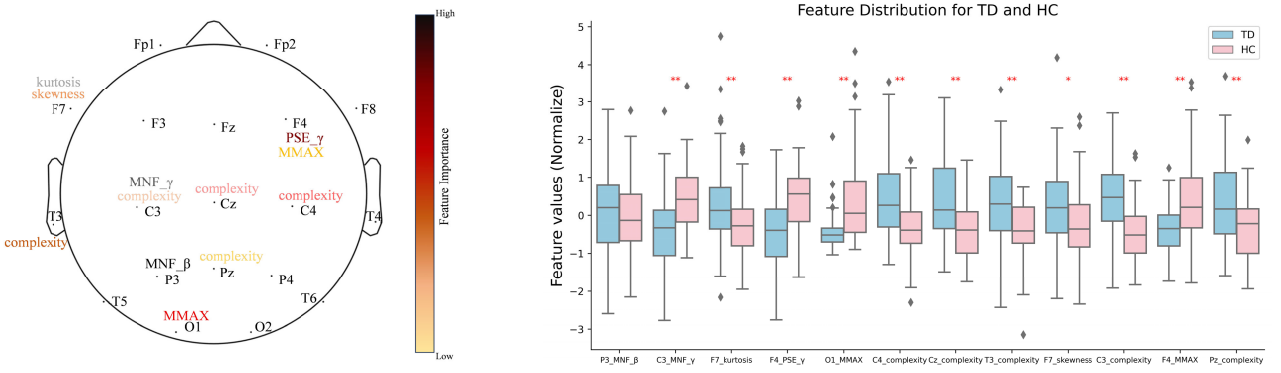
(a) (b) (c)

Fig. 4. Global interpretation at the feature level and local interpretation at the individual level were presented. (a) SHAP summary dot plot. The direction and strength of the influence of each feature on the model prediction, displayed in descending order. Features such as the higher MNF in the P3 channel β band, the older age, the lower MNF in the C3 channel γ band were associated with an increased likelihood of TD. (b) SHAP waterfall plot about the TD child. The model predicted a 99.9% probability that the child had TD. The skewness of the T3 channel promoted the increase of this prediction probability. (c) SHAP waterfall plot about the HC child. The predicted probability of TD for the child shown was only 8.1%. The PSE of the γ band in the F4 channel suppressed the prediction probability for TD.



(a) (b) (c)

Fig. 5. Quantification of importance based on the mean absolute SHAP value. (a) Feature importance. (b) Channel importance. (c) Brain region importance.



(a) (b)

Fig. 6. Positioning of features on the electrode and feature significance test plots. (a) Detailed layout of electrodes and feature importance plot (12 EEG features). Color variations were used to indicate the decreasing order of feature importance. (b) Significance test plot of different EEG characteristics of TD and HC (The importance of features decreased from left to right). ** denoted p<0.01, * denoted p<0.05.

the F4 channel suppressed the prediction probability for TD. Regarding age, older children were more likely to be correctly classified into TD based on EEG data, while younger children were more likely to be correctly classified as healthy. The result indicated that this boundary was roughly between 8 and 8.31 years of age. Additionally, the importance of features, calculated using the mean absolute SHAP value, was shown in Fig. 5(a). The significance of channels and brain regions were depicted in Figs. 5(b) and 5(c). We identified 13 key features and positioned electrodes on EEG features. The specific layout of the electrodes and the significance of the features were detailed in Fig. 6.

D. Comparison Experiment Result

We compared the direct use the feature set (30 features) screened by the RFECV method with the SFSW strategy.

TABLE V
SUMMARY OF PERFORMANCE COMPARISON WITH OTHER SIMILAR STUDIES

Authors (Year)	Number of subjects	Method used	Performance metrics (%)					
			ACC	SEN	SPE	PRE	AUROC	F-score
Duan et al. [7] (2020)	36 TS and 21 HC	Clustering coefficient, Characteristic path length, Global efficiency, Local efficiency and EEG	92.31	85.71	96.77	-	-	-
Greene et al. [8] (2016)	42 TS and 42 HC	RSFC and MRI	74.00	76.00	71.00	-	-	-
Liao et al. [9] (2016)	24 TS and 32 HC	VMHC and fMRI	91.00	88.00	94.00	-	-	-
Wen et al. [10] (2017)	44 TS and 48 HC	Nodal degree, Nodal efficiency, Nodal betweenness centrality and MRI	89.13	86.36	91.67	-	-	-
Wang et al. [11] (2024)	95 TD and 90 HC	ResNet and EEG	87.23	87.23	-	88.47	-	87.21
Proposed	90 TD and 88 HC	Time domain, Frequency domain, Time-frequency domain and EEG	**94.23**	**92.42**	**96.08**	**96.16**	**94.25**	**93.02**

