A Deep Learning Approach for Early Stress Detection Using Electrodermal Activity through Wearable Devices

Ashutosh Singh[†], Amit Kumar, Khushdeep Singh, and Santosh Kumar

Department of Data Science and Artificial Intelligence
IIIT Naya Raipur, Chhattisgarh, India, 493661
ashutoshs22102@iiitnr.edu.in, amit22102@iiitnr.edu.in,
khushdeep22102@iiitnr.edu.in, santosh@iiitnr.edu.in

Abstract. Stress is a critical health issue with significant implications for both physical and mental well-being. Early detection and accurate classification of stress levels are crucial for timely intervention and effective management. This study proposes a deep-learning framework for early stress detection using physiological data from wrist-worn devices. The framework utilizes Electrodermal Activity (EDA) signals, which are processed with advanced preprocessing techniques. It uses both extracted features and the complete feature set of the EDA signal in a custom deep-learning model to effectively distinguish between stress and non-stress states. The study also explores multimodal signal fusion by combining EDA with electrocardiogram (ECG) and photoplethysmography (PPG) signals to evaluate its effectiveness in early stress detection. However, the results show that EDA signals alone outperform the combined signals for early stress detection. The methodology was trained and tested on six publicly available EDA-based stress datasets: VerBIO, UTD, WESAD, SWELL-KW, MSD Nurses, and DriveDB. Experimental results demonstrate that the framework achieves 95% accuracy and a 90% F1-score in early stress detection. These findings highlight the potential of the proposed framework for real-world applications in stress management, providing a reliable, continuous, and non-invasive tool for early stress monitoring through wearable technology.

Keywords: Stress Detection, Electrodermal Activity, Deep Learning, Multimodal Fusion, Physiological Signals

1 Introduction

Early stress detection is gaining prominence due to its potential for fast diagnosis of stress severity levels and its application in analyzing stress associated with critical health conditions using artificial intelligence (AI) [1]. Stress, a complex psycho-physiological state, manifests as a response to various challenging situations or specific events encountered in daily life [2]. According to existing literature, stress can be categorized into several types: (1) acute stress, (2) chronic stress, (3) episodic acute stress, (4) eustress, and (5) distress. Acute stress is typically short-term and arises from immediate pressures, while chronic stress persists over time and often results from ongoing challenges. Episodic acute stress is observed in individuals with chaotic lifestyles, involving recurrent episodes of acute stress [3]. Eustress, a positive form of stress, occurs in situations perceived as beneficial, whereas distress is a harmful form of stress with negative impacts on physical and mental health [4]. Chronic or extreme stress can lead to serious health issues, including depression, cardiovascular problems, immune system dysfunction, digestive disorders, and other related concerns [5].

Stress can be measured through psychological or physiological data. Traditional methods primarily rely on psychological data for early stress detection, often involving interactive questionnaires; however, these methods frequently lack accuracy [6]. While statistical machine learning techniques are commonly employed for stress detection, they often fall short in providing comprehensive analysis when dealing with multimodal data. To overcome these limitations, AI has emerged as a pivotal tool for early stress classification. Researchers are increasingly leveraging AI-driven models to predict stress severity with greater accuracy, using data from smartphones, smartwatches, and human-computer interfaces (HCI) [7].

Various machine learning and deep learning approaches have been proposed for early stress diagnosis. Psychological data is typically collected through self-report devices and health monitoring tools, enabling stress analysis via verbal questionnaires or interviews [8]. Interactions between individuals and professionals allow for the expression and assessment of stress related experiences in diverse environments. Observations of behavioral changes, physiological alterations, psychological fluctuations, mood swings, sleep disturbances, and shifts in interpersonal relationships are crucial for interpreting psychological responses to stress [9]. For physiological data-based stress detection, wearable devices play a key role in monitoring indicators such as heart rate (HR), heart rate variability (HRV), blood pressure (BP), skin temperature, EDA, ECG, and PPG [10]. Smart devices equipped with these sensors can facilitate early stress detection.

This study investigates the feasibility of utilizing EDA signals for early stress detection using wrist-worn devices. EDA signals, which reflect skin conductance changes, can be tracked using smartwatches like Fitbit Sense and Empatica E4, which employ EDA sensors for stress and mood analysis. The study evaluates the effectiveness of three deep learning models, the Feedforward Neural Network (FNN), Convolutional Neural Network (CNN), and Wide and Deep Learning models, in classifying stress and non-stress states. Training and testing were performed using raw and extracted features of EDA to identify the most effective method for predicting stress severity. The results demonstrate that extracted features are more effective. Additionally, this work examines the potential of customized multimodal approaches for wrist-mounted wearables, identifying the modalities that offer the highest accuracy in stress classification.

The major contributions of this work are as follows:

1. Development of a deep learning framework for early stress detection using EDA signals from wrist-worn devices.
2. Implementation and comparison of multiple deep learning models, FNN, CNN, and Wide and Deep, for early stress classification.
3. Evaluation of stress classification accuracy with both raw and extracted EDA features on six publicly available datasets, demonstrating superior performance with extracted features.
4. Comparison of the proposed EDA-based framework with classifications using PPG, ECG, and their combinations, showing that EDA-based classification achieves the highest performance in stress assessment.

2 Related Work

This section reviews the literature on early stress detection, with a particular focus on the role of AI in analyzing physiological signals. Initial studies primarily utilized traditional machine learning techniques. For example, Moghimi et al. [11] employed Support Vector Machines (SVM) to classify stress based on signals such as Skin Conductance Response (SCR), Electroencephalogram (EEG), and Heart Rate (HR). Similarly, Maaoui et al. [12] used ensemble learning methods to classify stress based on multiple signals, including Blood Volume Pulse (BVP) and Electromyography (EMG). However, these approaches often faced challenges due to the complexity of multimodal data.

