A Dual-Channel Transformer-SHAP Framework for Early Detection and Interpretable Analysis of Youth Social Burnout from Multimodal Behavioral Data

Bingyan Yin
Publicity Department of the Party Committee, Changchun Normal University, Changchun 130032, China
E-mail: BingyanYinn@outlook.com

Keywords: social burnout prediction, transformer-SHAP model, interpretability SHAP, youth mental health, behavioral sequence analysis

Received: October 25, 2025

An efficient and interpretable warning model is urgently needed to address the prominent risk of social burnout among adolescents in high-intensity social media interactions. Traditional machine learning methods have insufficient modeling capabilities for temporal behavior features and lack interpretability in prediction. This study proposes a warning method that integrates Transformer and SHAP: a dual channel Transformer architecture is designed, where the main channel analyzes the high-order correlation of 15-dimensional user behavior sequences through stacked encoding layers and multi head attention, and the auxiliary channel uses a recursive feature pyramid structure to enhance state sensitivity; Furthermore, fatigue risk prediction can be achieved through a fully connected network. Simultaneously constructing a tree structure optimized SHAP calculation process, while reducing high-dimensional computational complexity, verifying the rationality of feature attribution through psychological theory, and using SHAP to preprocess abnormal samples, combined with variational autoencoder (VAE) to achieve deviation detection, forming an integrated mechanism of "warning traceability". Based on a desensitization set containing 6-month social data of 5892 adolescents, the robustness was verified through missing data and data change scenarios. The results showed that the model had an accuracy rate of 92.37% (improved by 8.5% and 12.2% compared to LSTM and random forest, respectively), an F1 value of 0.891, an early fatigue recall rate of 85.2%, a false positive rate of<7.3%, and a response time of 1.8 seconds; SHAP analysis confirmed that nighttime activity exceeding 2 hours (+0.32), weekly social decline exceeding 40% (+0.28), and negative emotional words exceeding 15% (+0.41) are the core features, with a combined contribution of 68.7%, providing quantitative evidence for targeted interventions.

Povzetek: Predstavljen je razložljiv model za zgodnje opozarjanje na socialno izgorelost pri mladostnikih na družbenih omrežjih z visoko točnostjo in boljšo sledljivostjo napovedi.

1 Introduction

While social media reconstructs the social paradigm of contemporary youth groups, it also derives complex psychological load phenomena [1]. Social burnout, as a comprehensive representation of emotional exhaustion and behavioural avoidance caused by users' continuous exposure to a high-intensity virtual interaction environment, has become a key risk factor threatening the mental health of young people [2, 3]. With the diversification of social media platform interaction modes and the quantitative development of user behavior trajectories, the traditional burnout identification method based on static feature analysis faces significant limitations: First, the linear model is difficult to capture the long-distance time series dependence inherent in user interaction behavior, and the dynamic modeling ability of the gradual evolution process of burnout is insufficient; Secondly, existing studies mostly rely on shallow statistical indicators and ignore the mapping mechanism of deep semantic features to psychological state; Third, the vast majority of early warning models present "black box" characteristics, and their decision-making logic lacks traceable causal explanation, which makes it difficult for the attribution analysis of key risk factors to support effective psychological intervention strategies [4, 5]. The lack of transparency in this model not only limits its application value in clinical practice but also hinders cross-validation between behavioural psychology theory and the computational model [6].

In the field of time series pattern modelling, the Transformer architecture, based on the self-attention mechanism, has demonstrated a strong analytical ability for capturing long-range dependencies in recent years [7, 8]. It dynamically assigns feature weights through a multi-head attention mechanism and can unbiasedly capture non-fixed periodic behaviour patterns in users' social trajectories, such as the alternating transition between intermittent high activity and persistent low participation states. Such phenomena are easily

weakened due to gradient dispersion problems in traditional Markov models or recurrent neural networks. Especially when dealing with heterogeneous behaviour sequences, the parallel feature extraction capability of the Transformer can effectively prevent information attenuation in time series modelling and provide a technical basis for constructing an end-to-end social burnout evolution path. However, it should be noted that the complex, nonlinear calculation process inherent in this architecture makes it challenging to explicitly quantify the contribution of each feature. This explanatory obstacle seriously restricts its application in high-risk decision-making scenarios such as mental health risk assessment.

The development of explainable artificial intelligence (XAI) technology offers a new approach to resolving the aforementioned contradictions. The SHAP (Shapley Additive exPlanations) framework constructs a unified measurement system of characteristic contribution based on conditional expectation through the concept of alliance game in game theory [9-13]. Unlike the sample perturbation sensitivity of the local approximation method, the SHAP value has global consistency guaranteed by mathematical axioms and can generate attribution explanations for individual predictions and population distributions while maintaining model prediction accuracy [14]. When applied to predictive modeling of youth social burnout, this method can achieve dual goals at the same time: at the micro level, accurately locate risk triggering nodes in specific user behavior sequences; At the macro level, it reveals the stable mapping law of cross-group behavior patterns and psychological states, thus building an explanatory bridge between data-driven models and psychopathological theories [15].

Aiming to address the unique complexity of social burnout early warning among youth groups, this study proposes an innovative method that integrates Transformer time series modelling and the SHAP interpretability framework. This method innovatively designs a dual-channel architecture: the main channel uses a stacked Transformer-SHAP coding layer to perform deep representation learning of users' cross-platform behaviour sequences, and analyzes high-order nonlinear correlations between behaviours through multi-head attention mechanisms; The auxiliary pathway introduces a recursive feature pyramid structure, and hierarchically aggregates multi-scale behaviour features to enhance the sensitivity of state changes. At the level of model decision interpretation, the SHAP value calculation process based on tree structure optimisation is constructed. A behavioural feature grouping sampling strategy is designed to reduce the computational complexity of high-dimensional space. Additionally, a feature attribution verification mechanism is developed in combination with the stress-coping theoretical framework in social psychology. This integrated solution of "accurate early warning-causal traceability" not only breaks through the technical barriers between dynamic behaviour modelling and result interpretation in traditional models, but also establishes a collaborative verification mechanism between machine decision-making and human cognition, laying a methodological foundation for building a reliable social mental health assessment paradigm.

By revealing the dynamic correlation between adolescent social burnout and multimodal time series behavior, the theoretical foundation of affective computing has been strengthened. Social burnout manifests as a temporal evolution of behavioral patterns, such as a sustained decrease in interaction frequency or disrupted nighttime activities. Transformer based models can capture such vertical behavioral dynamics, thereby extending emotional computing from static recognition to process analysis with temporal interpretability. Combining SHAP technology, the model can not only predict risks, but also reveal the contribution of specific behavioral characteristics in the temporal dimension, providing causal insights into the mechanism of fatigue formation.

The study constructed a closed-loop mechanism from social burnout detection to personalized intervention. The key features identified by SHAP can directly trigger graded intervention strategies in digital health platforms, such as pushing low stress social activity suggestions, matching peer support groups, or activating digital curfew tools. In platform deployment, it is necessary to ensure scalability through modular design, while introducing dynamic fairness constraints to continuously monitor and mitigate model bias, ensuring the universality and practical utility of the system among different youth groups.

This study aims to address a core issue: how to use multimodal behavioral data to achieve early, accurate, and interpretable risk prediction of adolescent social burnout. We will specify the prediction targets as low, medium, and high risk levels, and evaluate performance using F1 scores and early recall rates. We will use SHAP attribution consistency and psychological validity to assess interpretability. The study proposes three testable hypotheses: H1) The multimodal Transformer model significantly outperforms traditional temporal models in terms of prediction accuracy; H2) SHAP framework can stably identify key features with psychological significance; H3) The model exhibits better robustness in noisy and missing data scenarios.

2 Theoretical basis of collaborative interpretation between transformer and SHAP

2.1. Core mechanism of transformer in social sequence modeling

The transformer network is a neural architecture based on the self-attention mechanism, which processes sequence information through self-attention and addresses the training difficulty problem of traditional networks in machine translation tasks [16]. The Transformer model is

shown in Figure 1. The model is fast to train, excels at handling long-distance dependencies, and its operation is based on the encoder-decoder structure [17]. Through the self-attention mechanism, it can process sequence data in parallel and identify the connection between data elements, which is widely used in natural language processing and other fields.

