ARTICLE

Pressure Classification Analysis on CNN-Transformer-LSTM Hybrid Model

Peng Xia[1], Wu Zeng[2,*], Yin Ni[1] and Ye Jin[3]

[1]School of Electrical and Electronic Engineering, Wuhan Polytechnic University, Wuhan, 430023, China

[2]School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan, 430048, China

[3]School of Electromechanical and Intelligent Manufacturing, Huanggang Normal University, Huanggang, 438000, China

*Corresponding Author: Wu Zeng. Email: zengwu@whpu.edu.cn

ABSTRACT

Stress is defined as a subjective reflection of an internal psychological state of tension or arousal, manifesting as an interpretive, emotional, and defensive coping process within the body. Prolonged and sustained stress can significantly increase the risk of psychological and physiological disorders. Heart rate variability (HRV) is a key biomarker for assessing autonomic cardiac function, typically increasing during relaxation and decreasing under stress. Although measuring stress through physiological parameters like HRV is a common approach, achieving ultra-high accuracy based on HRV measurements remains a challenging task. In this study, the role of HRV features as biomarkers for stress detection was investigated, leading to the development of an advanced CNN-Transformer-LSTM multi-class stress detection model that leverages both the time-domain and frequency-domain characteristics of HRV. Specifically, the model incorporates a convolutional neural network (CNN) combined with a classifier that integrates long short-term memory (LSTM) networks and a Transformer architecture. The CNN effectively captures local patterns and features within time-series data, while the LSTM networks manage temporal dependencies. The Transformer component enhances the model's ability to understand the relationships between different time points in the HRV signal through self-attention mechanisms. HRV signals were extracted from the SWELL-KW database and cross-validated using the WESAD database. Experimental results demonstrate that the proposed method achieves higher accuracy compared to existing studies in the literature, even when using fewer HRV features. The obtained stress detection accuracy ranged between 98.84% and 99.31%. Additionally, this study employed marginal utility analysis to validate the effectiveness of HRV features in stress detection.

KEYWORDS

Stress detection; heart rate variability; transformer; CNN; long short-term memory network

1 Introduction

Psychological stress is widely regarded as one of the major contemporary issues affecting human health and productivity. According to statistics from the World Health Organization in 2022, approximately 350 million people globally suffer from depression, with an annual incidence rate of about 5% among adults. Stress, a response to various psychological, physiological, and emotional challenges, can disrupt an individual's psychological equilibrium [1].

Heart Rate Variability (HRV) serves as a crucial quantitative measure of autonomic nervous system activity. Research indicates that HRV can be an effective indicator of stress. It is defined as the fluctuation in the time intervals between consecutive heartbeats (RR intervals) [2,3]. HRV decreases under stress; typically, it is higher when the heart rate slows and *vice versa*. Consequently, there is an inverse relationship between heart rate and HRV, measured through electrocardiogram readings [4]. HRV varies over time depending on the level of activity and work-related stress. For instance, individuals with anxiety disorders tend to have consistently lower HRV [5].

Generally, stress is linked to negative thoughts and is considered a subjective experience that can affect both emotional and physical health. It can be defined as a psychological and physiological reaction to various stressors, including biological, chemical, and environmental factors, which may induce stress within an organism [6,7]. However, measuring stress poses challenges primarily due to its reliance on subjective reports from questionnaires, requiring participants to directly report their stress levels over time. This method has limitations, such as self-bias and the time commitment required for ongoing participation.

In recent years, utilizing HRV physiological parameters, various machine learning and deep learning algorithms have been developed to predict stress [8–10]. For instance, in 2007, Dinges et al. [11] conducted a study on stress detection using facial recognition. Lin et al. [12] proposed using physiological signals obtained from heart rate (HR) and blood pressure (BP) monitoring to assess mental stress. Yue et al. [13] filtered Photoplethysmography (PPG) signals, used Welch's spectral estimation to generate a spectrum, and calculated HRV, achieving an accuracy and specificity of 92.26% and 96.12%, respectively for psychological fatigue classification. Melillo et al. [14] employed Poincaré plots, approximate entropy, and detrended fluctuation analysis on short-term electrocardiogram records of HRV to conduct linear discriminant analysis, ultimately achieving a total classification accuracy of 90% for stress detection. Lee et al. [15] proposed using IMF energy features based on EMD, substituting the frequency domain information that is challenging to use in ultra-short terms, and assessed performance using LOSOCV, achieving a top accuracy score of 86.5%. Shikha et al. [16] employed a random forest to perform binary classification on HRV features derived from skin conductance response (GSR) and electrocardiogram (ECG) after feature selection, reaching a top accuracy score of 93.96%. Among the numerous publicly available stress detection datasets, the SWELL-KW HRV dataset developed in References [17,18] is one of the most extensively used. In multi-level stress classification, the methods employed in [19] achieved the highest accuracy of 87.6%; in [20], using 15 HRV features resulted in an accuracy of up to 96.50%, albeit with high computational costs and lengthy execution times. This underscores the ongoing need for new machine learning models that offer high accuracy and computational efficiency.

Existing studies on machine learning and deep learning based on the SWELL-KW dataset face two main issues. Firstly, the accuracy of multi-level stress classification is insufficient, limiting its effectiveness in practical applications. Secondly, while maintaining accuracy, the models are complex and require numerous parameters, which not only increase computational load but also prolong execution times. These issues indicate that, despite progress, there is significant room for improvement in creating efficient and accurate stress classification models.

In this study, we design and validate an innovative CNN-Transformer-LSTM hybrid model specifically for multivariate stress classification. This model uniquely integrates the strengths of CNN for effective local feature extraction from time-series data, Transformer for enhanced understanding of temporal relationships in HRV signals via self-attention mechanisms, and LSTM for capturing long-term dependencies crucial for interpreting dynamic changes in heart rate intervals. By combining

these architectures with residual connections to prevent gradient vanishing, our model ensures lossless information transmission across multiple network layers, demonstrating exceptional generalization capabilities and high accuracy in stress classification. Experimental results show that our approach achieves a classification accuracy of 99.01% based on seven HRV features, significantly outperforming existing state-of-the-art models on the SWELL-KW dataset in both predictive accuracy and computational efficiency.