Recent advancements have seen a shift towards deep learning, which has demonstrated improved accuracy in stress detection. Movahedi et al. [13] proposed a deep learning framework to extract features from EEG signals using wearable devices, though it faced challenges with overfitting. To address this, Yang et al. [14] developed a hierarchical neural network combined with ensemble learning to improve the robustness of stress classification. Electrodermal Activity (EDA) has proven particularly effective for stress detection. Studies by Choi et al. [15] and Anusha et al. [16] successfully used EDA signals to assess stress and anxiety. Additionally, Zontone et al. [17] combined EDA with ECG data to measure driver anxiety, while Liu et al. [18] employed Linear Discriminant Analysis (LDA) on EDA signals to classify stress during driving, achieving an 81.82% recognition rate.

The potential of deep learning combined with multimodal physiological data for stress detection is also becoming increasingly evident. Rim et al. [19] reviewed the application of deep learning to various physiological signals, including EMG, ECG, and EEG, for medical tasks. Similarly, Santamaria et al. [20] used a deep convolutional neural network (DCNN) with ECG and EDA data for emotion recognition. Thiam et al. [21] employed a CNN architecture with EDA, ECG, and EMG signals, achieving 84.57% accuracy in distinguishing between painful and non-painful states. These studies illustrate the growing use of deep learning and multimodal data fusion for early stress detection, with EDA emerging as a key physiological marker. A summary of recent work and a comparison with the proposed EDA method for stress detection are provided in Table-1.

Table 1. Comparison of Recent Work with the Proposed EDA Method for Stress Detection

Authors	Journals	Year	Methodology Used	Signal Used	Key Findings
Zhu et al.[22]	IEEE Journal of Biomedical and Health Informatics	2023	Support Vector Machine	EDA	92.9%
Rahma et al.[23]	Journal of Medical Signals & Sensors	2022	Extreme Learning Machine	EDA	91.0%
Nardelli et al.[24]	Computers in Biology and Medicine	2022	Machine Learning	EDA	89.7%
Greco et al.[25]	IEEE Transactions on Affective Computing	2021	Support Vector Machine	EDA	91.62%
Anusha et al.[16]	IEEE Journal of Biomedical and Health Informatics	2019	Supervised Machine Learning	EDA	85.06%
Ours	**A2IICPR**	**2024**	**Convolutional Neural Network**	**EDA**	**95.0%**

3 System Overview

As illustrated in Figure-1, the proposed early stress detection framework utilizes a wrist-worn device, such as a smartwatch, to capture various physiological signals, including EDA, ECG, PPG, and other signals. These signals are transmitted to a computer for analysis, where features are extracted and subjected to deep learning techniques for signal processing and classification. The deep learning approach is employed to analyze and integrate these signals, enabling accurate assessment of stress severity.

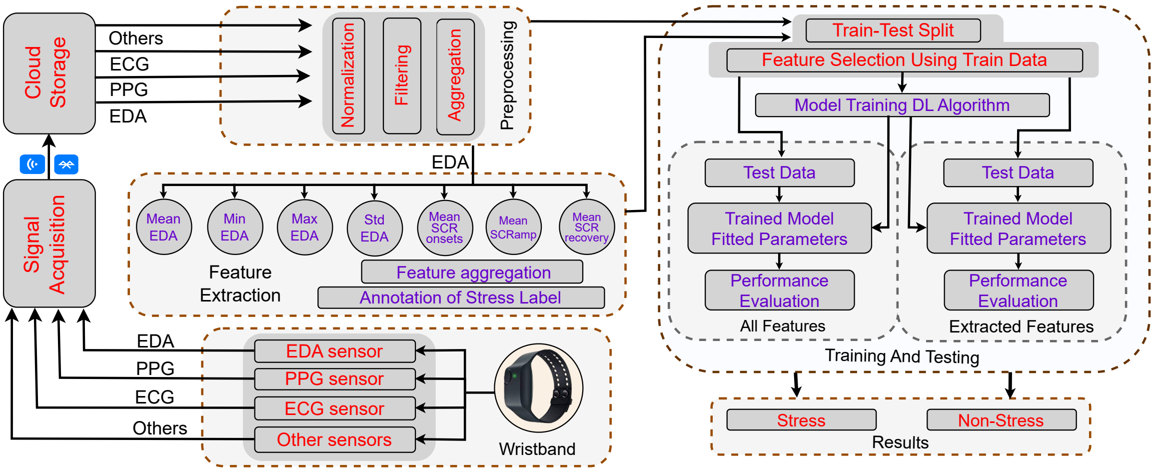

Fig. 1. Overview of the proposed method for early stress detection using electrodermal activity through wearable devices

3.1 Reason for Using EDA in Early Stress Detection

EDA measures variations in the skin electrical properties, which are closely related to sweat gland activity. EDA signals are characterized by two primary components: the tonic and phasic levels. The tonic level, indicated by the Skin Conductance Level (SCL), serves as a stable, slow-changing baseline that reflects skin moisture and autonomic function. On the other hand, the phasic level, measured by the Skin Conductance Response (SCR), exhibits more dynamic fluctuations, especially in reaction to stimuli. SCRs can appear as sharp peaks following specific events (event-related SCRs) or can occur spontaneously (non-specific SCRs). Essential features of event-related SCRs include latency (time from stimulus to response onset), peak amplitude (difference between onset and peak), rise time (duration from onset to peak), and recovery time (time from peak to full recovery) as shown in the Figure-2. These features provide valuable insights into emotional responses and engagement with stimuli, making EDA a powerful tool for real-time monitoring.

