

RESEARCH ARTICLE

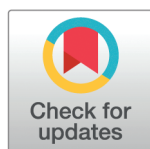
PhysioFormer: Integrating multimodal physiological signals and symbolic regression for explainable affective state prediction

Zhifeng Wang^{1*}, Wanxuan Wu¹, Chunyan Zeng^{2*}, Jialiang Shen¹

1 Department of Digital Media Technology, Central China Normal University, Wuhan, China,

2 High-Efficiency Utilization of Solar Energy and Operation Control of Energy Storage System, Hubei University of Technology, Wuhan, China

* zfwang@ccnu.edu.cn (ZW); cyzeng@hbut.edu.cn (CZ)



Abstract

As affective computing becomes increasingly crucial in health monitoring and psychological intervention, accurately identifying affective states is a key challenge. While traditional machine learning models have achieved some success in affective computation, their ability to handle complex, multimodal physiological signals is limited. Most affective computing tasks still rely heavily on traditional methods, with few deep learning models applied, particularly in multimodal signal processing. Given the importance of stress monitoring for mental health, developing a highly reliable and accurate affective computing model is essential. In this context, we propose a novel model—PhysioFormer, for affective state prediction using physiological signals. PhysioFormer model integrates individual attributes and multimodal physiological data to address inter-individual variability, enhancing its reliability and generalization across different individuals. By incorporating feature embedding and affective representation modules, PhysioFormer model captures dynamic changes in time-series data and multimodal signal features, significantly improving accuracy. The model also includes an explainability model that uses symbolic regression to extract laws linking physiological signals to affective states, increasing transparency and explainability. Experiments conducted on the Wrist and Chest subsets of the WESAD dataset confirmed the model's superior performance, achieving over 99% accuracy, outperforming existing SOTA models. Sensitivity and ablation experiments further demonstrated PhysioFormer's reliability, validating the contribution of its individual components. The integration of symbolic regression not only enhanced model explainability but also highlighted the complex relationships between physiological signals and affective states. Future work will focus on optimizing the model for larger datasets and real-time applications, particularly in more complex environments. Additionally,

OPEN ACCESS

Citation: Wang Z, Wu W, Zeng C, Shen J (2025) PhysioFormer: Integrating multimodal physiological signals and symbolic regression for explainable affective state prediction. PLoS One 20(10): e0335221. <https://doi.org/10.1371/journal.pone.0335221>

Editor: Amgad Muneer, The University of Texas, MD Anderson Cancer Center, UNITED STATES OF AMERICA

Received: July 9, 2025

Accepted: October 8, 2025

Published: October 31, 2025

Copyright: © 2025 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The dataset WESAD used in this paper is publicly available at <https://doi.org/10.1145/3242969.3242985>.

Funding: The research work in this paper were supported by the National Natural

Science Foundation of China (No. 62577027, 62177022, 61901165) and Self-determined Research Funds of CCNU from the Colleges' Basic Research and Operation of MOE (No. CCNU25ai011).

Competing interests: The authors have declared that no competing interests exist.

further exploration of physiological signals and environmental factors will help build a more comprehensive affective computing system, advancing its use in health monitoring and psychological intervention.

Introduction

With the rapid development of society and the increasing pace of life, the importance of emotions for individual physical and mental health has become increasingly evident. Emotions are spontaneously generated from daily experiences, stimulating the body to produce hormones and affecting various aspects, including bodily movements, facial expressions, and physiological characteristics. The accumulation of negative emotions can lead to depression. It is estimated that about one in five men and one in three women worldwide will experience major depression during their lifetime. Although other mental illnesses, such as schizophrenia and bipolar disorder, are less common, they still have a significant impact on individuals' lives [1]. The contemporary focus on mental and physical stress is growing, as the effects of stress on the human body are both widespread and profound, particularly in its impact on the brain, cardiovascular system, immune system, and metabolism. Chronic stress not only leads to mental health issues but may also alter gene expression through epigenetic mechanisms, further affecting physiological functions [2].

In existing research, physiological signals have been applied across various fields. First, in the domain of healthcare, monitoring physiological signals enables early warnings of chronic diseases and health management [3]. Second, in affective computation and mental health, analyzing signals such as electrodermal activity and electrocardiograms can identify individuals' affective states, providing effective tools for emotion management [4]. Additionally, in intelligent human-computer interaction, the application of physiological signals can optimize user experience by allowing devices to respond according to users' physiological and affective states [4]. Given the significant impact of negative emotions on daily life, monitoring affective states through physiological signals is particularly important. Moreover, the current global healthcare trend is shifting toward preventing chronic diseases and reducing treatment costs, with wearable devices playing a crucial role in this development [5]. These devices can monitor users' physiological signals in real time, providing personalized health recommendations and helping individuals detect potential health issues early, thus promoting a shift in healthcare systems from reactive treatment to proactive prevention.

Traditional affective computation has largely relied on questionnaires [6], which are often highly subjective. The reliability of the results is closely tied to the respondent's attentiveness and the seriousness with which they approach the questionnaire. Moreover, individuals are often unable to accurately perceive their own affective states. For instance, people who are under chronic stress may not be fully aware of the extent of the stress they are experiencing. Therefore, using questionnaires to identify affective states is undoubtedly time-consuming and inefficient. As a result,

subsequent

research has shifted towards utilizing machine learning and deep learning methods to predict affective states through physiological signals.

Numerous studies have demonstrated that physiological signals such as electrocardiograms (ECG), electrodermal activity (EDA), and electroencephalograms (EEG) play a crucial role in affective computation. The relevant researches and their limitations are shown in the Table 1 below.

Despite significant progress in the field of affective recognition, several key challenges and research gaps remain. First, in terms of reliability, while the complexity and diversity of physiological signals are widely acknowledged, many studies still inadequately collect and utilize multimodal physiological data. This shortfall not only limits the model's generalization capability under multimodal data conditions but also affects its stability in cross-individual affective recognition tasks. Due to the high variability of individual physiological signals, the generalizability and reliability of existing models in practical applications are weakened, and the lack of data support further restricts their applicability in diverse scenarios. Second, although models exhibit high accuracy in certain tasks, affective states are inherently dynamic processes. Existing methods significantly lack the ability to capture the continuity and temporal features of emotional changes, especially when dealing with complex physiological signals. This limitation reduces the model's ability to accurately recognize dynamic affective states. Finally, the issue of model explainability remains prominent. Most deep learning models still function as "black boxes," making it difficult to clearly explain their decision-making processes and feature selection. This lack of transparency not only reduces users' trust in model predictions but also limits the model's explainability and transparency in real-world applications. Therefore, developing an affective recognition model that emphasizes the integration and utilization of multimodal physiological data, enhances the ability to capture dynamic processes, and improves decision-making transparency could not only advance the accuracy and practicality of affective recognition but also drive progress in stress monitoring and mental health interventions.

To address the aforementioned issues, we developed the PhysioFormer model, which is capable of processing high-dimensional, complex, and sequential physiological signal data in parallel, further enhancing the ability to capture the timeliness and continuity of emotional changes. This section will introduce the contributions of the PhysioFormer model to the field of affective computation as below.

1. **Handling high-dimensional, complex temporal data and introducing individual attributes features.** Due to the high-dimensional, complex, and temporally continuous nature of physiological signal data, traditional machine learning models and existing deep learning models tend to exhibit instability. Therefore, in designing the PhysioFormer model, we incorporated three submodules: *ContribNet*, *AffectNet*, and *AffectAnalyser*, which not only process physiological signal data in parallel but also effectively capture the temporal dynamics and complexity of physiological signals. This enhances the model's ability to detect the timeliness and continuity of emotional changes. Additionally, we introduced an upper triangular matrix encoding in the *ContribNet* module, allowing the model to flexibly focus on

Table 1. Existing research and its limitations.

Literature	Research Content	Limitations
Han et al. (2017) [7]	Proposed a method combining Graph Signal Processing (GSP) and Convolutional Neural Networks (CNNs) to process EEG data, capturing spatial features for affective classification.	Focused only on EEG data; limited consideration of multimodal signals; insufficient handling of temporal continuity in affective states.
Song et al. (2020) [8]	Introduced a Dynamic Graph Convolutional Neural Network (DGCNN) for EEG-based affective recognition, modeling spatiotemporal dependencies among brain regions.	Still EEG-specific; lacks multimodal integration; insufficient explainability in decision-making.
Ashwin et al. (2022) [9]	Monitored HR, EDA, and body temperature for stress recognition, applying machine learning models such as KNN and SVM to classify features.	Shallow machine learning methods with limited capacity for high-dimensional sequential data; lacks robustness in cross-subject affective computation; weak interpretability.

<https://doi.org/10.1371/journal.pone.0335221.t001>

important information at different time points while also sensitizing it to the temporal dynamics of physiological signals. This addresses the limitations of traditional machine learning models and existing deep learning networks in effectively capturing the intrinsic relationships within temporally continuous data. Furthermore, to account for individual baseline physiological state differences, we incorporated features describing individual attributes into the dataset, such as age, gender, height, weight, smoking habits, and whether the individual exercised today. These individual attributes features provide rich contextual information, helping the model generalize better across different individuals and improving the accuracy and reliability of affective state prediction. The introduction of individual attributes features allows the model to better understand and explain physiological signals, thereby enhancing its adaptability.

2. **Improving reliability in cross-subject affective computation.** To enhance the reliability of cross-subject affective computation, we employed a Leave-One-Subject-Out (LOSO) cross-validation strategy. In each fold, all data from one participant were held out for testing, while the remaining participants' data were used for training. This protocol ensures strict subject independence and prevents potential data leakage caused by intra-subject correlations. By explicitly learning and adapting to the physiological signal characteristics of different individuals, this approach significantly reduces the risk of overfitting to specific participants, thereby improving the model's generalization ability and robustness in cross-subject affective computation.
3. **Development of a high-precision affective computation model.** The PhysioFormer model has demonstrated exceptional ability in capturing the intrinsic relationships within physiological signal data, significantly improving the accuracy of predicting individuals' affective states. Through rigorous experiments and in-depth analysis on the WEASD dataset, we validated the effectiveness of the PhysioFormer model. On both the Wrist and Chest sub-datasets, the PhysioFormer model achieved over 99% accuracy, far surpassing the performance of current state-of-the-art (SOTA) models. These results not only highlight the potential of PhysioFormer in the field of affective computation but also demonstrate its reliability and reliability in handling complex physiological signals, providing strong support for precise prediction of individual affective states.
4. **Enhancing model explainability.** To improve the model's explainability, we integrated an Explanation Model using symbolic regression to generate mathematical formulas that describe the relationship between input variables and outputs [10]. In our research, this model analyzed the influence of each physiological indicator on affective states, generating formulas that reveal the impact of signals like heart rate variability (HRV), EDA, and ECG on affective prediction. These formulas not only enhance transparency but also help users better understand the model's decision-making process, increasing both explainability and trust in the model's predictions.

The structure of this paper is arranged as follows. “[Related Work](#)” section provides a detailed overview of the related work in the field. “[Preliminary](#)” section defines the mathematical notation and model architecture used in this study. In “[Methods](#)”, we propose and describe the PhysioFormer model and the symbolic regression task in detail. “[Experiment Result and Analysis](#)” covers the dataset, baseline models, experimental setup, and evaluation metrics. Finally, “[Conclusions](#)” section summarizes the study and discusses future research directions.

Related work

Affective computation

Affective computing aims to analyze physiological signals, facial expressions, vocal patterns, and other behavioral data using sensors, algorithms, and machine learning techniques to predict or identify an individual's affective state [11]. In recent years, the research focus of affective computing has shifted from single-mode affective classification to the fusion of multimodal data. By integrating multiple sources of signals—such as physiological signals (e.g., heart rate, electrodermal activity), facial expressions, vocal features, and environmental factors—researchers can more accurately capture and

recognize an individual's affective state. This multimodal approach not only improves the accuracy of affective recognition but also enhances the timeliness and continuity of tracking affective state changes.

Bernhard et al. [12] proposed a deep learning-based method for text-based affective recognition, with key innovations including bidirectional processing using a Bidirectional Long Short-Term Memory (BiLSTM) network, combined with Dropout regularization and a weighted loss function to address small datasets and class imbalance. Additionally, the study introduced a transfer learning method called Sent2Affect, where the network is pre-trained on affective analysis tasks and then fine-tuned for affective recognition by adjusting the output layer. This approach improved the model's performance on small-scale datasets. Experiments conducted on six benchmark datasets demonstrated that this method significantly outperformed traditional machine learning approaches, achieving a 23.2% increase in F1 scores for classification tasks and an 11.6% reduction in mean squared error for regression tasks.

Ashwin et al. [9] induced stress states in participants through acute stress manipulation tasks (such as the Maastricht Acute Stress Task [13] and the Montreal Imaging Stress Test [14]) and recorded physiological signals, including HR, HRV, EDA, and respiratory rate, using wearable sensors. The resulting stress detection models achieved accuracy rates of 97% in controlled environments and 93% in everyday settings. This study confirmed the effectiveness of multimodal physiological signal fusion in stress detection and demonstrated the feasibility of wearable devices for stress monitoring across different environments. Similarly, Sarkar et al. [15] proposed an affective recognition method based on ECG signals using a self-supervised deep multi-task learning framework. By pre-training the network on signal transformation tasks and then transferring it to affective classification, the model showed significant performance improvements on the AMIGOS, DREAMER, WESAD, and SWELL datasets, outperforming traditional fully supervised methods. This validates the efficacy of multi-task learning in ECG-based affective recognition. Akre et al. [16] introduced a depression symptom detection framework using data collected from iPhones and Apple Watches. A gradient boosting classifier processed health data, including vital signs, activity levels, and sleep patterns. The model exhibited moderate predictive accuracy, with ROC AUC values ranging from 0.63 to 0.72, demonstrating the potential of personalized sensor data for depression detection. These studies collectively highlight the significant potential of multimodal physiological signals and personalized health data in the detection and recognition of stress, emotions, and depression symptoms.

Koldijk et al. [17] collected a multimodal dataset specifically designed for research on stress and user modeling, incorporating physiological signals (such as ECG, EMG, EEG), facial expressions, and behavioral data from mouse and keyboard usage. In a controlled experimental environment, participants performed various tasks (e.g., writing, reading, and meetings) to simulate real-world work scenarios. Stress levels were validated using a combination of questionnaires, interviews, and objective measurements. The experimental results indicated a significant correlation between induced stress conditions and both physiological and behavioral data, demonstrating the value of the SWELL dataset for developing robust stress detection models.

Siddharth et al. [18] explored affective computing using multimodal data, leveraging biosignals (EEG, ECG, GSR, HRV) and visual data (facial video) in experiments conducted on four publicly available multimodal affective datasets: DEAP, AMIGOS, MAHNOB-HCI, and DREAMER. These datasets include a range of physiological signals and visual information, with participants self-reporting affective states after watching videos. The results showed that combining biosignals with visual data significantly improved affective classification performance, with the model outperforming previous research particularly on the DEAP and MAHNOB-HCI datasets. Additionally, the study addressed issues such as discrepancies in sampling rates, sensor positions, and the number of signal channels through dataset fusion and transfer learning. Both studies emphasize the potential of multimodal data in stress and affective computing, offering valuable resources for the development of personalized stress management and affective recognition systems.

This study focuses exclusively on using physiological signals for affective computing, aiming to accurately identify participants' affective states through these signal features. In contrast to multimodal data fusion approaches, this research

emphasizes optimizing the processing and classification of single physiological signals, exploring their potential in affective recognition. This approach offers the possibility of simplifying affective computing systems while reducing the complexity of data collection and processing.

Explainability method

Symbolic regression is defined as the process of discovering symbolic expressions that fit data for an unknown function. Although this problem is theoretically considered NP-hard, in practice, many functions exhibit simplifying properties such as symmetry, separability, and compositionality, making the task feasible [10]. Symbolic regression can be applied across various fields, including economics, psychology, and biomedical engineering. It aids researchers in uncovering hidden mathematical models from experimental data, thus revealing the underlying dynamics of complex systems.

The SINDy (Sparse Identification of Nonlinear Dynamics) method proposed by Brunton et al. [19] combines symbolic regression with sparse regression to overcome the limitations of traditional symbolic regression in handling complex nonlinear equations, such as high computational cost and overfitting. SINDy reduces the candidate function space and uses convex optimization to generate concise and explainable equations, ensuring both efficiency and explainability in large-scale systems. This method has been successfully applied to complex systems such as fluid vortices and chaotic Lorenz systems, demonstrating broad applicability. Rogers et al. [20] further expanded the application of symbolic regression by integrating it with model-based design of experiments (MBoE). Symbolic regression generates explainable expressions, while MBoE optimizes experimental conditions, allowing for rapid differentiation between model candidates and significantly improving process optimization efficiency. This approach has performed exceptionally well in industrial processes like multiphase product synthesis. By incorporating physical knowledge constraints, the complexity of symbolic regression is controlled, enhancing both the explainability and practical value of the models. This combined method shows significant potential for applications in digital manufacturing, process engineering, and new product development, making it suitable for both laboratory research and industrial production optimization.