2 Related Work

Recent research in the field of electrocardiogram data analysis has focused on using machine learning and deep learning methods for stress classification, particularly in binary (presence or absence of stress) and multi-level stress state classifications. In recent studies, Kaya et al. [21] developed a method combining 1D-DS-LBP with LSTM to classify ECG signals, achieving 96.80% to 99.79% accuracy by analyzing neighboring point relationships. Bu et al. [22] analyzed differential RRI time series using Long Short-Term Memory (LSTM) networks. In a trinary classification of differential RRI data, they achieved a maximum accuracy score of 87.9% within 20 s, significantly enhancing the LSTM model's training and classification performance. Dalmeida et al. [6] classified HRV data into stressed and non-stressed states, evaluating various machine learning methods including Naïve Bayes, KNN, SVM, MLP, Random Forest, and Gradient Boosting, achieving a maximum recall rate of 80%. Another study [7] utilized both time-domain and frequency-domain features of HRV, employing Radial Basis Function (RBF) SVM, and achieved accuracies of 83.33% and 66.66%. Dimensionality reduction methods were also employed to identify the most representative time-domain and frequency-domain features in HRV [9]. Research using CNN for binary classification [10] reached an accuracy of 98.4% in stress detection (though this study focuses on multi-class classification). Jegan et al. [23] used Support Vector Machine (SVM) classifier techniques with ultra-short-term HRV analysis, achieving a mental stress classification accuracy of 91%. Coutts et al. [24] predicted using LSTM techniques from HRV data collected via wrist-wearable devices, achieving up to 85% classification accuracy. Gomathi et al. [25] used frequency-domain parameters of HRV, analyzed the power spectral density of ECG signals through the Welch method, and employed the K-Nearest Neighbors (KNN) classifier to analyze and classify individual stress levels, with a stress F1 score of 90%. Pajong et al. [26] proposed a time-series-based emotional classification algorithm and compared the emotional detection performance of five learning algorithms including Random Forest, Extreme Gradient Boosting (XGBoost), LSTM, Convolutional LSTM (CNN-LSTM), and Deep Convolutional LSTM (DeepConvLSTM), achieving a maximum accuracy of 80% and an Area Under the Curve (AUC) of 94%.

In studies on multi-level stress classification using the SWELL-KW dataset (such as no stress, interruptive stress, and time stress), Koldijk et al. [18] achieved a maximum accuracy of 90% using SVM. Sarkar et al. [10] reached an accuracy of 98.30% using CNN. Meanwhile, research [8] using the WESAD dataset [27] for multi-category (leisure, baseline, and stress) and binary (stress *vs*. non-stress) classifications, an accuracy of 84.32% was achieved in the three-class classification. Benita et al. [28] optimized a CNN with random forest-based feature selection, successfully classifying individuals into "stressed" or "not stressed" categories using HRV data from the WESAD dataset, achieving 98% accuracy.

In 2023, Mortensen developed a 1D-CNN multi-class stress discrimination model based on the SWELL-KW dataset [20], achieving 99% accuracy with 34 features, and 96.5% with 15 features, though this increased computational load and execution time. In comparison, our model achieved an

accuracy of 99.01% using only 7 features, surpassing existing advanced models in performance. When the number of features was increased to 14, the accuracy further improved to 100%. This achievement highlights that careful selection and utilization of a limited number of features can not only optimize model performance but also significantly enhance computational efficiency while precisely capturing stress states.

3 Framework Overview and Dataset Preprocessing

In this section, a framework for multivariate pressure detection is outlined.

3.1 Overview of the Framework

Fig. 1 presents the schematic diagram of the proposed stress level classification framework. In summary, the framework consists of the following steps:

- Data Acquisition and Dataset Preparation: HRV signals are gathered and split into training and test datasets. These datasets are used to design the model architecture and assess the model's effectiveness.

- Classification and Model Validation: The deep learning-based multi-class classifiers are trained, tested, and validated using critical features and annotations provided by medical professionals (e.g., no stress, interrupted conditions, and time stress).

- Testing: During the testing phase, the key features of the new test sample are extracted, and the class labels are predicted using the classification parameters.

- Performance Evaluation: The performance of the classifier is measured based on discriminant analysis metrics, such as accuracy, precision, recall, F1 score and MCC.
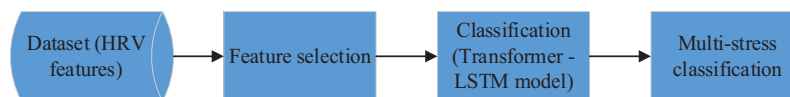

Figure 1: Framework of the proposed pressure state classification model

3.2 Introduction to the Dataset

The SWELL-KW dataset [17] and the WESAD dataset [27] were utilized for conducting intra-dataset testing on SWELL-KW and cross-dataset testing on WESAD. Model effectiveness was evaluated using various metrics, including accuracy, precision, recall, F1 score, and MCC.

The SWELL-KW dataset comprises HRV data intended for stress modeling and also records participants' subjective experiences related to task load, mental labor, emotion, and perceived stress. Data were collected from 25 individuals under various work conditions, such as report writing, presenting, reading emails, and information searching. Each participant was subjected to three distinct work environments, during this period, the participants' psychological and physiological state data were recorded. The annotations provided by medical experts are as follows:

- No stress: Participants could engage in activities as needed for a maximum duration of 45 min, although they were not informed of the task's time limit.

- Time pressure: The time to complete the same tasks was reduced to two-thirds of the normal duration under time pressure.

- Interruptions: Participants were interrupted upon receiving eight emails during an activity. Some emails necessitated specific actions related to their tasks, while others were unrelated to the ongoing activities.

The WESAD dataset is a multimodal dataset available to the public, aimed at investigating stress and emotion detection through wearable devices. It consists of physiological and motion data collected from 15 subjects in a laboratory setting using wrist and chest-worn devices. The dataset features sensor modalities such as blood volume pulse, electrocardiogram, electrodermal activity, electromyogram, respiration, body temperature, and triaxial acceleration. Furthermore, WESAD fills the gap in laboratory studies between stress and emotion, covering three different emotional states (neutral, stress, amusement), complemented by self-reports from participants acquired through established questionnaires.