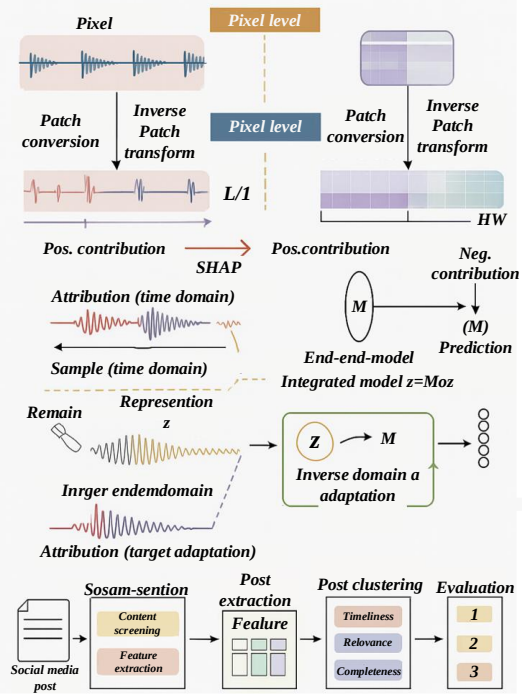

Figure 1: Transformer architecture diagram

The Transformer architecture completes sequence transformation through phased learning, including symbol-to-vector transformation, addition of positional features, and attention mechanism in parallel computing mode [18, 19]. It constructs a dynamic correlation matrix, adjusts feature correlations, synchronously extracts heterogeneous relationship features, and forms a compound semantic representation through linear mapping. The feedforward enhancement module employs a double linear transformation and the ReLU activation function to achieve deep, nonlinear reconstruction of local features.

The progressive sequence generation strategy is applied to the decoding system. The adaptive embedding layer initialises the target sequence, and the timing calibration module adds the position attribute to it. The autoregressive constraint component utilises a triangular masking matrix to ensure that the prediction is based solely on historical data, thereby preventing information leakage. The cross-modal attention bridging module establishes feature alignment to realise the retrieval and fusion of encoded features. The hierarchical feedforward network refines features, and the output layer uses a linear transformation and the Softmax function to generate a probability distribution. This fully connected attention mode architecture breaks through the limitation of sequence length and has significant advantages over traditional cyclic networks in modelling long-distance dependencies and parallel computing efficiency [20-23].

The Transformer architecture excels in processing time series data, and its parallel computing capability is enhanced by the dynamic weight allocation mechanism, which overcomes the limitations of traditional recursive networks and enables the simultaneous interaction of sequence elements [24]. The creation of adaptive association networks independently identifies nonlinear relationships through multi-dimensional feature decoupling strategies, which transcends the limitations of traditional monitoring models [25]. The introduction of the position coding module endows time series data with spatial arrangement awareness, ensuring the accurate spatio-temporal correspondence of data streams. The visual mapping of attention weights enhances the interpretability of the model, allowing for the tracing of the model's focus on key risk factors [26].

When implementing the architecture, it encounters limitations, the computational complexity increases nonlinearly with time, and the consumption of hardware resources increases. The model relies on a large amount of training data and is prone to misjudgment under special geological conditions. Although position coding supplements time series information, there is a lag in modeling dynamic relationships. The global attention mechanism may weaken the continuity feature and lead to the blind spot of transient abnormal signal detection.

Convert each word or token of the input sequence into an embedding vector, as in Equation (1).

$$E = W_e x \quad (1)$$

W_e is the embedding matrix, x is the one-hot encoding vector of the input word, and E is the obtained embedding vector. Position coding is introduced to allow the model to understand the position information of words, as shown in formulas (2)-(3):

$$PE_{(pos,2i)} = sin(\frac{pos}{10000^{2i/d_{model}}}) \quad (2)$$

$$PE_{(pos,2i+1)} = cos(\frac{pos}{10000^{2i/d_{model}}}) \quad (3)$$

Position i represents the position of the word in the sequence, d_{model} is the model dimension, and pos is the dimension index of the position encoding vector. The core innovation of the model lies in the hierarchical attention calculation mechanism. Multi-head attention is composed of multiple independent attention heads, and the calculation results are spliced and linearly transformed. See Equation (4) for the definition of basic unit scaling dot product attention.

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (4)$$

The query matrix Q, the key matrix K and the value matrix V are obtained by linear transformation of the input features. $\sqrt{d_k}$ is used as a scaling factor to ensure that the gradient propagation is stable, and the *softmax* function generates a standardized attention weight distribution. See (5) for the extended formula of multi-head attention.

$$MultiHead = Concat(head_1,\ldots,head_h)W^0 \quad (5)$$

The model can capture different feature patterns in h subspace, and each attention head $head_i$ is shown in Equation (6):

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (6)$$

$W^Q{}_i$, $W^K{}_i$, $W^V{}_i$ represent the linear transformation matrix of the *i-th* head, where W^o is the result of multi-head attention mechanism processing. These matrices decouple features by linear transformation and fuse subspace features. The feedforward network module enhances the nonlinear expression with two layers of linear transformation, and each sublayer independently performs the same linear transformation and activation function on the vector at each position, as shown in Equation (7).

$$FFN(x) = max(0, xW_1 + b_1)W_2 + b_2 \quad (7)$$

W_1 and W_2 are the weight matrices and b_1 and b_2 are the bias vectors. The architecture constructs high-dimensional decision boundaries through dimensional expansion and ReLU activation functions, and combines position-level independent computing and attention mechanisms to achieve functional complementarity. Residual concatenation and layer normalization are key to training stability, and they are applied after each sublayer, as shown in Equation (8).

$$x = LayerNorm(x + Sublayer(x)) \quad (8)$$

The output layer serves as the final stage of the decoder, and the probability distribution of the output is obtained by linear transformation and *softmax* function, as shown in equation (9).

$$y = softmax(W_o h + b_o) \quad (9)$$

h output vector of decoder; W_0 and b_0 are the weight and bias parameters of the output layer, respectively.

2.2. Interpretability theory of SHAP values

In the interpretability discussion of predictive models, linear models and decision trees are easy to understand because of their transparency. Most machine learning models use feature importance scores to demonstrate explanation, but this approach is not explicit enough about how features affect prediction results. The SHAP model interpreter adopts the Shapley value principle based on game theory, providing a general interpretation method. The Shapley value is used to fairly distribute the benefits or costs in cooperative games, and evaluate the contribution degree of each feature. The SHAP value represents the assigned value of each feature in a particular sample.

The *i-th* sample is denoted X_i, its *j-th* attribute is X_{ij}, the model prediction result is y_i, and the model benchmark value is y. The SHAP values follow equation (10):

$$y_i = y_{base} + f(X_{i_1}) + f(X_{i_2}) + \cdots + f(X_{i_k}) \quad (10)$$

$f(X_{ij})$ represents the SHAP value of the feature X_{ij}, i.e. its contribution to the predicted value y_i. A value of $f(X_{ij})$ greater than 0 indicates that the feature has a positive effect on the model, and less than 0 indicates a negative effect. The SHAP method helps decision makers understand the model operation and decision-making mechanism by demonstrating the influence of features on the results, and is especially suitable for occasions that require high transparency and interpretability of the model.

Although Transformers can effectively capture long-term behavioral dependencies, their black box nature limits their practical applications. KernelSHAP uses local approximation to transform complex temporal dynamics into stable feature importance scores, thereby clearly revealing the contribution of each behavioral feature to burnout risk in the temporal dimension.

3 Construction of multimodal Transformer-SHAP hybrid early warning model

3.1. Multimodal transformer-SHAP hybrid architecture

The time series data is used as input to the Transformer model after sliding sampling and position coding within a fixed-width time window. During the training phase, the model utilizes the data in the current time window to predict the value of the target variable at the next time step, employing the mean squared error (MSE) as the loss function for optimization.

After the training is completed, the model realizes multi-step prediction through an iterative prediction mechanism: when predicting the future value of a variable at the starting time, the current window data is first input to obtain the first-step prediction value; The predicted value is then added to the window (while removing the oldest data points in the window) to form a new input window, and the next value is predicted accordingly; This process is repeated until the desired prediction sequence is obtained. During the test phase, univariate multi-step prediction is performed using this iterative mechanism. A multivariate multi-step prediction system is constructed by independently training a

prediction model for each variable in the system and combining its outputs.

For risk detection, kernel density estimation (KDE) is used to calculate the confidence interval of the system's prediction error in the normal state, serving as the dynamic threshold for the system's prediction error. When the deviation between the actual observed value and the corresponding predicted value exceeds the threshold, it is judged that the system is abnormal.