The expanded application of symbolic regression in psychology and other fields demonstrates its significant interdisciplinary potential. Masato et al. [21] were the first to apply the symbolic regression tool AI-Feynman to intertemporal choice experiments in psychology, successfully generating seven candidate discount function models, some of which outperformed existing hyperbolic discount models. This study shows that symbolic regression can not only automatically uncover hidden patterns in psychology but also significantly improve the automation and precision of research, shifting away from traditional approaches reliant on human intuition and experience. The introduction of symbolic regression offers psychologists a novel data analysis method, enhancing the accuracy of theoretical modeling and the reproducibility of studies, thereby showcasing the unique value of applying AI technology to the social sciences. Liu et al. [22] further expanded the application of symbolic regression by integrating it with deep learning in the field of knowledge tracing. Their method automatically extracts algebraic expressions of learners' cognitive states, revealing the underlying patterns of skill acquisition. In experiments on the large-scale Lumosity training dataset, symbolic regression not only improved the model's fit but also discovered entirely new patterns of skill acquisition, verifying some existing theoretical findings. This approach provides theoretical support for analyzing large-scale behavioral data, particularly excelling in dynamic learning processes and naturally generated data, addressing challenges like the unobservability of cognitive states and the expanding search space of symbolic models. The application of symbolic regression in psychology, education, and related fields opens new avenues for research on skill acquisition, cognitive diagnostics, and personalized learning path modeling, demonstrating its broad potential in automated pattern discovery and model building.

This paper introduces symbolic regression into affective computing and combines it with deep learning models, offering a new approach for automated affective state inference. Traditional affective computing relies on black-box models, which, while accurate, lack explainability. Symbolic regression addresses this by extracting explainable algebraic expressions

from neural networks, enabling the analysis of complex relationships between physiological signals and behavioral features. This improves model transparency and provides a theoretical basis for the quantitative analysis of affective states.

Sequential model

Sequential models play a crucial role in affective computing, particularly when handling time-series data, where they have demonstrated exceptional performance. Traditional sequence models such as Hidden Markov Models (HMM) [23] and Conditional Random Fields (CRF) [24] have been widely applied in fields like natural language processing. However, with the rise of deep learning, more advanced models like Recurrent Neural Networks (RNN) [25] and Long Short-Term Memory Networks (LSTM) [26] have been introduced into affective computing to better capture temporal dependencies.

RNN are among the earliest deep learning models used for processing sequential data. They capture contextual information in time series through recurrently connected neurons and are suitable for various sequence tasks, such as speech recognition and machine translation. However, RNNs face challenges when dealing with long-term dependencies due to the vanishing and exploding gradient problems, making it difficult to retain distant dependencies [25]. To address these limitations, LSTM networks were developed. LSTMs use unique memory cells and gating mechanisms to effectively overcome the information loss problem in traditional RNNs when handling long sequences. Their memory cells retain important information over time, controlling what to keep and what to forget. LSTM networks excel at tasks involving long-term dependencies, particularly in affective computing, such as speech affective recognition [27] and affective state prediction based on physiological signals [28]. The memory cells in LSTM allow the model to capture and retain affective state information over extended periods, improving the accuracy of affective recognition. The combination of LSTM with CNNs further enhances the performance of multimodal affective recognition. CNNs are adept at processing spatial features, such as facial expressions in images or other static physiological signals, while LSTMs handle temporal features, such as dynamic changes in speech and heart rate. This combination allows multimodal affective computing to process both complex time-series data and integrate information from different modalities, leading to more accurate affective state predictions.

Recent studies have further advanced the application of Transformer-based sequence models in affective computing. Unlike LSTM, Transformer models process sequential data in parallel through the self-attention mechanism, overcoming the efficiency bottlenecks that RNNs and LSTMs face when handling long sequences. This makes Transformers particularly suitable for processing longer time-series data [29]. The self-attention mechanism flexibly captures global contextual information within a sequence without relying on sequential order, making Transformer models more effective in feature extraction and temporal modeling. Mittal et al. [30] proposed an M3ER model based on multimodal data, integrating facial expressions, text, and speech modalities. The model incorporates multiplicative fusion techniques to weigh the reliability of each modality, allowing it to automatically emphasize more reliable modalities while suppressing those with higher noise levels. Additionally, M3ER applies canonical correlation analysis (CCA) to filter out irrelevant signals from modalities and generate proxy features, enhancing the model's robustness against noise and missing data. Experimental results on the IEMOCAP and CMU-MOSEI benchmark datasets show that this method improved accuracy in affective computation tasks by approximately 5% compared to previous models.

Recent researchers have applied the aforementioned approaches to affective computation tasks. Kumar et al. [31] proposed two deep learning-based methods for speech affective recognition: CNN-LSTM and Vision Transformer (ViT). The study experimentally compared the performance of these two models in handling affective recognition tasks, focusing on the advantages of CNN-LSTM in audio feature extraction and affective classification, as well as the potential of ViT for processing speech signals through Mel-spectrograms. The CNN-LSTM model achieved an accuracy of 88.50% on the EMO-DB dataset, while the Vision Transformer model reached 85.36%. This work highlights the potential of deep learning technologies, particularly attention mechanisms and image processing techniques, in speech affective recognition, demonstrating the effectiveness of combining CNN and LSTM to extract affective features from speech signals.

This paper focuses on physiological signal sequence models, aiming to explore how to better utilize physiological signals for affective state prediction and enhance the accuracy and generalization of affective computing through multimodal approaches.

Preliminary

Notations and definitions

In this section, we formally define the computational modules involved in the affective computation tasks of this study and provide a set of mathematical notations used throughout the paper, as presented in the Table 2.

Definition 1 (ContribNet): *ContribNet* is a neural network model used to compute the contribution level of physiological indicator b_j , consisting of a batch normalization layer and two linear transformation layers. Specifically, the input data is first processed through batch normalization, followed by a linear transformation through the first weight matrix $\mathcal{W}_1^{b_j}$ with an added bias $\mathcal{B}_1^{b_j}$, then activated by a nonlinear activation function σ . The activated output is further linearly transformed by the second weight matrix $\mathcal{W}_2^{b_j}$, with an added bias $\mathcal{B}_2^{b_j}$, to obtain the final output, representing the contribution level of physiological indicator b_j to the prediction of individual p_i 's affective state. This process can be formally represented as:

$$\text{ContribNet}(\cdot) = \mathcal{W}_2^{b_j} \left(\sigma \left(\mathcal{W}_1^{b_j} \cdot \text{BN}(\cdot) + \mathcal{B}_1^{b_j} \right) \right) + \mathcal{B}_2^{b_j} \quad (1)$$

Where:

- $\mathcal{W}_1 \in \mathbb{R}^{H \times (k+m)}$ and $\mathcal{W}_2 \in \mathbb{R}^{1 \times H}$ represent the weight matrices for each layer;
- $\mathcal{B}_1 \in \mathbb{R}^H$ and $\mathcal{B}_2 \in \mathbb{R}$ represent the biases;
- H denotes the hidden layer dimension;
- σ denotes the activation function ReLU;
- BN represents the batch normalization operation.

Definition 2 (AffectNet): *AffectNet* is used to compute the affective state level reflected by physiological indicator b_j . It extracts and transforms input features progressively through a combination of linear transformations and nonlinear activation functions across multiple hidden layers, ultimately generating the final affective state prediction. Specifically, after

Table 2. Definitions of mathematical notation used in this paper.

Notations	Description
\mathcal{P}	A set containing N participants is denoted as $\mathcal{P} = p_1, p_2, \dots, p_N$
\mathcal{A}	A set of k individual attributes is denoted as $\mathcal{A} = \{a_1, a_2, \dots, a_k\}$, each element represents a feature.
\mathcal{B}	A set of M physiological indicators obtained from a monitoring device is denoted as $\mathcal{B} = \{b_1, b_2, \dots, b_M\}$.
D	Physiological monitoring data of N participants with M indicators.
e	The affective state of the participants at the current moment.
\mathcal{X}, X	The set of all windows after window segmentation, and the data contained within a specific window.
ξ	The number of window segments.
PF	Integrated features of A and B , denoted as $PF = A \oplus B$.
α^{b_j}	The contribution level of physiological indicator b_j
β^{b_j}	The encoded physiological indicator b_j .
γ^{b_j}	Integrated features of A and β^{b_j} , denoted as $\gamma^{b_j} = A \oplus \beta^{b_j}$.
θ^{b_j}	The affective state level reflected by the physiological indicator b_j .
u	Initial affective state level.
Θ	The affective state level reflected by all physiological indicators.
Φ	Affective state level from all indicators, adjusted by initial state.

<https://doi.org/10.1371/journal.pone.0335221.t002>

the input features are processed through batch normalization, they first undergo a linear transformation by the first layer weights $\mathfrak{W}_1^{b_j}$, with an added bias $\mathfrak{b}_1^{b_j}$, followed by nonlinear activation through the activation function σ . This process continues in subsequent layers until the final layer l , where the weights \mathfrak{W}_l produce the output. This output represents the affective state level reflected by physiological indicator b_j for participant p_i . The deep structure of the network allows it to capture complex feature interactions through layer-by-layer learning, enabling the estimation of the participant's affective state. This model can be formally expressed as:

$$AffectNet(\cdot) = \left(\mathfrak{W}_l^{b_j} \left(\sigma \left(\mathfrak{W}_{l-1}^{b_j} \left(\cdots \sigma \left(\mathfrak{W}_1^{b_j} \cdot BN(\cdot) + \mathfrak{b}_1^{b_j} \right) \cdots \right) + \mathfrak{b}_{l-1}^{b_j} \right) \right) + \mathfrak{b}_l^{b_j} \right) \quad (2)$$

Where:

- $\mathfrak{W}_1^{b_j}, \mathfrak{W}_2^{b_j}, \dots, \mathfrak{W}_l^{b_j}$ represent the weight matrices for each layer;
- $\mathfrak{b}_1^{b_j}, \mathfrak{b}_2^{b_j}, \dots, \mathfrak{b}_l^{b_j}$ represent the bias vectors for each layer;
- σ denotes the activation function ReLU;
- BN represents the batch normalization operation.

Definition 3 (AffectAnalyser): *AffectAnalyser* is a network model used to calculate an individual's affective state at the current moment. Each layer performs a linear transformation on the input features using weight matrices and bias vectors, progressively deriving the individual's affective prediction. Specifically, the input features undergo a linear transformation in the first layer, where a linear operation is performed using the weight matrix \mathcal{G}_1 and a bias c_1 is added to obtain an intermediate result. This is then followed by a second linear transformation, processed by the weight matrix \mathcal{G}_2 and bias vector c_2 , ultimately generating the affective state estimate for individual p_i . This model can be formally represented as:

$$AffectAnalyser(\cdot) = \mathcal{G}_2(\mathcal{G}_1 \cdot + c_1) + c_2 \quad (3)$$

Where:

- \mathcal{G}_1 and \mathcal{G}_2 represent the weight matrices of the first and second layers, respectively;
- c_1 and c_2 represent the bias vectors of the first and second layers, respectively.

Problem formulation

The objective of this study is to identify an individual's affective state using data obtained from human monitoring devices. The task is defined as a set $\mathcal{P} = \{p_1, p_2, \dots, p_N\}$ consisting of N individuals, where each individual has k corresponding features $\mathcal{A} = \{a_1, a_2, \dots, a_k\}$. It is assumed that a monitoring device can capture M types of physiological indicators, denoted as $\mathcal{B} = \{b_1, b_2, \dots, b_M\}$, where $\mathcal{B} \subseteq \{\text{ACC, BVP, ECG, EDA, EMG, RESP, TEMP}\}$. The physiological monitoring data of a set of participants can be represented as $D = \{d_{p_i, b_j} \mid p_i \in \mathcal{P}, b_j \in \mathcal{B}\}$. Based on these definitions, the task of this study can be defined as $e_{p_i} = \text{PhysioFormer}(A_{p_i} \parallel d_{p_i, b_j})$, where $e_{p_i} \in \{0, 1, 2\}$, with 0 representing a normal state, 1 representing an excited state, and 2 representing a stressed state.

Methods

The research task of this study can be divided into two parts. First, a PhysioFormer model is constructed to predict an individual's affective state at a given moment. Through the collaborative functioning of three submodules: feature embedding, affective representation, and affective state prediction. The model effectively transforms physiological data into predicted affective states. Second, symbolic regression is applied to fit various monitoring indicators, enabling a more precise capture and understanding of the relationships between these indicators and affective states. This approach not only

enhances the accuracy of affective state predictions but also provides a deeper analysis and explanations of the physiological indicators.

The PhysioFormer model

Model overview. In this section, our task is to predict an individual's affective state using physiological signal data. The structure of the PhysioFormer model proposed in this paper is shown in the Fig 1, and it can be divided into three submodules: (1) **Feature Embedding Module**, which is responsible for encoding the input features and generating feature representations containing physiological data; (2) **Affective Representation Module**, which models and represents the individual's affective state based on the feature representations output by the Feature Embedding Module, capturing and describing the user's affective state through neural networks; (3) **Prediction Module**, which uses the previously obtained features to predict the individual's affective state at the current moment.

Based on the aforementioned model architecture, we define the input and output of the entire model as follows:

- **Input:** The input data consists of various physiological signals, including but not limited to HR, HRV, and EDA. These physiological signals are captured in real time through wearable devices or other physiological monitoring tools and are provided to the model in the form of time series. Each input feature vector not only contains these physiological signals but also includes individual attributes (such as age, gender, etc.), enabling the model to better capture inter-individual differences.
- **Output:** The output of the model is a prediction of the individual's current affective state, providing a classification of the affective state as one of three categories: tense, calm, or excited.

After data preprocessing, the raw physiological signal data is first processed by merging individual attributes features with physiological features to form a comprehensive feature vector. The *ContribNet* is then constructed for each physiological indicator to calculate its contribution level to affective prediction, and an attention mechanism is introduced to re-encode the physiological indicators. The re-encoded physiological indicators are then concatenated with individual attributes again to form a new comprehensive feature vector. For each physiological indicator, we construct an *AffectNet* to calculate the affective state level. Finally, the affective states of all physiological indicators are fed into *AffectAnalyser*, which predicts the individual's affective state at the current moment. The detailed algorithm is shown in Algorithm 1.

Data preprocessing. The raw dataset containing physiological signals typically consists of sampling times and their corresponding measurements. Before using this data as input for the neural network, proper preprocessing is required. The following section will describe in detail the data preprocessing methods employed in this study to ensure data quality.

- (A) Denoising.** During the process of data monitoring, various types of noise are often present, such as power line interference, motion artifacts, and other environmental noise [32]. To address this, we employed a Butterworth low-pass filter to remove high-frequency noise. This is a smoothing filter with maximally flat response characteristics [33]. The filter is used to eliminate high-frequency noise from the signal data and perform signal smoothing, effectively reducing noise and improving the signal-to-noise ratio. The smoothed signal more accurately reflects the participant's actual physiological state, preventing false fluctuations caused by noise.
- (B) Segmentation.** Since physiological signal datasets are time series data, they need to be segmented into fixed-size windows to generate the corresponding time series dataset. The Fig 2 illustrates the process of data segmentation and window sliding used in this study. In the figure, Window1, Window2, and Window3 represent consecutive time windows. Each window contains time series data of the same length, and there is no partial overlap between the windows, allowing them to independently reflect physiological changes during each time period.