Fig. 2 displays the distribution of three distinct stress states within the SWELL and WESAD databases. In SWELL-KW, HRV indices were computed by extracting the inter-beat interval (IBI) signals from each participant's electrocardiogram peak signals. The duration of the experiments was about three hours per participant. In WESAD shows that original sensor data were recorded using a chest-worn device (RespiBAN), with all signals sampled at 700 Hz.
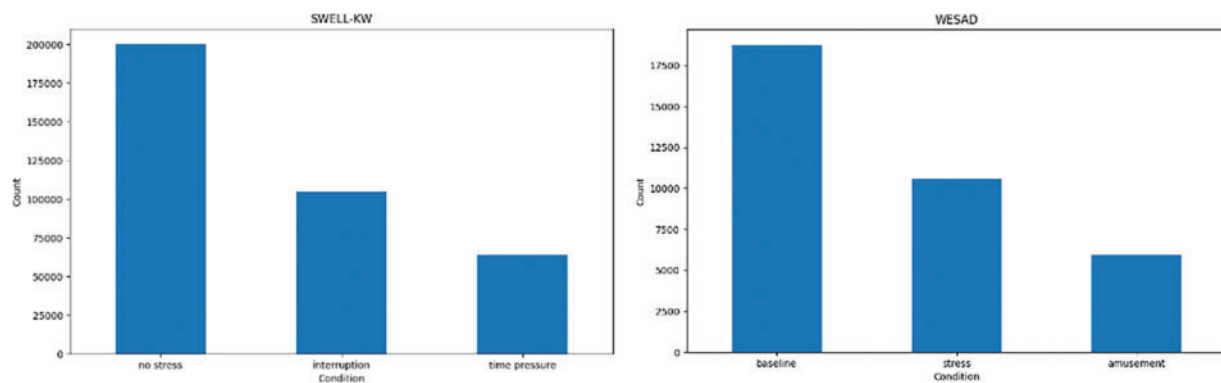

Figure 2: Distribution of different stress states in the SWELL-KW and WESAD databases

3.3 Feature Selection

Unlike the variance analysis method employed in [20], this study utilizes the Pearson correlation coefficient to conduct a linear analysis of all 34 features in the SWELL-KW dataset with respect to stress states, assessing and ranking the linear correlations between various features and stress. Fig. 3 presents a bar graph of the correlations between HRV features and stress states within the SWELL-KW dataset.

This paper selects 16 HRV parameters with correlations greater than 0.1, as shown in Table 1. The features ranked by the Pearson correlation coefficients were sequentially chosen to demonstrate the accuracy scores. The following two sections will introduce the developed CNN-Transformer-LSTM hybrid model and its performance metrics.
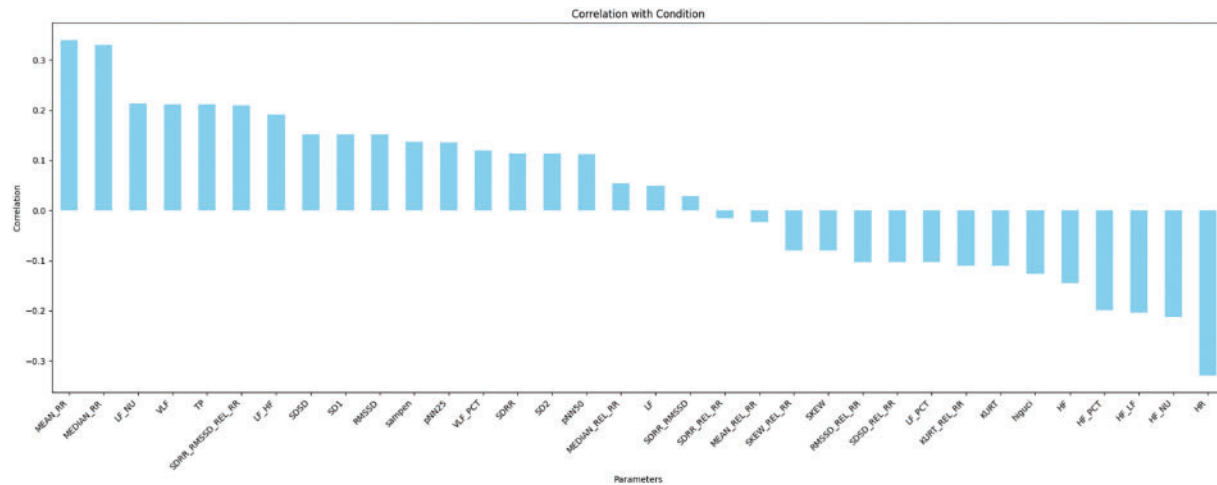
Figure 3: Bar chart of linear correlation using 34 HRV features

Table 1: Description of HRV parameters

No.	Feature	Meaning	Mean	Max	Min	StedDev
1	MEAN_RR	Mean of RR intervals	846.65	1322.01	547.49	124.60
2	MEDIAN_RR	Median of RR intervals	841.96	1653.12	517.29	132.32
3	LF_NU	Low frequency Normalized unit	95.56	99.98	69.88	4.12
4	VLF	Very low (0.003–0.04 Hz) frequency	2199.58	12617.98	159.48	1815.77
5	TP	Total HRV power spectrum	3185.35	13390.68	377.69	1923.22
6	SDRR_RMSSD _REL_RR	The ratio of SDRR_REL to RMSSD_REL	2.01	3.72	1.17	0.38
7	LF_HF	Ration of high to low frequency	115.97	7796.44	2.32	360.85
8	SDSD	Standard deviation of RR intervals	14.98	26.63	5.53	4.12
9	SD1	Poincaré plot Standard deviation perpendicular the line of identity	10.59	18.83	3.91	2.91
10	RMSSD	Root mean square of successive RR interval differences	14.98	26.63	5.53	4.12
11	SampEn	Sample entropy, which measures the regularity and complexity of a time series	2.06	2.23	0.43	0.21
12	pNN25	Percentage of consecutive RR intervals that differ by more than 25 ms	9.84	39.40	0.00	8.19
13	VLF_PCT	VLF percentage in HRV spectrum	64.29	97.74	19.03	16.77
14	SDRR	SD of RR intervals	109.35	563.49	27.23	77.12
15	SD2	Poincaré plot Standard deviation along the line of identity	154.18	796.85	38.31	109.17

