A Deep Learning Approach for Fear Recognition on the Edge Based on Two-Dimensional Feature Maps

Junjiao Sun, *Student Member, IEEE*, Jorge Portilla, *Senior Member, IEEE*, and Andres Otero, *Member, IEEE*

***Abstract*—Applying affective computing techniques to recognize fear and combining them with portable signal monitors makes it possible to create real-time detection systems that could act as bodyguards when users are in danger. With this aim, this paper presents a fear recognition method based on physiological signals obtained from wearable devices. The procedure involves creating two-dimensional feature maps from the raw signals, using data augmentation and feature selection algorithms, followed by deep learning-based classification models, taking inspiration from those used in image processing. This proposal has been validated with two different datasets, achieving, in WEMAC, WESAD 3-classes, and WESAD 2-classes, F1-score results of 78.13%, 88.07%, and 99.60%, respectively, and 79.90%, 89.12%, and 99.60% in accuracy. Furthermore, the paper demonstrates the feasibility of implementing the proposed method on the Coral Edge TPU device, prepared to make inferences on the edge.**

***Index Terms*—Affective computing, fear recognition, deep learning, feature selection, physiological signals, edge computing.**

I. Introduction

HUMANS, inherently emotional entities, experience fluctuations in daily performance due to both positive and negative emotional states [1]. These emotional responses, often arising spontaneously, manifest across various physiological systems, notably the brain and heart. Affective Computing (AC) seeks to decipher such emotions through an interdisciplinary approach, integrating insights from physiology, psychology, and cognitive sciences [2]. By equipping computers with the capacity to detect, comprehend, express, and respond to human emotions, a symbiotic relationship between humans and computers is fostered, enhancing computational intelligence. However, despite significant advancements in emotion recognition, the intricate interplay between physiological and psychological factors presents persistent challenges in achieving accurate and timely emotion identification.

Potential inputs for emotion recognition encompass both external human features and physiological signals. External features primarily include observable attributes like facial expressions, body movements, and vocal patterns. Conversely, physiological signals, obtained through specific sensors like electrocardiograms (ECGs) and electroencephalograms (EEGs), offer superior objectivity, as they emanate from involuntary physiological responses to emotions, making them less susceptible to conscious manipulation. Thus, physiological signals can more faithfully represent an individual's emotional state. Nevertheless, the collection and interpretation of these signals present challenges. Advances in technology, particularly the advent of smart wearable devices such as watches, fitness trackers, and sensor-embedded necklaces, are mitigating these issues by facilitating real-time monitoring of physiological signals [3], [4], [5].

Fear, a significant yet distressing emotion triggered by perceived threats, is vital for human survival. It leads to physiological modifications and elicits behavioral reactions such as aggression or impulsive flight from danger [6]. This emotion can originate from immediate threats, previous negative experiences, or the anticipation of future risks, serving as a fundamental protective mechanism. While inherently adverse, fear can yield beneficial effects on individuals by activating human survival instincts and fostering self-protection mechanisms [7]. Proper comprehension and management of fear can enable individuals to circumvent potential threats. However, self-defensive measures may prove inadequate in specific scenarios, such as during violent incidents like nighttime robberies, where precise fear detection could facilitate victim protection through automatic alerts to law enforcement agencies. Additionally, irrational fears, manifesting as phobias of public speaking, heights, or social interactions, often reflect deeper psychological issues requiring professional intervention. In these cases, fear recognition can significantly assist psychologists by providing insights into patients' conditions, thereby enabling the formulation of more

Manuscript received 23 January 2024; revised 21 March 2024 and 8 April 2024; accepted 16 April 2024. Date of publication 22 April 2024; date of current version 3 July 2024. This work has been funded by the Spanish Ministerio de Ciencia e Innovación through the project TALENT-HIPSTER under Grant PID2020-116417RB-C41. *(Corresponding author: Andres Otero.)*

The authors are with the Centro de Electrónica Industrial, Universidad Politécnica de Madrid, 28006 Madrid, Spain (e-mail: junjiao.s@upm.es; jorge.portilla@upm.es; joseandres.otero@upm.es).

The source codes related to the present work can be found at: https://github.com/des-cei/Empatia_DL

Digital Object Identifier 10.1109/JBHI.2024.3392373

effective strategies for the treatment of psychological trauma. However, current fear detection methods based on physiological signals need more reliability.

This paper proposes a fear recognition method based on Deep Learning (DL). It utilizes three physiological signals, Galvanic Skin Response (GSR), Skin Temperature (SKT), and Blood Volume Pulse (BVP), sourced from two datasets WEMAC [8] and WESAD [9], as input. After preprocessing the raw data, 123 features were extracted. Subsequently, these features were used to construct two-dimensional feature maps. Furthermore, Recursive Feature Elimination (RFE) was applied for feature selection to enhance accuracy. After that, Convolutional Neural Network (CNN) are utilized in the experiments to extract features from feature maps and provide the emotion detection results. This approach takes inspiration from image processing. The performance of the proposed method for the WEMAC, WESAD 3-classes, and WESAD 2-classes, respectively, reach 78.13%, 88.07%, and 99.60% in F1-score, and 79.90%, 89.12%, and 99.60% in accuracy. Additionally, this paper supplements the experiments to explore the applicability of this method on the edge, using devices with limited computing capabilities (Coral Edge TPU).

The main contributions of this paper are the following:

1) A method for generating two-dimensional feature maps from physiological signals, combined with a feature selection technique.
2) A DL approach based on the concept of feature map for processing physiological signals with CNNs, providing subject-independent fear detection capabilities.
3) A validation with two different datasets, which proofs the generalizable nature of the proposal: one from the state-of-the-art (SOTA) and another one specific for fear detection.
4) An implementation using hardware devices with limited resources to prove its applicability at the edge.

The remainder of the paper is organized as follows. Section II discusses the SOTA of fear detection algorithms and AC solutions with wearable devices. Section III introduces the datasets used in the paper and the data preparation process. Section IV details the proposed fear recognition method. In Section V, all the experimental results and comparisons of different functions are provided. Finally, conclusions and expected future works are presented in Section VI.

II. Related Work

In AC systems, input signals such as physiological signals, facial expressions, and voice are utilized [2]. This research emphasizes physiological signals due to their objectivity, as they are less influenced by individual control compared to external expressions [1]. The most widely used physiological signals include electrocardiogram (ECG), GSR, SKT, and BVP. Emotion recognition via these signals has been validated as a reliable method for emotion tracking, illustrated by studies like [10], which engages in real-time human emotion analysis using IoT for remote monitoring, and [11], a pioneering work in the SOTA for analyzing fear-related emotions in women. An essential advantage of physiological signals is their capacity for continuous emotional status monitoring. This research specifically focuses on fear detection, enabling the development of alert systems to protect individuals in aggressive or violent contexts [12]. The application of Internet-of-Things (IoT) technology allows for the real-time monitoring of emotions and their transmission to a connected device, such as a smartphone or tablet, offering protection to vulnerable groups like women, children, and the elderly by enabling swift contact with security forces [13].

In terms of targeted emotions, this study focuses on fear identification. A few existing works in the SOTA have been proposed targeting fear detection, providing specific datasets and algorithms to detect this emotion. The most relevant ones are the works related to WEMAC [8], a dataset specifically designed to collect physiological data from women when they experience fear. This dataset was created by the *UC3M4Safety* research group at the University Carlos III de Madrid, aimed at advancing technologies to shield women from Gender-based Violence (GBV). This initiative has addressed a significant void in AC research by supplying a dataset centered on fear-related emotions and has further encouraged the creation of wearable devices tailored for fear detection. Another key development is Bindi [3], an autonomous multimodal system derived from the WEMAC dataset, designed to combat GBV. Bindi demonstrates a binary classification accuracy (distinguishing fear from non-fear) of 64.63% using physiological data. This system underscores the potential for IoT devices in real-time fear detection, propelling forward the research into fear-specific detection technologies.