The total length of the monitoring data is L seconds, and the length of each window is set to T seconds. Therefore, the monitoring data of participant p_i in the dataset can be divided into $\xi_{p_i} = \left\lfloor \frac{L}{T} \right\rfloor$ non-overlapping time windows. The set

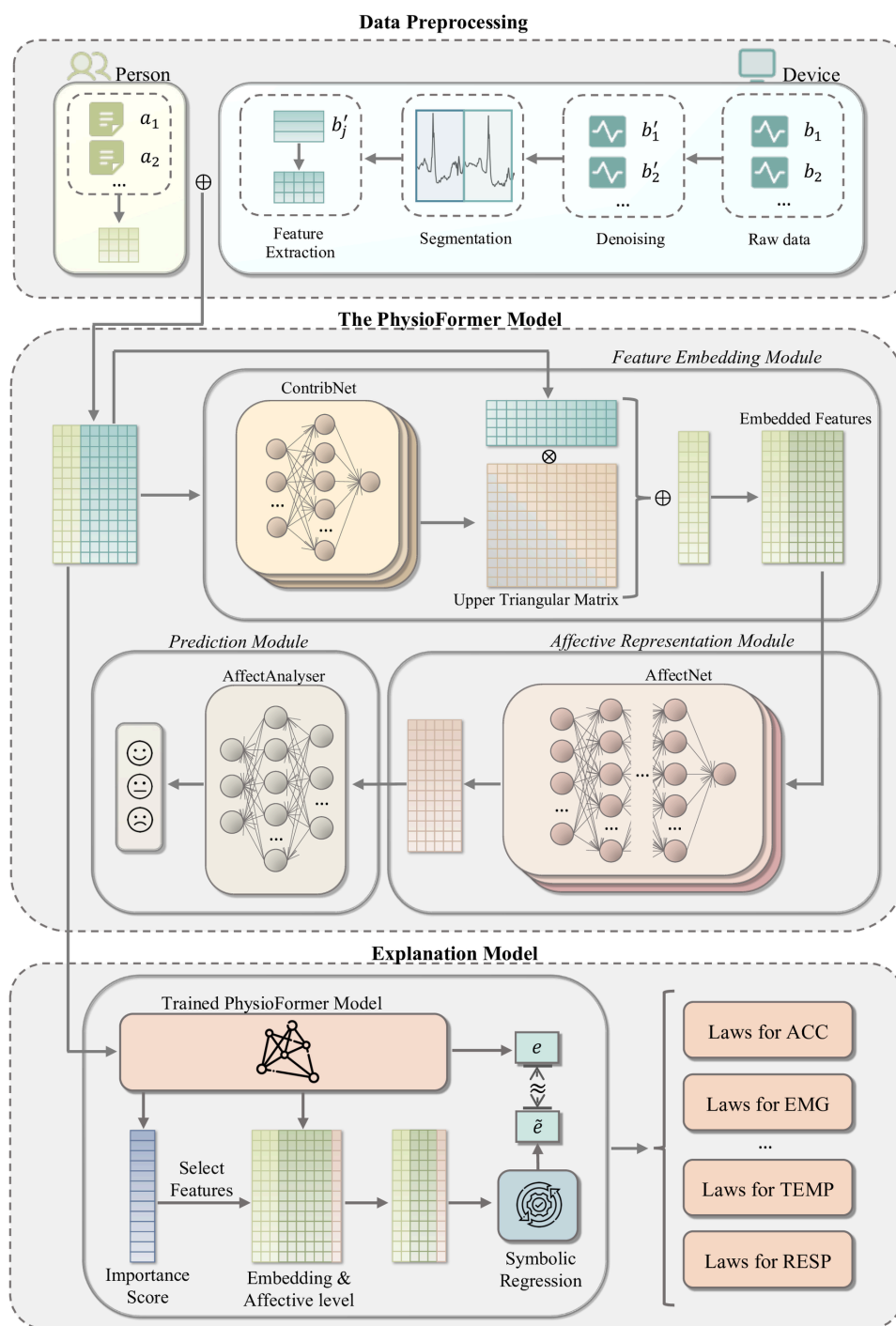


Fig 1. PhysioFormer model architecture consists of three submodules: the Feature Embedding Module, Affective Representation Module, and Prediction Module. The Feature Embedding Module encodes physiological data, the Affective Representation Module builds on these encoded features, and the Prediction Module forecasts the individual's current affective state. The Explanation model analyzes data within the trained model, generating feature importance scores and selecting key features, followed by symbolic regression to derive formulas that explain and quantify the influence of physiological indicators on affective states.

<https://doi.org/10.1371/journal.pone.0335221.g001>

Algorithm 1 The proposed PhysioFormer model for affective computation.

Input: \mathcal{X} which is dataset after window segmentation and feature extraction

Output: Current moment affective categorization result e

```

1: Overall features of subject  $p_i$  under all windows:  $PF_{p_i} \leftarrow \text{Concat}(A_{p_i}, B_{p_i})$ 
2: repeat
3:   The contribution level  $\alpha_{p_i}^{b_j}$  of a physiological indicator  $b_j$ :
      $\alpha_{p_i}^{b_j} \leftarrow \text{ContribNet}(\text{BN}(PF_{p_i}))$ 
4:   Physiological indicator  $b_j$  after embedding:  $\beta_{p_i}^{b_j} \leftarrow B_{p_i}^{b_j} \cdot \text{Triu}$ 
5:   Features after embedding under all windows:  $\gamma_{p_i}^{b_j} \leftarrow \text{Concat}(A_{p_i}, \beta_{p_i}^{b_j})$ 
6:   Affective level reflected in physiological indicator  $b_j$ :  $\theta_{p_i}^{b_j} \leftarrow \text{AffectNet}(\text{BN}(\gamma_{p_i}^{b_j}))$ 
7: until Reaching the steps number equal to physiological indicators number, each get  $\theta_{p_i}^{b_j}$ 
8: Splice all  $\theta_{p_i}^{b_j}$  to obtain  $\Theta_{p_i}$ 
9: Sum with initial affective level  $u_{p_i}$ :  $\Phi_{p_i} \leftarrow \Theta_{p_i} + u_{p_i}$ 
10: Affective categorization result  $e_{p_i} \leftarrow \text{AffectAnalyser}(\Phi_{p_i})$ 
11: return  $e_{p_i}$ 

```

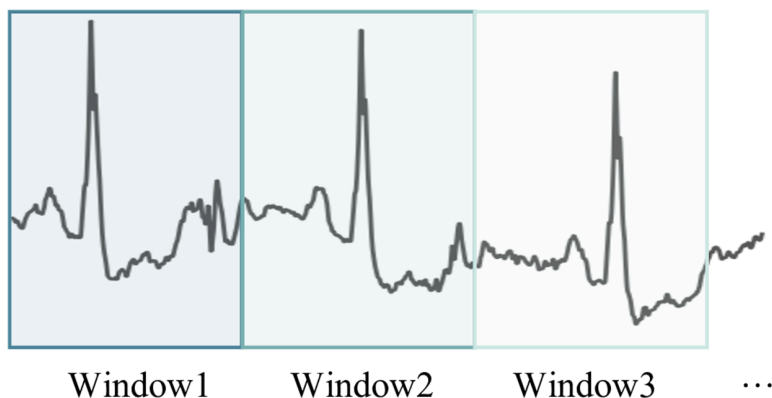


Fig 2. Example of time window segmentation in the dataset. This figure illustrates the time window segmentation process for the physiological signal dataset. The monitoring data is divided into multiple consecutive time windows, with each window being independent and having no overlapping parts.

<https://doi.org/10.1371/journal.pone.0335221.g002>

of all windows can be represented as:

$$\mathcal{X}_{p_i} = \{X_{1_{p_i}}, X_{2_{p_i}}, \dots, X_{\xi_{p_i}}\} \quad (4)$$

The data contained in the n -th window can be represented as:

$$X_{q_{p_i}} = \{x_{p_i}(t) : t \in [(q-1)T, qT]\} \quad (5)$$

Here, $x_{p_i}(t)$ represents the monitoring data at time t . Each window $X_{q_{p_i}}$ is independent, meaning there is no overlap between windows, which helps to avoid data overlap issues during feature extraction and subsequent analysis.

(C) **Feature Extraction.** Physiological signal data from the human body is complex and diverse, making it difficult to directly extract feature information. Therefore, after segmenting the dataset according to window size, feature extraction is necessary to transform the dataset into a format that can reveal affective states and be more easily processed by neural network models. To better understand and handle this data, we analyzed each type of physiological indicator and described the feature extraction methods for each indicator within each window after the fixed-size segmentation.

For each indicator, we calculated its mean, standard deviation, maximum, and minimum values within each window, collectively referred to as the basic statistical features of the signal.

- **ACC:** The ACC data contains three dimensions (x, y, z), and the basic statistical features were calculated for each dimension. Additionally, the sum of the acceleration across the three dimensions was computed, followed by calculating the corresponding basic statistical features.
- **EDA:** A low-pass filter was first applied to the raw data, and the *cvxEDA* [34] algorithm was used to compute the relevant statistical features. From these, three features that reflect both short-term affective responses and long-term affective states, as well as patterns of autonomic nervous system activity, were selected, and their corresponding basic statistical features were calculated.
- **EMG:** A low-pass filter was first applied to the raw data, and then the basic statistical features were calculated within each window.
- **BVP:** For this indicator, the peak frequency within each window was computed. As this indicator reflects heart activity, HRV was also calculated within the window using the *NeuroKit2* tool [35].
- **ECG:** The *NeuroKit2* tool was used to compute HRV within the window, and HRV features such as standard deviation (SDNN) and root mean square of successive differences (RMSSD) were extracted.
- **TEMP:** TEMP is a numerical signal, and its basic statistical features were calculated within each window. Additionally, the slope of TEMP within the window was computed to reveal temperature trends over a specific time period.
- **RESP:** RESP is a numerical signal, and the basic statistical features of the respiratory rate within the window were calculated.

For detailed information on the specific features, please refer to Appendix A, which provides the variable definitions and descriptions for each feature.

Feature embedding module. The core task of feature embedding module is to extract features from the data obtained through sliding windows. This step not only effectively captures the temporal characteristics of physiological signals but also enhances the model's adaptability and reliability to individual differences by incorporating individual attributes. By merging individual attributes with physiological features to form a comprehensive feature vector, the model can gain a more holistic understanding of each individual's unique physiological responses in different affective states. The contribution level of physiological indicators is calculated using *ContribNet*, and an attention mechanism is introduced to re-encode the physiological indicators. This allows the model to dynamically adjust its focus on various physiological signals, ensuring that the most informative features are fully utilized in affective prediction. This refined feature processing method not only improves the accuracy of affective computation but also enhances the model's ability to capture complex affective states, laying a solid foundation for subsequent affective representation and affective computation.

Specifically, the dataset is divided into multiple fixed-length windows, each containing a processed segment of continuous time-series data, which represents the overall features of each physiological indicator. This section uses two types of data to represent feature embedding. One part is the individual attributes features, containing basic information about the individual, such as age, gender, and height. Here, the individual attributes features of participant p_i are denoted as A_{p_i} . The other part is the windowed monitoring data, consisting of the feature values calculated from the physiological indicator data within a given window. The monitoring data for participant p_i is denoted as B_{p_i} . These feature data include both discrete and continuous data. For discrete data, one-hot encoding is used to represent categorical information in a format

suitable for model processing. Continuous data are directly represented by their raw values to preserve their true quantitative information. By integrating individual attributes features A_{p_i} with monitoring data B_{p_i} , the overall features of the participant across all windows can be represented as:

$$PF_{p_i} = A_{p_i} \oplus B_{p_i} \quad (6)$$

Here, \oplus denotes the feature concatenation operation, and $PF_{p_i} \in \mathbb{R}^{k+m}$, where m represents the dimension of the features calculated from the physiological indicator data.

Given that the overall static features PF_{p_i} consist of multiple dimensions and exhibit complex linear relationships, this study constructs a *ContribNet* network to process these high-dimensional features. The network possesses nonlinear modeling capabilities and can effectively capture and represent the complex relationships between input features through a multi-layer structure. To accommodate the characteristics and requirements of different indicators, we built an independent *ContribNet* for each type of physiological indicator. In summary, the contribution level of physiological indicator b_j can be represented as:

$$\alpha_{p_i}^{b_j} = \text{ContribNet}(PF_{p_i}) \quad (7)$$

In the feature encoding process, special attention must be paid to the temporal variation of physiological indicators. One of the key characteristics of time series data is its sequential nature, meaning that the physiological state at a given moment is often influenced by prior moments, with this influence being progressive. Therefore, accurately capturing these dynamic temporal changes is critical during feature extraction and encoding. In the feature encoding process, two types of data are used: individual attributes features and physiological indicators. Individual attributes features can be considered static features that remain constant over the entire time period, such as age and gender, and thus do not require temporal feature fusion. However, physiological indicators vary over time, and it is essential to capture their dynamic characteristics to more accurately reflect the changes in an individual's physiological state over time. To model temporal dependencies, we constructed an upper triangular matrix. This matrix ensures that the features at each time point depend only on the current and prior moments, without relying on future moments. This approach ensures that the model is sensitive to the temporal dynamics of physiological signals, allowing it to capture more subtle emotional changes. Based on the process described above, the encoded physiological indicator b_j for participant p_i can be represented as:

$$\beta_{p_i}^{b_j} = \left((B_{p_i}^{b_j})^T \cdot \text{Triu}(\alpha_{p_i}^{b_j} \cdot \mathbf{1}_{1 \times \xi}) \right)^T \quad (8)$$

Here, *Triu* denotes the upper triangular matrix operation, and ξ represents the number of windows.

Based on the aforementioned encoding process, the overall features of physiological indicator b_j for individual p_i across all windows can be represented as:

$$\gamma_{p_i}^{b_j} = A_{p_i} \oplus \beta_{p_i}^{b_j} \quad (9)$$

Affective representation module. The core task of affective representation module is to map the overall features of an individual across all time windows to affective state levels, which is crucial for accurately predicting affective states. By aggregating features from the time series, the affective representation module can identify and model the temporal patterns of affective states. The *AffectNet* network is used here to map high-dimensional features into a lower-dimensional affective state space, simplifying data complexity while retaining key affective information and handling subtle affective changes. By establishing contextual relationships within the time series, the affective representation module captures the continuity and evolution of affective states, thus more accurately reflecting an individual's actual affective experience, laying the foundation for final affective classification and state prediction.

To achieve this task, we constructed an *AffectNet* network, which has strong feature integration capabilities and can effectively fuse and process multiple types of feature information. This method allows for more precise capture and reflection of the dynamic changes in an individual's affective state, enhancing the accuracy and reliability of affective state predictions. When building the model, an individual *AffectNet* network was constructed for each type of physiological indicator, allowing it to process its corresponding physiological data independently and dynamically adjust its parameters in response to variations in different physiological indicators. This enables more precise feature extraction and affective state prediction. In summary, the affective state level reflected by physiological indicator b_j can be represented as:

$$\theta_{p_i}^{b_j} = \text{AffectNet}(\gamma_{p_i}^{b_j}) \quad (10)$$

Here, $\Theta_{p_i} = \{\theta_{p_i}^{b_j} | i \in \mathbf{N}, j \in \mathbf{M}\}$ is used to represent the set of affective state levels mapped from all physiological indicators.

Additionally, to improve overall prediction accuracy and model reliability while reducing error accumulation, we introduced the initial affective state as a baseline, denoted as u_{p_i} , representing the initial affective state level before training. Thus, Θ_{p_i} is adjusted to:

$$\Phi_{p_i} = \Theta_{p_i} + u_{p_i} \quad (11)$$

Thus, the model is not only able to capture affective state changes caused by variations in physiological indicators but also takes into account the individual's initial state, providing a more comprehensive affective state assessment.

Prediction module and model training. The core task of the Prediction module is to use the individual's affective state levels to predict their current affective state. This module analyzes and processes the affective state levels output by the affective representation module and applies the *AffectAnalyser* network to map the affective state levels to specific affective categories.

We constructed an *AffectAnalyser* network to accomplish this task. By mapping the input feature vector to a high-dimensional feature space, it captures the complex relationships between input features, enabling effective classification. Based on the adjusted affective state levels Φ_{p_i} obtained earlier, the final affective prediction is represented as:

$$e_{p_i} = \text{AffectAnalyser}(\Phi_{p_i}) \quad (12)$$

During the training process, it is essential to account for the complexity and diversity of physiological data. Therefore, we chose to use the cross-entropy loss function. The cross-entropy loss function is well-suited for classification problems, as it measures the difference between the predicted probability distribution and the true label distribution, providing a natural probabilistic explanations and exhibiting favorable gradient properties. Additionally, we introduced a regularization term into the loss function, which helps constrain the size of the model parameters, smoothens them, prevents overfitting, and improves the model's generalization ability.

The complete loss function for the model is as follows:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N \log \left(\frac{\exp(\hat{e}_{p_i})}{\sum_{k=1}^3 \exp(e_{p_i}^{(k)})} \right) + \lambda \cdot \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{j=1}^M (\alpha_{p_i}^{b_j} - 1)^2 \quad (13)$$

The above function consists of two parts: the first part is the Cross Entropy loss function, where e_{p_i} represents the predicted affective state of the i -th individual, and \hat{e}_{p_i} represents the true affective state of the i -th individual. This function is

used to calculate the difference between the predicted and actual values in the classification task. The second part is the regularization term, where λ is the regularization coefficient that controls the strength of the regularization term.

Based on the aforementioned loss function, we selected the **RMSprop** optimizer to optimize the model. RMSprop dynamically adjusts the learning rate for each parameter by taking the moving average of the squared gradients, thus providing an adaptive learning rate for each parameter. Given that the loss function includes both the cross-entropy loss and the regularization term, this may result in a non-stationary objective function, meaning that statistical data may exhibit time-varying and time-dependent properties during different periods [36]. The RMSprop optimizer calculates the moving average of the squared gradients using exponential decay, effectively handling this non-stationary objective function and significantly improving the stability of the optimization process. Additionally, since RMSprop automatically adjusts the learning rate, it typically converges faster compared to fixed learning rate optimizers such as Stochastic Gradient Descent (SGD).

In this study, the model's training mode adopts a cross-validation strategy. Specifically, in each epoch, the model sequentially uses data from each individual for training, with the goal of explicitly learning and adapting to the physiological signal characteristics of different individuals. This method allows the model to progressively capture the unique physiological features of each individual, enhancing its ability to generalize across individuals. Specifically, in each epoch, the model is trained through the following process.