4 Artificial Intelligence Models for Assessing Stress

The CNN-Transformer-LSTM model proposed in this paper (as illustrated in Fig. 4) represents an innovative improvement over the traditional Transformer model, specifically designed to enhance the accuracy of Heart Rate Variability (HRV) signal classification. The model consists of two main components: initially, a CNN encoder for feature extraction, followed by a classifier that combines Transformer and LSTM elements. The model utilizes CNN to extract local features from HRV data, employs the Transformer to model global dependencies, and leverages LSTM to manage long-term dependencies in time series, incorporating residual connections to boost overall performance.
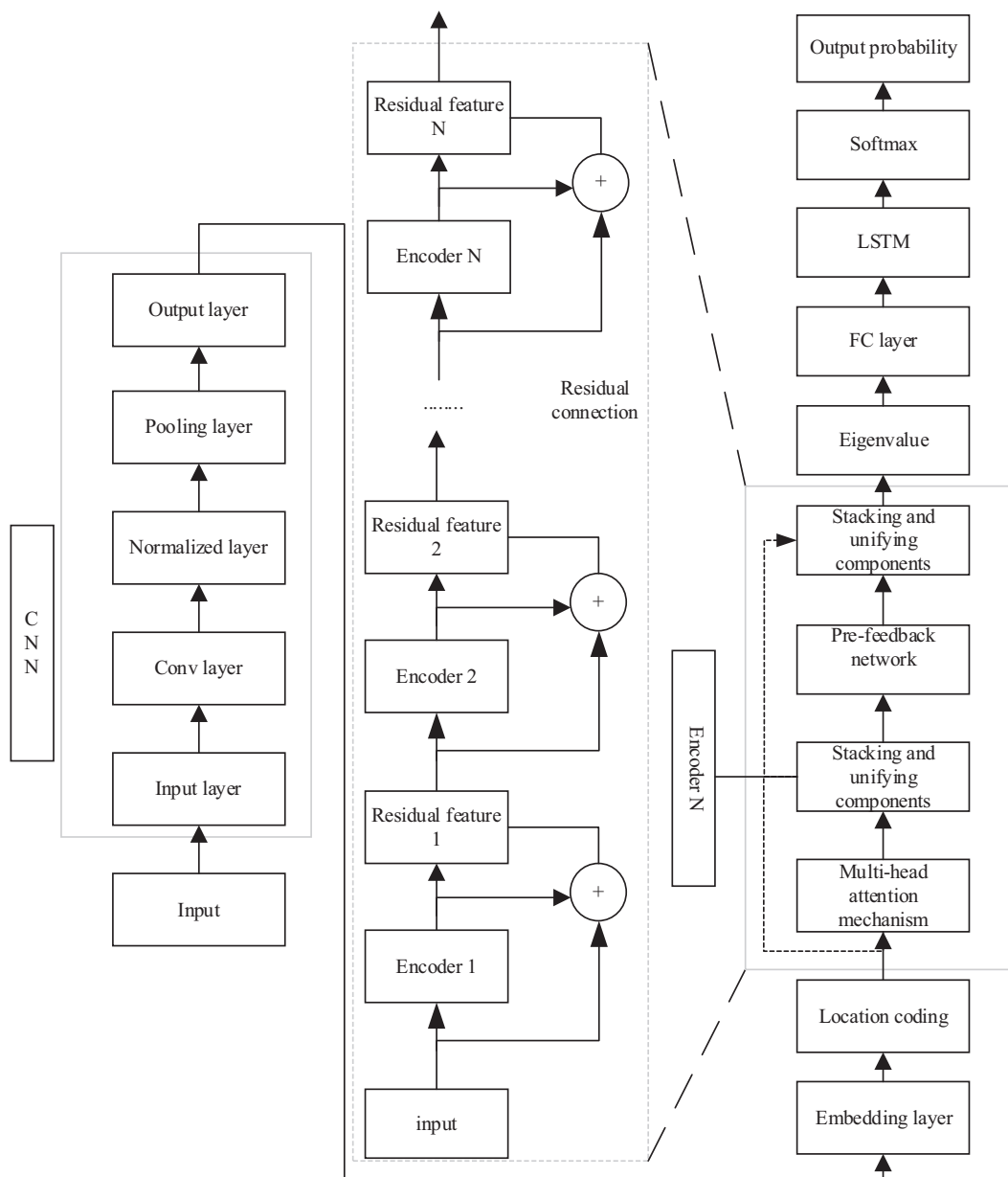
Figure 4: CNN-Transformer-LSTM model

During the data preprocessing and feature extraction stage, the model applies a one-dimensional Convolutional Neural Network (CNN) to extract local features from the input HRV signals. This step effectively identifies key features within the data, which are then relayed to the encoder part of the Transformer classifier. The Transformer captures dependencies between different positions in the input sequence through its self-attention mechanism. The self-attention mechanism computes the weights for each position by linear transformations of Query, Key, and Value, expressed by the formula:

$$Attention\,(Q, K, V) = soft \max \left(\frac{QK^T}{\sqrt{d_k}} \right) V \tag{1}$$

In the Transformer, the multi-head attention mechanism allows multiple attention heads to compute features in parallel and combines these outputs through a linear transformation, further enhancing the model's ability to capture complex dependencies. Additionally, each Transformer encoder layer includes a Feed-Forward Network (FFN), which independently processes the features at each position. The formula for the FFN is:

$$FFN\,(x) = \max\,(0, xW_1 + b_1)\,W_2 + b_2 \tag{2}$$

In the formula, W_1, W_2 are the weight matrices for the linear transformations, and $b_1 b_2$ are the bias terms. Residual connections and layer normalization mechanisms are employed to preserve input information and ensure the stability of the deep network. The formula for this integration is:

$$LayerNorm\,(x + sub - layer\,(x)) \tag{3}$$

The sequence encoded by the Transformer is subsequently passed to an LSTM network. The LSTM is tasked with learning temporal dynamic features within the input sequence and is capable of capturing long-term dependencies in time series data. The core of the LSTM consists of forget gates, input gates, and output gates, which dynamically update and store sequence information. The formulas for updating the state are as follows:

$$h_t = o_t * \tanh\,(C_t) \tag{4}$$

The current hidden state is represented by h_t, and the state of the memory cell is denoted as C_t.

To enhance the stability of the model, a residual connection mechanism combines the output of the LSTM with the output from the Transformer encoder. This residual structure allows for effective information flow, mitigates the issue of vanishing gradients, and ensures stable performance across multiple layers of the neural network. The formula for the residual connection is:

$$h_{lstm} = h_{lstm} + skip\_x \tag{5}$$

Overall, the hybrid CNN-Transformer-LSTM model effectively integrates the local feature extraction capabilities of CNNs, the global attention mechanism of Transformers, and the memory capabilities of LSTMs for time series, providing a comprehensive understanding of HRV signals. The CNN is tasked with extracting local features from raw HRV data, the Transformer captures long-distance dependencies through its self-attention mechanism, and the LSTM handles the dynamic variations in the time series. This triple architecture not only enhances the model's comprehension of complex time series data but also offers improved classification accuracy and stability.

5 Experimental Validation and Analysis

In this section, we detail the experimental results. Various methods, including CNN, SVM, LSTM, and Transformer, were evaluated. Given the random splitting process used, five-fold cross-validation was conducted to average the performance metrics of each classifier, ensuring comparability of the results. Additionally, this study specifically focused on the marginal improvement effects associated with increasing the number of parameters. This analysis helps to understand the potential benefits of maintaining or enhancing model accuracy while increasing complexity.