Additionally, numerous datasets exist containing comprehensive physiological data applicable to fear recognition model research despite not being explicitly designed for this purpose. One prominent example is the WESAD [9] dataset, which subjects individuals to three distinct emotional states: baseline, stress, and amusement, while collecting corresponding physiological data from both chest and wrist sensors. This dataset has been employed to validate the effectiveness of Machine Learning (ML) techniques in emotion recognition, achieving accuracy rates of 80% and 93% for three-class (baseline, stress, amusement) and binary (stress vs. non-stress) classifications, respectively, when utilizing full data sets. When only wrist data is used, the accuracy drops slightly to 75.21% and 87.12% for three-class and binary classifications, respectively. Other studies targeting the WESAD dataset include research outlined in [14], where authors developed a Deep Neural Network (DNN) named StressNAS specifically for stress detection via smartwatches. This model, focusing solely on wrist-generated data, accomplished accuracy rates of 81.78% and 92.87% in three-class and binary classifications, respectively.

In the realm of algorithms for emotion recognition, the primary methodologies encompass traditional ML and DL. ML-based strategies necessitate specialized knowledge for feature extraction from physiological signals, presenting a significant challenge. In contrast, DL methods have revolutionized AC and emotion recognition by their capacity to autonomously learn and abstract features from data, enhancing the understanding of physiological signals. A notable example of DL application is

presented in [15], where authors combined a CNN with Long Short-term Memory (LSTM) networks to analyze EEG and GSR signals. This approach involves applying convolution operations to signal waveforms for initial information extraction, subsequently processed by LSTM to learn sequence-based features. In the EnvBodySens dataset [16], the combined CNN-LSTM model achieved an accuracy rate of 87.3%. Another innovative approach involves using graph neural networks for EEG analysis, aiding Parkinson's disease (PD) diagnosis as detailed in [17]. These networks have proven highly effective in feature extraction and integration from physiological signals. In [18], a 1-D CNN structure was used to extract information from ECG and GSR data, and a Fully Connected (FC) network was employed for emotion classification. This method achieved a 75% accuracy rate on the Amigos dataset [19], a known public database for multimodal affective research. Furthermore, in [20], an LSTM-based architecture was specifically designed for emotion recognition within the DEAP dataset [21]. This model's uniqueness lies in its ability to capture long-term dependencies across temporal and spatial dimensions of brain signals, culminating in an accuracy of 80.64%. These advancements highlight the dynamic and evolving landscape of emotion recognition techniques, particularly through the lens of DL methodologies.

III. Data Preparation

Data is critical for artificial intelligence research, especially for DL, which demands substantial, high-quality datasets. The success of DL models is contingent on the availability of suitable data. This section outlines the utilization of WEMAC and WESAD datasets and delineates the feature extraction methodology for generating two-dimensional feature maps for DL model inputs.

A. Description of the DataSets

WEMAC is a multi-modal dataset designed explicitly to collect women's fear data [8]. It includes BVP, GSR, SKT as physiological data, and speech data collected through wearable sensors. In this study only physiological signals are used. The sample rates are 200 Hz, 200 Hz, and 10 Hz, respectively, for BVP, GSR, and SKT. During the data collection, individuals were exposed to 14 audio-visual stimuli in a virtual reality environment to elicit the target emotions. After watching the videos and recording the corresponding physiological signals, every participants were asked to finish an interactive self-report, which present them with different discrete possible emotion labels from which they were to select the primary emotion felt. The dataset comprises ten emotions, including Fear, Joy, Hope, Surprise, Anger, Tedium, Tenderness, Calm, Disgust and Sadness. The proportion occupied by fear is 44.4%. This study focuses on developing a method for fear recognition from physiological signals. Hence, the WEMAC dataset was binarized into two categories: fear and other emotions, for classification purposes. Fear is categorized as a negative emotion (label '1'), while all other emotions are considered positive (label '0').

WEMAC, with its extensive collection of physiological data related to fear, serves as an ideal dataset for this paper's focus. However, it exclusively comprises physiological data from females, which limits the method's generalizability. Consequently, it is imperative to incorporate an additional dataset for emotion-physiological signal analysis to validate the approach universally. Thus, WESAD has been chosen as a complementary dataset for concurrent experimental verification in this study. WESAD [9] is an available public dataset for detecting stress in human beings. It includes multi-modal sensory data from 15 participants, including males and females. WESAD has been used in several research works [14], [22], [23], proving its robustness. Stress can be regarded as a strongly correlated negative emotion with fear [24]. This is also one of the reasons why WESAD was chosen as the experimental dataset because, as far as we know, there is no other dataset specifically established for fear recognition based on physiological signals. WESAD contains physiological signals and emotional labels recorded from the wrist-worn Empatica E4 and chest-worn RespiBAN. Only BVP, GSR and SKT captured by Empatica E4 were used in this study because they are more similar to the signals available in WEMAC. Their sample rates are respectively 64 Hz, 4 Hz, and 4 Hz. There are three types of emotions recorded in this dataset, including baseline, stress, and amusement. In Section V, the training and validation for WESAD are divided into two stages: three classes (baseline, stress, amusement) and two classes (stress, non-stress). Here, non-stress comprises the union of baseline and amusement.

B. Feature Extraction

In previous works [3], [11], many features have been proposed for the data training process, such as mean value and standard deviation (*std*). Besides, considering that the main advantage of DNNs is their deep structures to handle complex data, it is possible to introduce more biomarkers and features. In this work, a total of 123 features have been implemented, including 34 features for GSR, 84 for BVP, and five for SKT. The names of the features can be found in Appendix A.

In the case of BVP, the features are obtained based on a series of biomarkers from [25] which are:

1) *PP:* Distance between consecutive Peak biomarkers.
2) *HRV:* Heart Rate Variability.
3) *PW:* Distance between consecutive Onset biomarkers.
4) *PDT:* Distance between a biomarker Onset and its previous Peak.
5) *PRT:* Distance between a Peak biomarker and its previous.
6) *PA:* Amplitude, measured from the value of the signal in the Onset to the value of the signal value in the Peak that corresponds to it.
7) *PWR:* Distance between a Dicrotic biomarker and its previous Onset.
8) *LF/HF Quotient:* Typically used to measure the relative balance between sympathetic and parasympathetic nervous system activity. The low frequency (LF) component

is often considered to reflect sympathetic nervous system activity, while the high frequency (HF) component reflects parasympathetic nervous system activity

Moreover, based on these biomarkers, this work designed a series of features including: first calculation of the *std* (sd1), second calculation of the *std* (sd2), transversal length (T), longitudinal length (L), cardiac sympathetic index (CSI), modified cardiac sympathetic index (MCSI), and cardiac vagal index (CVI) from [26]. In addition for HRV, NN50 and PNN50 were also designed, which represent the number of elements in the HRV vector with a distance between them of less than 50 ms, and the percentage of NN50 divided by the total number of NN intervals. In (1) to (7), the computation of BVP features from the biomarkers is detailed, using PP as an example, Num_{PP} represents the total number of PP measures.

$$\vec{sd1} = PP(n) - PP(n+1), 1 \leq n \leq (Num_{PP} - 1)$$

$$sd1 = std\left(\frac{\sqrt{2}}{2} * \vec{sd1}\right) \tag{1}$$

$$\vec{sd2} = PP(n) + PP(n+1), 1 \leq n \leq (Num_{PP} - 1)$$

$$sd2 = std\left(\frac{\sqrt{2}}{2} * \vec{sd2}\right) \tag{2}$$

$$T = 4 * sd1 \tag{3}$$

$$L = 4 * sd2 \tag{4}$$

$$CSI = L/T \tag{5}$$

$$MCSI = L^2/T \tag{6}$$

$$CVI = L * T \tag{7}$$

In the case of GSR, this research utilized Skin Conductance Level (SCL) and Skin Conductance Response (SCR) from [27], Power Spectral Density (PSD) from [28], and Delineated vector containing the slope of a signal (FD) as its biomarkers. Specifically, PSD was divided into 10 different bands based on the signal's frequency. FD was analyzed using Average For Negative (afn), Proportion of Negative (pon), Average For Positive (afp), Proportion of Positive (pop), mean, and standard deviation.