The results were presented in Tables III and IV. The SFSW strategy proposed in TD diagnosis had improved the classification performance, with an average ACC and SPE increase of 1.2% and 2.5%, respectively. In the severity classification, this method improved the average ACC, SEN, and SPE by 21.4%, 21.4%, and 14.7%. To assess the performance difference between the proposed SFSW strategy and the original SHAP-based feature selection method [35], a comparative experiment was conducted under identical experimental conditions. In the TD diagnosis task, the control method achieved an average ACC of 91.0%, a SEN of 88.7%, and a SPE of 93.4%. In comparison, our method improved model performance, with an average ACC, a SEN, and a SPE reaching 94.2%, 92.4%, and 96.1%, respectively. At the same time, this method improved the average ACC, SEN, and SPE in the severity classification task by 31.4%, 31.4%, and 22.1%, respectively. These findings demonstrated the effectiveness of the SFSW strategy in enhancing both feature selection precision and classification accuracy.

This study further selected three recent state-of-the-art deep learning models that have been widely applied in recent years to EEG classification tasks, in order to more comprehensively evaluate the effectiveness of our proposed model. Specifically, we included ResNet [11], CNN [15], and DNN [16] as baseline methods. Among them, ResNet has been applied to TD diagnosis, while CNN and DNN have been used in the detection of epileptic seizure, making them highly relevant and appropriate for comparison. In our experiments, we faithfully reproduced the network architectures and parameter settings reported in the original studies. To ensure fairness and comparability, we adopted the same data preprocessing and feature extraction procedures across all models. Each model was trained independently on the same data for 100 epochs to avoid any performance bias due to undertraining or overfitting. Performance was evaluated using several metrics, including ACC, F-score, AUROC, and training time, to provide a comprehensive assessment of classification performance and computational efficiency. The experimental results were presented in Tables III and IV. The results indicated that the proposed model outperformed these comparison methods across all evaluation metrics. Notably, it demonstrated superior SEN and SPE, reflecting greater classification stability and generalization capability. This enabled more accurate identification of children with TD while reducing the likelihood of misclassification. The diagnostic framework developed in this study achieved higher accuracy and robustness in TD identification and showed promising potential for clinical application.

V. DISCUSSION

This study proposed a two-stage progressive diagnosis framework that integrated the individual-based feature-weighted integration module and the SFSW strategy to accurately identify and assess the severity of TD based on a relatively large EEG dataset (90 TD and 88 HC). This study was based on individual-level modeling, and the integration module effectively addressed the challenge caused by varying numbers of EEG valid segments across subjects, improving the robustness and generalizability of the model. The SFSW strategy further enhanced the model performance and interpretability through a refined feature weighting strategy. Table V presented the performance comparison between the proposed method and other related studies in the field of TD diagnosis. The results showed that the proposed framework outperformed existing methods in key metrics such as ACC and SEN. Moreover, by introducing a two-stage modeling method, this study enabled end-to-end analysis, from initial diagnosis to severity classification, thereby enhancing the clinical relevance and practical value of the model.

Compared with recent deep learning methods such as Resnet [11], CNN [15], and DNN [16], the method proposed in this study achieved better prediction performance. This advantage could be attributed to our two-stage feature selection and weighting strategy, which extracted more discriminative core features from the original EEG signals and captured the essential characteristics of TD with a relatively small number of features. Due to the low-dimensional nature of the extracted features, complex deep learning networks may struggle to fully utilize the available information, resulting in suboptimal performance. In contrast, our proposed model was better suited to handle such refined feature sets, allowing it to achieve stronger classification results.