For the i -th participant p_i , the prediction of their affective state can be expressed as:

$$e_{p_i} = \text{PhysioFormer}(A_{p_i} || d_{p_i, b_j}) \quad (14)$$

Here, A_{p_i} represents the individual attributes features, and d_{p_i, b_j} represents the physiological data features.

For each participant p_i , the loss function \mathcal{L} is used to measure the difference between the predicted value e_{p_i} and the true value $hate_{p_i}$:

$$\mathcal{L}(e_{p_i}, \hat{e}_{p_i}) = \mathcal{L}(\text{PhysioFormer}(A_{p_i} || d_{p_i, b_j}), \hat{e}_{p_i}) \quad (15)$$

In each epoch, the model sequentially uses data from each participant for training and updates the model parameters using the gradient descent method. For the i -th participant in the t -th epoch, the parameter update formula is:

$$\varphi^{(t)} \leftarrow \varphi^{(t-1)} - \eta \cdot \nabla_{\varphi} \mathcal{L}(e_{p_i}, \hat{e}_{p_i}) \quad (16)$$

Here, $\varphi^{(t)}$ represents the model parameters at the end of the current epoch, and $\varphi^{(t-1)}$ represents the model parameters at the end of the previous epoch.

In each epoch, the model is trained on data from all N participants. The entire training process can be represented as:

$$\text{For each epoch } t : \text{For each } i \in \{1, 2, \dots, N\} : \varphi^{(t)} \leftarrow \varphi^{(t-1)} - \eta \cdot \nabla_{\varphi} \mathcal{L}(e_{p_i}, \hat{e}_{p_i}) \quad (17)$$

Here, $\mathcal{L}(e_{p_i}, \hat{e}_{p_i})$ represents the loss function for each participant, and η represents the learning rate.

Symbolic laws extraction

Extraction process overview. The physiological signal data involved in affective computing tasks are complex and multidimensional, often containing highly nonlinear relationships. Although the PhysioFormer model can capture these intricate dynamics, its “black box” nature tends to limit explainability, making it difficult to clearly understand the relationship between inputs and outputs. In this task, we aim to uncover the specific roles of various physiological indicators in affective computation, as well as the complex interactions between physiological signals. This will not only help us better

understand how multiple physiological indicators jointly influence affective states, but also enhance the trust of users and researchers in the model's predictions.

To achieve this goal, we designed an Explanation Model that combines symbolic distillation with deep learning to improve both the explainability and performance of the model. The detailed process structure is shown in the Fig 3. First, the input data is processed through the Feature Embedding and Affective Representation modules, generating features and state information. The trained PhysioFormer model generates feature importance scores, based on which key features are selected. Using symbolic regression, predicted values \tilde{e} are generated and compared to the model's predicted values e . This process extracts symbolic laws for each physiological indicator, making the internal mechanisms of the model more transparent and providing explanations and quantifications of the influence of different physiological indicators on affective states.

Symbolic distillation for physiological indicators. To enhance the explainability of the internal mechanisms of the PhysioFormer model, we introduced a Explanation Model, transforming the influence of various physiological indicators on affective state prediction into symbolic expressions. In this study, we employed symbolic regression techniques with the aim of finding the optimal mathematical formula that fits the given data, thereby revealing the model's intrinsic decision logic and the interactions between physiological indicators. Let P represent the network model to be analyzed, S represent the symbolic regression model, and $X \in \mathbb{R}^{\xi \times (k+m)}$ represent the input data to the model. According to the objective of symbolic regression, we can define the optimization goal of symbolic regression as:

$$\varepsilon = \arg \min_{\varepsilon} |P(X) - S(\varepsilon, X)| \quad (18)$$

Here, ε represents the mathematical formula calculated by symbolic regression, X denotes the range of input data, ξ represents the number of windows into which the individual's physiological data is segmented, k denotes the feature

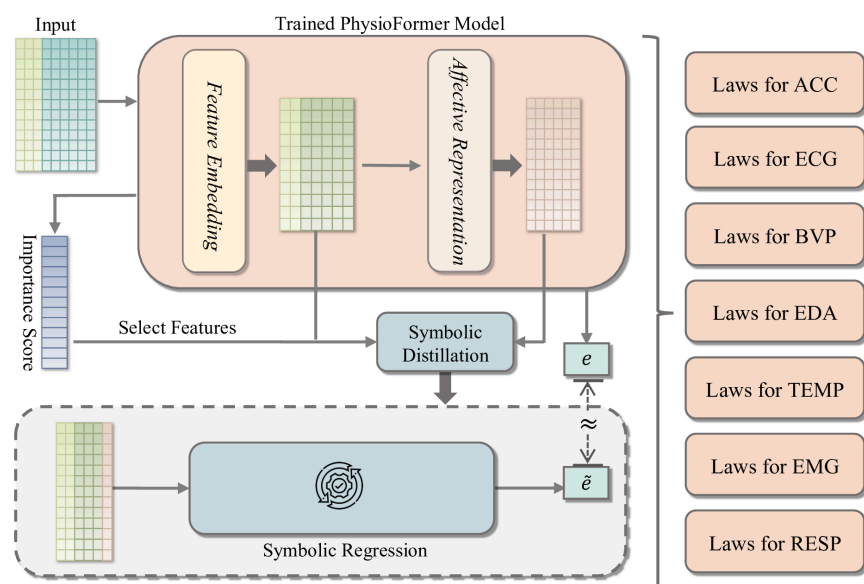


Fig 3. The figure shows the process of Explanation model. First, the input data undergoes feature extraction and affective state representation through the Feature Embedding and Affective Representation modules. The processed features and state information are used to symbolic distillation, where feature importance scores generated by the PhysioFormer model are used to select key features. Next, symbolic regression is employed to generate the predicted value \tilde{e} , which is compared with the model's predicted value e thereby extracting and generating symbolic laws for the physiological indicators.

<https://doi.org/10.1371/journal.pone.0335221.g003>

dimension of the individual attributes features, and m represents the dimension of the features computed from the physiological data.

Since symbolic regression is a typical combinatorial optimization problem, the solution space grows exponentially with the number of variables. Using all the features for computation would incur enormous computational costs and reduce the model's efficiency and stability. To address this issue, we applied a gradient-based feature importance estimation method to select a subset of features, which were then used for the symbolic regression task. Specifically, this method calculates the gradient of the intermediate outputs of the model with respect to the input features. These gradients reflect how small changes in the input features influence the output [37]. The resulting gradients may be either positive or negative, but we are concerned with the magnitude of change rather than the direction. Therefore, the absolute value of the gradients is taken to represent feature importance, and these absolute values are normalized to obtain relative importance scores. This process can be expressed by the following formula:

$$S = \sum_{i=1}^{\xi} y_i \quad (19)$$

Here, y_i represents the intermediate output of the model.

The gradient of the output S with respect to the input features X is calculated and then the absolute value is taken:

$$\left| \frac{\partial S}{\partial X} \right| = \left\| \left[\frac{\partial S}{\partial X_{11}}, \frac{\partial S}{\partial X_{12}}, \dots, \frac{\partial S}{\partial X_{\xi(k+m)}} \right] \right\| \quad (20)$$

The importance score for each feature is calculated:

$$I_j = \frac{\sum_{i=1}^{\xi} G_{ij}}{\sum_{j=1}^{(k+m)} \sum_{i=1}^{\xi} G_{ij}} \quad (21)$$

Here, I_j represents the importance score of the j -th feature.

Based on the calculated feature importance scores, a subset of features is selected for symbolic regression, denoted as $\rho(X)$. Based on the aforementioned process, the optimization objective of symbolic regression can be defined as:

$$\varepsilon = \arg \min_{\varepsilon} |P(X) - S(\varepsilon, \rho(X))| \quad (22)$$

During the model optimization process, the absolute error loss function was used to optimize the model. Since each indicator's *ContribNet* and *AffectNet* operate independently, these components can perform symbolic regression in parallel. However, considering the coupling relationships and sequential dependencies between the components, the analysis process must be carried out step-by-step according to a predefined order to ensure overall model consistency and optimal performance.

Experiments

This section outlines the experimental design for comparing the performance of the PhysioFormer model with nine existing affective computation networks on two sub-datasets, Wrist and Chest, from the publicly available WESAD dataset. The goal is to evaluate the effectiveness of the proposed framework in affective computation tasks. Through this experiment, we aim to address five core research questions (RQ), providing empirical evidence to thoroughly understand the

advantages and contributions of PhysioFormer and further validate its potential application in the field of affective computation.

- RQ1: How does the performance of the PhysioFormer model compare to existing SOTA models in affective computation tasks? This research question seeks to evaluate the advantages of the PhysioFormer model in processing multimodal physiological signals.
- RQ2: Does the choice of different window sizes during data preprocessing significantly affect the model's prediction performance? This question explores how the window size, a critical preprocessing parameter, impacts the model's effectiveness and accuracy.
- RQ3: How does the number of hidden layer neurons at different scales affect the performance of the PhysioFormer model during training? This research question aims to assess the balance between the number of neurons in the model's structure, model complexity, and performance.
- RQ4: Does the feature embedding module have a significant impact on the model's prediction performance? This question seeks to uncover the role of feature embedding in extracting useful information and enhancing model performance, providing insights into its importance.
- RQ5: Does the inclusion of individual attributes features (e.g., age, gender) significantly improve the model's prediction performance? This question investigates the role of individual characteristics in affective computation tasks and further explains their impact on the model's generalization ability.

Dataset

The dataset used in this study is the publicly available WESAD dataset, designed for affective computation. It contains physiological and motion data recorded by two devices: one from a chest-worn device (RespiBAN), which includes six monitoring indicators collected at a frequency of 700Hz; and another from a wrist-worn device (Empatica E4), which includes four monitoring indicators with different sampling frequencies [38]. The physiological indicators included in this dataset are introduced below:

- Accelerometer, ACC: ACC data records changes in acceleration in three-dimensional space (x, y, z axes), reflecting the intensity, direction, and frequency of body movement.
- Electrodermal Activity, EDA: EDA data records changes in skin conductance, which are often closely related to psychological states.
- Electromyography, EMG: EMG data captures the electrical activity of muscle fibers during contraction and relaxation, providing information on muscle function and nervous system control.
- Blood Volume Pulse, BVP: BVP data records changes in blood volume through peripheral vessels with each heartbeat, which can be used to monitor cardiovascular health and analyze heart rate variability (HRV).
- Electrocardiogram, ECG: ECG data records the electrical activity generated by the heart during each heartbeat, providing insights into heart health.
- Temperature, TEMP: TEMP data records measurements of the body's core temperature, reflecting the individual's temperature status at specific time points.
- Respiration, RESP: RESP data captures the respiratory cycle and airflow changes during breathing, reflecting an individual's respiratory rate and depth.

Compared method

In this subsection, we provide an overview of the nine models used in the comparison experiments, which include both traditional machine learning methods and deep learning approaches. Specifically, the machine learning methods include

Random Forest, SVM, KNN, AdaBoost, Decision Tree, and Linear Discriminant Analysis (LDA); the deep learning methods include Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM). These models have been widely applied in previous affective computation studies and have achieved significant results. To ensure fairness and reliability in the comparison experiments, we reconstructed these models in this study and conducted experiments on the two sub-datasets (Wrist and Chest) of the WESAD dataset. Through this process, we are able to evaluate the performance differences between the PhysioFormer model and existing models under the same data conditions, thereby validating the effectiveness of the proposed approach and its advantages in affective computation tasks.

- Bobade et al. [39] proposed a stress detection system based on multimodal physiological data, utilizing machine learning to identify individuals' stress states. The models used for classification tasks include KNN, LDA, Random Forest, Decision Tree, AdaBoost, and SVM, along with a simple feedforward neural network (ANN). The classification tasks were divided into two categories: a three-class task (pleasant, neutral, and stress) and a binary task (stress vs. non-stress). The experimental results showed that in the three-class task, the best accuracy achieved by machine learning methods was 81.65%, while in the binary task, the accuracy reached 93.20%. With deep learning methods, the accuracy for the three-class and binary tasks increased to 84.32% and 95.21%, respectively. Additionally, the study explored different feature extraction and preprocessing steps, including Principal Component Analysis (PCA) and normalization techniques, to optimize classification performance. Ultimately, the deep learning-based ANN model outperformed traditional machine learning methods in both classification tasks, becoming the best model. In my research, I have adopted the Decision Tree, SVM, and AdaBoost algorithms from this study as baseline models for comparison experiments, in order to further evaluate the performance of the PhysioFormer model in multimodal affective computation tasks.
- Siirtola et al. [40] conducted continuous stress detection using sensors from a commercial smartwatch, focusing on how stress recognition can be achieved through physiological signals such as skin temperature (ST), BVP, and HR without relying on EDA signals or user dependence. The study utilized the WESAD dataset and conducted experimental comparisons using three classifiers: LDA, Quadratic Discriminant Analysis (QDA), and Random Forest. The experimental results showed that the LDA classifier combined with ST, BVP, and HR signals performed best in the stress recognition task, achieving a balanced accuracy of 87.4%. In this study, I drew on the work of Siirtola et al. by using Random Forest and LDA classifiers as the basis for comparison experiments. I further expanded the experiments by testing the impact of different combinations of physiological signals on the model's performance.
- Ferdinando et al. [41] aimed to process EDA signals using the cvxEDA method to extract key features for affective recognition and used the KNN classifier to address a three-class affective recognition problem on the MAHNOB-HCI dataset. The experimental results showed that under subject-dependent conditions, the recognition accuracies for valence and arousal reached 74.6% and 77.3%, respectively. In my research, I adopted the KNN model from this study as part of the comparison experiments and combined different physiological signal sets to validate the performance of my model.
- Yu et al. [42] conducted an in-depth exploration of affective recognition based on EDA signals, focusing on the structure and performance of three deep neural network (DNN) models: ResNet, LSTM, and a hybrid model combining ResNet and LSTM. The study was based on the MAHNOB-HCI dataset and experimented on a three-class affective recognition task for valence and arousal dimensions. The experimental results showed that the ResNet model outperformed both LSTM and the hybrid model in affective recognition tasks, achieving an accuracy of 86.73% and an F1 score of 85.71% for valence recognition, and 86.92% accuracy and 85.96% F1 score for arousal recognition. In my research, I adopted the CNN (ResNet) model, which performed exceptionally well in capturing both static and dynamic features. Inspired by LSTM's strength in handling time-series data, I further introduced a RNN as a comparison model. By combining different physiological signal sets and evaluating the performance of ResNet, LSTM, and RNN models on the Wrist and Chest

sub-datasets of the WESAD dataset, I conducted an in-depth analysis of the performance differences between these models in multimodal affective computation tasks.

Evaluation metrics and basic parameterization

In this experiment, we used three metrics to evaluate the model's performance: accuracy (ACC), F1-Score, and Mean Squared Error (MSE). To calculate ACC and F1-Score, we first need to compute the confusion matrix. The formula is as follows:

$$\text{Confusion Matrix} = \begin{pmatrix} TP_A & FP_B & FP_C \\ FN_A & TP_B & FP_C \\ FN_A & FN_B & TP_C \end{pmatrix} \quad (23)$$

Here, TP represents the number of samples correctly predicted as the correct class, FP represents the number of samples incorrectly predicted as the correct class, and FN represents the number of samples incorrectly predicted as the wrong class.

ACC calculates the proportion of correctly predicted samples among all test samples. Based on the confusion matrix, the formula for calculating ACC is as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (24)$$

F1-Score is the harmonic mean of Precision and Recall, and it can effectively handle situations with imbalanced data distributions. For each class $i \in \{A, B, C\}$, the formula for calculating F1-Score is as follows:

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (25)$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (26)$$

$$F1 - Score_i = 2 \cdot \frac{Precision_i \cdot Recall_i}{Precision_i + Recall_i} \quad (27)$$

MSE calculates the average of the squared differences between the predicted values and the actual values, providing a measure of the degree of difference between the model's predictions and the true values. The formula for MSE is:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (28)$$

Here, y_i and \hat{y}_i represent the predicted value and the true value, respectively.

The code framework for the experiments in this paper was implemented using Python 3.10 and PyTorch 2.2.2. The experiments were conducted on a Mac with an Apple M1 Pro CPU (8 cores) with 16GB of memory and a Windows machine equipped with a GeForce GTX 1650 GPU with 4GB of memory.

The basic configuration parameters are as follows:

- The maximum number of epochs was set to 150, with an early stopping mechanism to prevent overfitting.
- The initial learning rate was set to 1e-4, and a StepLR scheduler was used to dynamically adjust the learning rate.
- The batch size was adjusted depending on the dataset and window size.

Experiment result and analysis

Performance evaluation results (RQ1). In this experiment, we used a 30-second window to segment the dataset and adopted a Leave-One-Subject-Out cross-validation strategy to ensure subject-independent evaluation. In each fold, the data from one participant were reserved for testing, while the data from all remaining participants were used for training. This procedure effectively avoids potential data leakage caused by intra-subject autocorrelation and provides a more reliable assessment of cross-subject generalization. We then conducted comparative experiments between the PhysioFormer model and various traditional machine learning and deep learning models. The traditional machine learning models include Random Forest, SVM, KNN, AdaBoost, Decision Tree, and LDA, while the deep learning models include CNN, RNN, and LSTM. The specific experimental results are shown in Table 3:

The experimental results show that the PhysioFormer model outperformed all other models across all metrics. Its ACC on the Wrist and Chest datasets reached 99.54% and 99.21%, respectively, with F1-scores of 99.54 and 99.17, and the lowest MSE of 0.025 and 0.057, respectively. This demonstrates that PhysioFormer exhibits exceptionally high predictive accuracy and stability when processing these datasets.