In the case of SKT, only five features are applied to this signal, including: mean, std, pow, temp-t0, and temp-t1. Here, t0 and t1 respectively represent the mean skin temperature of the participant at the starting and ending moments.

IV. Proposed Methodology

The methodology proposed in this paper for extracting emotional information from physiological data involves generating two-dimensional feature maps from this data, which serve as inputs to the DL models. This approach enables the application of established, highly effective image processing models for emotion classification.

Three physiological signals, including GSR, BVP, and SKT, which are widely used, are applied to fear recognition. Fig. 1 outlines the overall summary of the proposed methodology.

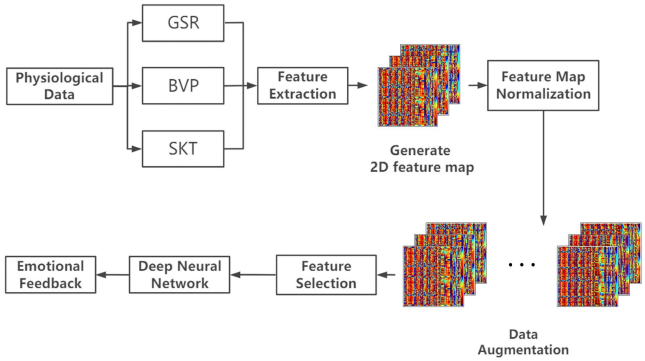


Fig. 1. Process of predicting from input physiological signals into corresponding emotional outputs following the methodology proposed in this work. The two-dimensional feature map in the figure is taken from a participant in the used dataset.

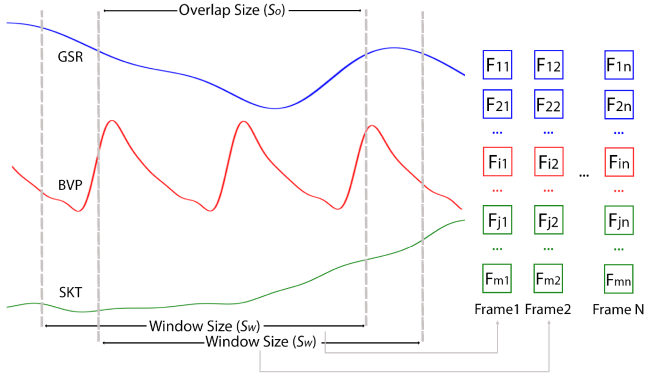


Fig. 2. Process of generating the feature frames. Blue, red, and green lines represent GSR, BVP, and SKT signals. S_w and S_o represent window size and overlap size. F_{1x} - F_{mx} represent the corresponding features generated from x^{th} window, being m the number of feature. The color of each feature represents the corresponding physiological signal from which it was extracted. The features extracted during the same window compose a frame.

Each stage in the figure is explained in detail in the following subsections.

A. Feature Map Generation

Fig. 2 illustrates the procedure for generating features from raw physiological signals, using data from a single participant as an example. The process is contingent upon the window size (S_w) and the overlap size (S_o). The window size denotes the quantity of signal samples utilized to derive all pertinent features for that segment, as detailed in Section III. The overlap size indicates the count of samples shared between successive windows, aimed at maintaining data integrity and optimizing data utilization. Both window and overlap sizes are crucial hyperparameters in our method, with specific values to be detailed in the experimental section. In this study, the aggregation of all features extracted from different physiological signals within a given time window is termed a *frame*.

Feature maps are produced by arranging these feature frames in a two-dimensional structure, as shown in Fig. 3. It is important

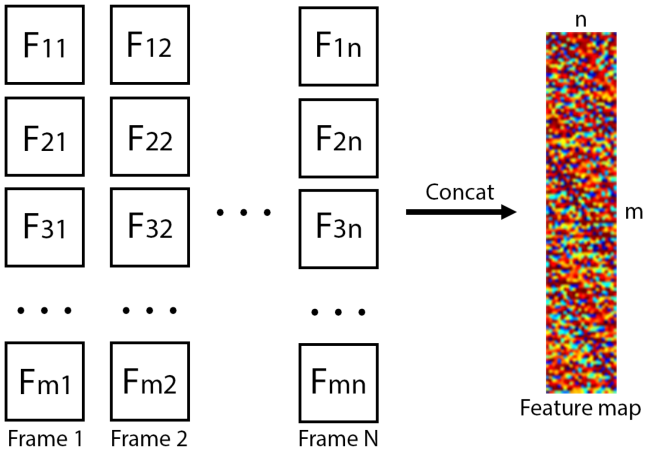


Fig. 3. Process of generating a feature map. m represents the feature number in each frame, and n is the frame number in each map.

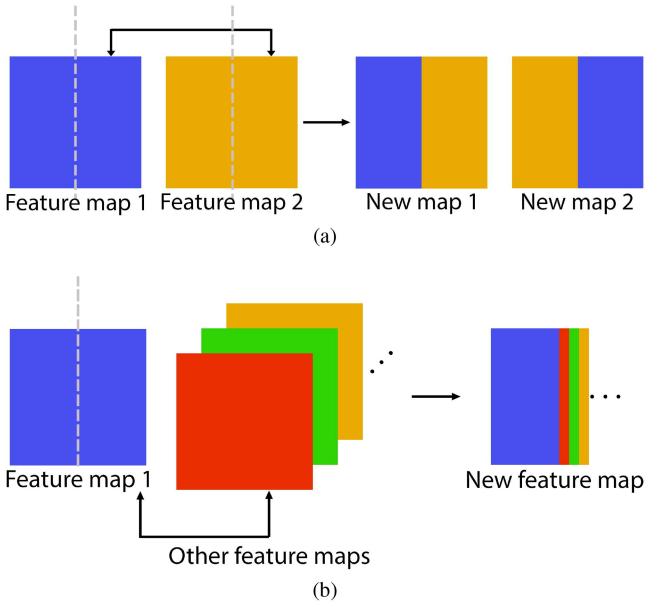


Fig. 4. Two proposed feature map augmentation functions. (a) represents Half and Half, where different color means different feature maps. (b) represents Half and Random. The new feature map is the compilation of half of feature map 1 and frames from others randomly sampled from different subjects.

to note that one feature map should only contain data associated with one feeling, which means every frame contained in one feature map should have the same label.

Equation (8) indicates the number of feature frames (N_f) that can be generated within T_{all} seconds of an experiment. In the formula, f_p represents the sampling frequency of the physiological signal in hertz. S_w and S_o represent the window and overlap sizes, respectively (as in Fig. 2). The values of T_{all}, S_w, and S_o are hyperparameters that have been determined during the experimental process.

$$N_f = \frac{T_{all} \times f_p - S_w}{S_w - S_o} + 1 \tag{8}$$

B. Feature Map Normalization

During the feature maps generation, notable disparities in feature magnitudes are observed, stemming from the different computational methods used for feature extraction. Consequently, normalization of all features is essential when constructing a feature map. To address this, Feature-wise-normalization (FWN) [29] was employed. This method amalgamates various normalization techniques and utilizes Ant Lion Optimization (ALO) to select the optimal normalization method for each feature, based on classification performance. K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) served as lightweight evaluation classifiers during this phase. It is important to note that these classifiers are solely for optimizing the normalization process of each feature and are not employed in the final emotion classification stage, which relies on more sophisticated DL-based models.