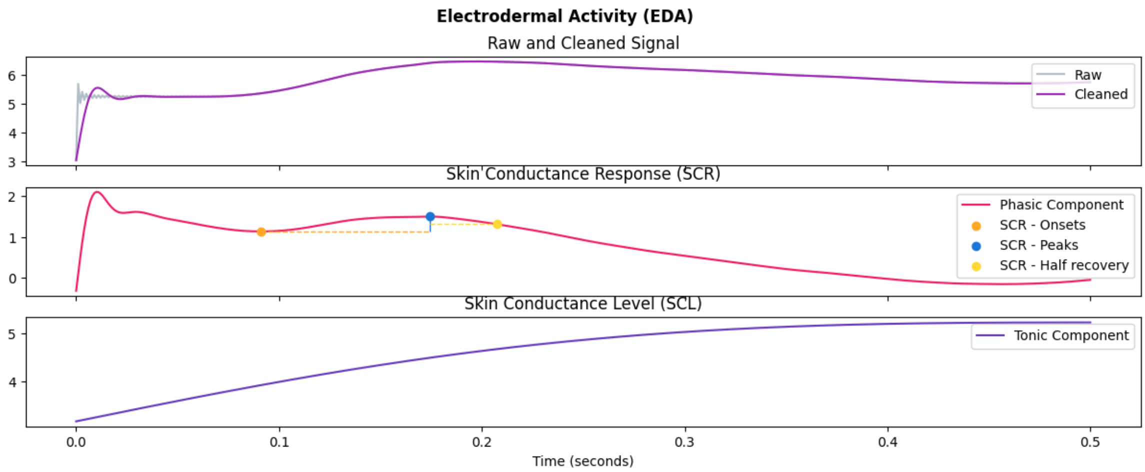

Fig. 2. Components and key features of EDA signals: tonic (SCL) and phasic (SCR) levels, with SCR features including latency, peak amplitude, rise time, and recovery time.

3.2 Dataset Description

This study leverages six publicly available datasets, each containing EDA signals, along with additional physiological signals ECG, PPG, and other signals. These datasets were selected for their diversity in signal types, coverage across different skin areas, and variation in activities and age groups of participants. The data was collected using various devices, each tailored to the specific objectives of the respective studies. Figure-3 illustrates the different devices used for data collection. The datasets utilized in this study are VerBIO[26], SWELL-KW[27], WESAD[28], UTD[29], MSD Nurse[30], and DriveDB[31]. Table-2 provides a comprehensive overview of these datasets, with further details discussed below. The primary reason for selecting these datasets is the availability of EDA signals recorded using different devices during various activities, which enhances the effectiveness of stress severity prediction models.

VerBIO The VerBIO dataset includes physiological signals recorded during public speaking sessions, with data from 18 individuals captured using Empatica E4 devices. A total of 344 speeches were delivered to either real or virtual audiences. The EDA data was recorded at 4 Hz and labeled with self-reports collected both before and after each session.

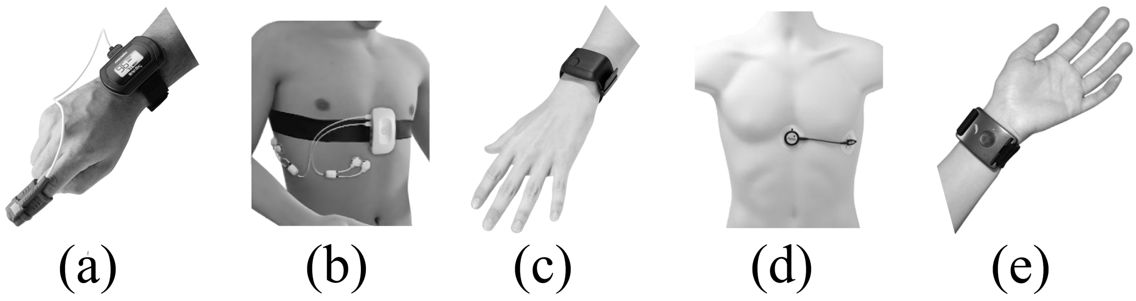

Fig. 3. Wearable devices used in physiological data acquisition (a) Nonin Wristox2, (b) RespiBAN, (c) Empatica E4, (d) Actiwave Cardio, (e) Affectiva-Q

Table 2. EDA-based datasets used in this study

Dataset	Subjects	Device	Activities	Modalities
VerBIO [26]	55	Empatica E4, Actiwave Cardio	Public speaking	EDA, ECG, BVP, Temperature, 3-axis accelerometer
SWELL-KW [27]	25	Empatica E4, Actiwave Cardio	Public speaking	EDA, ECG, Video (Face and upper body), Posture, Computer performances
WESAD [28]	15	RespiBAN, Empatica E4	Reading, Video Watching, Public Speaking, Meditation, Mental Arithmetic task	EDA, ECG, BVP, Temperature, Accelerometer
UTD [29]	20	Affectiva Q, Curve, Nonin, WristOx2	Watching video, Stroop test, Walking, Counting	EDA, HR, BVP, Accelerometer, Temperature
MSD Nurses [30]	15	Empatica E4	Regular duties in a hospital setting during COVID-19 outbreak	EDA, HR, ST, BVP, Accelerometer, IBI
Drive DB [31]	9	Multiple single devices	Rest, Highway driving, City driving	ECG, EMG, GSR, Respiration

UTD The UTD dataset, collected by the University of Texas at Dallas, examines responses to cognitive, emotional, and physical stress among 20 university students. It includes data from seven staged activities, with physiological signals recorded using wrist-worn devices. These signals include EDA, temperature, acceleration, HR, and blood oxygen saturation (SpO2), providing valuable insights into the dynamics of stress and relaxation.

WESAD The WESAD dataset focuses on emotional state recognition through physiological signals, including EDA, ECG, EMG, respiration, temperature, and acceleration. The dataset features recordings from 15 participants engaged in various activities, such as watching videos, public speaking, and meditating. EDA data was collected at a sampling rate of 4 Hz using Empatica E4 devices.

SWELL-KW The SWELL-KW dataset includes data from 25 participants performing knowledge work tasks under stressors like email interruptions and time pressure. The dataset records multiple data types, including computer logging, facial expressions, and physiological signals like heart rate variability and skin conductance, offering insights into stress management in professional environments.