In contrast, SVM performed exceptionally well among traditional machine learning algorithms, particularly on the Chest dataset, where both ACC and F1-score exceeded 98%, and MSE remained low at 0.134 and 0.128, respectively. This indicates that SVM handles high-dimensional data very effectively. The Random Forest model also performed well on both the Wrist and Chest datasets, with ACC and F1-scores approaching 97%, and MSEs of 0.151 and 0.122, respectively, demonstrating its high accuracy and low error when processing physiological data. KNN's performance on the Wrist dataset was close to that of Random Forest, but it underperformed slightly on the Chest dataset, particularly in terms of F1-score and MSE. This could be related to KNN's sensitivity to data distribution. AdaBoost and Decision Tree models performed slightly worse than the previously mentioned models, particularly in terms of ACC and F1-score. For example, AdaBoost achieved ACC values of 90.50% and 91.74% on the Wrist and Chest datasets, respectively, while Decision Tree performed noticeably better on the Chest dataset (95.41%) than on the Wrist dataset (88.27%). LDA performed the worst, especially on the Chest dataset, where ACC and F1-score were 72.93% and 68.69%, respectively, and the error was high, indicating that LDA struggles with this type of complex data.

Among deep learning models, CNN performed relatively well, particularly on the Wrist dataset, where ACC reached 98.33%, F1-score was 97.95%, and MSE was low at 0.065, showing CNN's strength in handling physiological data with spatial features. However, CNN's performance significantly dropped on the Chest dataset, with ACC at only 81.46%, F1-score falling to 75.07%, and MSE increasing to 0.185. This performance degradation may be attributed to the relatively

Table 3. The performance comparison between the proposed PhysioFormer Model and the state-of-the-art (SOTA) methods in terms of AUC (%), F1-Score and MSE on WESAD dataset. The performance results reported in the table correspond to the best accuracy achieved for each model under its optimal hyperparameter settings.

	ACC		F1-Score		MSE	
	Wrist	Chest	Wrist	Chest	Wrist	Chest
Random Forest	97.46	97.71	97.40	97.71	0.151	0.122
SVM	97.88	98.62	97.88	98.58	0.134	0.128
KNN	97.45	96.79	97.62	96.20	0.138	0.146
AdaBoost	90.50	91.74	90.53	91.76	0.175	0.179
Decision Tree	88.27	95.41	88.29	95.39	0.205	0.146
LDA	85.20	72.93	84.51	68.69	0.303	0.298
CNN	98.33	81.46	97.95	75.07	0.065	0.185
RNN	84.93	79.26	84.98	80.03	0.391	0.437
LSTM	96.87	87.00	97.02	87.39	0.184	0.400
PhysioFormer	99.54	99.21	99.54	99.17	0.025	0.057

<https://doi.org/10.1371/journal.pone.0335221.t003>

limited amount of available data in the Chest subset, which restricted the CNN's ability to fully capture spatial representations. Consequently, CNN demonstrated weaker adaptability under conditions of data scarcity compared to other models. LSTM showed stability in handling time-series data, achieving 96.87% ACC, 97.02% F1-score, and 0.184 MSE on the Wrist dataset. Although its performance declined on the Chest dataset (ACC: 87.00%, F1-score: 87.39%), it was still more stable than RNN. RNN showed relatively poor performance on both datasets, especially on the Chest dataset, where ACC and F1-score were 79.26% and 80.03%, respectively, with a high MSE of 0.437, indicating that RNN may have limitations in capturing long-term dependencies and handling more complex physiological signals.

Based on the experimental results, it is evident that the PhysioFormer model addresses many of the shortcomings present in traditional machine learning and deep learning models, demonstrating its unique advantages and outstanding performance. Firstly, PhysioFormer model, through its advanced architecture design, effectively overcomes the limitations of traditional machine learning models in handling multimodal physiological signals. By integrating multi-level feature embedding and affective representation modules, it extracts more representative and distinctive features from multimodal signals, excelling across diverse data types. Secondly, PhysioFormer model excels in addressing the stability issues found in deep learning models. Its adaptive feature processing mechanism not only captures dynamic changes in time-series data but also effectively fuses information across different modalities, ensuring consistent performance across various datasets. This stability gives PhysioFormer model a significant advantage in handling complex real-world scenarios. Finally, PhysioFormer model's low MSE indicates minimal prediction error, which is attributed to its innovative design in multimodal fusion and time-series modeling. By introducing a feature embedding module, PhysioFormer model is better equipped to handle long-term dependencies and sequential information, making it far superior to other models in affective computation tasks.

Convergence rate. For each dataset, we trained the PhysioFormer model and recorded its loss values during the training process. The variation in the loss values over time during training is shown in the Fig 4.

The results show that all models exhibit a rapid decline in loss values during the initial stages of training, indicating that the models quickly learn effective features, significantly reducing prediction error. More importantly, all models demonstrate good convergence in the later stages of training, with the final loss values stabilizing, suggesting that the PhysioFormer model is capable of achieving convergence when processing both the Wrist and Chest datasets with different time windows.

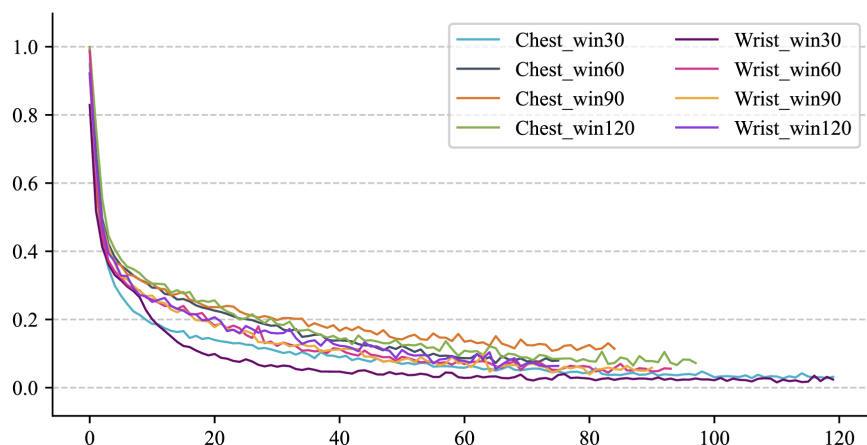


Fig 4. Convergence trends of the PhysioFormer model across datasets by splitting the WESAD dataset through windows of different sizes. Although convergence trends are shown on all datasets, there are differences in the speed of convergence and the magnitude of losses.

<https://doi.org/10.1371/journal.pone.0335221.g004>

Notably, the datasets with different time windows show some variation in convergence speed and final loss values. Although models with all time windows eventually converge, the dataset with a 30-second window exhibits faster convergence and lower final loss values. This indicates that the model can reach an optimal state more quickly with shorter time windows, while also reducing the risk of overfitting during training. In contrast, as the time window length increases, the model's convergence slows down slightly, and the final loss value is slightly higher. This may be attributed to the longer time windows introducing more temporal dependencies; while such information helps the model capture long-term trends, it also increases optimization difficulty, requiring more time for the model to converge.

Overall, the experimental results indicate that the PhysioFormer model consistently converges across datasets with different time windows, and it shows faster convergence and lower final loss values with shorter time windows. This further validates the efficiency and reliability of the PhysioFormer model in processing multimodal physiological signals.

Sensitivity tests. When designing and optimizing complex models, sensitivity tests is a critical step. Through sensitivity experiments, we can systematically evaluate the model's response to different parameters, revealing how these parameters impact the model's performance. This not only helps to understand the internal mechanisms of the model but also provides a scientific basis for optimizing model parameters. Specifically, in this experiment, we explored the effects of adjusting the window size and the number of hidden layer neurons in the PhysioFormer model on its performance across different datasets. This allowed us to identify the optimal model configuration.

The effect of window size. (RQ2) This experiment aimed to evaluate the impact of different window sizes on the model's performance. We selected four different window lengths: 30 seconds, 60 seconds, 90 seconds, and 120 seconds, and tested them on the sensor data from both the Wrist and Chest positions. For each window size, we calculated the model's ACC, F1-Score, and MSE to comprehensively assess its performance. The experimental results are shown in the Table 4.

In this experiment, we evaluated the impact of different window sizes (30 seconds, 60 seconds, 90 seconds, and 120 seconds) on the performance of the PhysioFormer model. The results show that window size significantly affects the model's predictive performance. With a 30-second window, the model performed best, achieving ACC of 99.54% and 99.21%, and F1-Score of 99.54 and 99.17 on the Wrist and Chest datasets, respectively, with the lowest MSE of 0.025 and 0.057. This suggests that shorter time windows can more effectively capture subtle variations in physiological signals, resulting in higher predictive accuracy and stability.

However, as the window size increased, the model's performance declined. With a 60-second window, although the model maintained relatively high accuracy and F1-Score, MSE increased, indicating that handling longer time sequences introduced more temporal dependencies, adding complexity and increasing prediction error. For the 90-second and 120-second windows, particularly for the 90-second Chest dataset, accuracy and F1-Score dropped significantly, and MSE increased further, suggesting that excessively long windows may make it difficult for the model to handle complex temporal dependencies, affecting prediction performance.

In summary, the 30-second window provides the optimal balance, capturing sufficient information while avoiding redundancy and complexity, ensuring higher predictive accuracy and lower error.

Table 4. The performance comparison of the proposed PhysioFormer model on different datasets divided by different window sizes in terms of AUC (%), F1-Score and MSE.

	ACC		F1-Score		MSE	
	Wrist	Chest	Wrist	Chest	Wrist	Chest
30s	99.54	99.21	99.54	99.17	0.025	0.057
60s	97.61	97.52	97.59	97.53	0.108	0.117
90s	97.87	95.45	97.86	95.38	0.081	0.203
120s	97.70	96.85	97.68	96.50	0.089	0.141

<https://doi.org/10.1371/journal.pone.0335221.t004>

The effect of the number of neurons in the hidden layer. (RQ3) This experiment aimed to evaluate the impact of the number of hidden layer neurons in the *ContribNet* and *AffectNet* models within the PhysioFormer framework on overall model performance. To achieve this, we selected four different scales of neuron counts (50, 100, 200, 500) and conducted experiments on both the Wrist and Chest datasets to assess the performance of the *ContribNet* and *AffectNet* models within PhysioFormer. The experimental results are shown in the [Tables 5 and 6](#).

On the Wrist dataset, the performance of the model fluctuated with an increase in the number of neurons in the *ContribNet* and *AffectNet* models. For instance, when the number of neurons in *ContribNet* was fixed at 50, the combination with 500 neurons in *AffectNet* performed best, achieving an accuracy of 99.03%. However, when the number of neurons in *ContribNet* was increased to 100, the model reached its peak performance, with an accuracy of 99.54% when *AffectNet* also had 100 neurons. Overall, the results from the Wrist dataset indicate that increasing the number of neurons beyond a certain point did not continue to improve the model's accuracy. In some cases, performance even declined, which could be related to overfitting due to the increased complexity of the model.

On the Chest dataset, the results similarly showed the impact of neuron count on model performance. When the number of neurons in both *ContribNet* and *AffectNet* was set to 50, the model performed relatively well, achieving an accuracy of 99.08%. The best performance, with an accuracy of 99.21%, was observed when both *AffectNet* and *ContribNet* had 100 neurons. As the neuron count increased, particularly with configurations of 500 neurons, the model's performance stabilized but did not significantly surpass the performance with fewer neurons.

Overall, the results suggest that while increasing the number of hidden layer neurons can improve the model's performance, this improvement is limited. Too many neurons can lead to increased model complexity, which raises the risk of overfitting, especially with relatively limited data. The best model performance often occurred with moderate neuron configurations, such as 100 or 200 neurons. In conclusion, setting the number of neurons in both *ContribNet* and *AffectNet* to 100 offers optimal performance for the PhysioFormer model.

Ablation studies. Ablation studies are used to assess the contribution of various components or features within a model. They help evaluate the actual impact of each model component. In this experiment, we gradually removed specific modules from the model to evaluate their influence on overall performance, thereby validating their effectiveness. Specifically, we designed two ablation experiments: one to verify the importance of the Feature Embedding module and another

Table 5. The performance comparison of the proposed PhysioFormer model on the Wrist dataset with different numbers of hidden layer neurons in the *ContribNet* and *AffectNet* models in terms of AUC (%).

Wrist					
AffectNet	ContribNet				
		50	100	200	500
	50	96.97	97.39	97.87	97.45
	100	97.51	99.54	96.97	98.00
	200	98.06	98.30	97.69	98.87
	500	99.03	98.56	98.84	98.78

<https://doi.org/10.1371/journal.pone.0335221.t005>

Table 6. The performance comparison of the proposed PhysioFormer model on the Wrist dataset with different numbers of hidden layer neurons in the *ContribNet* and *AffectNet* models in terms of AUC (%).

Chest					
AffectNet	ContribNet				
		50	100	200	500
	50	99.08	96.78	96.96	97.24
	100	98.80	99.21	97.33	97.61
	200	98.52	97.98	99.08	98.80
	500	99.08	98.99	98.90	98.90

<https://doi.org/10.1371/journal.pone.0335221.t006>

to assess the role of individual attributes in the model. The results of these experiments will provide scientific evidence for model optimization and improvement.

The Validity of Feature Embedding. (RQ4) In this experiment, we evaluated the performance differences between the model with the feature embedding module and the model without it on the Wrist and Chest datasets. To assess the effectiveness of the feature embedding module, we conducted two sets of comparison experiments: one using the complete model with the feature embedding module, and the other by removing it. ACC was used as the primary evaluation metric. The experimental results are shown in the Fig 5.

According to the experimental results, the model with the feature embedding module achieved an accuracy of 99.54% on the Wrist dataset, compared to 97.65% for the model without it, showing an improvement of approximately 1.89%. On the Chest dataset, the model with the feature embedding module achieved an accuracy of 99.21%, while the model without the feature embedding module had an accuracy of 89.87%, reflecting an improvement of about 9.34%.

These results demonstrate that the model with the physiological data feature embedding module outperforms the one without it on both the Wrist and Chest datasets, with a particularly notable improvement on the Chest dataset. This indicates that the feature embedding module effectively quantifies and highlights features that significantly contribute to physiological signal prediction, better capturing key characteristics in the data and enhancing the model's predictive capability.

The Validity of Individual Attributes Features. (RQ5) In this experiment, we evaluated the performance differences between models that include individual attributes features (such as age, gender, etc.) alongside physiological monitoring data and models that only use physiological monitoring data on the Wrist and Chest datasets. The experiment was designed in two parts: one model used the complete input data, including individual attributes features and physiological monitoring data, while the other model excluded the individual attributes features and used only the physiological monitoring data for training and testing. This comparative experiment allowed us to assess the impact of individual attributes features on the model's performance. ACC was used as the primary evaluation metric. The experimental results are shown in the Fig 6.

The model including individual attributes features data achieved an accuracy of 99.54% on the Wrist dataset, compared to 97.61% for the model without individual attributes features, reflecting an improvement of approximately 1.93%. On the Chest dataset, the model with individual attributes features reached an accuracy of 99.21%, while the model without these features achieved 98.85%, showing an improvement of about 0.36%.

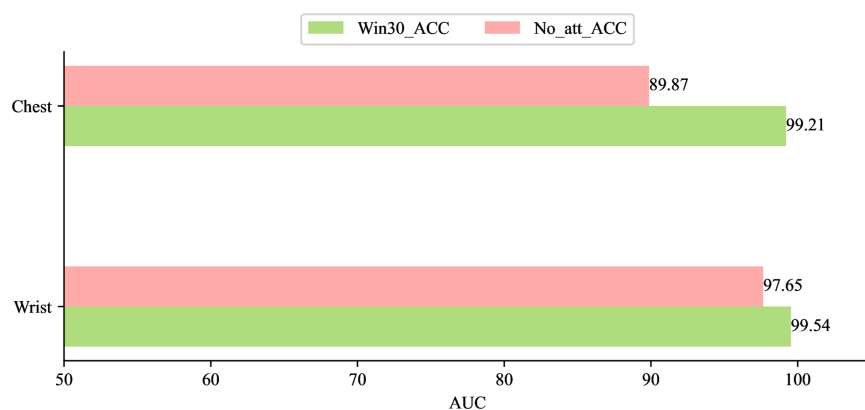


Fig 5. The role of feature embedding in affective computation tasks on both datasets, with the results presented in terms of ACC. *No_att_ACC* represents the performance of the PhysioFormer model without using the feature embedding module, while *Win30_ACC* indicates the performance of the PhysioFormer model after applying the feature embedding module.