5.1 Performance of the Eight HRV Parameters before Use

In our study, the performance of the CNN-Transformer-LSTM hybrid model, developed using the SWELL-KW dataset, was assessed for multi-class stress classification. The discriminative metrics used included Precision, Recall, Accuracy, F1 score, Matthew's Correlation Coefficient (MCC), and the confusion matrix. The confusion matrix, a two-dimensional table is used to compare actual *vs.* predicted categories, containing four elements: True Positives (TP) are the correctly predicted positive instances; True Negatives (TN) are the correctly predicted negative instances; False Positives (FP) are the incorrectly predicted positive instances; and False Negatives (FN) are the incorrectly predicted negative instances. Thus, the performance metrics for a given class are represented as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$\text{F1} - \text{score} = \frac{2 \times Recall \times \text{Precision}}{Recall + \text{Precison}} \tag{9}$$

The higher these metrics, the better the performance of the model. Additionally, the model's bias and variance are crucial considerations. Bias refers to errors caused by incorrect assumptions in the algorithm, while variance arises from sensitivity to minor fluctuations in the training data. High bias can lead to underfitting, whereas high variance can lead to overfitting. While accuracy and the F1 score are crucial, they fail to capture the proportions of all categories within the confusion matrix. In contrast, the MCC offers a more holistic measure by considering the balance among all four categories (TP, TN, FP, FN) in the confusion matrix. The formula for MCC is as follows:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{10}$$

The advantage of the MCC is that it remains unaffected by the classification of positive and negative classes, making it more reliable than the F1 score, as it eliminates the problem of misclassifying positive classes.

Additionally, during the model training process, the cross-entropy loss function was employed to measure the discrepancy between the predicted probability distribution of the model's output and the

actual class labels. The formula for cross-entropy loss is:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log (p_{i,c}) \tag{11}$$

In this formula, N represents the number of samples, and C denotes the number of categories. The term $y_{i,c}$ indicates the true label of sample i for category c, while $p_{i,c}$ refers to the model's predicted probability for category c. By minimizing this loss function, the model can effectively optimize its predictions and enhance classification accuracy.

5.2 Performance of the 7 HRV Parameters before Use

The CNN-Transformer-LSTM model developed for the SWELL-KW dataset categorizes emotional stress states into three classes: no stress, time pressure, and interruptions. This model achieved exceedingly high accuracy, surpassing existing methods documented in the literature. Fig. 5a presents the confusion matrix generated by the CNN-Transformer-LSTM model using the SWELL-KW dataset. As shown in the figure, the classifier demonstrates high accuracy in predicting the correct labels. with all three categories showing less than 1% error when utilizing 7 heart rate variability features. In Fig. 5b, when 14 heart rate variability features are used, the accuracy reaches 100%.
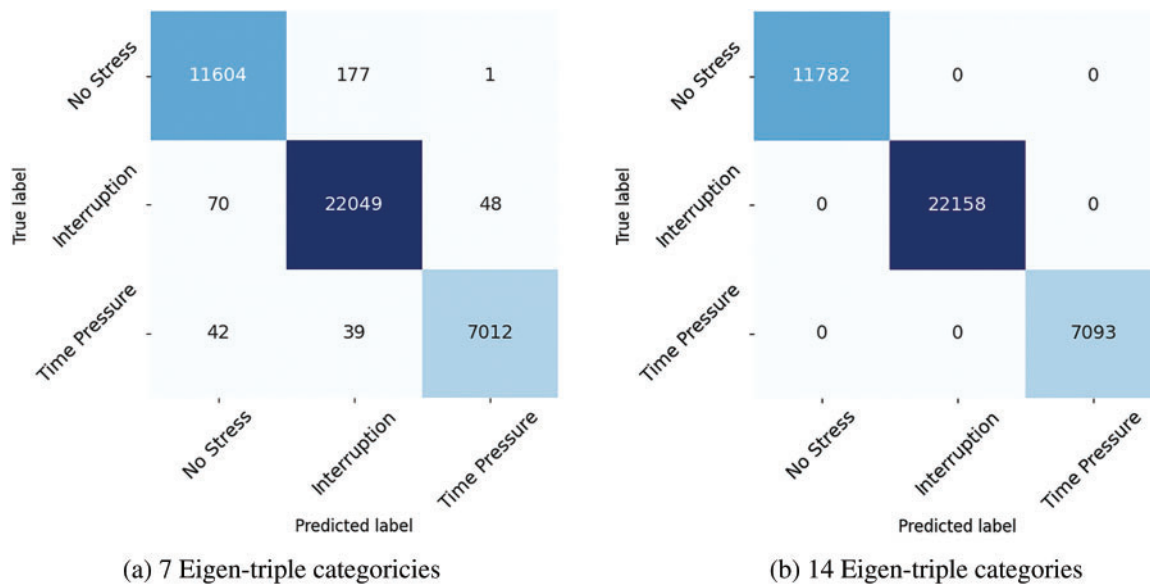

(a) 7 Eigen-triple categoricies (b) 14 Eigen-triple categories

Figure 5: Confusion matrix for three-class classification

Moreover, validation results indicate that the model did not experience overfitting during the training process. Fig. 6a shows the training and validation accuracy when utilizing 7 HRV features, with both metrics maintaining nearly identical levels. This consistency demonstrates the model's stability across training and validation sets, highlighting its robust generalization capabilities.

In Fig. 6b, the cross-entropy loss function is used to calculate loss values, further corroborating the model's stability throughout the training and validation phases. Experimental results reveal that the validation loss is slightly higher than the training loss, indicating that the model has not overfitted and meets the criteria for a well-fitted model.
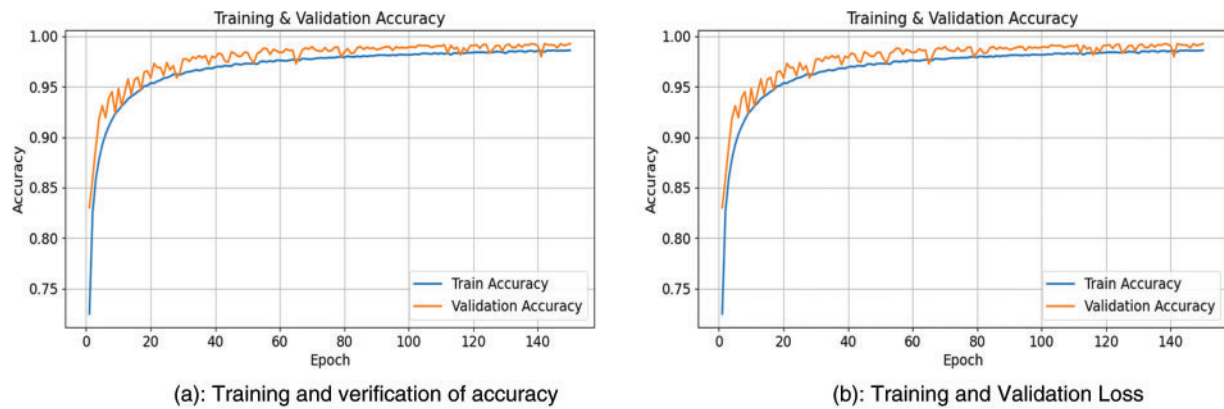
(a): Training and verification of accuracy   (b): Training and Validation Loss