MSD Nurses The MSD Nurse dataset from the University of Louisiana at Lafayette contains data from 15 female nurses during their regular hospital shifts, totaling 1,250 hours. Physiological signals, including EDA, heart rate, skin temperature, and accelerometer data, were recorded using Empatica E4 wearables. This dataset was collected to support the well-being of nurses by enabling real-time stress detection. Exclusion criteria, such as pregnancy and chronic diseases, were applied to ensure the integrity of the data.

DriveDB The DriveDB dataset, collected by Healey and Picard, includes data from 17 drivers, with complete recordings for nine participants. It features physiological signals such as ECG, EMG, GSR, and respiration, recorded during rest periods, highway driving, and city driving. The dataset is segmented into 1-minute intervals labeled as either relaxed or stressed, providing valuable data for road safety research and the development of stress detection systems.

4 Methodology

The method for predicting stress severity involves mapping input physiological signals to an output indicating stress levels. Our approach focuses on early diagnosis and accurate prediction of stress severity through a binary classification task, distinguishing between stress with severity measures and non-stress states. Let $E = [e_1, e_2, e_t, \cdots, e_n]$ represent a sample of physiological signal input, the EDA data sample. Here, e_t signifies the EDA measurement obtained using Siemens at discrete time t. The stress detection problem is formulated as follows:

$$[s, \epsilon] = C(E_w, \theta) \tag{1}$$

Where $s = \{0, 1\}$ represents the severity stress state, with $s = 1$ indicating the stress state (Unhealthy) and $s = 0$ indicating the non-stress state (Healthy). $E_w \in E$ represents a discriminatory subset of EDA physiological signal features based on the segmentation method, where $w < n$. C denotes the classifier, and θ represents the corresponding coefficients. A deep learning algorithm learns the classifier based on physiological signals and associated discriminatory feature coefficients.

4.1 Data Preprocessing

Data preprocessing is a critical step in this study, as it ensures the quality and consistency of the EDA signals and other signals before they are used for stress severity prediction. This process involves four main steps, as illustrated in Figure-4.

Data Segmentation The collected EDA data represents various everyday activities of subjects, with acquisition times ranging from five minutes to more than two hours, and sometimes even a week. However, handling extensive amounts of EDA data poses challenges regarding computational cost and data consistency. To address these challenges, we segmented the collected EDA data into distinct lengths, aiming to create manageable data chunks that still capture essential features [32]. This segmentation reduces processing time and computational costs. Second-order statistical measures were employed for clustering the segmented data, and labels were annotated by minute using a non-overlapping sliding window approach (Hamming Window Method) for further processing [33]. The datasets used include Verbio, WESAD, UDT, MSD-Nurse, and DriveDB, encompassing activities across multiple domains. To ensure a fair comparison and enhance dataset consistency, data related to physical activity from UTD were excluded. Specifically, data labeled as "PhysicalStress," obtained when participants were standing or walking, was removed. The Hamming window-based approach is used to taper the physiological signal windows to zero at the start and end of each cluster of the discriminatory set of features for accurate stress prediction. The window function $w(n)$ is defined as follows:

$$x[n] = x[n] \times w(n) \quad \text{for } 0 \leq n \leq N - 1 \tag{2}$$

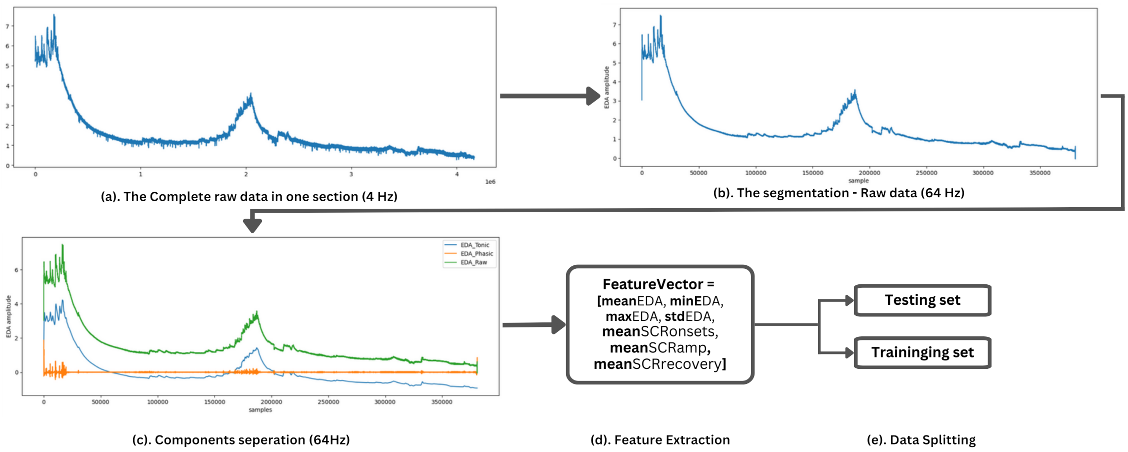

Fig. 4. Pipeline for Raw Data Processing: (a) Data undergoes four steps: (b) segmentation, (c) component separation, (d) feature extraction, and (e) data splitting.

$$w[n] = 0.54 - 0.46 \times \cos\left(\frac{2\pi n}{N-1}\right) \quad \text{for } 0 \leq n \leq N-1 \tag{3}$$

After applying the Hamming sliding window method, the segmentation results for all datasets are presented in Table-3.