Additionally, our proposed method offered greater interpretability. Our interpretability analysis revealed specific differences in the P3 channel β band and the C3 channel γ band between children with TD and HC, which

significantly impacted the diagnosis of TD. These findings were aligned with previous studies. For example, Loo et al. [38] reported that children with chronic tic disorder (CTD) exhibited increased γ band spectral power in the left dorsolateral prefrontal cortex, posterior cingulate, and additional motor area. Significant differences in the power spectrum of inferior parietal cortex β band were also observed. This suggested a strong connection between brain electrical activity in specific frequency bands and the pathological state of TD. Kurtosis and complexity, as two important statistical parameters, played a significant role in the diagnostic process. Previous studies had highlighted the predictive value of kurtosis in identifying the outcome of pediatric epilepsy seizures [39]. An increase in kurtosis in EEG data often reflected abnormal changes in brain electrical activity, potentially linked to heightened neuronal synchrony. Furthermore, the Hjorth complexity parameter had demonstrated strong diagnostic performance in children with Autism Spectrum Disorders (ASD) [40]. These findings suggested that both kurtosis and complexity may have valuable potential for application in the diagnosis of TD. This study also highlighted the importance of the P3 and F7 channels, as well as the central and parietal regions, in the diagnosis of TD. The central region played an important role in body movement and cognitive control, and the primary motor cortex (precentral gyrus) in this region regulated voluntary movements of the body [41]. Dysfunction in this area could impair motor control, resulting in involuntary twitches and abnormal movements in patients with TD [42]. The frontal and parietal cortex were key regions of the sensorimotor cortex, which regulated facial, oral, and laryngeal muscle activity. Children with TS had cortical thinning in these regions, correlated with increased facial tics [43]. Furthermore, our findings indicated that older children were more likely to be classified as TD, suggesting that age influenced the pathogenesis of TD. TD was a neurodevelopmental disorder. Early tic symptoms may be difficult to distinguish from involuntary movements associated with immature brain development. With increasing age, brain complexity increased and the persistence of TD further indicated functional coordination abnormalities within brain tissue, resulting in distinct EEG patterns compared to healthy individuals. Previous studies had shown that TD usually appeared in childhood (around 6 or 7 years) and peaked in severity between ages 8 and 12 [44]. This was consistent with our findings.

Despite the promising results of this study, there were several limitations to consider. First, models based solely on EEG features may not fully capture the clinical significance and characteristics that clinicians considered during evaluations. This limitation could result in the inability of the model to fully explain the complex pathophysiological mechanisms of TD. Future research should consider integrating additional biomarkers (such as behavioral data and genetic information) to provide a more holistic evaluation. Second, further cohort validation should be necessary. Different datasets may introduce biases and different characteristics, which could influence the performance of the model. We are collecting additional data from multiple centers to better assess the generalizability of the proposed model.

REFERENCES

[1] American Psychiatric Association, "Diagnostic and statistical manual of mental disorders: DSM-5T (5th ed.)," *Codas*, vol. 25, no. 2, p. 191, 2013.

[2] S. Wang, B. Ma, R. Li, and P. Xu, "Research progress of tic disorder in children," *Pediatr. Integr. Tradit. Chin. West. Med. China*, vol. 13, no. 4, pp. 297–301, Aug. 2021.

[3] A. Pérez-Vigil et al., "Association of Tourette syndrome and chronic tic disorders with objective indicators of educational attainment: A population-based sibling comparison study," *JAMA Neurol.*, vol. 75, no. 9, pp. 1098–1105, 2018.

[4] J. E. Swain and J. F. Leckman, "Tourette syndrome and tic disorders: Overview and practical guide to diagnosis and treatment," *Psychiatry (Edgmont)*, vol. 2, no. 7, pp. 26–36, Jul. 2005.

[5] C. E. Gill and K. Kompoliti, "Clinical features of Tourette syndrome," *J. Child Neurol.*, vol. 35, no. 2, pp. 166–174, Feb. 2020.

[6] R. J. Felling and H. S. Singer, "Neurobiology of Tourette syndrome: Current status and need for further investigation: Table 1," *J. Neurosci.*, vol. 31, no. 35, pp. 12387–12395, Aug. 2011.

[7] K. Duan et al., "Discrimination of Tourette syndrome based on the spatial patterns of the resting–state EEG network," *Brain Topography*, vol. 34, no. 1, pp. 78–87, Jan. 2021.

[8] D. J. Greene et al., "Multivariate pattern classification of pediatric Tourette syndrome using functional connectivity MRI," *Develop. Sci.*, vol. 19, no. 4, pp. 581–598, Jul. 2016.