FWM ensures the normalization of each feature independently, maintaining the integrity of the information from each raw signal. This approach, FWM, has been validated as a method that significantly enhances the quality of the ML process. Furthermore, the experimental procedure may encounter sporadic outliers, typically deviating from the expected data range. To address this, this study employs Tukey's test for anomaly detection, aiding in the reduction of outlier-induced disruptions during normalization. Details on the normalization method applied to each feature are delineated in Appendix A.

C. Data Augmentation

A recognized limitation in applying DL for emotion recognition is the small size of available datasets. This restriction is due to the complex process of generating them, involving emotion elicitation and labeling using volunteers, typically supervised by psychologists or other experts in human emotions. The feature map generation technique outlined previously yields roughly 370 feature maps for the WEMAC dataset and 600 for WESAD, which need to be more for training a DL model. Therefore, augmenting the dataset was considered necessary. The apparent solution would be to apply augmentation directly to the raw signals. However, some limitations were found. Firstly, the original dataset comprises three types of raw data, each with different amplitude ranges and physiological natures, which prevents direct combining signals. Besides, raw signals include too much redundant information, making data augmentation less effective at this point. Furthermore, the time needed to load and process all raw data in each experiment is relatively high, which would be worse if data augmentation were applied before extracting the signal biomarkers. As an alternative, data augmentation methods based on two-dimensional feature maps are considered, taking inspiration from the techniques used in the state-of-the-art for the image processing domain. Data augmentation methods in image processing mainly include image manipulation [30], image mix [31], and image erasing [32]. Inspired by them, this study introduces two data augmentations to generate additional feature maps from the existing ones, as detailed in Fig. 4. It must

also be highlighted that data augmentation occurs separately in the training and test sets, divided according to subjects, to ensure the method is unbiased. This means no participant's data exists simultaneously in the training and test sets.

Half and Half: New feature maps are derived from original maps coming from two different participants. This means selecting two feature maps from different participants with the same label. Separate them from the middle, exchange the halves, and recombine them, as shown in Fig. 4(a).

Half and Random: New feature maps were derived from data provided by several participants. For each feature map, the first 50% of the frames are provided by a single participant. The remaining 50% is randomly sampled, frame by frame, from the maps of other participants, as shown in Fig. 4(b). Extracted frames are concatenated with the previous data until the required number of frames is reached. Coherency of labels is also required, which is guaranteed by constraining the random selection procedure to use only frames with the same labels.

All feature maps have been resized into squares based on their larger dimension, using Bilinear Interpolation, facilitating the application of existing CNN topologies during the classification stage.

D. Classification Algorithms

Following data preparation, a set of two-dimensional feature maps alongside their respective emotion labels was produced. Consequently, the task of emotion recognition can be approached as an image classification issue. CNNs are extensively used in computer vision for their exceptional ability, as they can autonomously extract image features without requiring manual intervention. The convolution process helps in isolating the most relevant features, thereby minimizing network redundancy and facilitating the enhancement of network depth.

A series of mainstream deep learning models have been preselected for evaluation: VGG [33], ResNet [34], GoogLeNet [35], DenseNet [36], EfficientNet [37], MobileNet [38], and ResNeXt [39]. These models are sequentially evaluated in the experimental section to find the most suitable one for each problem. Model selection is the final step in the proposed methodology.

The training results of these DL models are presented in Section V. For image classification, the input data is often three-channel images, i.e., RGB. In this study, two-dimensional feature maps are single-channel. Therefore, it is necessary to modify the input layer of the model to adapt to the data type. This article seeks to establish a subject-independent method to accurately recognize emotions across different subjects via physiological signals. Consequently, segregating the training and test sets based on individual subjects during the training phase ensures that no single subject's data is included in both sets. This research employs 10-fold cross-validation in its training methodology, with the average of all training iterations considered as the final result. Additionally, the standard deviation is documented to serve as an indicator of the model's robustness.

Algorithm 1: RFE (Fisher Score).

Required: X, L, p, h
  $M_z \leftarrow X$
  for $z = N;\ z \geq p;\ z--$ **do**
    $FSV \leftarrow fisherScore(M_z, L)$
    $R \leftarrow sortFisher(FSV, ascending)$
    $Acc_{best} \leftarrow 0$
    $Mark \leftarrow$ **null**
    for $j = 1;\ j \leq h;\ j++$ **do**
      $F_{now} \leftarrow R[j]$
      $M_{now} \leftarrow M_z.remove(F_{now})$
      $Acc_{now} \leftarrow evaluateModel(M_{now}, L)$
      if $Acc_{now} > Acc_{best}$ **then**
        $Acc_{best} \leftarrow Acc_{now}$
        $Mark \leftarrow F_{now}$
      end if
    end for
    $M_z \leftarrow M_z.remove(Mark)$
  end for

E. Feature Selection

All 123 features are employed in generating the two-dimensional feature maps. However, due to interactions among features, some may be redundant or detrimental to the classification task, making feature selection critical for enhancing the proposed method. Unlike traditional machine learning (ML), the training process of deep learning (DL) is often regarded as a 'black box', making it difficult to directly interpret the significance of parameters or weights produced during training. To address this, the Recursive Feature Elimination (RFE) algorithm is integrated into the training process to facilitate the systematic removal of irrelevant features [40]. RFE works by iteratively eliminating features one by one, retraining the model with the remaining features, and retaining the set that yields the best performance, continuing until the feature count meets a predetermined threshold. Consequently, although the training parameters remain opaque, this approach enables our study to isolate the most effective feature combination for maximizing classification accuracy.

The utility of RFE primarily lies in its capacity to eliminate non-contributory features. However, the challenge in this study arises from the presence of 123 features, which leads to prohibitive training costs for the underlying DL model at each iterative step. To address this issue, the Fisher Score metric has been introduced to determine which features should be removed initially. The Fisher Score evaluates the importance of each feature based on their means and variances, thereby generating a ranked list of features. This ranking facilitates the early removal of features with lower relevance, thereby substantially reducing the training time. By implementing this method, the research effectively manages the large feature set and streamlines the feature selection process. Importantly, the Fisher Score is recalculated after the removal of each feature during training, because that feature correlations–and thus their relevance–can shift as the feature composition alters.

TABLE I
FEATURE SELECTION PARAMETERS SUMMARY

Parameters	Description
X	Data from input dataset
L	Labels from input dataset
p	Threshold of feature elimination. RFE stops when the number of features reaches it.
h	Number of features evaluated during each iteration
N	Number of feature before RFE.
M_z	Data generated in each removing process
FSV	Vector of fisher scores
R	Ascending order of FSV
Acc_{best}	The best performance of each iteration
$Mark$	Corresponding feature of Acc_{best}
M_{now}	Data generated to test features in one iteration

TABLE II
HYPERPARAMETERS OF DIFFERENT SIGNALS AND DATABASE BASED ON THE TIME LENGTH T_{all}

	WEMAC			WESAD		
	GSR	BVP	SKT	GSR	BVP	SKT
T_{all} (s)	60	60	60	60	60	60
f_s (Hz)	200	200	10	4	64	4
S_w	3000	3000	150	40	640	40
S_o	2500	2500	125	32	512	32
N_f	19	19	19	26	26	26

Algorithm 1 presents the entire process of feature selection, and a summary of each parameter is described in Table I. The input dataset (X, L) and hyperparameters (p, h) are initially required. The specific (p, h) values are determined in Section V. Subsequently, scores for each feature are computed based on FC (line 4). The minor relevant features are removed individually, and the processed data is evaluated until testing the first h features (lines 5 and 9-11). At each iteration, the worst feature is recorded and permanently removed (lines 12-17). Consequently, the best group of features and its performance are obtained.