Table 3. Overview of segmentation for each dataset

Dataset	Selected Subjects	No. of Segments	
		Stress	Non-stress
UTD	20	12,470	19,192
VerBIO	18	99	141
WESAD	15	3546	27702
SWELL-KW	25	264	1,296
MSD Nurses	15	4,785	7,660
DriveDB	9	50	18

Normalization Normalization is crucial for mitigating data redundancy in raw EDA signals, laying the groundwork for effective feature extraction and classification. Initially, the data is scaled into the [0,1] range, followed by the application of the moving average filtering method to minimize the effects of noise and other specific artifacts in the normalized data [34]. The z-score method is then employed to ensure standardized data for robust analysis, defined as follows:

$$z = \frac{x - \mu}{\sigma} \tag{4}$$

where μ is the mean of a sample, σ is the standard deviation of the sample, z is the standard score, and x is the observed value.

Filtering To mitigate noise and unwanted frequency components in raw EDA data, a Moving Average (MA) based filtering method is employed. This, combined with a low-pass filter, smooths signals while preserving meaningful peaks [35]. The evaluation of $x[N]$ using M prior values ensures effective noise suppression. Smoothing signals ($X[N]$) while preserving peaks is essential for the accurate representation of physiological responses. A low-pass filter of order 5, combined with MA, enhances signal clarity, which is crucial for deeper insights into emotional arousal and stress severity prediction.

$$X[N] = \frac{1}{M} \sum_{j=0}^{M} x[N-j] \tag{5}$$

Component Separation To address the imbalance in the proposed dataset and other related datasets, an equal number of segments from each dataset were selected. Following data segmentation, normalization, and filtering, the raw EDA data still contained redundant information. To enhance the results, additional data processing was necessary. Artifacts in the EDA data, caused by sensor movement on the skin and changes in skin moisture, posed challenges for accurate stress assessment. Therefore, it was crucial to extract the SCR and SCL components and apply artifact removal techniques for more accurate analysis. The cvxEDA model [36], which is based on maximum a posteriori (MAP) estimation, sparsity, and convex optimization, was employed to decompose the SCR and SCL components, thereby refining the accuracy of stress severity prediction models.

Feature Extraction Feature extraction is vital for building feature vectors for training the proposed framework. Since training with all feature vectors in physiological EDA signals is computationally expensive, we perform data clustering of selected discriminatory features to reduce computational costs. Additionally, statistical characteristics and SCR attributes from recorded physiological signals are considered for a detailed analysis of stress severity for early diagnosis [28]. The feature vectors consist of different discriminatory components of the physiological signals, as shown below:

$$\begin{aligned}\text{FeatureVector} = [&\text{mean}_{\text{EDA}}, \text{min}_{\text{EDA}}, \text{max}_{\text{EDA}}, \text{std}_{\text{EDA}},\\ &\text{mean}_{\text{SCRonsets}}, \text{mean}_{\text{SCRamp}}, \text{mean}_{\text{SCRrecovery}}]\end{aligned} \tag{6}$$

Here, meanEDA represents the average EDA value, minEDA and maxEDA represent the minimum and maximum values, respectively, and stdEDA quantifies the variability of EDA values. The dataset containing all features, without feature extraction, is utilized for training the deep learning models, which are evaluated based on their classification performance against feature vectors obtained through extraction techniques.

Splitting of Physiological Dataset After preprocessing, the data is partitioned into two segments: (1) training and (2) testing, based on either collected samples or subjects. During the training phase, the preprocessed data is converted into feature vectors to train the models. This training phase helps mitigate interference factors or latent variables inherent in individual physiological signals, including the relationship between emotional states and EDA data. Consequently, subject-independent analyses are more advantageous compared to subject-dependent methods. During the testing phase, the performance of the proposed framework is validated using 5-fold cross-validation. In this cross-validation approach, the test sample datasets are divided into different folds, with each fold typically consisting of 10% of the subjects and corresponding physiological data designated as the testing set. In contrast, the remaining subjects are allocated for training.

4.2 Classification Using Deep Learning Algorithms

The classifier is constructed to categorize physiological data features using three customized deep learning algorithms FNN Model, the Wide and Deep Model, and CNN model. Employing deep learning, the proposed framework effectively distinguishes between stress and non-stress classes by leveraging extracted features of all features, and through multimodal fusion see Figure-5, the model structure of the deep learning model for stress classification is showcased. A detailed description of the deep learning techniques is provided in the following subsection.

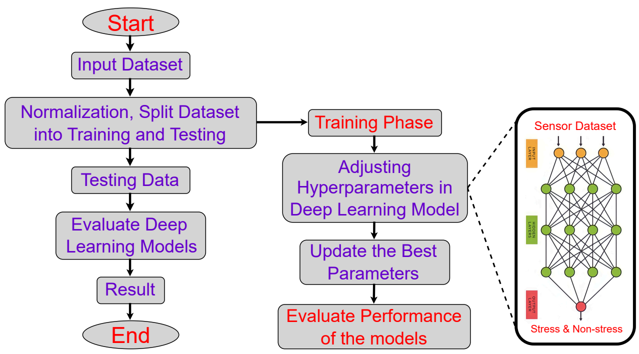

Fig. 5. Structure of deep learning model for classification

Feedforward Neural Network (FNN) A Feedforward Neural Network (FNN) is a fundamental deep learning architecture composed of interconnected layers for input, hidden, and output. It processes data in a single path, employing activation functions to generate nonlinearity and capture complicated patterns. The weighted sum $v_j = \sum_i w_{ij} x_i$, the output $y_j = \phi(v_j)$, and the error $\mathcal{E} = \frac{1}{2}\sum_j e_j^2$ are the main components of a feedforward neural network (FNN). We used FNNs to assess physiological signals obtained from wearable sensors. By sending physiological signal data into the FNN's input layer, the network learned to extract stress-related properties. The FNN's hidden layers processed the input data, changing it into representations that may be used to forecast stress levels. Based on the processed input data, the output layer made predictions about the stress severity level.