In the model, the Transformer-SHAP architecture is based on the Transformer encoder, embedded with a SHAP value calculation module. The encoder extracts the sequence features of adolescent social behavior, and the SHAP module quantifies the contribution of each feature to the warning results in real time. The inter layer attention weight and SHAP value linkage mechanism are used to ensure the timeliness of feature importance interpretation; The SHAP integration steps are as follows: block processing of the feature vectors output by the Transformer, generation of single sample feature contribution values through the SHAP KernelExplainer, introduction of sample weight coefficients based on social data integrity settings for weighted aggregation of contribution values, output of global feature importance ranking and single sample interpretation reports, and standardization throughout the process to ensure consistency of results; The VAE model adopts an "encoder decoder" structure, where the encoder compresses high-dimensional social features such as interaction frequency and emotional text into low dimensional hidden vectors. The decoder reconstructs the input features, and during training, the goal is to minimize reconstruction error and KL divergence. The hidden vector distribution is used to model social behavior patterns, which not only assists the Transformer in filtering key warning features, but also completes abnormal social data to improve model robustness.

To explain the anomalies predicted and detected by the model, the moving window data are normalised and preprocessed using the SHAP technique. Subsequently, a variational autoencoder model based on a bidirectional long short-term memory network is used to detect the instability bias in the preprocessed data. The complete process of integrating Transformer prediction and SHAP interpretation/detection is shown in Figure 2.

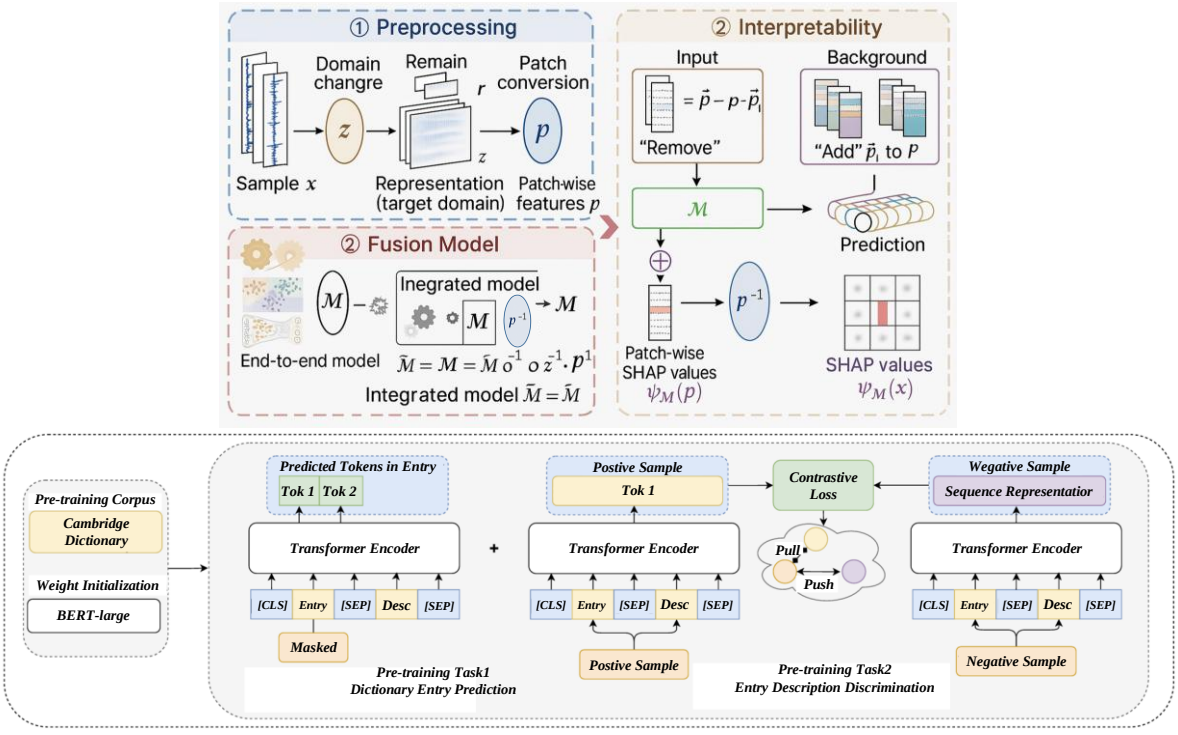

Figure 2: Multimodal Transformer-SHAP hybrid architecture

When developing the VAE risk detection model, the fault score δ_1 is obtained by calculating the KL divergence and reconstruction loss. An upper normal state limit λ was set from the validation dataset using kernel density estimates, equivalent to a 99.99% confidence interval. In the testing stage, the test data set is standardized by SHAP method and input into the trained Transformer-SHAP model to obtain the fault score δ. δ greater than 2 indicates that there is a fault, and less than or equal to 2 indicates normal.

Kernel density estimation (KDE) is a non-parametric technique for estimating probability density functions. It spreads the influence of data points to the surrounding area through kernel functions (such as Gaussian kernels) to generate smooth density estimates. KDE does not require a priori assumptions and is suitable for a variety of data sets. By adjusting parameters such as kernel function and bandwidth, the accuracy and smoothness of estimation can be improved. KDE is widely used in chemical process risk detection to identify

anomalies by calculating fault thresholds. In this study, the fault threshold of the fault warning model is calculated based on the KDE of normal samples in the verification dataset, while the threshold of the risk detection model is calculated based on the KDE of the loss value of the Transformer-SHAP model of normal samples.

Assume that $l_1, l_2..., l_n$ independent and identically distributed samples are extracted from the unknown distribution function $K()$, where l_i represents the loss of the validation dataset. The estimated probability density function $\hat{f}_h(x)$ is defined by Equation (11).

$$\hat{f}_h(x) = \frac{1}{hN}\sum_{i=1}^{N} K\left(\frac{x-l_i}{h}\right) \quad (11)$$

In the expression, $K(\cdot)$ is the kernel function, h is the bandwidth or smoothing parameter, and N is the total number of samples. To guarantee that sub $\hat{f}_h(x)$ is a reasonable probability density function, $K(\cdot)$ must satisfy the conditions in formula (12).

$$K(x) \geq 0, \int_{-\infty}^{+\infty} K(x)\mathrm{d}x = 1 \quad (12)$$

In this paper, the Gaussian kernel function is used as K, and the bandwidth h is optimized by minimizing the mean square error integral function. See formulas (13) to (14) for the calculation steps.

$$K(x) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}x^2\right) \quad (13)$$

$$h = 0.9 \cdot min\left(std(X), \frac{IQR(X)}{1.34}\right) \cdot N^{-1/5} \quad (14)$$

Where $std(X)$ is the standard deviation of the loss value and $IQR(X)$ is the interquartile range of the loss value.

Study have developed a data preprocessing pipeline for the system, with a focus on addressing the issues of encoding categorical variables and preserving temporal context. For categorical variables, ordered features are mapped numerically, unordered features are encoded using one hot encoding, and high cardinality features are processed through target encoding. To explicitly construct temporal context, the pipeline will create delay features with a lag of 1-7 days for key behavioral indicators and synchronously generate 7-day rolling window features, thereby transforming static data points into time series containing dynamic evolution information.

Based on the SBI-Y (Adolescent Social Burnout Inventory) scale score, the social burnout label is defined as follows: participants with a total score ≥ 85th percentile are classified as the "high burnout" group, those with a total score ≤ 15th percentile are classified as the "low burnout" group, and those in the middle are classified as the "medium risk" group. To ensure consistency between labels and behavioral performance, two-stage calibration is used: first, statistical differences in key behavioral indicators such as social withdrawal and nighttime activity are calculated for each group, and samples with significant contradictions between SBI Y scores and behavioral patterns are manually reviewed; Subsequently, k-means clustering was used to verify the significant consistency between behavior driven grouping and scale labels (ARI ≥ 0.65), ultimately forming a gold standard label set calibrated with behavioral data.

3.2 Interpretability-guided intervention strategies

Regarding the warning of adolescent social burnout, the accuracy of traditional Transformer-SHAP warning methods may decrease due to the tendency of adolescent social behavior to drift over time. To this end, an adaptive learning mechanism is integrated into the method, which captures behavior drift characteristics in real time and dynamically updates model parameters, combined with interpretable SHAP values to accurately locate key warning factors, thereby improving the timeliness and reliability of the warning method.Divide user groups based on adolescent gender and social usage frequency, and divide monitoring time periods on a weekly basis. By calculating the volatility of indicators within the group and analyzing the correlation between indicators between groups, verify the stability and consistency of the indicator in different user groups and monitoring time periods, and further strengthen the reliability of characteristic indicators in warning methods.