[9] W. Liao, Y. Yu, H.-H. Miao, Y.-X. Feng, G.-J. Ji, and J.-H. Feng, "Interhemispheric intrinsic connectivity as a neuromarker for the diagnosis of boys with Tourette syndrome," *Mol. Neurobiol.*, vol. 54, no. 4, pp. 2781–2789, May 2017.

[10] H. Wen et al., "Multi-threshold white matter structural networks fusion for accurate diagnosis of Tourette syndrome children," *Proc. SPIE*, vol. 10134, pp. 436–448, Mar. 2017.

[11] C. Wang et al., "Deep learning-assisted non-invasive pediatric tic disorder diagnosis using EEG features extracted by residual neural networks," *J. Radiat. Res. Appl. Sci.*, vol. 17, no. 4, Dec. 2024, Art. no. 101151.

[12] P. Durongbhan et al., "A dementia classification framework using frequency and time-frequency features based on EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 826–835, May 2019.

[13] D. E. Hernández, L. Trujillo, E. Z-Flores, O. M. Villanueva, and O. Romo-Fewell, "Detecting epilepsy in EEG signals using time, frequency and time-frequency domain features," *Comput. Sci. Eng.-Theory Appl.*, vol. 143, pp. 167–182, Jan. 2018.

[14] S. K. Khare and U. R. Acharya, "An explainable and interpretable model for attention deficit hyperactivity disorder in children using EEG signals," *Comput. Biol. Med.*, vol. 155, Mar. 2023, Art. no. 106676.

[15] H. Kode, K. Elleithy, and L. Almazaydeh, "Epileptic seizure detection in EEG signals using machine learning and deep learning techniques," *IEEE Access*, vol. 12, pp. 80657–80668, 2024.

[16] A. Romney and V. Manian, "Comparison of frontal-temporal channels in epilepsy seizure prediction based on EEMD-ReliefF and DNN," *Comput.*, vol. 9, no. 4, p. 78, Sep. 2020.

[17] J. F. LECKMAN et al., "The Yale global tic severity scale: Initial testing of a clinician-rated scale of tic severity," *J. Amer. Acad. Child Adolescent Psychiatry*, vol. 28, no. 4, pp. 566–573, Jul. 1989.

[18] F. Perrin, J. Pernier, O. Bertrand, and J. F. Echallier, "Spherical splines for scalp potential and current density mapping," *Electroencephalogr. Clin. Neurophysiol.*, vol. 72, no. 2, pp. 184–187, Feb. 1989.

[19] J. Levman et al., "Hold-out validation for the assessment of stability and reliability of multivariable regression demonstrated with magnetic resonance imaging of patients with schizophrenia," *Int. J. Develop. Neurosci.*, vol. 81, no. 7, pp. 655–662, Nov. 2021.

[20] J. Ma et al., "Unleashing the strengths of unlabeled data in pan-cancer abdominal organ quantification: The FLARE22 challenge," 2023, *arXiv:2308.05862*.

[21] Y. Chen, R. Chang, and J. Guo, "Effects of data augmentation method borderline-smote on emotion recognition of EEG signals based on convolutional neural network," *IEEE Access*, vol. 9, pp. 47491–47502, 2021.

[22] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, Jul. 2014.

[23] M. J. A. M. van Putten, T. Kind, F. Visser, and V. Lagerburg, "Detecting temporal lobe seizures from scalp EEG recordings: A comparison of various features," *Clin. Neurophysiol.*, vol. 116, no. 10, pp. 2480–2489, Oct. 2005.

[24] M. Baygin, "An accurate automated schizophrenia detection using TQWT and statistical moment based feature extraction," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102777.

[25] N. A. Alzahab, M. Baldi, and L. Scalise, "Efficient feature selection for electroencephalogram-based authentication," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2021, pp. 1–6.

[26] V. S. Swisher et al., "Sleep disorders, sleep medication use, and predictors of sleep disturbance in children with persistent tic disorders," *Children's Health Care*, vol. 53, no. 1, pp. 23–40, Jan. 2024.

[27] S. Gudmundsson, T. P. Runarsson, S. Sigurdsson, G. Eiriksdottir, and K. Johnsen, "Reliability of quantitative EEG features," *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2162–2171, Oct. 2007.