Figure 6: Accuracy and loss curve

Table 2 provides a detailed overview of the performance of the developed CNN-Transformer-LSTM model in horizontal classification. Evidently, the CNN-Transformer-LSTM model crafted in this study achieved a peak accuracy score of 99.01% across all three classification levels.

Table 2: Performance for three-level categorization with the first eight ANOVA ranking features

Categorization	Precision	Recall	F1 score
No stress	0.9818	0.9868	0.9842
Time pressure	0.9915	0.9918	0.9916
Disruptions	0.9922	0.9994	0.9908

5.3 K-Fold Cross Validation

In this study, the default setting of 5-fold cross-validation was employed, with each iteration involving the training and performance evaluation of the model based on the test dataset, encompassing metrics such as precision, recall, accuracy, F1 score, and Matthew's Correlation Coefficient (MCC). As demonstrated in Table 3, the experimental results reveal that the model achieved an average accuracy of 99.09% on the test dataset. These results strongly suggest that the proposed model can categorize samples into their appropriate groups with remarkable precision.

Table 3: Five-fold cross-validation and average results

Features	MCC	Accuracy	Precision	Recall	F1 score	Cross-entropy loss
1	98.17%	98.91%	98.91%	98.91%	98.91%	0.0247
2	98.54%	99.13%	99.13%	99.13%	99.13%	0.0188
3	98.03%	98.84%	98.84%	98.84%	98.84%	0.0231
4	98.09%	98.86%	98.86%	98.86%	98.86%	0.0294

(Continued)

Table 3 (continued)

Features	MCC	Accuracy	Precision	Recall	F1 score	Cross-entropy loss
5	98.83%	99.31%	99.31%	99.31%	99.31%	0.0184
Average	98.33%	99.01%	99.01%	99.01%	99.01%	0.0192

5.4 Marginal Improvement

In data analysis and machine learning, marginal improvement refers to the impact of adding or changing a certain amount of input on model performance or outcomes. The marginal improvement effect illustrates that as the number of parameters in the model increases, so do performance metrics such as average precision, average recall, average accuracy, average F1 score, and average MCC (Matthew's Correlation Coefficient). From 7 to 16 parameters, the incremental changes in these metrics can be observed with each additional parameter.

In this study, when the number of parameters increased from 7 to 8, there was a 0.51% improvement in average precision, average recall, and average accuracy, and a 0.51% increase in average F1 score, while the average MCC rose by 0.85%. As the number of parameters further increased, varying degrees of gains were noted. By the 14th parameter, the values for precision, recall, accuracy, F1 score, and MCC all reached 100%. Upon adding the 15th parameter, a slight decline in performance was observed, but with the 16th parameter, performance again improved. These marginal improvement calculations are presented in Table 4.

Table 4: Marginal lift for different parameters

Arity	Marginal improvement MCC	Marginal improvement accuracy	Marginal improvement recall	Marginal improvement F1 score
7	0.00%	0.00%	0.00%	0.00%
8	0.89%	0.53%	0.53%	0.53%
9	0.07%	0.04%	0.04%	0.04%
10	0.10%	0.06%	0.06%	0.06%
11	0.46%	0.27%	0.27%	0.27%
12	0.06%	0.03%	0.03%	0.03%
13	0.03%	0.02%	0.02%	0.02%
14	0.06%	0.04%	0.04%	0.04%
15	−0.03%	−0.02%	−0.02%	−0.02%
16	0.03%	0.02%	0.02%	0.02%

As shown in Fig. 7, the accuracy improves as more features are used in model training. Specifically, the proposed model achieves over 99% accuracy using fewer than half of the features, ranked by the Pearson correlation coefficient.
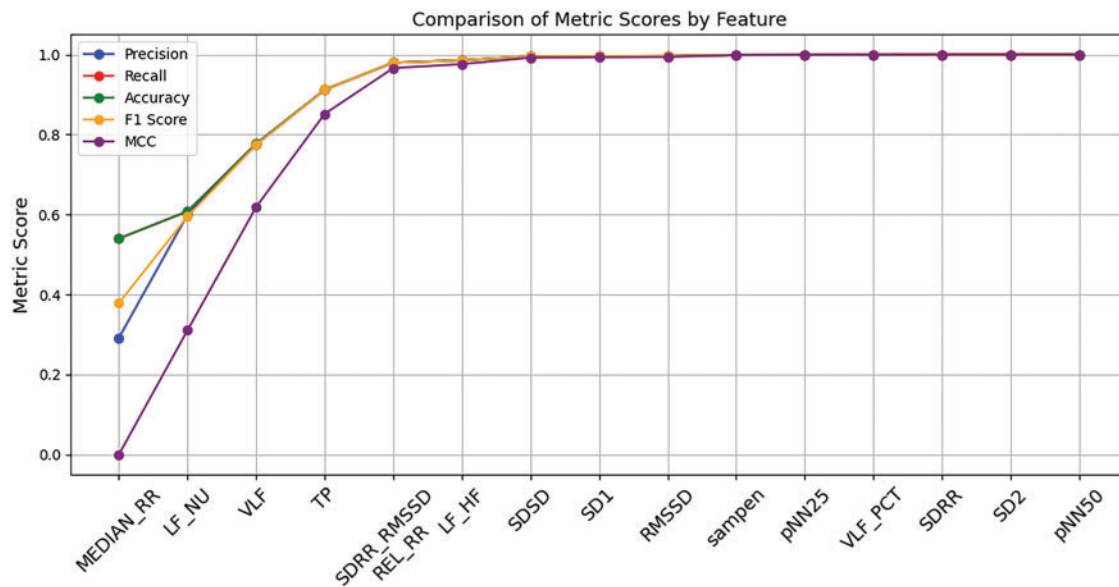
Figure 7: The impact of different HRV feature counts on classification accuracy

5.5 Quantitative Comparison with Existing Studies

In this study, the developed CNN-Transformer-LSTM hybrid model demonstrated superior classification accuracy with fewer features compared to existing stress classification models [9,10,17,18,20]. As illustrated in Table 5, utilizing only 7 HRV features, the model achieved a classification accuracy of 99.01%, significantly surpassing the 1D-CNN model that used 15 features as reported in [20]. This approach not only simplifies the complexity of feature processing but also greatly enhances computational efficiency. When the number of features was increased to 14, the classification accuracy further improved to 99.99%. Although the 1D-CNN model based on 34 HRV features showed higher accuracy, its substantial parameter requirements and prolonged execution cycles added to the computational burden.