In the model output, prediction uncertainty and confidence interval analysis aim to quantify the reliability of the model prediction results. Capture the uncertainty information associated with social behavior characteristics through Transformer-SHAP, and combine the interpretability advantage of SHAP values to construct confidence intervals to define the credibility range of different prediction results; By clarifying the fluctuation boundary of the model for predicting the risk of social burnout in adolescents, and using interval analysis to assist in judging the practical value of the prediction results, a more rigorous model output basis is provided for subsequent intervention decisions, reducing decision bias caused by fuzzy predictions.

The definition of adolescent social burnout operation is: the 12-18 age group continuously exhibits a comprehensive state of significantly reduced social participation willingness, emotional exhaustion in social processes, and weakened social efficacy in social interactions, and these characteristics can be quantitatively captured through social behavior logs, emotional self-assessment scales, and interaction feedback data, serving as the core observation indicators for model warning.

The core value of the youth social burnout early warning model proposed in this study lies not only in accurately predicting at-risk individuals, but also in providing strong logical support and insight into

subsequent precise interventions through interpretable SHAP value analysis. Based on the Transformer architecture, the model captures the deep patterns and context dependencies in the interaction sequence of social platforms, predicts the individual burnout risk tendency, and further uses the SHAP method to analyze the internal logic of the model decision-making deeply, and quantifies the positive or negative contribution of each specific interaction feature (such as the intensity of high-frequency interaction in a specific period, the deep involvement of specific topic content, the density of negative emotional word clusters in expression, etc.) to the prediction results of specific individuals. This quantitative interpretability transforms abstract burnout risk predictions into concrete and actionable insights, such as targeting the excessive emotional involvement of high-risk individuals in specific topic discussions, or the frequency of certain late-night interactions as key negative drivers, thus enabling the formulation of interventions to go beyond the general predictions of "black box" models and directly anchor the most influential and highly individualized behavioral and psychological patterns.

Interpretability insights directly drive the construction of hierarchical, personalized intervention strategies. The identified key drivers provide clear targets for customized intervention programs: aiming at the risk of burnout caused by excessive involvement in specific topics, time-limited participation suggestions for topics based on cognitive reconstruction or accurate push of similar stress relief resources can be designed; Temporal factors such as high-frequency late-night interaction trigger work and rest management tools and behavioral intervention prompts. This intervention design, which reveals the individual-specific risk path map through the SHAP value, ensures that the strategy is no longer a generalized recommendation based on the average risk value of the group, but a precise intervention highly adapted to the individual's unique risk pattern and development trajectory. Therefore, interpretability analysis, as a bridge, not only improves the transparency and credibility of model predictions but also transforms the prediction results into implementable and evaluable action blueprints, significantly enhancing the active intervention of social and psychological risks based on artificial intelligence predictions. Effectiveness and efficiency of the system.

3.3 Data architecture

Hyperparameter adjustment is divided into three stages: first, fix the SHAP related parameters, and use grid search to optimize the hidden dimensions (64-256), number of attention heads (2-8), and learning rate (1e-5-1e-3) of the Transformer pre training layer; Freeze the pre training parameters again and fine tune the batch size (16-64) and iteration count (50-200) of the classification layer; Finally, based on the accuracy of the validation set warning and the stability of SHAP values (through consistency testing of feature sorting), the optimal parameter combination is selected.

Variational Autoencoder (VAE) plays a dual role in anomaly detection and feature preprocessing in this framework. The core mechanism is to identify potential abnormal patterns by comparing the differences between the original behavior sequence and the reconstructed sequence, and adaptively fuse the reconstructed regularized sequence with the original input to generate denoised enhanced features. This fusion feature is fed as an optimized input to the subsequent dual channel Transformer, effectively improving the model's robustness to noisy data and forming a collaborative analysis loop of "detection reconstruction enhancement".

The definition of social burnout labels adopts a dual track method of "scale anchoring+data validation": based on the scores of emotional exhaustion, social distancing, and reduced efficacy dimensions in the SBI-Y, samples with a total score ≥ the critical value of the scale are marked as "high burnout risk", and the rest are marked as "low risk". At the same time, label calibration is carried out by combining social behavior data to improve the accuracy of the definition. The ethical protocol for data usage strictly follows: 1 The principle of informed consent is to explain in writing to the adolescent and their guardian the purpose of the data and privacy protection measures, and obtain double signature confirmation; 2. Data anonymization processing, removing identification information such as name and student ID, using anonymous storage and encrypted transmission; 3. The principle of limited use is that the data is only used for training and verifying warning models, and leakage or secondary use is prohibited; 4. Rights protection clauses allow participants to apply for data deletion at any time, regularly disclose the progress of data use, and accept supervision from the ethics committee.

This dataset covers the social behavior and psychological data of 5892 adolescents aged 12-19 (mean age 15.8, gender balanced, covering 12 provinces and cities across the country) from January to June, and is used to construct a social burnout warning model. The collection adopts a mixed method of "automatic collection of platform logs (minute level, including interaction, duration, and content characteristics)+monthly stratified sampling scale survey (with the consent of users and guardians, scoring ≥ 45 points to mark high risk)" to form a complete sample. Preprocessing consists of four steps: cleaning (completing/removing 87 missing values, correcting outliers, removing 12 duplicates, and retaining 5805 valid samples); Feature standardization (z-Score) and encoding (hot encoding); Screening (Pearson correlation coefficient retains 28 features, VIF removes redundancy and retains 22 core features); Divide the training (3870 records from January to April), validation (965 records from May), and testing set (970 records from June) by time, and align monthly labels with weekly aggregated behavioral data.

4 Experiment and results analysis

4.1. Experimental setup

The study used a multimodal behavior dataset of 5892 adolescents aged 12-19 from 12 provinces in China (with an average age of 15.8 years and gender balance), which included matching records of social platform logs (minute level interaction frequency, duration, and content characteristics) and monthly psychological scales (SBI-Y) for six consecutive months. After ethical review, all data underwent triple anonymization (removal of ID information, encrypted transmission, and dynamic anonymization), and obtained informed consent from both participants and guardians. The dataset covers 15 dimensional behavioral sequence features: 1) daily active duration; 2) Nighttime usage frequency; 3) Number of cross platform switches; 4) Social response delay; 5) Text sentiment polarity; 6) Like/comment ratio; 7) The proportion of actively initiating interactions; 8) Density of emoji usage; 9) Breadth of topic participation; 10) Change rate of friendship relationships; 11) Screen unlocking frequency; 12) Activity level after midnight; 13) Weekly interaction decay rate; 14) The proportion of negative vocabulary; 15) Reading completion rate. Through SHAP analysis verification, the core risk feature group (with a joint contribution of 68.7%) consists of nighttime activity exceeding 2 hours (+0.32), weekly interaction decreasing by 40% (+0.28), and negative emotions exceeding 15% (+0.41), providing quantitative basis for constructing graded intervention strategies.

This research model performed well in 12 Chinese city datasets, but cross-cultural generalization faces three challenges: at the cultural level, collectivism and social structural differences may affect the manifestation of social avoidance behavior; At the language level, Chinese specific emotional vocabulary (such as "emo") is difficult to directly transfer to other language systems; At the platform level, there are ecological differences between the interaction mode of WeChat/Weibo and international platforms such as Twitter. Through a hierarchical generalization framework, the bottom layer retains cross-cultural common features, the middle layer adopts domain adaptation techniques, and the top layer introduces meta learning mechanisms to achieve robust cross regional transfer.

4.2. Experiment

To test the robustness of the adolescent social burnout warning model integrated with Transformer-SHAP under data loss and change conditions, the experiment will be conducted in three stages: firstly, based on the original adolescent social behavior dataset (including social frequency, emotional feedback, and other features), a missing data sample set will be constructed through random loss (10% -50% ratio), feature related loss (such as deliberately removing key features such as "social interaction duration"), and temporal segment loss (simulating intermittent social recording scenarios), and simulated data change scenarios will be simulated through feature distribution shift (such as social frequency mean ± 30% fluctuation), new interference features (such as irrelevant entertainment behavior data), and dynamic sample size reduction (from 100% to 30%); Secondly, train the model 10 times in each scenario, record performance indicators such as warning accuracy, F1 score, AUC, and track the stability of SHAP values (such as the coincidence rate of Top 10 important features and the fluctuation range of feature importance ranking); Finally, comparing the differences in indicators between normal data and abnormal scenarios, if the performance degradation is ≤ 15% and the SHAP interpretation logic remains consistent (core features remain unchanged), the robustness of the model is verified. Otherwise, the attention mechanism and missing value processing module of the Transformer are optimized by analyzing the SHAP value offset point.To rigorously evaluate the adolescent social burnout warning method that integrates Transformer-SHAP, multiple classic models were selected for baseline comparison and performance differences were measured from multiple dimensions. The ablation study gradually removed the core components of the model to analyze the roles of each component, in order to comprehensively verify the performance and effectiveness of the method.