<https://doi.org/10.1371/journal.pone.0335221.g005>

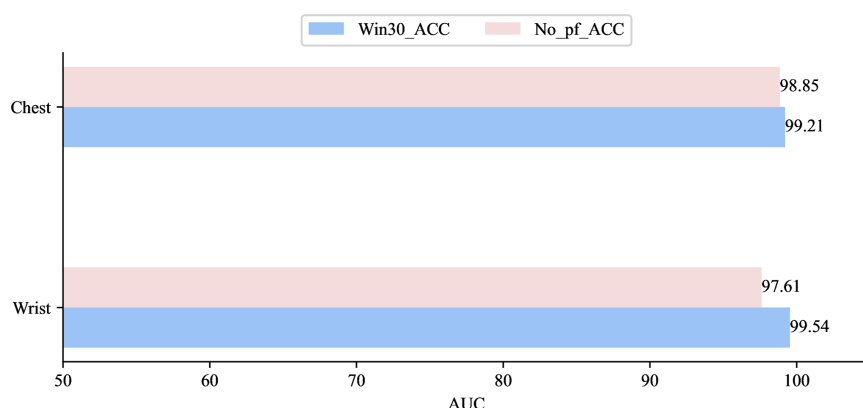


Fig 6. The role of individual attributes features in affective computation tasks on both datasets, with the results presented in terms of ACC. *No_pf_ACC* represents the performance of the PhysioFormer model without combining individual attributes features, while *Win30_ACC* indicates the performance of the PhysioFormer model after combining individual attributes features.

<https://doi.org/10.1371/journal.pone.0335221.g006>

The results indicate that the model incorporating individual attributes features data outperforms the one without it on both the Wrist and Chest datasets. Although the improvement on the Chest dataset is relatively modest, it still demonstrates the positive impact of individual attributes features data on model performance. This suggests that individual attributes features data contributes to enhancing the model's accuracy by providing additional context, helping the model capture and predict physiological signal changes more precisely. By combining individual attributes features with physiological monitoring data, the model gains a more comprehensive understanding of physiological states, resulting in improved overall prediction performance.

Discovered laws and analysis

Features distribution and importance. The Wrist dataset contains four types of physiological indicators, while the Chest dataset contains six types of physiological indicators. The distribution of the number of features computed from the physiological indicators in both datasets is shown in the Fig 7.

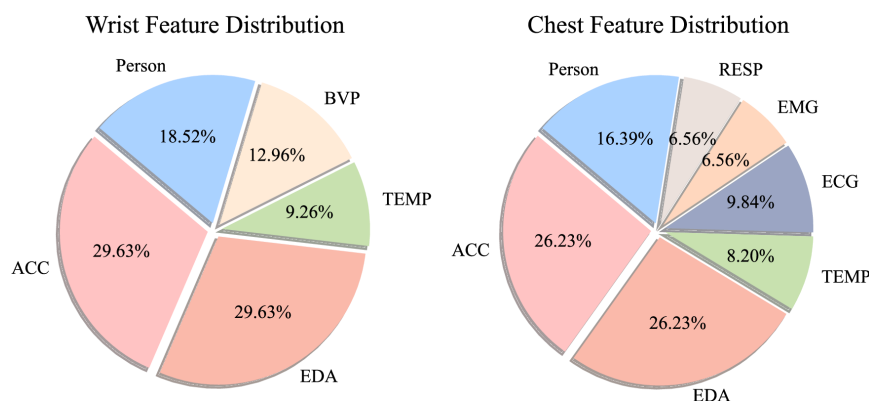


Fig 7. Distribution map of the number of features. The figure shows the distribution of the number of features calculated from various physiological indicators in the Wrist dataset and the Chest dataset, respectively.

<https://doi.org/10.1371/journal.pone.0335221.g007>

From the distribution chart, it is clear that ACC and EDA are the primary features in the Wrist dataset, each accounting for nearly one-third of the total features. Individual attributes also make up a significant portion, while BVP and TEMP have smaller proportions. In the Chest dataset, ACC and EDA similarly occupy major positions, each representing more than a quarter of the total features. Individual attributes also hold a large share, while ECG and TEMP occupy a notable portion, with EMG and RESP having smaller proportions. In summary, ACC and EDA are the dominant features in both datasets, but the Chest dataset includes a greater variety of physiological indicators.

In the PhysioFormer model, the importance scores of feature data for the four physiological indicators in the Wrist dataset (ACC, EDA, BVP, and TEMP) are shown in the Fig 8. By analyzing the figure, it is evident that ACC-related features significantly influence all physiological indicators. Specifically, features like *ACC_y_mean* and *Net_ACC_mean* not only have high importance for ACC itself but also show strong correlations with EDA, BVP, and TEMP. ACC signals, which capture both gravitational and movement-induced acceleration, are a standard physiological modality in wearable-based affective computing and are widely employed to characterize body dynamics associated with stress. Rather than excluding motion-related features, our study deliberately incorporates them into the multimodal framework to evaluate their contribution alongside other physiological signals. This suggests that physical activity intensity plays a key role in monitoring and predicting physiological states, likely due to the effects of movement on cardiovascular activity, electrodermal activity, and body temperature, making it important across multiple physiological indicators.

Additionally, individual attributes features (such as smoking status and weight) play a crucial role in TEMP and EDA. Smoking status shows high importance for TEMP, indicating a significant impact of smoking on temperature regulation. Similarly, weight has a notable influence on EDA, possibly due to its effect on skin conductance. Furthermore, features like temperature change rate and standard deviation are highly important for the TEMP indicator, reflecting the dynamic changes involved in the process of temperature regulation.

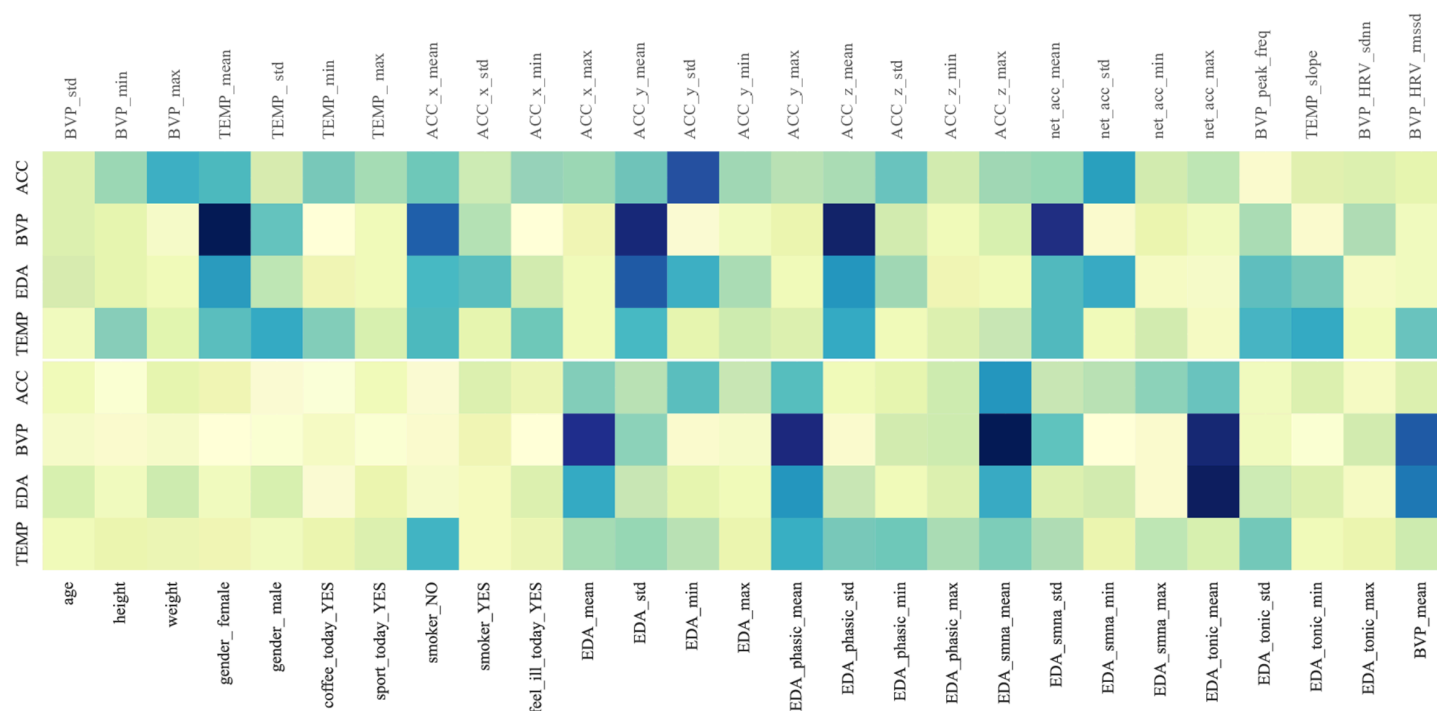


Fig 8. The visualization of the importance scores for all features in the Wrist dataset, with darker colors indicating higher importance scores. In the subsequent symbolic distillation task, the top ten features with the highest importance scores will be selected for further analysis and modeling.

<https://doi.org/10.1371/journal.pone.0335221.g008>

The importance scores of feature data for the six physiological indicators in the Chest dataset (ACC, EDA, ECG, TEMP, EMG, and RESP) are shown in the Fig 9. In the subsequent symbolic regression task, the top 10 features with the highest importance scores were extracted for further analysis, as detailed in Appendix B. ACC-related features still show a dominant influence across all six physiological indicators, further underscoring the critical role of physical activity in the monitoring and prediction of various physiological states.

Additionally, in the Chest dataset, the impact of individual attributes on physiological features becomes more pronounced. Specifically, weight significantly affects ECG and RESP indicators, while height has a notable impact on TEMP and ECG. Weight may influence the function of the cardiovascular and respiratory systems; changes in weight can lead to variations in blood pressure and heart rate, which are reflected in ECG features. Furthermore, weight gain can lead to breathing difficulties, which can be captured by the RESP indicator. Height may affect the ratio of body surface area to volume, influencing heat dissipation and thermoregulation mechanisms, which can be observed in TEMP features.

Discovered laws expression. In this section, symbolic regression was used to model each physiological indicator in both the Wrist and Chest datasets. The symbolic regression algorithm employed was implemented using the open-source symbolic regression library PySR, with the predefined operators including “+”, “−”, “×”, “sin”, “cos”, “log”, “exp”, and “pow”.

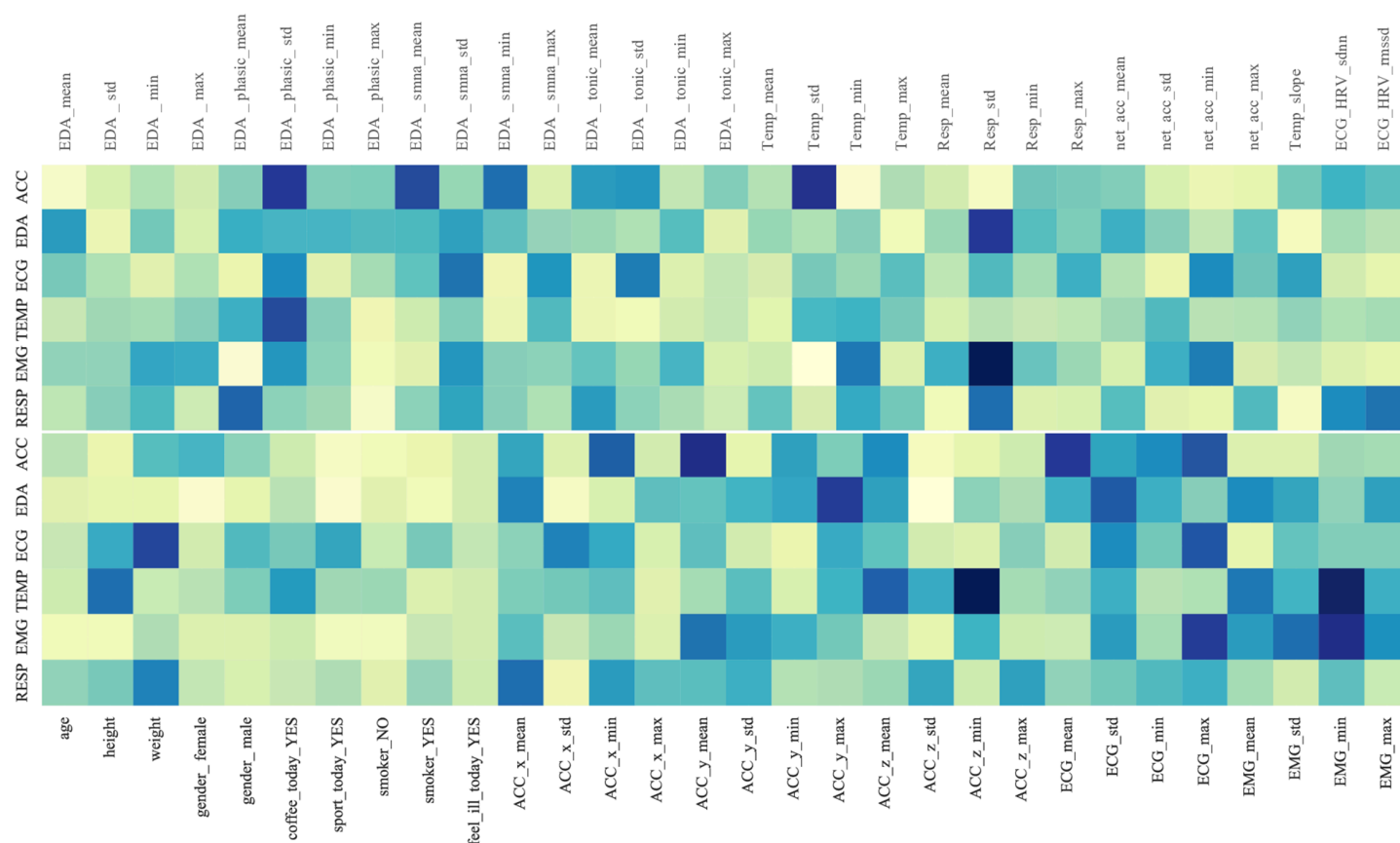


Fig 9. The visualization of the importance scores for all features in the Chest dataset, with darker colors indicating higher importance scores. In the subsequent symbolic distillation task, the top ten features with the highest importance scores will be selected for further analysis and modeling.

<https://doi.org/10.1371/journal.pone.0335221.g009>

Through symbolic regression, we obtained several formulas (with complexity less than 15). The next step is to select an optimal formula to explain the physiological indicator. In the selection process, we first focus on formulas that have converged, as these indicate that the model reached a stable state during training, with the loss value no longer significantly decreasing, suggesting that the model has well-fitted the data and found a relatively optimal parameter combination.

Among the converged formulas, we selected the one with the fewest variables. This is because formulas with fewer variables simplify the structure to some extent, improving the explainability of the model. Although these formulas may still be complex, they are able to capture subtle patterns and complex relationships in the data, which is critical for accurate modeling and prediction. Additionally, choosing formulas with fewer variables helps reduce the risk of overfitting, enhancing the model's generalization ability, and making it more stable and reliable when applied to new data.

Based on the training process and the formula selection method described above, the Complexity-Loss curve and the selected optimal formula are shown in the Figs 10 and 11, with the red dot indicating the chosen formula. The specific formulas generated through symbolic regression can be found in Appendix C.

From the experimental results, these curves exhibit a typical pattern: as the complexity of the formulas increases, the loss value rapidly decreases, but after reaching a certain level of complexity, the downward trend slows, showing convergence or slight fluctuations. Specifically, across all physiological indicator curves, as the formula complexity increases from low to moderate (usually in the complexity range of 2 to 6), the loss value drops rapidly. This indicates that increasing complexity in the initial stages significantly improves the model's ability to fit the data, allowing it to better capture the relationship between physiological signals and affective states. When the formula complexity reaches a moderate level (typically between 8 and 10), the loss value continues to decrease, but the rate of decline slows. For example, in the ACC and EDA curves, the loss value stabilizes as complexity reaches 8 to 10. This phase suggests that increasing formula complexity has diminishing returns in reducing the loss, and the model begins to converge. As the complexity increases further (usually beyond 12), the curve flattens, with minimal changes in the loss value, and in some cases, slight fluctuations are observed. For instance, in the BVP and TEMP curves, the loss value shows little to no significant reduction beyond complexity 12, indicating that the model's complexity has reached a sufficient level. Further increases in complexity may introduce the risk of overfitting, without significant improvement in model performance.

Additionally, the complexity-loss curves reveal that formulas for certain physiological indicators converge more slowly as complexity increases. This phenomenon may be attributed to several factors: First, some physiological indicator data may contain higher levels of noise and variability, making it more challenging for the model to accurately capture patterns in the data. Second, interactions between different physiological indicators may involve multiple physiological mechanisms, requiring more complex mathematical expressions to explain these relationships. Finally, limitations in the search space and computational resources may also affect the ability to efficiently find optimal formulas.