Wide and Deep Learning Model The wide and deep learning approach combines a linear model for feature interactions with a deep neural network to handle complicated patterns. The linear component records specific feature associations, but the deep layers generalize using hierarchical representations to improve prediction accuracy. The combined output y is $y = f(x;\theta) + h(x)$, where x represents input features, θ represents linear model parameters, $f(x;\theta)$ is the broad component, and $h(x)$ is the deep component. For the broad component, $f(x;\theta) = \sum_{i=1}^{n} \theta_i x_i$, where n represents the number of features. The deep component, $h(x)$, consists of many layers with weights W_i and biases b_i, using an activation function σ, generally ReLU. The Wide and Deep Learning Models were used to assess physiological information captured by wearable sensors. By combining both shallow and deep data representations, the model efficiently caught both subtle and complicated stress-related patterns. The wide component identified specific feature correlations in physiological signals, and the deep component learned hierarchical representations to generalize across stress contexts.

Convolutional Neural Network (CNN) 1D Convolutional Neural Networks (CNNs) are specialized deep learning models for processing sequential data like signals and structured grid data. They are made up of convolutional layers to extract hierarchical characteristics from input signals effectively. Convolutional layers, activation functions (such as ReLU), pooling layers, and fully connected layers make up the foundation of the CNN architecture. The convolutional operation is defined as follows:

$$(f * g)(t) = \sum_{a} f(a)g(t-a) \tag{7}$$

Here,f represents the input physiological signals acquired from EDA sensors, g denotes the filter or kernel with dimensions of 3×3, and t represents the position within the output feature map. During the convolutional operation, the filter g is slid over the input image f, centered at different positions. At each position, the filter g interacts with the input data around the location t, computing the convolution operation to produce a single value in the output feature map. Pooling layers lower spatial dimensions, whereas fully connected layers classify them. In our stress prediction study, we used CNNs to examine physiological information acquired from EDA sensors. The customized CNNs learned the relevant stress-related properties by considering the physiological signal data as a time-series sequence. Convolutional layers discovered temporal patterns in the signals, while pooling layers reduced noise and extracted essential characteristics. Fully linked layers used the derived information to forecast stress levels. The above models are trained with a learning rate of 0.001, using the Adam optimizer, a dropout rate of 0.2, and L2 regularization with a value of 0.001.

4.3 Performance Measure

The performance of the proposed method is evaluated using standard statistical evaluation methods employed to analyze classification performances. These include accuracy, recall, precision, and F1-score.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

5 Results and Discussion

This study evaluates binary classification models for distinguishing between stress and non-stress states using three deep learning models trained on six publicly available datasets. The training process utilized both extracted features and the complete set of features from EDA data. Additionally, a multimodal approach was applied, integrating ECG, PPG, and EDA data to predict stress severity.

Binary classification tasks were performed to predict stress severity across seven diverse datasets using both (a) all available features and (b) extracted features of EDA. Table-4 provides a comprehensive comparison of the performance of three deep learning algorithms across these datasets. Overall, the results indicate that the Wide and Deep model consistently outperformed others when using all features, particularly in the UTD, MSD-Nurse, and DriveDB datasets. The FNN model achieved the highest accuracy on the VerBIO and SWELL datasets, while the CNN model demonstrated superior accuracy on the WESAD dataset.When focusing on extracted features, the CNN model outperformed other models in predicting stress severity across the VerBIO, UTD, WESAD, and MSD-Nurse datasets. Meanwhile, the FNN model effectively predicted stress severity for the SWELL and DriveDB datasets.

Table 4. Performance Comparison of Classifiers for Stress Detection Using All Features and Extracted Features

Dataset	Data	Metrics	Models		
			FNN	CNN	Wide & Deep
VerBIO	All Features	F1-score (%)	54.5	61.5	42.3
		Accuracy (%)	76.9	73.0	69.2
	Extracted Features	F1-score (%)	95.6	**96.0**	93.0
		Accuracy (%)	95.8	**96.2**	93.7
UTD	All Features	F1-score (%)	90.5	90.9	93.0
		Accuracy (%)	92.4	92.3	93.3
	Extracted Features	F1-score (%)	90.5	**97.5**	96.0
		Accuracy (%)	92.4	**98.1**	95.9
WESAD	All Features	F1-score (%)	29.9	58.4	88.4
		Accuracy (%)	88.9	90.2	87.1
	Extracted Features	F1-score (%)	98.7	**99.1**	98.1
		Accuracy (%)	96.1	**97.3**	67.6
SWELL	All Features	F1-score (%)	35.7	46.3	64.0
		Accuracy (%)	81.7	81.0	65.7
	Extracted Features	F1-score (%)	**68.4**	32.2	77.0
		Accuracy (%)	**92.3**	86.5	44.2
MSD Nurses	All Features	F1-score (%)	72.7	39.4	79.2
		Accuracy (%)	61.9	61.3	75.6
	Extracted Features	F1-score (%)	72.4	**86.6**	78.0
		Accuracy (%)	76.7	**89.4**	78.4
DriveDB	All Features	F1-score (%)	72.7	39.4	71.2
		Accuracy (%)	61.9	61.3	65.1
	Extracted Features	F1-score (%)	**88.8**	88.8	70.0
		Accuracy (%)	**88.7**	85.7	69.9

5.1 Evaluation of the Effect of Multimodal Fusion

We further refined our classification model by integrating multiple signal types obtained from a wrist-worn wearable device, aiming to explore the potential impact of different physiological signals on stress severity prediction. Our focus encompassed ECG, PPG, and EDA, given their availability in wearable devices. ECG captures the timing and diverse electrical signals associated with heartbeats, while PPG utilizes non-invasive technology to measure changes in blood volume within living tissue. Both heartbeats and vasoconstriction, which respond to stimulation, can influence ECG and PPG signals. Consequently, the amalgamation of ECG and PPG data holds promise for yielding valuable insights into effective stress severity prediction.