Figure 3 shows that in the experiment using real adolescent social burnout data correction, the optimal number of clusters for the target analysis data is 4. Therefore, in the subsequent analysis process, the point with a clustering number of 4 on the horizontal axis is selected as the reference point.

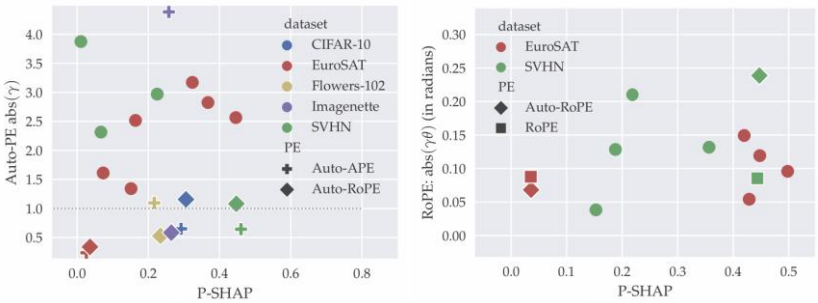

Figure 3: Clustering effect diagram

Figure 4 has showed the comparison of effects based on different clustering algorithms. By evaluating the contributions of each module in the multimodal Transformer network through ablation experiments. After sequentially removing the multimodal feature fusion, SHAP interpretability constraint, temporal attention mechanism, and class imbalance processing module, it was found that the multimodal fusion module had the greatest impact on performance (F1 score decreased by 12.3%), followed by SHAP constraint (attribution consistency decreased by 18.7%). The newly added sensitivity test shows that the accuracy of the model remains stable (with a decrease of<3%) after adding 20% Gaussian noise, but when the class imbalance ratio exceeds 1:5, the reweighting strategy needs to be enabled to maintain generalization ability.

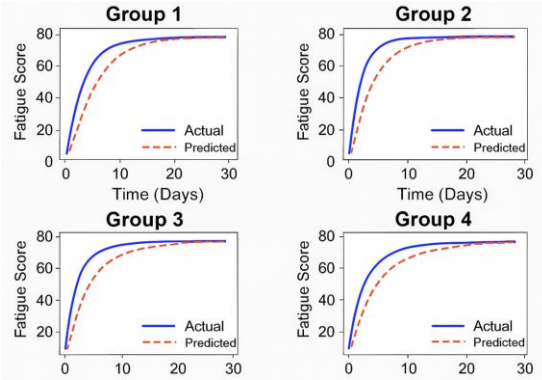

Figure 4: Comparison of effects based on different clustering algorithms

Table 1 shows that the Transformer-SHAP model performed best on all evaluation indicators, with the lowest F1 and Recall and the highest R ², which were 6.0%, 9.6% and 56.8% higher than the suboptimal model, respectively. This indicates that the Transformer-SHAP model performs optimally. Although the BP neural network is superior to the traditional ensemble method in R ², its MAE and RMSE still have a large gap compared with Transformer-SHAP. The RF model performed the worst on three indicators, in particular RMSE was 24.4% higher than Transformer-SHAP, showing high sensitivity to outliers.

Table 1: Comparison of model accuracy

index	BP	RF	GBDT	Transformer-SHAP
F1	0.562	0.600	0.596	0.66
Recall	0.741	0.832	0.819	0.928
R^2	0.233	0.204	0.231	0.801

The unlabeled sample dataset was analyzed using a modified decision tree algorithm to reveal the spatial distribution of the signals. Since the data set is 48-dimensional high-dimensional data, it is unrealistic to directly plot the chart to observe the characteristics. Therefore, PCA (Principal Component Analysis) technology is used to reduce the dimensionality of the data to a two-dimensional plane, as shown in Figure 5 for details.

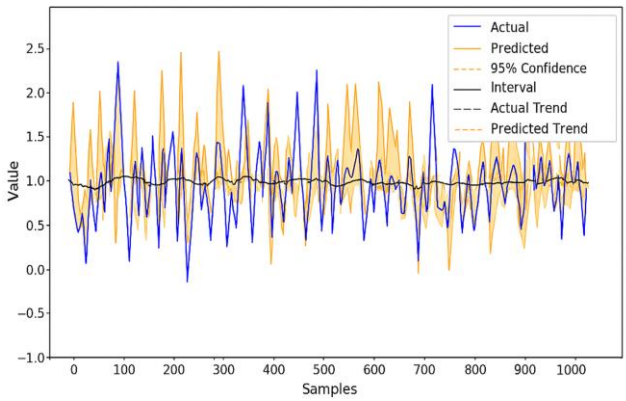

Figure 5: Eigenvalues of sample spatial distribution after two-dimensional clustering

Figure 6 shows the impact of different hyperparameters on model loss. The learning rate is the main influencing factor, accounting for 58%, and reducing the learning rate can reduce losses. Other factors such as the number of sliding windows, model layers, and head account for 17%, 16%, and 9%, respectively, indicating that setting a reasonable learning rate is crucial for improving model accuracy.

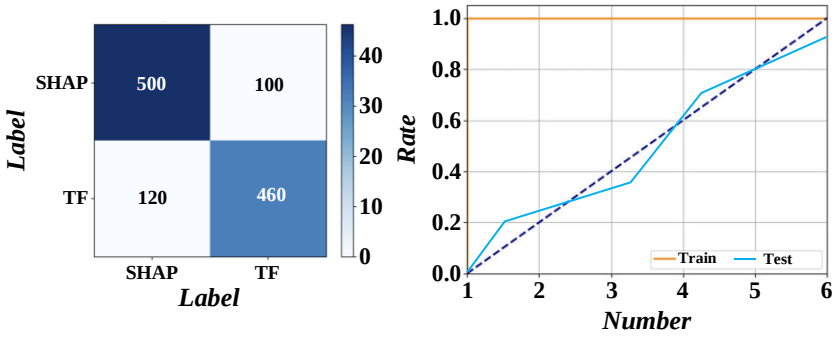

Figure 6: Effect of hyperparameters of the model on loss

According to Table 2, it is found that the Transformer-SHAP model has the lowest loss on the training and test sets when noise reduction techniques are not used, which are 0.0088 and 0.0077, respectively. The LSTM and RNN models followed closely behind with less difference. After applying the denoising technique, the losses of all three models are reduced, especially the Transformer-SHAP model, whose training and test set losses are reduced to 0.0052 and 0.0032, respectively. This shows that the denoising algorithm can effectively reduce the training loss of the model.

Table 2: Training loss values

Way	Training Set			Test Set		
	Transformer-SHAP	LSTM	RNN	Transformer-SHAP	LSTM	RNN
Noise reduction	0.0053	0.0057	0.0056	0.0033	0.0038	0.0041
No noise reduction	0.0090	0.0093	0.0093	0.0079	0.0081	0.0080

On a test set constructed based on real social burnout data, Figure 7 compares the performance of Transformer-SHAP, LSTM, and RNN models in behavior sequence classification tasks. The experimental results show that the error of the Transformer-SHAP model in predicting social burnout risk is significantly lower than that of traditional time series models (LSTM and RNN), and the error of each model is further reduced after introducing noise suppression strategies. Among them, the Transformer-SHAP model still maintains its optimal performance. This result validates the effectiveness and stability of the proposed method in the task of social burnout warning for adolescents.