Overall, these curves illustrate the typical trade-off between model complexity and performance. Increasing complexity initially leads to significant improvements in model accuracy, but after reaching a certain level, the gains from further complexity diminish and may even result in overfitting. Therefore, selecting an appropriate complexity level for the final model formula ensures a low loss value while avoiding unnecessary complexity.

Based on the selected formulas, we randomly chose four individuals' affective indicators. Using the selected formulas and the output from the PhysioFormer model, we plotted the fitting curves for each physiological indicator across both datasets, as shown in the Figs 12 and 13.

To evaluate the fit of each formula, we used the R^2 metric, which measures the goodness of fit between the predicted values and the actual values. The closer the R^2 value is to 1, the stronger the model's ability to explain the data. We calculated the R^2 values for all selected formulas, and the results are shown in the Table 7.

According to the results shown in the table, the high-fitting formulas include ACC and BVP from the Wrist dataset, as well as ACC, TEMP, and ECG from the Chest dataset, with R^2 values close to or reaching 0.98 to 0.99. This indicates that these formulas predict the corresponding data very well, and the model is able to explain the variability in the data accurately. These physiological indicators are well predicted and explained under the current model and formula combinations.

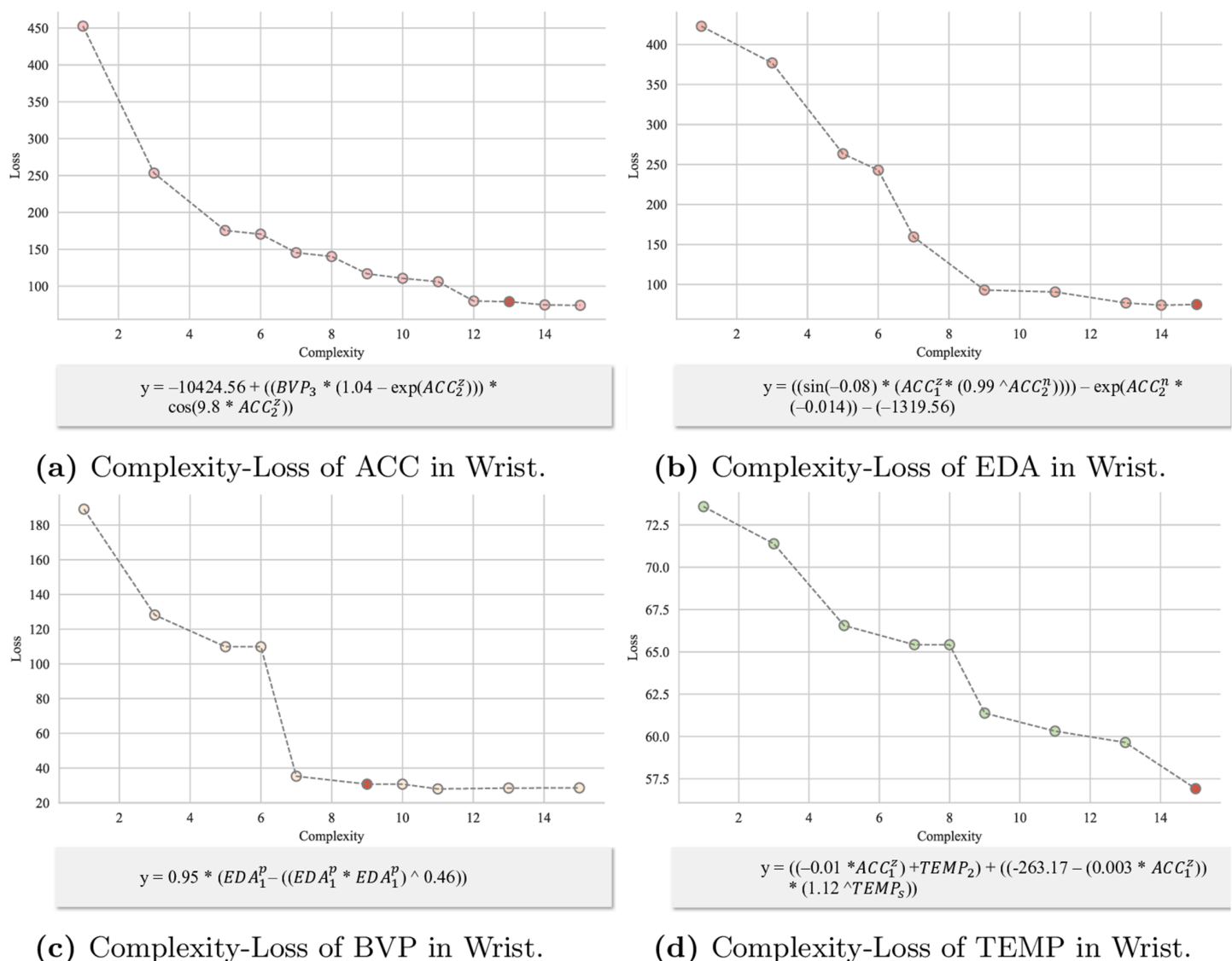


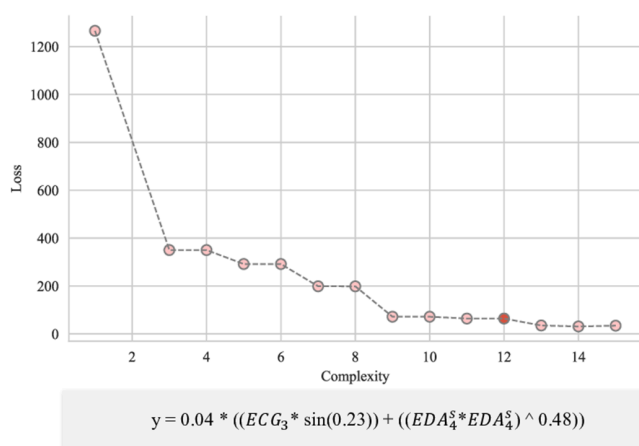
Fig 10. Complexity-Loss curve in Wrist dataset. The red dot indicate the selected formula, and the corresponding formula are shown below the charts.

<https://doi.org/10.1371/journal.pone.0335221.g010>

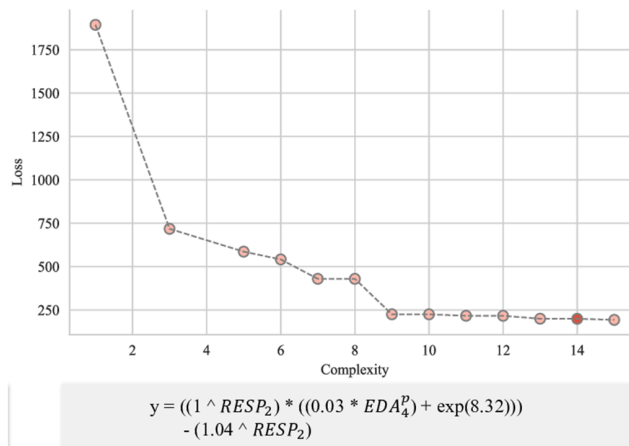
Moderate-fitting formulas include TEMP from the Wrist dataset and EDA and RESP from the Chest dataset, with R^2 values ranging from 0.39 to 0.76. These formulas perform reasonably well but still have room for improvement. While they capture some important patterns in the data, they do not fully explain the data's variability.

Low-fitting formulas include EDA from the Wrist dataset and EMG from the Chest dataset, both with low R^2 values of 0.15, indicating poor predictive performance for these formulas. The model struggles to capture the patterns in the data for these indicators, which may be due to the model's insufficient complexity or interference from noise and outliers in the data.

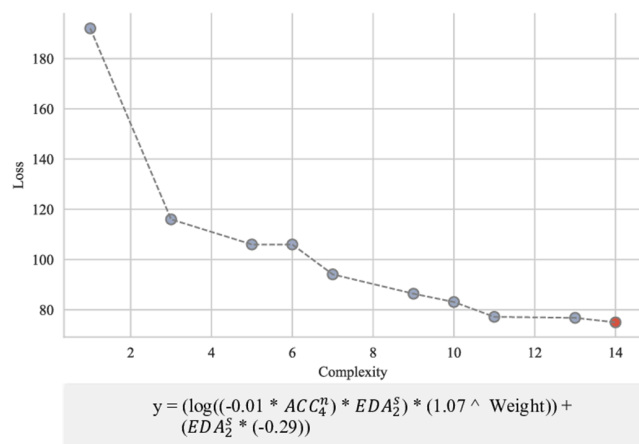
The results suggest that the model performs satisfactorily for most physiological indicators, but its predictive ability needs improvement for certain indicators. Future work can focus on the following areas: first, introducing more relevant



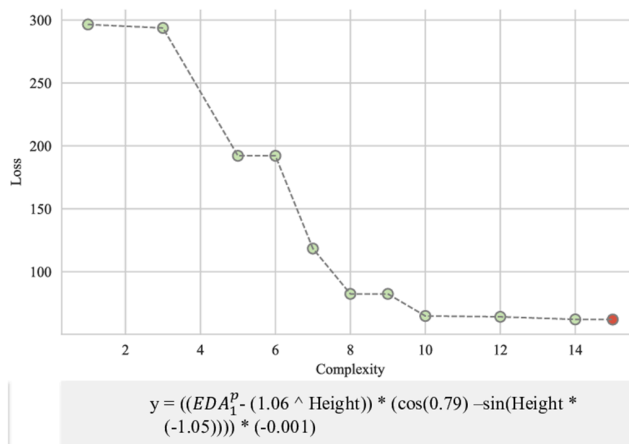
(a) Complexity-Loss of ACC in Chest.



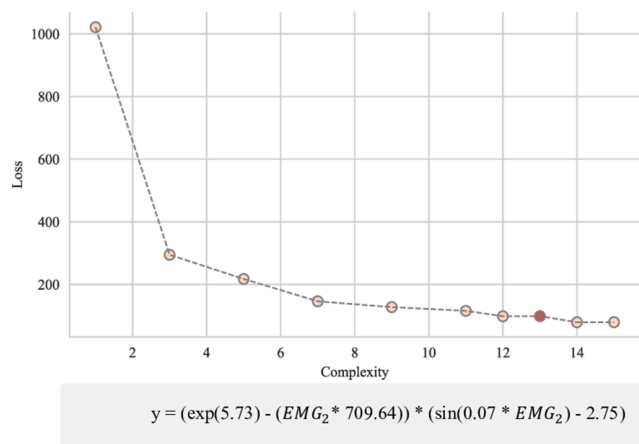
(b) Complexity-Loss of EDA in Chest.



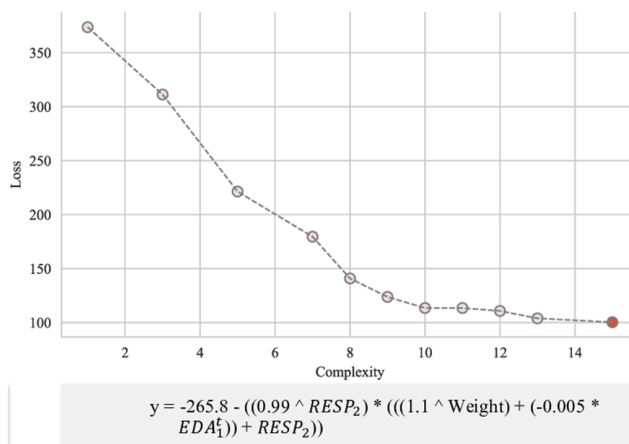
(c) Complexity-Loss of ECG in Chest.



(d) Complexity-Loss of TEMP in Chest.



(e) Complexity-Loss of EMG in Chest.



(f) Complexity-Loss of RESP in Chest.

Fig 11. Complexity-Loss curve in Chest dataset. The red dot indicate the selected formula, and the corresponding formula are shown below the charts.

<https://doi.org/10.1371/journal.pone.0335221.g011>

Wrist

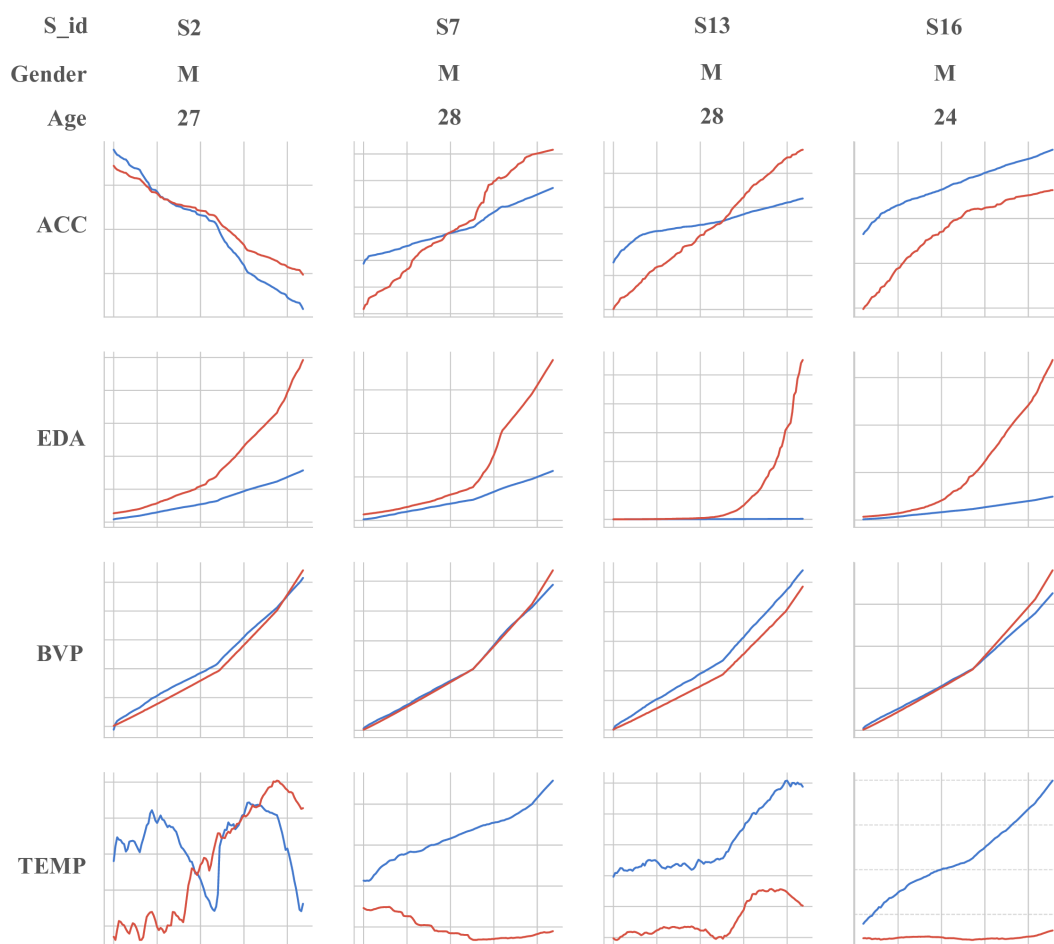


Fig 12. The comparison of the fitted curves for affective indicators based on the selected formulas for four randomly chosen individuals. Using the selected formulas and the output from the PhysioFormer model, the fitting curves for each affective indicator in the Wrist dataset were plotted. In the figure, the blue curve represents the model's output, while the red curve represents the results calculated from the formulas.

<https://doi.org/10.1371/journal.pone.0335221.g012>

features to bring additional information and enhance the model's predictive ability; second, conducting data cleaning and preprocessing for indicators with low R^2 values to remove noise and outliers, thereby improving data quality.