Multimodal signals are fused at three levels: sensor-level fusion merges raw data from multiple sensors, feature-level fusion integrates extracted features, and decision-level fusion combines outputs for final inference. In this study, signals from three different sensors were concatenated to create a new feature vector for classification. Features derived from ECG and PPG data were utilized, with principal component analysis (PCA) applied to ECG data to extract significant elements from heart rate (HR) and heart rate variability (HRV) time series. Additionally, PPG features such as maximum HR after stimulus onset (PPG-Rate-Max), minimum HR after stimulus onset (PPG-Rate-Min), mean HR after stimulus onset (PPG-Rate-Mean), and standard deviation of heart rate after stimulus onset (PPG-Rate-SD) were incorporated. We examined the Verbio and WESAD datasets, which contain EDA, ECG, and PPG signals, for multimodal stress severity prediction. Tables-5 and Table-6 display the detection accuracy based on various modalities and different combinations among them.

Table 5. Comparison of model performance for VerBIO

VerBIO	Classification		
	FNN	CNN	Wide & Deep
EDA	9.58	9.62	9.37
ECG	6.25	8.33	8.75
PPG	7.08	8.33	9.16
EDA+ECG	8.33	9.61	9.27
EDA+PPG	9.16	8.75	9.37
EDA+ECG+PPG	9.14	8.94	9.25

Table 6. Comparison of model performance for WESAD

WESAD	Classification		
	FNN	CNN	Wide & Deep
EDA	9.60	9.73	6.76
ECG	8.78	8.54	7.46
PPG	8.67	9.34	7.23
EDA+ECG	9.21	8.94	7.12
EDA+PPG	8.69	8.79	7.91
EDA+ECG+PPG	8.78	8.71	6.71

6 Conclusion and Future Directions

In this work, we proposed a multimodal framework using three deep learning techniques to evaluate performance based on six publicly available EDA datasets for stress classification. The proposed framework and different classifiers were trained on both extracted and all features, including extracted EDA features, separately. Additionally, multimodal fusion was conducted on three datasets, VerBIO and WESAD, utilizing combinations of EDA, ECG, and PPG signals.

Overall, our proposed framework demonstrated strong performance across all features. The Wide and Deep model yielded the highest accuracy in the UTD, MSD-Nurse, and DriveDB datasets, achieving accuracies of 93.3%, 75.6%, and 65.1%, respectively. The FNN model achieved the highest accuracy in the VerBIO and SWELL datasets, with accuracies of 76.9% and 81.7%, respectively. The CNN model performed well on the WESAD dataset, achieving an accuracy of 90.2%. When focusing on extracted features, our CNN framework excelled in four datasets, achieving high accuracies in VerBIO, UTD, WESAD, and MSD-Nurse, with accuracies of 96.2%, 98.1%, 97.3%, and 89.4%, respectively. The FNN model demonstrated the highest accuracy for the SWELL and DriveDB datasets, with accuracies of 92.3% and 88.7%, respectively, in stress severity prediction. These findings underscore the effectiveness of our multimodal framework and highlight the importance of feature selection and fusion techniques in stress severity prediction using wearable sensor data.

In future endeavors, we plan to extend our multimodal framework to enable early detection of stress across diverse datasets encompassing signals, speech, and facial images. Our efforts will concentrate on aggregating extensive datasets to facilitate cross-validation of experimental outcomes across various benchmark settings. By leveraging multimodal deep learning techniques, we aim to achieve early diagnosis and precise prediction of stress in both patients and general populations, thereby advancing the field of stress management and healthcare.

Data and Code Avialbility

We provide open course code via Github repository [1]

Declarations

Conflict of Interest: The authors declared that they have no conflict of interest in this work.

References

1. A. N. Edes, D. E. Crews, Allostatic load and biological anthropology, American journal of physical anthropology 162 (2017) 44–70.
2. P. D. Bliese, J. R. Edwards, S. Sonnentag, Stress and well-being at work: A century of empirical trends reflecting theoretical and societal influences., Journal of Applied psychology 102 (3) (2017) 389.
3. J. Bienertova-Vasku, P. Lenart, M. Scheringer, Eustress and distress: neither good nor bad, but rather the same?, BioEssays 42 (7) (2020) 1900238.
4. G. Brulé, R. Morgan, Working with stress: Can we turn distress into eustress, Journal of Neuropsychology & Stress Management 3 (4) (2018) 1–3.
5. B. S. McEwen, Stress, adaptation, and disease: Allostasis and allostatic load, Annals of the New York academy of sciences 840 (1) (1998) 33–44.
6. A. Singh, K. Singh, A. Kumar, A. Shrivastava, S. Kumar, Machine learning algorithms for detecting mental stress in college students, in: 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), IEEE, 2024, pp. 1–5.
7. A. Singh, K. Singh, A. Kumar, A. Shrivastava, S. Kumar, Optimizing well-being: Unveiling eustress and distress through machine learning, in: 2024 International Conference on Inventive Computation Technologies (ICICT), IEEE, 2024, pp. 150–154.
8. A. D. Crosswell, K. G. Lockwood, Best practices for stress measurement: How to measure psychological stress in health research, Health psychology open 7 (2) (2020) 2055102920933072.
9. K. Kyriakou, B. Resch, G. Sagl, A. Petutschnig, C. Werner, D. Niederseer, M. Liedlgruber, F. Wilhelm, T. Osborne, J. Pykett, Detecting moments of stress from measurements of wearable physiological sensors, Sensors 19 (17) (2019) 3805.
10. R. K. Nath, H. Thapliyal, Smart wristband-based stress detection framework for older adults with cortisol as stress biomarker, IEEE Transactions on Consumer Electronics 67 (1) (2021) 30–39.
11. M. Moghimi, R. Stone, P. Rotshtein, Affective recognition in dynamic and interactive virtual environments, IEEE Transactions on Affective Computing 11 (1) (2017) 45–62.
12. C. Maaoui, A. Pruski, F. Abdat, Emotion recognition for hhman-machine communication, in: 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2008, pp. 1210–1215.
13. F. Movahedi, J. L. Coyle, E. Sejdić, Deep belief networks for electroencephalography: A review of recent contributions and future outlooks, IEEE journal of biomedical and health informatics 22 (3) (2017) 642–652.
14. Y. Yang, Q. J. Wu, W.-L. Zheng, B.-L. Lu, Eeg-based emotion recognition using hierarchical network with subnetwork nodes, IEEE Transactions on Cognitive and Developmental Systems 10 (2) (2017) 408–419.
15. B. Choi, H. Jebelli, S. Lee, Feasibility analysis of electrodermal activity (eda) acquired from wearable sensors to assess construction workers' perceived risk, Safety science 115 (2019) 110–120.
16. A. Anusha, P. Sukumaran, V. Sarveswaran, A. Shyam, T. J. Akl, S. Preejith, M. Sivaprakasam, et al., Electrodermal activity based pre-surgery stress detection using a wrist wearable, IEEE journal of biomedical and health informatics 24 (1) (2019) 92–100.
17. P. Zontone, A. Affanni, R. Bernardini, L. Del Linz, A. Piras, R. Rinaldo, Emotional response analysis using electrodermal activity, electrocardiogram and eye tracking signals in drivers with various car setups, in: 2020 28th European Signal Processing Conference (EUSIPCO), IEEE, 2021, pp. 1160–1164.