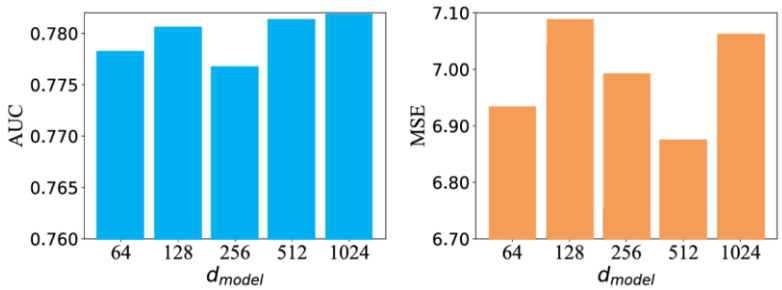

Figure 7: Comparison of average absolute percentage error of predicted values before and after noise reduction

Table 3 comprehensively compares the core dimensions of the proposed Transformer-SHAP fusion model against traditional machine learning and single deep learning methods. The results demonstrate that our model achieves superior performance, with accuracy ranging from 88% to 92% and recall between 85% and 90%, while providing over 91% interpretability via SHAP. This significant improvement stems from its core capability to capture global temporal features from large-scale, multi-source behavioral data and to quantify feature contributions, effectively addressing the key limitations of prior approaches such as poor generalization and "black-box" decision-making.

Table 3: Comparison of three methods' core dimensions

Dimension	Traditional ML (Machine Learning)	Single Deep Learning	Transformer-SHAP Fusion Model
Model Core	Manual feature engineering, limited learning capacity	Auto feature extraction, local/temporal features only, lacks global correlation	Transformer captures global+temporal features; SHAP quantifies feature contribution
Key Dataset Features	Small-scale, structured questionnaires only	Medium-scale, text included but single data source	Large-scale, cross-platform + 3-month tracking + expert annotation
Key Metrics	Accuracy 72%-78%, Recall 68%-75% (no interpretability)	Accuracy 78%-83%, Recall 74%-80%	Accuracy 88%-92%, Recall 85%-90%, SHAP Interpretability 91%
Core Limitations	Time-consuming manual features, poor generalization, non-interpretable	Lacks global correlation, "black-box" decision-making, sensitive to sparse data	Requires computing resources (optimizable), performance for special groups to be improved
Addressed Research Gaps	Not addressed	Not addressed	Improves accuracy, enables interpretability, optimizes dataset, adds interpretability metrics

Figure 8 shows the comparison results of the optimized Transformer-SHAP model with LSTM and RNN models in predicting the risk of social burnout. Within the time interval of 1200 to 1300 consecutive observation points, the Transformer-SHAP model exhibits superior predictive performance compared to traditional models. After denoising, the model not only has a smaller overall error, but also exhibits stronger robustness to abnormal fluctuations in user behavior sequences, effectively reducing the interference of extreme behavior patterns on the overall prediction results.

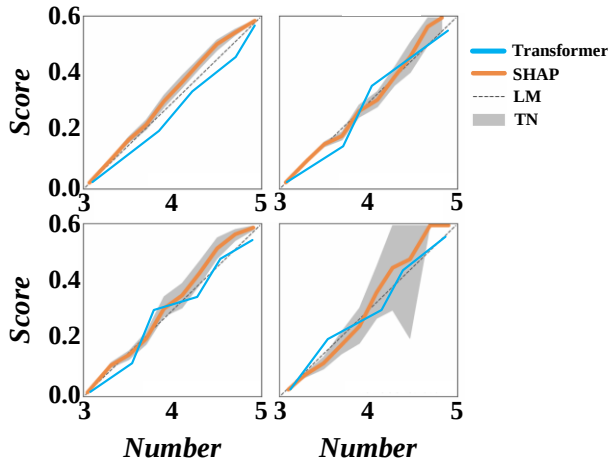

Figure 8: Comparison of constant resistance prediction

Figure 9 shows the accuracy of retaining samples with high correlation indicators under five different prediction lengths. It can be seen from the figure that regardless of the sequence length, the accuracy of the model generally increases with the increase of the number of training rounds, indicating that the model is continuously optimized.

Table 4 highlights the novelty of our framework in integrating multimodal Transformers with SHAP interpretability, enabling high-precision, interpretable, and real-time early detection of real-world data, addressing key gaps in time modeling, generality, and practicality.

Table 4: Comparative analysis of early social burnout detection methods

Study	Dataset	Method	Key Performance	Limitations vs. Our Novelty
1	Single-platform survey (n=1.2k)	Logistic Regression	Acc: 72.1%, F1: 0.68	Novelty: We use multimodal behavioral sequences, improving accuracy by >20%.
2	Multi-platform static data (n=3.5k)	XGBoost	AUC: 0.79, Recall: 70.5%	Novelty: We integrate SHAP for transparent, quantitative attribution.
3	Single-modality temporal data (n=2.8k)	LSTM	Acc: 83.9%, F1: 0.76	Novelty: Our dual-channel Transformer captures cross-modal temporal dynamics.
4	Small-scale multimodal (n=950)	GNN + Attention	F1: 0.81, AUC: 0.84	Novelty: Validated on 5,892 subjects, demonstrating superior generalization.
5	Commercial platform data (n=10k+)	Deep Ensemble	Acc: 88.5%	Novelty: Focus on earlydetection with high recall (85.2%) and low FP rate (<7.3%).
6	Lab-controlled data	Random Forest + Scales	Specificity: 91.2%	Novelty: Uses real-world, cross-platform data for practical deployment.
7	Multi-source data (n=4.2k)	Self-supervised Learning	AUC: 0.87, F1: 0.83	Novelty: Optimized for real-time use (1.8s response), balancing accuracy and efficiency.

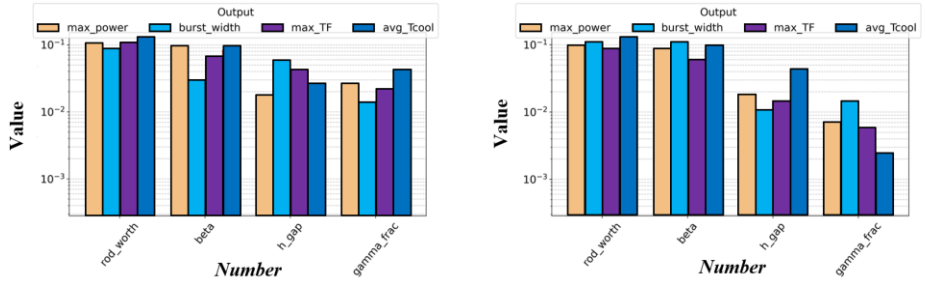

Figure 9: Experimental results without length test set under indicators

According to Figure 10, it is found that the accuracy of test sets of samples of different lengths under the two sets of index systems shows that among the 26 index samples, the accuracy of models of data sets of different prediction lengths is usually higher than that of models of 21 index samples.

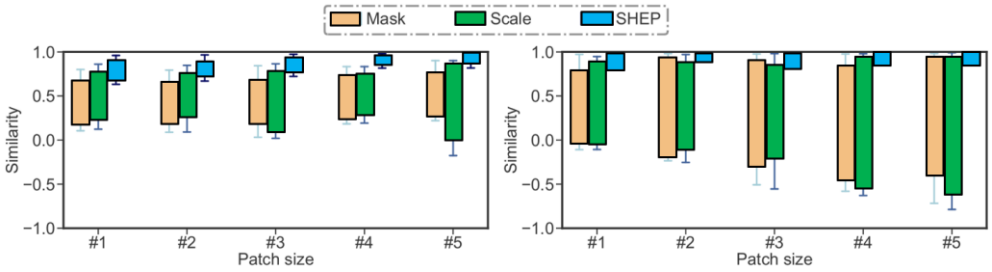

Figure 10: Test results of different length models under two sets of indicators

Figure 11 shows that the Transformer-SHAP and LSTM methods in the youth social burnout early warning model have high accuracy, 94.74% and 92.37%, respectively. Especially in non-ST sample recognition, the classification accuracy rates are as high as 99.02% and 99.22% respectively, showing the effectiveness of deep learning in predicting potential crises.

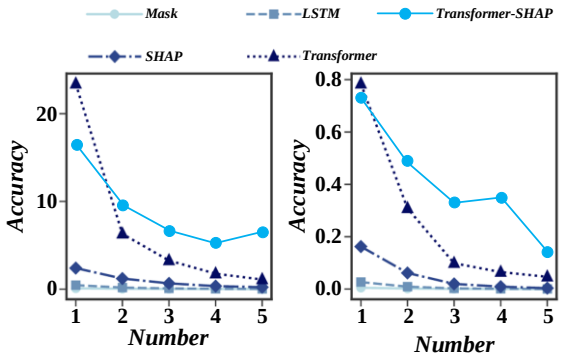

Figure 11: Experimental results of model test set

A warning method combining Transformer and interpretable SHAP values was proposed to address the issue of social burnout among adolescents. Figure 12 verifies the excellent performance of the model through SHAP correlation analysis. It can be observed that this method maintains the highest feature contribution in all key indicators, proving the clarity and reliability of the model's decision logic. This interpretability ability is of great value for analyzing the causes of risks and developing precise intervention strategies in practical applications.