V. EXPERIMENTAL SETUP AND RESULTS

This section details experiments validating the effectiveness and stability of the proposed methodology. It involves the setup of experimental environment and presenting the test results. Afterwards, feature selection is applied to improve model performance. In the final subsection, the algorithm is deployed on embedded hardware for edge emotion inference testing.

A. Experimental Setup

First, the determination of the hyperparameters of the algorithm was carried out. This means obtaining S_w and S_o for each signal and the total time length T_{all} required to create one feature map. The selected configuration parameters are exhibited in Table II, from which N_f was calculated according to (8). In this table, f_s represents the sampling frequency. N_f corresponds to the number of frames generated within T_{all} seconds for S_w and S_o, which is the width of the two-dimensional feature map. S_w is set to 3000, 150, 40, and 640, depending on the sampling rate of each signal, to guarantee that all the windows contain the required number of data points to extract a feature frame. In WEMAC and WESAD, S_o is respectively set to 83% and 80% of S_w. These values were optimized experimentally, guaranteeing that these overlap proportions provide sufficient data and make better use of sequential information in the continuous physiological signals. The sampling rates of the three physiological signals in the two datasets were already mentioned in Section III. Regarding T_{all}, firstly, a length of 60 seconds was chosen following the conclusion of [41]. This study suggests that 60 seconds is optimal for balancing the inclusiveness of emotional information and time. This part has also tested the performance within different time lengths as shown in Section V-E. The hyperparameters S_w and S_o were empirically obtained after several experiments, ensuring data utilization and generating an appropriate number of feature maps.

Different DL algorithms were evaluated as part of the experimental environment. Pytorch 2.0.0 was used to conduct the experiments on a server equipped with an AMD EPYC 7513 32-core Processor as the CPU, 64 GB RAM, and an NVIDIA A30 24-GB GPU. The initial optimizer chosen was Adam, a commonly used option in DL research, with a learning rate of 0.001. Several high-performing CNN models, including VGG [33], ResNet [34], GoogLeNet [35], DenseNet [36], EfficientNet [37], MobileNet [38], and ResNeXt [39], were selected for the verification process, as was described in Section IV-D. The fully-connected layer of each model was modified based on the number of classes to adapt the models for the classification task.

The results are analyzed separately for WEMAC and WESAD datasets. In the case of WESAD, two different scenarios are evaluated: binary classification (stress and non-stress) and triple classification (stress, amusement, and baseline), denoted as WESAD-2 C and WESAD-3 C, respectively. Accuracy and F1-score metrics are used as quantitative measures to evaluate the performance and make comparisons. A 10-fold cross-validation procedure has been employed to ensure the robustness of the results. The average and standard deviation are used to demonstrate the stability of the models during cross-validation.

B. Comparison With SOTA Works

The effectiveness of the method proposed in this paper was validated by comparing it with the SOTAs. The comparison mainly involved the results from Bindi [3] in WEMAC and [14] in WESAD. Instead of using the normal training function in DL, these works used the leave-half-subject-out (LASO) and leave-one-subject-out (LOSO) methods, respectively. Therefore, LASO and LOSO were adopted in this work's training process to ensure compatibility and generate comparable results. Table III presents the performance of the DL models and the SOTA results (copied from their respective papers). It is evident that while some selected models did not show convincing results, significant improvements were observed in the proposed work using ResNet, EfficientNet, and ResNeXt. Additionally, the values of *std* are moderated, indicating the advantageous robustness of the proposed method.

LOSO and LASO are pertinent methodologies for application and comparison within the SOTA. However, LASO necessitates the creation of multiple models for each participant to ensure

TABLE III
COMPARISON WITH SOTAS

		VGG	ResNet	GoogLeNet	DenseNet	EfficientNet	MobileNet	ResNeXt	SOTA
WEMAC - LASO									[3]
F1-score	Mean	62.50	77.90	64.29	63.18	**82.58**	66.32	77.70	66.67
	Std	15.56	5.56	5.10	5.90	7.24	3.69	5.98	17.31
Accuracy	Mean	71.43	79.12	71.21	66.36	**83.29**	68.82	79.24	64.63
	Std	16.98	6.98	6.08	7.52	7.58	5.77	7.31	16.56
WESAD-3C - LOSO									[14]
F1-score	Mean	54.63	72.56	66.51	73.64	79.66	62.57	**84.09**	-
	Std	11.03	6.08	4.19	9.44	7.13	4.88	7.65	-
Accuracy	Mean	58.97	74.42	67.90	76.59	82.85	68.49	**85.49**	81.78
	Std	9.78	3.84	7.28	8.65	6.11	4.32	6.84	-
WESAD-2C - LOSO									[14]
F1-score	Mean	76.89	90.26	66.03	81.34	**93.81**	70.12	86.97	-
	Std	8.96	7.12	9.79	11.02	7.17	3.67	9.98	-
Accuracy	Mean	80.36	91.08	71.13	82.89	**94.53**	75.38	89.71	92.87
	Std	7.93	6.88	6.32	7.80	6.64	4.41	7.20	-

TABLE IV
RESULTS OF GENERAL TRAINING

		VGG	ResNet	GoogLeNet	DenseNet	EfficientNet	MobileNet	ResNeXt
WEMAC								
F1-score	Mean	44.94	64.54	44.17	48.09	**67.14**	46.1	63.76
	Std	6.06	4.11	8.01	5.89	5.48	3.9	5.98
Accuracy	Mean	59.86	66.76	59.14	62.42	**68.95**	60.86	65.67
	Std	4.93	2.88	6.58	4.74	3.87	3.21	3.42
WESAD-3C								
F1-score	Mean	60.55	**78.62**	56.32	75.24	75.88	43.18	76.48
	Std	19.18	8.66	8.25	10.71	7.87	0.58	9.35
Accuracy	Mean	69.28	**79.06**	59.66	78.89	78.23	58.50	77.69
	Std	13.70	6.60	10.14	9.31	7.28	0.48	11.22
WESAD-2C								
F1-score	Mean	84.84	86.95	74.16	94.00	**94.97**	61.94	90.21
	Std	12.19	12.31	8.92	5.22	3.89	0.73	4.66
Accuracy	Mean	86.74	89.10	72.59	94.44	**94.91**	73.27	90.18
	Std	9.31	7.86	5.88	4.47	3.96	0.54	4.58

model accuracy. Additionally, LASO's approach of incorporating half of a subject's data into the training set can introduce bias in validating DL classification models. This method leads to the presence of a specific subject's emotional features in both the test and training sets, potentially inflating the results beyond what would be achieved using standard DL performance metrics. LOSO in [14], despite ensuring that the training and testing sets do not intersect at the subject level, the test set only contains data from one participant. This discrepancy results in a significant imbalance between the sizes of the training and testing sets, adversely impacting the training process. So, to validate the proposal in a more realistic and less optimistic scenario, the general train-test split strategy has also been applied, as described in the next section.