[28] A. Phinyomark, S. Thongpanja, H. Hu, P. Phukpattaranont, and C. Limsakul, "The usefulness of mean and median frequencies in electromyography analysis," in *Computational Intelligence in Electromyography Analysis: A Perspective on Current Applications and Future Challenges*. Rijeka, Croatia: IntechOpen, Oct. 2012, pp. 195–220.

[29] A. Zhang, B. Yang, and L. Huang, "Feature extraction of EEG signals using power spectral entropy," in *Proc. Int. Conf. Biomed. Eng. Informat.*, May 2008, pp. 435–439.

[30] J. Wang and M. Wang, "Review of the emotional feature extraction and classification using EEG signals," *Cognit. Robot.*, vol. 1, pp. 29–40, Jan. 2021.

[31] P. Boonyakitanont, A. Lek-Uthai, K. Chomtho, and J. Songsiri, "A review of feature extraction and performance evaluation in epileptic seizure detection using EEG," *Biomed. Signal Process. Control*, vol. 57, Mar. 2020, Art. no. 101702.

[32] M. Murugappan, N. Ramachandran, and Y. Sazali, "Classification of human emotion from EEG using discrete wavelet transform," *J. Biomed. Sci. Eng.*, vol. 3, no. 4, pp. 390–396, 2010.

[33] K. Lweesy, N. Khasawneh, M. Fraiwan, H. Wenz, H. Dickhaus, and L. Fraiwan, "Classification of sleep stages using multi-wavelet time frequency entropy and LDA," *Methods Inf. Med.*, vol. 49, no. 3, pp. 230–237, 2010.

[34] A. Z. Mustaqim, S. Adi, Y. Pristyanto, and Y. Astuti, "The effect of recursive feature elimination with cross-validation (RFECV) feature selection algorithm toward classifier performance on credit card fraud detection," in *Proc. Int. Conf. Artif. Intell. Comput. Sci. Technol. (ICAICST)*, Jun. 2021, pp. 270–275.

[35] W. E. Marcilio and D. M. Eler, "From explanations to feature selection: Assessing SHAP values as feature selection mechanism," in *Proc. 33rd SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, 2020, pp. 340–347.

[36] B. Garcia-Delgar 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 Adolescent Psychiatry*, vol. 31, no. 10, pp. 1539–1548, Oct. 2022.

[37] T. Jochmann, M. S. Seibel, E. Jochmann, S. Khan, M. S. Hämäläinen, and J. Haueisen, "Sex-related patterns in the electroencephalogram and their relevance in machine learning classifiers," *Hum. Brain Mapping*, vol. 44, no. 14, pp. 4848–4858, Oct. 2023.

[38] S. K. Loo et al., "Neural activation and connectivity during cued eye blinks in chronic tic disorders," *NeuroImage, Clin.*, vol. 24, Jan. 2019, Art. no. 101956.

[39] T. Akiyama et al., "High kurtosis of intracranial electroencephalogram as a marker of ictogenicity in pediatric epilepsy surgery," *Clin. Neurophysiol.*, vol. 123, no. 1, pp. 93–99, Jan. 2012.

[40] Z. J. Peya, B. Zaman, M. A. H. Akhand, and N. Siddique, "Autism spectrum disorder detection from EEG through Hjorth parameters and classification using neural network," in *Proc. Int. Conf. Mach. Intell. Emerg. Technol.*, Jan. 2023, pp. 31–40.

[41] L. Q. Uddin, J. Kinnison, L. Pessoa, and M. L. Anderson, "Beyond the tripartite cognition–emotion–interoception model of the human insular cortex," *J. Cognit. Neurosci.*, vol. 26, no. 1, pp. 16–27, Jan. 2014.

[42] A. B. Bruce et al., "Altered frontal-mediated inhibition and white matter connectivity in pediatric chronic tic disorders," *Exp. Brain Res.*, vol. 239, no. 3, pp. 955–965, Mar. 2021.

[43] E. R. Sowell et al., "Thinning of sensorimotor cortices in children with Tourette syndrome," *Nature Neurosci.*, vol. 11, no. 6, pp. 637–639, Jun. 2008.

[44] J. F. Leckman et al., "Course of tic severity in Tourette syndrome: The first two decades," *Pediatrics*, vol. 102, no. 1, pp. 14–19, Jul. 1998.