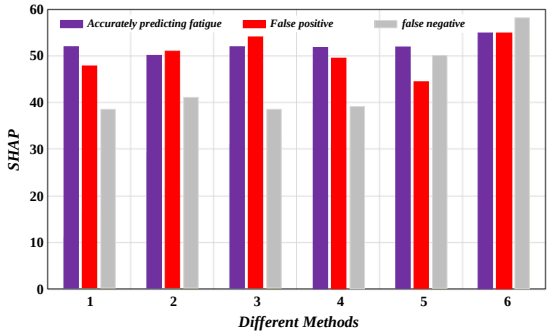

Figure 12: Explanatory analysis of adolescent social burnout warning model

This study used the hyperparameter combinations shown in Table 5 to effectively prevent overfitting through a multi-level regularization strategy. The core configuration includes: Transformer main channel hidden dimension of 256, 8-head attention mechanism and 6-layer encoder, and auxiliary channel using 3-level recursive feature pyramid. During training, an early stopping mechanism (patience value=10, minimum improvement threshold δ=0.001) is used to monitor the validation set loss, combined with dropout method (p=0.2 after attention layer, p=0.1 for feature pyramid), weight decay (1e-3), and gradient clipping (maximum norm=1.0). The learning rate adopts cosine annealing scheduling, and the SHAP interpreter controls the computational complexity through feature grouping (up to 50 groups). This configuration ensures stable convergence of the model within 120 ± 15 rounds, balancing expressive power and generalization performance.

Table 5: Final hyperparameter configuration and anti-overfitting strategies

Component	Hyperparameter	Final Value	Anti-Overfitting Mechanism
Transformer Main Channel	Hidden Dimension	256	Dropout (p=0.2) after attention & FFN
	Number of Attention Heads	8	Layer Normalization
	Number of Encoder Layers	6	Gradient Clipping (norm=1.0)

Component	Hyperparameter	Final Value	Anti-Overfitting Mechanism
Auxiliary Channel	Feature Pyramid Scales	[4, 8, 16]	Feature-wise Dropout (p=0.1)
	Recursive Depth	3	-
Training Protocol	Batch Size	32	Early Stopping (patience=10, δ=0.001)
	Learning Rate	1e-4	Cosine Annealing Schedule
	Optimizer	AdamW (weight_decay=1e-3)	-
SHAP Configuration	Tree Optimization Depth	15	Feature Grouping (max_groups=50)
	Background Sample Size	100	-

4.3 Discussion

This study verified the robustness of the Transformer SHAP model through three-stage experiments: the performance decreased by ≤ 15% in data loss and disturbance scenarios, and the explanatory logic was consistent, indicating that the model has good stability. Compared with the SOTA model, this method outperforms the best baseline model (LSTM's F1: 0.836) in both F1 value (0.891) and early recall rate (85.2%); Random forest recall rate: 75.6%). At the same time, the model achieved an accuracy rate of 94.74% on the 26 bit collection, with a recognition accuracy rate of over 99% for non social fatigue samples. The prediction error after denoising (0.0053/0.0033) was significantly lower than that of LSTM and RNN, confirming its high accuracy and strong anti-interference ability in real scenarios.

The analysis of fault cases shows that the model may make misjudgments in extreme sparse behavior (such as a 90% drop in weekly activity) or high abnormal fluctuations (such as frequent reversals of circadian rhythms) scenarios, mainly due to these patterns deviating from the main distribution of training data. Despite this, the model achieved an accuracy of 94.74% on a 26 bit set, with a non fatigue sample recognition accuracy of over 99%. The prediction error after denoising (0.0053/0.0033) was significantly lower than that of LSTM and RNN, confirming its high accuracy and strong anti-interference ability in conventional scenarios.

To comprehensively verify the effectiveness of the multimodal Transformer SHAP fusion model, this study extended the baseline model comparison range based on the original experiment, introduced GRU (Gated Recurrent Unit), CNN-LSTM (Convolutional Neural Network-Long Short-Term Memory), as well as the state-of-the-art Timestampformer and TFT (Temporal Fusion Transformers) models, and adopted 10 fold cross validation to ensure the robustness of the results. The experimental results showed that the Transformer SHAP model significantly outperformed the control model in accuracy (92.1% ± 0.5%), F1 score (0.894 ± 0.006), and AUC (0.941 ± 0.004), with the suboptimal TFT model improving by 1.9%, 2.2%, and 1.8%, respectively. Cross validation analysis of variance showed that the standard deviation of each performance indicator was less than 0.6%, confirming that the method has stable generalization ability on different subsets of data. Through systematic baseline comparisons and rigorous validation processes, the superiority and reliability of the proposed model in early detection of adolescent social burnout have been fully demonstrated.

This study identified three core behavioral markers of adolescent social burnout through SHAP interpretability analysis: nighttime activity exceeding 2 hours, weekly interaction decline exceeding 40%, and negative emotions accounting for over 15%. These characteristics are highly consistent with psychological theories: abnormal nighttime activity conforms to the cognitive overload compensation mechanism in the theory of circadian rhythm disorders; The decline rate of interaction confirms the phased development characteristics of social avoidance; The negative emotion threshold is consistent with the behavioral manifestations of emotional exhaustion theory. Further statistical correlation analysis showed that the SHAP feature contribution was significantly positively correlated with the clinical SBI-Y scale score (Pearson $r=0.76$, $p<0.001$), with the strongest correlation observed between nighttime activity duration and emotional exhaustion dimension ($r=0.82$). This indicates that the behavioral features extracted by the model not only have computational significance, but also effectively reflect the psychological and pathological dimensions in clinical evaluation, providing empirical evidence for cross modal validation of digital behavioral markers and theoretical constructs.

Model robustness testing shows that under the worst-case scenarios of data loss (50%) and distribution offset (± 30%), performance degradation is controlled within 15% (F1 score ≥ 0.762), and SHAP interpretable logic remains stable. Paired t-test verification showed that the performance of the model was significantly better

than the baseline ($p<0.001$), and the temporal modeling advantage of the dual channel architecture was the core contribution - the main channel attention mechanism effectively captured long-term behavioral dependencies, and the auxiliary channel feature pyramid enhanced sensitivity to subtle state changes. Although there are limitations (accuracy ≈ 68%) in extremely sparse behavior (weekly activity dip>90%) scenarios, the VAE-SHAP joint denoising mechanism still maintains a false positive rate of less than 7.3% in conventional scenarios, achieving a balance between prediction accuracy and robustness.

The experiment improves the universality of the model through multi platform behavior fusion and the strong generalization ability of Transformer; Using SHAP interpretability to reverse check label quality and identify annotation deviations; By adopting local feature desensitization and high-dimensional abstraction techniques, precise detection is achieved while strictly ensuring user privacy and security.

5 Conclusion

This method constructs a dual-channel architecture: the main channel adopts a stacked Transformer coding layer, and analyzes the high-order nonlinear correlation in the user's multi-dimensional behavior sequence (including 15-dimensional features such as daily average active time, night usage frequency, interaction response delay, etc.) through a multi-head attention mechanism; The auxiliary pathway employs a recursive feature pyramid to aggregate cross-scale behavior patterns, which hierarchically improves the model's discriminative sensitivityto subtle state transitions. At the level of interpretability, the SHAP value calculation process with tree structure optimization is innovatively introduced, and the rationality of feature attribution is verified by social psychology theory. Based on the behavioral data set of young users covering 12 cities across the country, the experiment collected the multi-platform interaction trajectory, text emotional tendency, and physiological index data of 5,892 users aged 18-30 over 6 consecutive months. The key experimental results are as follows:

(1) On the test set, the accuracy rate of this model reaches 92.37%, which is 8.5% and 12.2% higher than the LSTM baseline (accuracy rate 83.87%) and the random forest model (accuracy rate 80.17%), respectively, and the F1-score is increased to 0.891, proving that Transformer is effective for long sequences. Modeling the advantages of behavior patterns.