Conclusions

In this study, we proposed and implemented the PhysioFormer model for affective computation based on multimodal physiological signals. PhysioFormer model effectively addresses the variability in physiological signals across individuals by integrating individual attributes and multimodal signals. This integration allows the model to demonstrate high reliability and generalization in cross-individual affective computation tasks, ensuring its stability across different affective computation tasks, making it highly applicable and reliable in real-world scenarios. By incorporating feature embedding and affective representation modules, PhysioFormer is able to capture the temporal dependencies and multimodal features of physiological signals, significantly enhancing its accuracy. Extensive experiments on the Wrist and Chest subsets

Chest

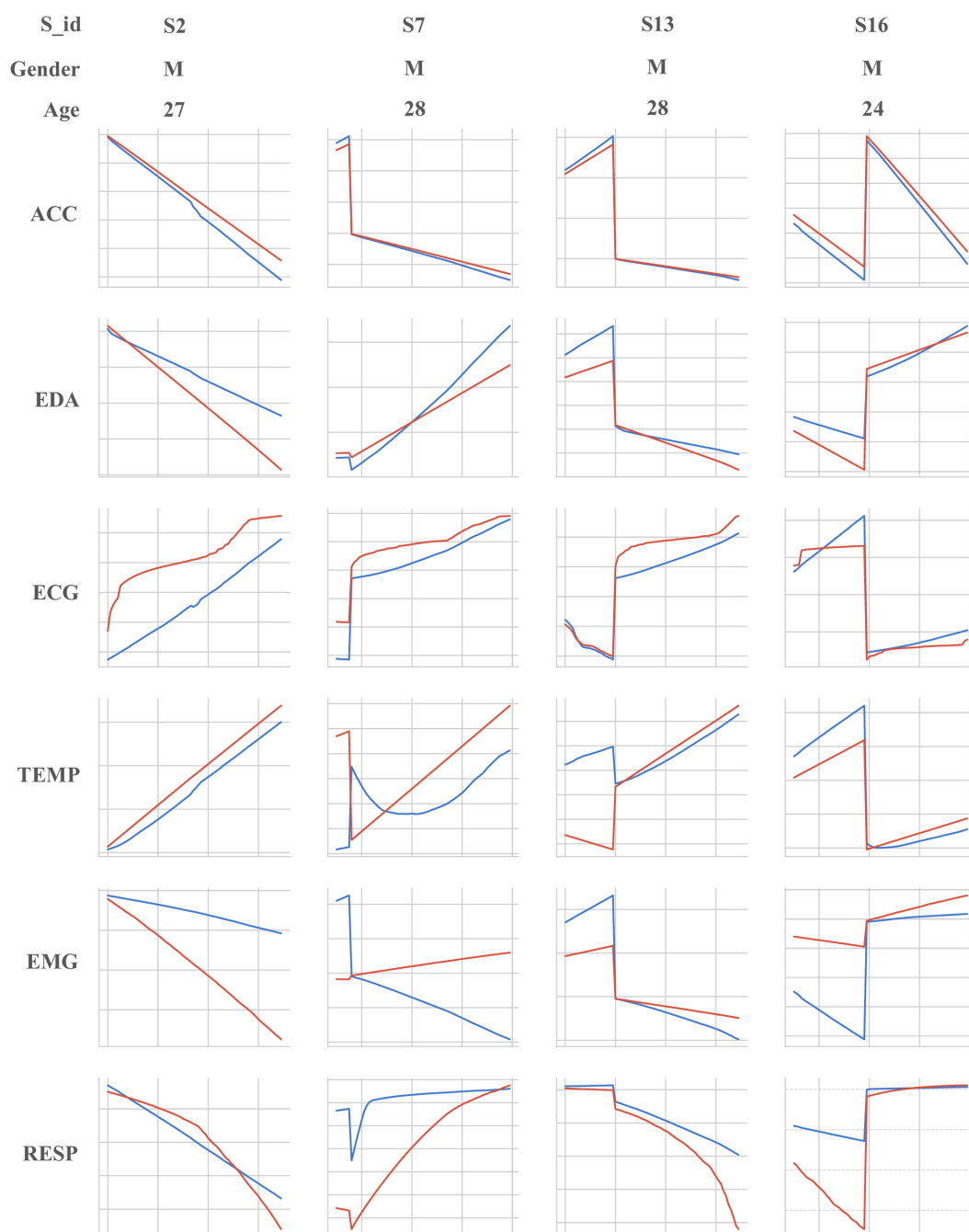


Fig 13. The comparison of the fitted curves for affective indicators based on the selected formulas for four randomly chosen individuals. Using the selected formulas and the output from the PhysioFormer model, the fitting curves for each affective indicator in the Chest dataset were plotted. In the figure, the blue curve represents the model's output, while the red curve represents the results calculated from the formulas.

<https://doi.org/10.1371/journal.pone.0335221.g013>

Table 7. The table presents the evaluation results of the selected formulas using the R^2 metric, which measures the goodness of fit between the predicted values and the actual values. The R^2 values for all the selected formulas are listed, providing a clear view of the fitting performance of each formula.

	ACC	EDA	TEMP	BVP	ECG	EMG	RESP
Wrist	0.98	0.15	0.39	0.99	—	—	—
Chest	0.98	0.63	0.76	—	0.73	0.15	0.67

<https://doi.org/10.1371/journal.pone.0335221.t007>

of the WESAD dataset demonstrated PhysioFormer’s superior performance in affective computation tasks. The experimental results showed that PhysioFormer achieved over 99% accuracy on both subsets, outperforming the current SOTA models. Additionally, we introduced an Explanation model to the PhysioFormer model, utilizing symbolic regression techniques to extract symbolic laws that reveal the relationships between physiological signals and affective states. Additional visualization results are presented in [S1–S15 Figs](#) of the Supporting Information. This enhancement improves the model’s explainability, offering new insights into the intrinsic connections between physiological signals and affective states. The significance of this study lies in its potential to enhance affective computing applications in practical domains such as personalized mental health care, intelligent tutoring systems, stress management in workplace environments, and adaptive human–computer interaction. By enabling accurate, generalizable, and interpretable recognition of affective states, PhysioFormer can provide valuable support for early psychological intervention, continuous health monitoring, and the design of more empathetic and adaptive AI-driven systems. These applications highlight the model’s contribution not only to advancing research but also to addressing pressing societal needs in healthcare and education.

Nevertheless, despite the promising results, this study highlights several critical challenges in real-world applications. First, the applicability of PhysioFormer to large-scale datasets and practical scenarios requires further validation, particularly in balancing computational efficiency and recognition accuracy in real-time affective recognition systems to ensure smooth operation on mobile and wearable devices. Second, to enhance the model’s adaptability across diverse populations, future research should optimize the feature embedding module to improve stability and generalization under individual variability, ensuring reliability in applications such as health monitoring and psychological interventions. Furthermore, integrating additional physiological signals (e.g., blood pressure, EEG) and environmental factors (e.g., noise, light intensity) could further enhance the model’s adaptability to complex real-world environments, increasing its practical value in stress detection, affective health assessment, and related domains.

Future research will involve more fine-grained feature ablation experiments at the sensor level to further clarify the contributions of different physiological features. In addition, future research can also expand and refine the symbolic distillation approach to extract more explainable mathematical formulas, thereby improving model explainability and transparency. This is particularly important for applying the PhysioFormer model in emotion-driven real-world applications such as health monitoring and psychological interventions. Moreover, further exploration of how symbolic regression can be used to explain the relationship between multimodal physiological signals and complex affective states will provide new directions and inspiration for future research in affective computation.

Supporting information

Features description

S1 Fig. Individual features table, which describes all individual attributes features, their types, and the meanings they represent.

(TIFF)

S2 Fig. Features table for Wrist dataset, which describes all features and the meanings they represent.
(TIFF)

S3 Fig. Features table for Chest dataset, which describes all features and the meanings they represent.
(TIFF)

Top 10 features in terms of features importance tables

S4 Fig. Results on wrist dataset.
(TIFF)

S5 Fig. Results on chest dataset.
(TIFF)

Laws tables

This part shows the indicators law table of WESAD dataset. The highlight formula indicates the selected formula.

S6 Fig. Indicators law table of ACC in Wrist.
(TIFF)

S7 Fig. Indicators law table of EDA in Wrist.
(TIFF)

S8 Fig. Indicators law table of BVP in Wrist.
(TIFF)

S9 Fig. Indicators law table of TEMP in Wrist.
(TIFF)

S10 Fig. Indicators law table of ACC in Chest.
(TIFF)

S11 Fig. Indicators law table of EDA in Chest.
(TIFF)

S12 Fig. Indicators law table of ECG in Chest.
(TIFF)

S13 Fig. Indicators law table of TEMP in Chest.
(TIFF)

S14 Fig. Indicators law table of EMG in Chest.
(TIFF)

S15 Fig. Indicators law table of RESP in Chest.
(TIFF)

Author contributions

Conceptualization: Zhifeng Wang.

Data curation: Jialiang Shen.

Formal analysis: Zhifeng Wang.

Funding acquisition: Zhifeng Wang, Chunyan Zeng.

Investigation: Wanxuan Wu.

Project administration: Chunyan Zeng.

Supervision: Zhifeng Wang.

Writing – original draft: Zhifeng Wang, Wanxuan Wu.

Writing – review & editing: Zhifeng Wang.

References

1. Dattani S, Rodés-Guirao L, Ritchie H, Roser M. Mental Health. Our World in Data; 2023.
2. McEwen BS. Brain on stress: how the social environment gets under the skin. *Proc Natl Acad Sci U S A*. 2012;109(Suppl 2):17180–5. <https://doi.org/10.1073/pnas.1121254109> PMID: 23045648
3. Ertin E, Stohs N, Kumar S, Raji A, al'Absi M, Shah S. AutoSense. In: Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems. 2011. p. 274–87. <https://doi.org/10.1145/2070942.2070970>
4. De Nadai S, D'Inca M, Parodi F, Benza M, Trotta A, Zero E, et al. Enhancing safety of transport by road by on-line monitoring of driver emotions. In: 2016 11th System of Systems Engineering Conference (SoSE). 2016. <https://doi.org/10.1109/sysose.2016.7542941>
5. Hosseini E, Fang R, Zhang R, Rafatirad S, Homayoun H. Emotion and stress recognition utilizing galvanic skin response and wearable technology: a real-time approach for mental health care. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2023. p. 1125–31. <https://doi.org/10.1109/bibm58861.2023.10386049>
6. Andreou E, Alexopoulos EC, Lionis C, Varvogli L, Gnardellis C, Chrousos GP, et al. Perceived stress scale: reliability and validity study in Greece. *Int J Environ Res Public Health*. 2011;8(8):3287–98. <https://doi.org/10.3390/ijerph8083287> PMID: 21909307
7. Mei H, Xu X. EEG-based emotion classification using convolutional neural network. In: 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC). 2017. p. 130–5. <https://doi.org/10.1109/spac.2017.8304263>
8. Song T, Zheng W, Song P, Cui Z. EEG emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans Affective Comput*. 2020;11(3):532–41. <https://doi.org/10.1109/taffc.2018.2817622>
9. Ashwin VH, Jegan R, Rajalakshmy P. Stress detection using wearable physiological sensors and machine learning algorithm. In: 2022 6th International Conference on Electronics, Communication and Aerospace Technology. 2022. p. 972–7. <https://doi.org/10.1109/iceca55336.2022.10009326>
10. Udrescu S-M, Tegmark M. AI Feynman: a physics-inspired method for symbolic regression. *Sci Adv*. 2020;6(16):eaay2631. <https://doi.org/10.1126/sciadv.aay2631> PMID: 32426452
11. García-Hernández RA, Luna-García H, Celaya-Padilla JM, García-Hernández A, Reveles-Gómez LC, Flores-Chaires LA, et al. A systematic literature review of modalities, trends, and limitations in emotion recognition, affective computing, and sentiment analysis. *Applied Sciences*. 2024;14(16):7165. <https://doi.org/10.3390/app14167165>
12. Kratzwald B, Ilic S, Kraus M, Feuerriegel S, Prendinger H. Decision support with text-based emotion recognition: deep learning for affective computing. *CoRR*. 2018.
13. Quaedflieg CWEM, Meyer T, Smeets T. The imaging Maastricht Acute Stress Test (iMAST): a neuroimaging compatible psychophysiological stressor. *Psychophysiology*. 2013;50(8):758–66. <https://doi.org/10.1111/psyp.12058> PMID: 23701399
14. Dedovic K, Renwick R, Mahani NK, Engert V, Lupien SJ, Pruessner JC. The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *J Psychiatry Neurosci*. 2005;30(5):319–25. <https://doi.org/10.1139/jpn.0541> PMID: 16151536
15. Sarkar P, Etemad A. Self-supervised ECG representation learning for emotion recognition. *IEEE Trans Affective Comput*. 2022;13(3):1541–54. <https://doi.org/10.1109/taffc.2020.3014842>
16. Akre S, Cohen ZD, Welborn A, Zbozinek TD, Balliu B, Flint J, et al. Detection of symptoms of depression using data from the iPhone and Apple Watch. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2023. p. 1818–23. <https://doi.org/10.1109/bibm58861.2023.10385797>
17. Koldijk S, Sappelli M, Verberne S, Neerincx MA, Kraaij W. The SWELL knowledge work dataset for stress and user modeling research. In: Proceedings of the 16th International Conference on Multimodal Interaction, 2014. p. 291–8. <https://doi.org/10.1145/2663204.2663257>
18. Siddharth, Jung T, Sejnowski TJ. Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing. *CoRR*. 2019.
19. Brunton SL, Proctor JL, Kutz JN. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc Natl Acad Sci U S A*. 2016;113(15):3932–7. <https://doi.org/10.1073/pnas.1517384113> PMID: 27035946

20. Rogers AW, Lane A, Mendoza C, Watson S, Kowalski A, Martin P, et al. Integrating knowledge-guided symbolic regression and model-based design of experiments to automate process flow diagram development. *Chemical Engineering Science*. 2024;300:120580. <https://doi.org/10.1016/j.ces.2024.120580>
21. Miyazaki M, Ishikawa K-I, Nakashima K, Shimizu H, Takahashi T, Takahashi N. Application of the symbolic regression program AI-Feynman to psychology. *Front Artif Intell*. 2023;6:1039438. <https://doi.org/10.3389/frai.2023.1039438> PMID: 36776421
22. Liu S, Li Q, Shen X, Sun J, Yang Z. Automated discovery of symbolic laws governing skill acquisition from naturally occurring data. *Nat Comput Sci*. 2024;4(5):334–45. <https://doi.org/10.1038/s43588-024-00629-0> PMID: 38811819
23. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE*. 1989;77(2):257–86. <https://doi.org/10.1109/5.18626>
24. Lafferty JD, McCallum A, Pereira FCN. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the Eighteenth International Conference on Machine Learning*. 2001. p. 282–9.
25. Mikolov T, Karafiát M, Burget L, Černocký J, Khudanpur S. Recurrent neural network based language model. In: *Interspeech 2010*. 2010. p. 1045–8.
26. Frinken V, Zamora-Martínez F, España-Boquera S, Castro-Bleda MJ, Fischer A, Bunke H. Long-short term memory neural networks language modeling for handwriting recognition. In: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. 2012. p. 701–4.
27. Dhavale M, Bhandari ProfS. Speech Emotion Recognition Using CNN and LSTM. In: *2022 6th International Conference on Computing, Communication, Control And Automation (ICCUBEA)*. 2022. p. 1–3. <https://doi.org/10.1109/iccubea54992.2022.10010751>
28. Sateesh SK, S BK, U D. Decoding human emotions: Analyzing multi-channel EEG data using LSTM networks. 2024. <https://arxiv.org/abs/2408.10328>
29. Hazmoune S, Bougamouza F. Using transformers for multimodal emotion recognition: taxonomies and state of the art review. *Engineering Applications of Artificial Intelligence*. 2024;133:108339. <https://doi.org/10.1016/j.engappai.2024.108339>
30. Mittal T, Bhattacharya U, Chandra R, Bera A, Manocha D. M3ER: multiplicative multimodal emotion recognition using facial, textual, and speech cues. *AAAI*. 2020;34(02):1359–67. <https://doi.org/10.1609/aaai.v34i02.5492>
31. Kumar CSA, Maharana AD, Krishnan SM, Hanuma SSS, Lal GJ, Ravi V. Speech emotion recognition using CNN-LSTM and vision transformer. In: *Abraham A, Bajaj A, Gandhi N, Madureira AM, Kahraman C, editors. Innovations in Bio-Inspired Computing and Applications*. Cham: Springer; 2023. p. 86–97.
32. Setz C, Arnrich B, Schumm J, La Marca R, Troster G, Ehlert U. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Trans Inform Technol Biomed*. 2010;14(2):410–7. <https://doi.org/10.1109/titb.2009.2036164>
33. Bianchi G, Sorrentino R. *Electronic Filter Simulation & Design*. McGraw-Hill Education; 2007.
34. Greco A, Valenza G, Lanata A, Scilingo EP, Citi L. cvxEDA: a convex optimization approach to electrodermal activity processing. *IEEE Trans Biomed Eng*. 2016;63(4):797–804. <https://doi.org/10.1109/TBME.2015.2474131> PMID: 26336110
35. Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, et al. NeuroKit2: a python toolbox for neurophysiological signal processing. *Behav Res Methods*. 2021;53(4):1689–96. <https://doi.org/10.3758/s13428-020-01516-y> PMID: 33528817
36. Liu Y, Li C, Wang J, Long M. Koopa: learning non-stationary time series dynamics with koopman predictors. In: *NIPS, Red Hook, NY, USA*, 2024. p. 1–9.
37. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. In: *Proceedings of the 34th International Conference on Machine Learning*. 2017. p. 3319–28.
38. Schmidt P, Reiss A, Duerichen R, Marberger C, Van Laerhoven K. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In: *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. 2018. p. 400–8. <https://doi.org/10.1145/3242969.3242985>
39. Bobade P, Vani M. Stress detection with machine learning and deep learning using multimodal physiological data. In: *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India. 2020. p. 51–7.
40. Siirtola P. Continuous stress detection using the sensors of commercial smartwatch. In: *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019. p. 1198–201. <https://doi.org/10.1145/3341162.3344831>
41. Ferdinando H, Alasaarela E. Emotion recognition using cvxEDA-based features. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*. 2018;10(2-3):19–23.
42. Yu D, Sun S. A systematic exploration of deep neural networks for EDA-based emotion recognition. *Information*. 2020;11(4):212. <https://doi.org/10.3390/info11040212>