C. General Training

For general training, the dataset was divided using an 8:2 split into training and test sets, with the division based on subjects to ensure 80% were allocated to training and the remaining 20% to testing. This split ensures that an individual's data is exclusively used either for training or testing, avoiding overlap. Throughout the training phase, 10-fold cross-validation is applied, with evaluation metrics including mean accuracy and F1-score across all folds. Specifically, within the WEMAC dataset, the training set includes data from 22 subjects, and the test set from five subjects. In the case of WESAD, the training set consists of data from 12 subjects, with the remaining three subjects' data forming the test set. Each cross-validation iteration employs different subject combinations in training and testing phases. Table IV documents the performance metrics of the evaluated DL models on these datasets. Additionally, the low standard deviation values across different subjects suggest minimal impact of individual differences on the training outcomes, supporting the subject-independent nature of this method.

While the generation of two-dimensional feature maps and their subsequent classification using DL models have demonstrated effectiveness in fear recognition, opportunities for enhancement remain to better translate these findings to real-world applications. Consequently, feature selection has been incorporated into the training process to further improve performance, as detailed in the subsequent subsection.

D. Feature Selection Training

As indicated in Table IV, EfficientNet emerges as the superior model across the evaluated datasets, based on its accuracy and standard deviation metrics. The feature selection training

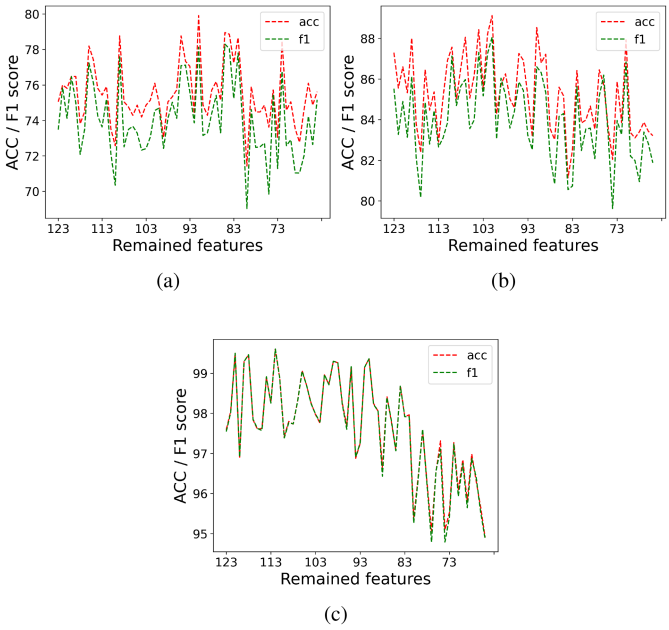


Fig. 5. Number of remaining features and corresponding precision. Red line represent accuracy whereas green represent F1 score. (a), (b) and (c) respectively show the trends of WEMAC, WESAD 3 C and WESAD 2 C.

process is notably time-intensive, requiring individual feature evaluation. Consequently, EfficientNet has been chosen exclusively for the implementation of the proposed feature selection strategy to assess its impact on performance. In Algorithm 1, the methodology for feature selection using RFE and the Fisher Score is outlined. To maintain a balance between training costs and optimal outcomes, the parameters p and h are configured to 63 and 15, respectively, which denotes that 15 features are evaluated per iteration, leading to the removal of 60 features in total. This approach adheres to the same 10-fold cross-validation scheme used in the baseline assessments, ensuring the consistency and reliability of the results.

The performance and the corresponding number of features in the three processes are displayed in Fig. 5, showing an approximate rising trend followed by a downward trend. Table V presents the results of training EfficientNet with RFE and Fisher score. Compared to the baseline process, the performance has significantly improved with RFE.

Based on these results, it can be concluded that RFE with Fisher score is an effective approach to enhance the method's performance. During feature removal, accuracy generally ascends until reaching the highest point, after which it declines. Moreover, the highest points differ, indicating that different data may require distinct feature groups and the most appropriate number of features.

E. Trade-Off Training

As outlined in the Experimental Setup, using 60 seconds of data has been deemed adequate for capturing emotional information necessary for training. Nevertheless, different time

TABLE V
RESULTS RFE-FISHER TRAINING

		Baseline	RFE-fisher
		WEMAC	
F1-score	Mean	67.14	**78.13**
	Std	5.48	6.52
Accuracy	Mean	68.95	**79.90**
	Std	3.87	4.16
		WESAD-3C	
F1-score	Mean	75.88	**88.07**
	Std	7.87	5.27
Accuracy	Mean	78.23	**89.12**
	Std	7.28	4.95
		WESAD-2C	
F1-score	Mean	94.97	**99.60**
	Std	3.89	3.25
Accuracy	Mean	94.91	**99.60**
	Std	3.96	4.17

EfficientNet works as baseline.

length were also tested to identify the optimal balance between time efficiency and accuracy. In this segment of the experiment, EfficientNet continued to serve as the training model. The durations for T_{all} in (8) were adjusted to 15 seconds, 30 seconds, and 45 seconds to facilitate comparisons with the 60-second benchmarks used in prior tests. The training outcomes for each time interval are detailed in Table VI. Notably, the 60-second window consistently delivered superior performance, affirming the premise that a greater aggregation of emotional data contributes to enhanced accuracy. The data trends reveal an increase in accuracy corresponding with longer time frames, particularly noting a significant improvement transitioning from 15 to 30 seconds. This indicates that while the system effectively captures emotional features, the sequential nature of the data is pivotal for accurate emotion recognition, underscoring the value of even brief temporal segments.

F. Inference Results on the Edge

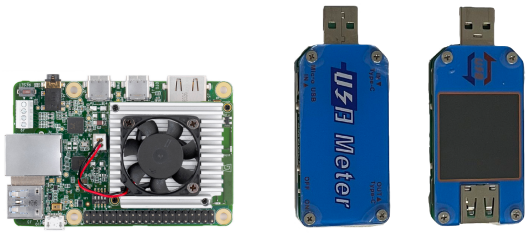
The prior experiments have illustrated that constructing two-dimensional feature maps from physiological signals and applying DL models for classification significantly enhances performance in AC. However, it's recognized that DL models demand greater computational resources compared to traditional ML approaches. Consequently, integrating DL models onto hardware platforms presents more complexities than ML, which is a crucial consideration for real-world applications. In this context, the Coral Edge TPU Dev Board was chosen as the experimental platform to assess the feasibility of deploying the proposed methodology at the edge.

Coral Edge TPU Dev Board [42] is a single-board computer designed for prototyping applications that demand fast on-device DL inference at the computing edge. As the picture shows in Fig. 6(a), the TPU comprises an on-board System-on-Module (SoM), which provides a platform that supports the equipment of a Linux system. It also has a Google Edge TPU coprocessor, the DL accelerator. The Edge TPU can execute deep neural networks such as CNNs, but only with the format of TensorFlow Lite (TF).

The performance and comparison are illustrated in Table VII. EfficientNet is still used as the training structure. Regarding

TABLE VI
TRADE-OFF TRAINING ABOUT TIME AND PERFORMANCE

		Baseline(15s)	RFE-fisher(15s)	Baseline(30s)	RFE-fisher(30s)	Baseline(45s)	RFE-fisher(45s)	Baseline(60s)	RFE-fisher(60s)
WEMAC									
F1-score	Mean	57.83	59.34	61.27	67.19	64.21	70.45	67.14	**78.13**
	Std	3.22	2.37	2.63	6.35	3.91	7.10	5.48	6.52
Accuracy	Mean	61.49	65.78	63.67	71.32	66.67	75.16	68.95	**79.90**
	Std	4.17	4.23	3.08	7.62	4.44	4.91	3.87	4.16
WESAD-3C									
F1-score	Mean	57.20	46.60	61.67	71.37	74.19	81.07	78.62	**88.07**
	Std	5.36	2.74	5.47	8.01	3.44	7.29	8.66	5.27
Accuracy	Mean	56.83	61.01	63.01	75.27	76.22	84.62	79.06	**89.12**
	Std	4.12	3.58	6.57	10.25	3.54	5.88	4.98	4.95
WESAD-2C									
F1-score	Mean	65.30	66.43	83.28	88.61	89.04	92.15	94.97	**99.60**
	Std	3.68	2.98	3.76	11.80	5.83	9.64	3.89	3.25
Accuracy	Mean	70.18	74.92	84.59	91.83	88.96	94.64	94.91	**99.60**
	Std	4.35	2.44	2.86	7.99	5.87	6.59	3.96	4.17



(a) (b)

Fig. 6. Hardware modules used in this section and the monitor of consumption.