Table 5: Quantitative comparison of results with other state-of-the-art models

Reference	Dataset	Binary/Multilevel	Features	Model	Accuracy	Recall	F1 score
[9]	SWELL-KW	Binary	17	SVM	92.75%	N.A	N.A
[10]	SWELL-KW AMIGOS [7]	3 class	N.A	CNN	98.30%	96.30%	95.80%
[18]	SWELL-KW	3 class	N.A	SVM	90.00%	N.A	N.A
[8]	WESAD [27]	Binary/3 class	7	ML	84.32% 95.21%	N.A	78.71%/94.24%
[24]	From experiment	3 class	12	LSTM	85.00%	N.A	N.A
[29]	From experiment	Binary/3 class	7	MLP	92.85% 64.28%	N.A	N.A
[19]	SWELL-KW	Binary	34	MLP	88.64%	92.68%	82.75%
[20]	SWELL-KW	3 class	34/15	1D-CNN	99.99%/96.50%	100%/97%	100%/96%

(Continued)

Table 5 (continued)

Reference	Dataset	Binary/Multilevel	Features	Model	Accuracy	Recall	F1 score
Our study	WESAD	3 class	7	Our model	96.88%	96.87%	96.88%
Our study	SWELL-KW	3 class	7	Our model	99.01%	99.01%	99.01%
Our study	SWELL-KW	3 class	14	Our model	99.99%	99.99%	100%

Moreover, compared to traditional models, the CNN-Transformer-LSTM hybrid model combines the rapid feature extraction capability of CNNs, the global dependency handling of Transformers, and the sequential dependency capture of LSTMs. This integrated approach effectively boosts the model's ability to process time-series data, successfully addressing the shortcomings of single models in managing long-term dependencies and positional information.

Tests on the WESAD dataset further verified the model's generalization capability, where it achieved a classification accuracy of 97.68%, surpassing the models discussed in [8,26]. These results underscore the potential of the designed hybrid model in handling complex biomedical time-series signals, excelling not only in traditional classification tasks but also in effectively managing long-term dependencies and positional information.

6 Concluding Remarks

In this study, an innovative CNN-Transformer-LSTM hybrid model was developed and validated for classifying stress levels based on physiological parameters. This model integrates Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and Transformer structures, significantly enhancing stress classification accuracy both within and outside the database. Experimental results indicate that using just 7 HRV features, the model achieved a classification accuracy of 99.01%. Additionally, the CNN in the model architecture effectively captures local patterns in time-series data through convolution and pooling operations; LSTMs process time-series data to capture long-term dependencies, crucial for understanding the dynamics of inter-beat intervals; meanwhile, the Transformer component enhances the model's comprehension of the relationships at different time points in HRV signals through its self-attention mechanism. Residual connections ensure lossless information transfer across multiple layers, preventing the issue of vanishing gradients. By integrating the strengths of these three models, both classification accuracy and model generalization capabilities were significantly improved. Furthermore, variance analysis for feature selection further reduced the model's dimensionality, enhancing computational efficiency, enabling efficient real-time monitoring both in traditional computing environments and on resource-constrained edge devices.

Despite its high accuracy in experiments, this model highly depends on the size and quality of datasets, particularly when trained with a limited number of HRV features. Moreover, the model's complexity requires substantial computational resources, which may limit its application on low-power or computationally weak devices.

In the future, plans include utilizing non-invasive remote photoplethysmography (rPPG) to extract HRV from facial videos, further reducing reliance on specialized equipment and making stress monitoring more accessible. Additionally, exploration of model compression and optimization algorithms is underway to enable rapid, accurate stress state detection on low-power devices, providing robust technical support for real-time mental health monitoring and further enhancing the practicality, accuracy, and reliability of the model.

Author Contributions: The authors confirm their contributions to this paper as follows: study conception and design: Peng Xia, Wu Zeng; data collection: Ye Jin, Yin Ni; analysis and interpretation of results: Peng Xia, Yin Ni, Wu Zeng; draft manuscript preparation: Peng Xia, Ye Jin. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in SWELL-KW at https://www.kaggle.com/datasets/qiriro/swell-heart-rate-variability-hrv (accessed on 18 November 2024); The data that support the findings of this study are openly available in WESAD at https://ubicomp.eti.uni-siegen.de/home/datasets/icmi18/ (accessed on 18 November 2024).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] S. Reisman, "Measurement of physiological stress," in *Proc. IEEE 23rd Northeast Bioeng. Conf.*, Durham, NH, USA, 1997, pp. 21–23.

[2] X. Li, W. Zhu, X. Sui, A. Zhang, L. Chi and L. Lv, "Assessing workplace stress among nurses using heart rate variability analysis with wearable ECG device—A pilot study," *Front. Public Health*, vol. 9, 2021, Art. no. 810577.

[3] E. Cayir, T. Cunningham, R. Ackard, J. Haizlip, J. Logan and G. Yan, "The effects of the medical pause on physiological stress markers among health care providers: A pilot randomized controlled trial," *West J. Nurs. Res.*, vol. 44, no. 11, pp. 1036–1046, 2022. doi: 10.1177/01939459211027657.

[4] D. Muhajir, F. Mahananto, and N. A. Sani, "Stress level measurements using heart rate variability analysis on Android based application," *Procedia Comput. Sci.*, vol. 197, no. 15, pp. 189–197, Jan. 2022. doi: 10.1016/j.procs.2021.12.200.

[5] J. Held, A. Vîslă, C. Wolfer, N. Messerli-Bürgy, and C. Flückiger, "Heart rate variability change during a stressful cognitive task in individuals with anxiety and control participants," *BMC Psychol.*, vol. 9, no. 1, Mar. 2021, Art. no. 44. doi: 10.1186/s40359-021-00551-4.