[1] `https://github.com/ashutosh22102/Early_Stress_Detection_Using_Deep_Learning_and_EDA_Through_Wearable_Devices`

18. Y. Liu, S. Du, Psychological stress level detection based on electrodermal activity, Behavioural brain research 341 (2018) 50–53.
19. B. Rim, N.-J. Sung, S. Min, M. Hong, Deep learning in physiological signal data: A survey, Sensors 20 (4) (2020) 969.
20. L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-Gonzalez, E. Abdulhay, N. Arunkumar, Using deep convolutional neural network for emotion detection on a physiological signals dataset (amigos), IEEE Access 7 (2018) 57–67.
21. P. Thiam, P. Bellmann, H. A. Kestler, F. Schwenker, Exploring deep physiological models for nociceptive pain recognition, Sensors 19 (20) (2019) 4503.
22. L. Zhu, P. Spachos, P. C. Ng, Y. Yu, Y. Wang, K. Plataniotis, D. Hatzinakos, Stress detection through wrist-based electrodermal activity monitoring and machine learning, IEEE Journal of Biomedical and Health Informatics 27 (5) (2023) 2155–2165.
23. O. N. Rahma, A. P. Putra, A. Rahmatillah, Y. S. K. A. Putri, N. D. Fajriaty, K. Ain, R. Chai, Electrodermal activity for measuring cognitive and emotional stress level, Journal of Medical Signals & Sensors 12 (2) (2022) 155–162.
24. M. Nardelli, A. Greco, L. Sebastiani, E. P. Scilingo, Comeda: A new tool for stress assessment based on electrodermal activity, Computers in Biology and Medicine 150 (2022) 106144.
25. A. Greco, G. Valenza, J. Lazaro, J. M. Garzon-Rey, J. Aguilo, C. de la Cámara, R. Bailón, E. P. Scilingo, Acute stress state classification based on electrodermal activity modeling, IEEE Transactions on Affective Computing 14 (1) (2021) 788–799.
26. M. Yadav, M. N. Sakib, E. H. Nirjhar, K. Feng, A. H. Behzadan, T. Chaspari, Exploring individual differences of public speaking anxiety in real-life and virtual presentations, IEEE Transactions on Affective Computing 13 (3) (2020) 1168–1182.
27. S. Koldijk, M. Sappelli, S. Verberne, M. A. Neerincx, W. Kraaij, The swell knowledge work dataset for stress and user modeling research, in: Proceedings of the 16th international conference on multimodal interaction, 2014, pp. 291–298.
28. P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, K. Van Laerhoven, Introducing wesad, a multimodal dataset for wearable stress and affect detection, in: Proceedings of the 20th ACM international conference on multimodal interaction, 2018, pp. 400–408.
29. J. Birjandtalab, D. Cogan, M. B. Pouyan, M. Nourani, A non-eeg biosignals dataset for assessment and visualization of neurological status, in: 2016 IEEE International Workshop on Signal Processing Systems (SiPS), IEEE, 2016, pp. 110–114.
30. S. Hosseini, R. Gottumukkala, S. Katragadda, R. T. Bhupatiraju, Z. Ashkar, C. W. Borst, K. Cochran, A multimodal sensor dataset for continuous stress detection of nurses in a hospital, Scientific Data 9 (1) (2022) 255.
31. J. A. Healey, R. W. Picard, Detecting stress during real-world driving tasks using physiological sensors, IEEE Transactions on intelligent transportation systems 6 (2) (2005) 156–166.
32. C.-P. Hsieh, Y.-T. Chen, W.-K. Beh, A.-Y. A. Wu, Feature selection framework for xgboost based on electrodermal activity in stress detection, in: 2019 IEEE International Workshop on Signal Processing Systems (SiPS), IEEE, 2019, pp. 330–335.
33. J. Shukla, M. Barreda-Angeles, J. Oliver, G. C. Nandi, D. Puig, Feature extraction and selection for emotion recognition from electrodermal activity, IEEE Transactions on Affective Computing 12 (4) (2019) 857–869.
34. M. Viana-Matesanz, C. Sánchez-Ávila, Adaptive normalization and feature extraction for electrodermal activity analysis, Mathematics 12 (2) (2024) 202.
35. C. Tronstad, O. M. Staal, S. Sælid, Ø. G. Martinsen, Model-based filtering for artifact and noise suppression with state estimation for electrodermal activity measurements in real time, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015, pp. 2750–2753.
36. A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, L. Citi, cvxeda: A convex optimization approach to electrodermal activity processing, IEEE transactions on biomedical engineering 63 (4) (2015) 797–804.