TABLE VII
PERFORMANCE IN TPU CORAL

	Baseline		RFE-fisher	
	TPU	GPU	TPU	GPU
WEMAC				
F1-score	63.11	67.14	73.10	**78.13**
Accuracy	66.08	68.95	74.38	**79.90**
MTC(ms)	90.4	7.82	90.1	6.88
MPC(W)	2.09	-	2.10	-
WESAD-3C				
F1-score	73.73	78.62	83.84	**88.07**
Accuracy	76.54	79.06	85.85	**89.12**
MTC(ms)	90.1	6.78	90.4	6.83
MPC(W)	2.00	-	1.97	-
WESAD-2C				
F1-score	84.49	94.97	86.49	**99.60**
Accuracy	88.53	94.91	90.16	**99.60**
MTC(ms)	90.1	6.92	90.3	6.74
MPC(W)	2.01	-	1.97	-

MTC and MPC represent mean time consumption and mean power consumption respectively.

accuracy and f1 score, the performance in TPU is lower than GPU(baseline). That is mainly because Coral TPU can only process 8-bit data in training, whereas the original data is 32-bit in GPU training. Although it is a considerable data difference, the drop is still within an acceptable range. Taking the F1-score as the evaluation standard, the average reduction ratios in baseline and RFE-fisher are around 8%. That illustrates that the method in this paper has good robustness when dealing with data of varying quality.

In this experiment, besides assessing performance, time and power consumption are also monitored. Fig. 6(b) displays the USB power meter, which is used to observe the power consumption of the TPU in real-time by connecting it between the TPU and its power source. Monitoring hardware consumption is crucial for determining the practicality of the method proposed in this paper for real-world applications. According to Table VII, although the hardware execution time is longer compared to that on a GPU, an approximate 90 ms is still considered acceptable for real-world scenarios. Moreover, the power consumption levels do not significantly impede application viability. Consequently, the methodology presented in this paper is feasible for implementation on hardware platforms for edge computing, suggesting the potential for future development of products capable of real-time emotion monitoring.

VI. CONCLUSION AND FUTURE WORK

This study introduced an end-to-end method for fear recognition, processing physiological raw signals into two-dimensional feature maps for subsequent classification via deep learning to ascertain the presence of fear in subjects. Utilizing three physiological signals–GSR, BVP, and SKT–the research delineated 123 features for the creation of feature maps. Deep neural networks were then applied for emotion recognition analysis. To optimize performance, recursive feature elimination with Fisher Score was incorporated within the training framework. The methodology underwent validation and assessment utilizing the WEMAC and WESAD datasets, which are established benchmarks for fear and stress detection. During the experimental stage, various deep learning models were evaluated for their efficacy in classifying two-dimensional feature maps. Outcomes indicate that the proposed method attains accuracy levels of 79.90% on WEMAC, and 89.12% and 99.60% on WESAD, with a three-classes and two-classes classification respectively.

Trade-off experiments were conducted to explore the correlation between the duration of emotional information and classification accuracy, with feature map durations varied from 15 to 60 seconds. These experiments confirmed a positive relationship between the extent of emotional data and the accuracy of classification. Lastly, to assess the applicability of this method in

embedded systems, the Coral Edge TPU Dev Board operating on a Linux system was selected to evaluate the performance, power, and time consumption of the classification method. The findings indicated that despite a reduction in classification accuracy attributed to diminished data quality, the method's performance on hardware remains within acceptable limits.

At the same time, this study presents certain limitations and opportunities for enhancement. The research utilizes GSR, BVP, and SKT as input signals. However, with technological advancements, an increasing array of physiological signals can now be captured by wearable devices. Introducing new signals may enhance accuracy but could potentially increase computational costs. Therefore, exploring how to utilize new signals and improve method accuracy while limiting computational consumption will be worth investigating; this study employs two data augmentation strategies, both based on pre-generated two-dimensional feature maps. Future research could investigate addsitional augmentation methods, particularly those applied directly to raw data; The trade-off experiments underscored the significance of sequential features. Moving forward, it may be beneficial to employ algorithms capable of deeper sequential information learning, such as RNN, to enhance performance.

APPENDIX A

Appendix shows 123 features extracted from three physiological signals, GSR, BVP, and SKT, and the chosen normalization function for each feature using the FWN method. The detailed description is in Section III-B.

Acronyms	Description
PSD [28]	Power Spectral Density
FD [28]	Delineated vector containing the slope of a signal
SCL [27]	Skin Conductance Level
SCR [27]	Skin Conductance Response
PP [25]	Distance between consecutive Peak biomarkers
HRV [25]	Heart Rate Variability
PW [25]	Distance between consecutive Onset biomarkers
PDT [25]	Distance between a biomarker Onset and its previous Peak
PRT [25]	Distance between a Peak biomarker and its previous
PA [25]	Amplitude, measured from the value of the signal in the Onset to the value of the signal value in the Peak that corresponds to it
PWR [25]	Distance between a Dicrotic biomarker and its previous Onset

Signal	Feature name	Normalization	Feature name	Normalization
GSR	signal-mean	TH	signal-std	VTH
	PSD-band1	HT	PSD-band2	TH
	PSD-band3	VTH	PSD-band4	TH
	PSD-band5	HT	PSD-band6	TH
	PSD-band7	TH	PSD-band8	VTH
	PSD-band9	HT	PSD-band10	VTH
	peaks-max-max	VTH	peaks-max-number	TH
	peaks-min-min	TH	peaks-min-number	HT
	peaks-min-range	TH	peaks-min-distance	VTH
	FD-afn	HT	FD-pon	HT
	FD-afp	HT	FD-pop	HT
	FD-mean	VTH	FD-std	TH
	SCL-mean	HT	SCL-std	HT
	SCL-range	PT	SCL-distance	TH
	SCR-mean	VTH	SCR-std	TH
	SCR-max	PT	SCR-min	HT
	SCR-range	HT	SCR-distance	VTH