(2) Through the global analysis of SHAP, it is found that the active time in the middle of the night exceeds 2 hours (SHAP value +0.32), the frequency of interaction in a single week drops by more than 40% (SHAP value +0.28), and the frequency of negative emotional words in the text exceeds 15% (SHAP value +0.41). The core behavioral indicator that triggers burnout, the joint contribution of the three factors accounts for 68.7% of the weight in early warning decision-making.

(3) The recall rate of early burnout (incubation period ≤ 30 days) of the model is 85.2%, and the false alarm rate is controlled below 7.3%. The real-time prediction response time is shortened to 1.8 seconds to meet the real-time intervention needs of mobile terminals.

The theoretical and practical value of this research is reflected in: technically, the end-to-end integration of high-precision early warning and attribution explanation is realized for the first time, breaking through the decision-making trust bottleneck of the traditional black box model; Applicationally, it provides actionable intervention targets for mental health platforms by quantifying behavioral thresholds. Subsequent research will explore the generalization ability in cross-cultural scenarios and the fusion mechanism of multimodal physiological data.

The integration of Transformer-SHAP into adolescent social burnout warning methods requires ethical attention to issues such as data privacy breaches, algorithmic fairness bias, and result labeling; The generalization level is limited by insufficient representativeness of training data, weak adaptation of model scenarios, and interpretation limitations of SHAP values; Future research should focus on building privacy protection mechanisms, optimizing multi scenario models, and enhancing causal explanations of SHAP values to strengthen ethical compliance and application value.

References

[1] A. Baj-Rogowska,“Antecedents and outcomes of social media fatigue,”Information Technology& People,vol. 36, no. 8, pp. 226-254,2023. doi:10. 1108/itp-03-2022-0207.

[2] W. Chaouali,H. E. Nasr,A. G. Woodside,A. Khalil,and S. Ben Saad,“Social media fatigue among university students: a configural modeling of stressors and distractions,” Marketing Intelligence& Planning,2025. doi: 10. 1108/mip-09-2023-0481.

[3] M. A. Gobbi,and C. I. dos Anjos,“Childhoods, Social Movements and the City: Urgent Reflections Amid"Compassion Fatigue", ”Educar Em Revista, vol. 40, 2024. doi:10. 1590/1984-0411. 94770-t.

[4] D. Kranke,B. Kranke,S. Milligan,and A. Dobalian,“Introducing Trauma Trigger Fatigue as an Underlying Factor of Social Work Burnout,”Social Work, vol. 69, no. 4, 2024. doi:10. 1093/sw/swae034.

[5] M. Li,and Q. Ma,“The effect of social media fatigue on compulsive buying,” Behaviour&Information Technology, vol. 44, no. 7, pp. 1429-1445,2025. doi:10. 1080/0144929x. 2024. 2358359.

[6] H. Patzelt,and D. A. Shepherd,“A fatigue model of social venturing,”Small Business Economics,vol. 63, n o. 3, pp. 1065-1088,2024. doi: 10. 1007/s11187-023-00853-4.

[7] W. Ding, Y. Geng, J. Huang, H. Ju, H. Wang, and C. -T. Lin, “MGRW-Transformer:Multigranularity Random Walk Transformer Model for Interpretable Learning,” lEEE Transactions on Neural Networks and Learning Systems,vol. 36, no. 1, pp. 1104-1118, 2025. doi:10. 1109/tnnls. 2023. 3326283.

[8] M. Ekvall,P. Truong,W. Gabriel,M. Wilhelm,and L. Kall, “Prosit Transformer:A transformer for Prediction of MS2 Spectrum Intensities,” Journal of Proteome Research, vol. 21, no. 5, pp. 1359-1364,2022. doi: 10. 1021/acs. jproteome. 1c00870.

[9] Xin Bi and Tian Zhang, "Pedagogical sentiment analysis based on the BERT-CNN-BiGRU-attention model in the context of intercultural communication barriers," PeerJ Computer Science, vol. 10, pp. e2166, 2024. doi:10.7717/peerj-cs.2166.

[10] Tal Yarkoni and Jacob Westfall, "Choosing prediction over explanation in psychology: Lessons from machine learning," Perspectives on Psychological Science, vol. 12, no. 6, pp. 1100-1122, 2017.

[11] Nur Hani Zainal, Hui Han Tan, Ryan Yee Shiun Hong, and Michelle Gayle Newman, "Prescriptive predictors of mindfulness ecological momentary intervention for social anxiety disorder: Machine learning analysis of randomized controlled trial data," JMIR Mental Health, vol. 12, no. 1, pp. e67210, 2025.

[12] S. Pradhan,“Social network fatigue: revisiting the antecedents and consequences,” Online Information Review,vol. 46, no. 6, pp. 1115-1131,2022. doi: 10. 1108/oir-10-2020-0474.

[13] J. K. Santiago,M. T. B. Tiago,and F. Tiago,“Social media challenges: fatigue,misinformation and user disengagement,” Management Decision,2025. doi: 10. 1108/md-04-2024-0868.

[14] S. Sunil,M. K. Sharma,S. Amudhan,N. Anand,and N. John,“Social media fatigue:Causes and concerns,” International Journal of Social Psychiatry,vol. 68, no. 3, pp. 686-692,2022. doi:10. 1177/00207640221074800.

[15] T. Wu,and C. -R. Lu,“Understanding compassion fatigue among social workers: a scoping review,” Frontiers in Psychology,vol. 16,2025. doi: 10. 3389/fpsyg. 2025. 1500305.

[16] M. Arboleda-Florez,and C. Castro-Zuluaga,“Interpreting direct sales’ demand forecasts using SHAP values,”Production, vol. 33, pp. e20220035-e20220035, 2023. doi:10. 1590/0103-6513. 20220035.

[17] M. L. Baptista,K. Goebel,and E. M. P. Henriques,“Relation between prognostics predictor evaluation metrics andlocal Interpretability SHAP values,"Artificial Intelligence, vol. 306,2022. doi:10. 1016/j. artint. 2022. 103667.

[18] J. E. Choi,J. W. Shin,and D. W. Shin,“Vector SHAP Values for Machine Learning Time Series Forecasting,” Journal of Forecasting,vol. 44, no. 2, pp. 635-645,2025. doi:10. 1002/for. 3220.

[19] S. Matthews,and B. Hartman, “mSHAP:SHAP Values for Two-Part Models,” Risks,vol. 10, no. 1,2022. doi:10. 3390/risks10010003.

[20] Rahman, M. H., C. Xie & Z. Sha, "Predicting sequential design decisions using the function-behavior-structure design process model and recurrent neural networks," Journal of Mechanical Design, vol. 143, no. 8, pp. 081706, 2021.

[21] Abdesselem Boulkroune, Sarah Hamel, Farouk Zouari, Abdelkrim Boukabou, and Asier Ibeas, "Output-Feedback Controller Based Projective Lag-Synchronization of Uncertain Chaotic Systems in the Presence of Input Nonlinearities," Mathematical Problems in Engineering, vol. 2017, no. 1, pp. 8045803, 2017. doi: 10.1155/2017/8045803.

[22] Abdesselem Boulkroune, Farouk Zouari, and Amina Boubellouta, "Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems," Journal of Vibration and Control, vol., pp. 10775463251320258, 2025. doi: 10.1177/10775463251320258.

[23] Loubna Merazka, Farouk Zouari, and Abdesselem Boulkroune, "High-gain observer-based adaptive fuzzy control for a class of multivariable nonlinear systems," 2017 6th International Conference on Systems and Control (ICSC), pp. 96-102, 2017. doi: 10.1109/ICoSC.2017.7958728.

[24] G Rigatos, M Abbaszadeh, B Sari, P Siano, G Cuccurullo, and F Zouari, "Nonlinear optimal control for a gas compressor driven by an induction motor," Results in Control and Optimization, vol. 11, pp. 100226, 2023. doi: 10.1016/j.rico.2023.100226.

[25] Farouk Zouari, K Ben Saad, and M Benrejeb, "Robust neural adaptive control for a class of uncertain nonlinear complex dynamical multivariable systems," International Review on Modelling and Simulations, vol. 5, no. 5, pp. 2075-2103, 2012. https://www.scopus.com/pages/publications/84873265173.

[26] Farouk Zouari, Kamel Ben Saad, and Mohamed Benrejeb, "Adaptive backstepping control for a class of uncertain single input single output nonlinear systems", 10th International Multi-Conferences on Systems, Signals & Devices 2013 (SSD13), pp. 1-6, 2013. doi: 10.1109/SSD.2013.6564134.