[6] K. M. Dalmeida and G. L. Masala, "HRV features as viable physiological markers for stress detection using wearable devices," *Sensors*, vol. 21, no. 8, Apr. 2021, Art. no. 2873. doi: 10.3390/s21082873.

[7] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A dataset for affect, personality and mood research on individuals and groups," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 479–493, Apr./Jun. 2021. doi: 10.1109/TAFFC.2018.2884461.

[8] P. Bobade and M. Vani, "Stress detection with machine learning and deep learning using multimodal physiological data," in *Proc. 2nd Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Coimbatore, India, Jul. 2020, pp. 51–57.

[9] S. Sriramprakash, V. D. Prasanna, and O. V. R. Murthy, "Stress detection in working people," *Procedia Comput. Sci.*, vol. 115, no. 8, pp. 359–366, Dec. 2017. doi: 10.1016/j.procs.2017.09.090.

[10] P. Sarkar and A. Etemad, "Self-supervised learning for ECG-based emotion recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.(ICASSP)*, Barcelona, Spain, May 2020, pp. 3217–3221.

[11] D. F. Dinges *et al.*, "Optical computer recognition of facial expressions associated with stress induced by performance demands," *Aviation Space Environ. Med.*, vol. 76, pp. 172–182, 2005.

[12] Q. Lin, T. Li, P. M. Shakeel, and R. D. J. Samuel, "Advanced artificial intelligence in heart rate and blood pressure monitoring for stress management," *J. Ambient Intell. Hum. Comput.*, vol. 12, no. 3, pp. 3329–3340, 2021. doi: 10.1007/s12652-020-02650-3.

[13] Y. Yue, D. Liu, S. Fu, and X. Zhou, "Heart rate and heart rate variability as classification features for mental fatigue using short-term PPG signals via smartphones instead of ECG recordings," in *2021 13th Int. Conf. Commun. Softw. Netw. (ICCSN)*, Chongqing, China, 2021, pp. 370–376.

[14] P. Melillo, M. Bracale, and L. Pecchia, "Nonlinear heart rate variability features for real-life stress detection. Case study: Students under stress due to university examination," *BioMed. Eng. OnLine*, vol. 10, no. 1, 2011, Art. no. 96. doi: 10.1186/1475-925X-10-96.

[15] S. Lee *et al.*, "Mental stress assessment using ultra short term HRV analysis based on non-linear method," *Biosensors*, vol. 12, no. 7, Jun. 2022, Art. no. 465. doi: 10.3390/bios12070465.

[16] A. Shikha, L. Arya, and D. Sethia, "HRV and GSR as viable physiological markers for mental health recognition," in *2022 14th Int. Conf. COMmun. Syst. NETw. (COMSNETS)*, Bangalore, India, 2022, pp. 37–42.

[17] S. Koldijk, M. Sappelli, S. V. erberne, M. A. Neerincx, and W. Kraaij, "SWELL knowledge work dataset for stress and user modeling research," in *Proc. 16th Int. Conf. Multimodal Interact.*, Nov. 2014, pp. 291–298.

[18] S. Koldijk, M. A. Neerincx, and W. Kraaij, "Detecting work stress in offices by combining unobtrusive sensors," *IEEE Trans. Affect. Comput.*, vol. 9, no. 2, pp. 227–239, Apr. 2018. doi: 10.1109/TAFFC.2016.2610975.

[19] M. Albaladejo-González, J. A. Ruipérez-Valiente, and F. G. Mármol, "Evaluating different configurations of machine learning models and their transfer learning capabilities for stress detection using heart rate," *J. Ambient Intell. Human. Comput.*, pp. 1–11, Aug. 2022. doi: 10.1007/s12652-022-04365-z.

[20] J. A. Mortensen, M. E. Mollov, A. Chatterjee, D. Ghose, and F. Y. Li, "Multi-class stress detection through heart rate variability: A deep neural network based study," *IEEE Access*, vol. 11, no. 2, pp. 57470–57480, 2023. doi: 10.1109/ACCESS.2023.3274478.

[21] Y. Kaya, F. Kuncan, and R. Tekin, "A new approach for congestive heart failure and arrhythmia classification using angle transformation with LSTM," *Arab. J. Sci. Eng.*, vol. 47, no. 8, pp. 10497–10513, 2022. doi: 10.1007/s13369-022-06617-8.

[22] N. Bu, M. Fukami, and O. Fukuda, "Pattern recognition of mental stress levels from differential RRI time series using LSTM networks," in *2021 IEEE 3rd Global Conf. Life Sci. Technol. (LifeTech)*, Nara, Japan, 2021, pp. 408–411.

[23] R. Jegan, S. Mathuranjani, and P. Sherly, "Mental stress detection and classification using SVM classifier: A pilot study," in *2022 6th Int. Conf. Devices, Circuits Syst. (ICDCS)*, Coimbatore, India, 2022, pp. 139–143.

[24] L. V. Coutts, D. Plans, A. W. Brown, and J. Collomosse, "Deep learning with wearable based heart rate variability for prediction of mental and general health," *J. Biomed. Inform.*, vol. 112, 2020, Art. no. 103610. doi: 10.1016/j.jbi.2020.103610.

[25] P. Gomathi Shankari, D. A. Hrithik, D. Ravikumar, D. Chaitanya, S. P. Pushpa Mala and S. Kokila, "Heart rate variability and machine learning for stress analysis," in *2022 4th Int. Conf. Circuits, Control, Commun. Comput. (I4C)*, Bangalore, India, 2022, 73–78. doi: 10.1109/I4C57141.2022.10057650.

[26] W. Pajong, P. Eiamcharoen, K. Srisomboon, and W. Lee, "Time series based emotion classification algorithm exploiting deep learning," in *2023 Res., Invention, Innov. Congress: Innov. Electr. Electron. (RI2C)*, Bangkok, Thailand, 2023, pp. 151–154.

[27] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. V. Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interact.*, Oct. 2018, pp. 400–408.

[28] D. S. Benita, A. S. Ebenezer, L. Susmitha, M. S. P. Subathra, and S. J. Priya, "Stress detection using CNN on the WESAD dataset," in *2024 Int. Conf. Emerg. Syst. Intell. Comput. (ESIC)*, Bhubaneswar, India, 2024, pp. 308–313.

[29] A. Arsalan, M. Majid, A. R. Butt, and S. M. Anwar, "Classification of perceived mental stress using a commercially available EEG headband," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 6, pp. 2257–2264, Nov. 2019. doi: 10.1109/JBHI.2019.2926407.