Signal	Feature name	Normalization	Feature name	Normalization
BVP	PP-mean	VTH	PP-LF-quotient	VTH
	PP-HF-quotient	VTH	PP-LF-HF-sum	VTH
	PP-LF-HF-quotient	HT	PP-gauss	VTH
	PP-RR-HF-ponderated	TH	PP-sd1	VTH
	PP-sd1-T	VTH	PP-sd2	VTH
	PP-sd2-L	TH	PP-sd2-csi	VTH
	PP-sd2-mcsi	VTH	PP-sd2-cvi	HT
	HR-mean	HT	HR-std	VTH
	HRV-mean	HT	HRV-std	HT
	HRV-rms	HT	HRV-nn50	HT
	HRV-pnn50	VTH	-	-
	PW-mean	HT	PW-LF-quotient	TH
	PW-HF-quotient	VTH	PW-LF-HF-sum	HT
	PW-LF-HF-quotient	HT	PW-gauss	HT
	PW-RR-HF-ponderated	VTH	PW-sd1	TH
	PW-sd1-T	HT	PW-sd2	TH
	PW-sd2-L	VTH	PW-sd2-csi	TH
	PW-sd2-mcsi	HT	PW-sd2-cvi	VTH
	PDT-mean	HT	PDT-LF-quotient	VTH
	PDT-HF-quotient	VTH	PDT-LF-HF-sum	HT
	PDT-LF-HF-quotient	HT	PDT-gauss	HT
	PDT-RR-HF-ponderated	VTH	PDT-sd1	HT
	PDT-sd1-T	TH	PDT-sd2	TH
	PDT-sd2-L	VTH	PDT-sd2-csi	VTH
	PDT-sd2-mcsi	VTH	PDT-sd2-cvi	TH
	PRT-mean	PT	PRT-LF-quotient	VTH
	PRT-HF-quotient	TH	PRT-LF-HF-sum	VTH
	PRT-LF-HF-quotient	TH	PRT-gauss	PT
	PRT-RR-HF-ponderated	VTH	PRT-sd1	TH
	PRT-sd1-T	HT	PRT-sd2	VTH
	PRT-sd2-L	VTH	PRT-sd2-csi	TH
	PRT-sd2-mcsi	VTH	PRT-sd2-cvi	VTH
	PA-mean	HT	PA-LF-quotient	VTH
	PA-HF-quotient	HT	PA-LF-HF-sum	HT
	PA-LF-HF-quotient	VTH	PA-gauss	VTH
	PA-RR-HF-ponderated	TH	-	-
	PWR-mean	HT	PWR-LF-quotient	VTH
	PWR-HF-quotient	VTH	PWR-LF-HF-sum	TH
	PWR-LF-HF-quotient	HT	PWR-gauss	HT
	PWR-RR-HF-ponderated	VTH	PWR-sd1	VTH
	PWR-sd1-T	VTH	PWR-sd2	VTH
	PWR-sd2-L	HT	PWR-sd2-csi	VTH
	PWR-sd2-mcsi	VTH	PWR-sd2-cvi	HT

Signal	Feature name	Normalization	Feature name	Normalization
SKT	SKT-mean	VTH	SKT-std	HT
	SKT-pow	HT	SKT-temp-t0	VSS
	SKT-temp-t1	VTH	-	-

REFERENCES

[1] L. Shu et al., "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, 2018, Art. no. 2074.

[2] J. Tao and T. Tan, "Affective computing: A review," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, 2005, pp. 981–995.

[3] J. A. M. Calero et al., "Bindi: Affective Internet of Things to combat gender-based violence," *IEEE Internet Things J.*, vol. 9, no. 21, pp. 21174–21193, Nov. 2022.

[4] S. Saganowski et al., "Consumer wearables and affective computing for wellbeing support," in *Proc. 17th EAI Int. Conf. Mobile Ubiquitous Syst.: Comput., Netw. Serv.*, 2020, pp. 482–487.

[5] S. Saganowski, B. Perz, A. Polak, and P. Kazienko, "Emotion recognition for everyday life using physiological signals from wearables: A systematic literature review," *IEEE Trans. Affect. Comput.*, vol. 14, no. 3, pp. 1876–1897, Jul.-Sep. 2023.

[6] A. Öhman, "The role of the amygdala in human fear: Automatic detection of threat," *Psychoneuroendocrinology*, vol. 30, no. 10, pp. 953–958, 2005.

[7] O. Bălan, G. Moise, A. Moldoveanu, M. Leordeanu, and F. Moldoveanu, "Fear level classification based on emotional dimensions and machine learning techniques," *Sensors*, vol. 19, no. 7, 2019, Art. no. 1738.

[8] J. A. Miranda et al., "Wemac: Women and emotion multi-modal affective computing dataset," 2022, *arXiv:2203.00456*.

[9] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing wesad, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interact.*, 2018, pp. 400–408, doi.: 10.1145/3242969.3242985.

[10] M. Awais et al., "LSTM-based emotion detection using physiological signals: IoT framework for healthcare and distance learning in COVID-19," *IEEE Internet Things J.*, vol. 8, no. 23, pp. 16863–16871, Dec. 2021.

[11] J. A. Miranda, M. F. Canabal, L. Gutierrez-Martin, J. M. Lanza-Gutierrez, M. Portela-García, and C. López-Ongil, "Fear recognition for women using a reduced set of physiological signals," *Sensors*, vol. 21, no. 5, 2021, Art. no. 1587.

[12] J. A. Miranda, M. F. Canabal, J. M. Lanza-Gutiérrez, M. P. García, and C. López-Ongil, "Toward fear detection using affect recognition," in *Proc. IEEE 34th Conf. Des. Circuits Integr. Syst.*, 2019, pp. 1–4.

[13] L. Gutiérrez-Martín et al., "Fear detection in multimodal affective computing: Physiological signals versus catecholamine concentration," *Sensors*, vol. 22, no. 11, 2022, Art. no. 4023.

[14] L. Huynh, T. Nguyen, T. Nguyen, S. Pirttikangas, and P. Siirtola, "Stressnas: Affect state and stress detection using neural architecture search," in *Proc. Adjunct Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput. Proc. ACM Int. Symp. Wearable Comput.*, 2021, pp. 121–125.

[15] E. Kanjo, E. M. Younis, and C. S. Ang, "Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection," *Inf. Fusion*, vol. 49, pp. 46–56, 2019.

[16] E. Kanjo, E. M. Younis, and N. Sherkat, "Towards unravelling the relationship between on-body, environmental and emotion data using sensor information fusion approach," *Inf. Fusion*, vol. 40, pp. 18–31, 2018.

[17] H. Chang, B. Liu, Y. Zong, C. Lu, and X. Wang, "EEG-based Parkinson's disease recognition via attention-based sparse graph convolutional neural network," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 11, pp. 5216–5224, Nov. 2023.

[18] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-Gonzalez, E. Abdulhay, and N. Arunkumar, "Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS)," *IEEE Access*, vol. 7, pp. 57–67, 2019.

[19] 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.

[20] V. M. Joshi and R. B. Ghongade, "Eeg based emotion detection using fourth order spectral moment and deep learning," *Biomed. Signal Process. Control*, vol. 68, 2021, Art. no. 102755.

[21] S. Koelstra et al., "DEAP: A database for emotion analysis; using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan.–Mar. 2012.

[22] D. Bajpai and L. He, "Evaluating KNN performance on WESAD dataset," in *Proc. IEEE 12th Int. Conf. Comput. Intell. Commun. Netw.*, 2020, pp. 60–62.

[23] A. Liapis, E. Faliagka, C. Katsanos, C. Antonopoulos, and N. Voros, "Detection of subtle stress episodes during ux evaluation: Assessing the performance of the wesad bio-signals dataset," in *Proc. 18th IFIP TC 13 Int. Conf. Hum.-Comput. Interact.*, 2021, pp. 238–247.

[24] S. Maren and A. Holmes, "Stress and fear extinction," *Neuropsychopharmacology*, vol. 41, no. 1, pp. 58–79, 2016.

[25] V. Montesinos, F. Dell'Agnola, A. Arza, A. Aminifar, and D. Atienza, "Multi-modal acute stress recognition using off-the-shelf wearable devices," in *Proc. IEEE 41st Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2019, pp. 2196–2201.

[26] M. Toichi, T. Sugiura, T. Murai, and A. Sengoku, "A new method of assessing cardiac autonomic function and its comparison with spectral analysis and coefficient of variation of R–R interval," *J. Autonomic Nervous Syst.*, vol. 62, no. 1/2, pp. 79–84, 1997.

[27] F. Hernando-Gallego, D. Luengo, and A. Artes-Rodriguez, "Feature extraction of galvanic skin responses by nonnegative sparse deconvolution," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 5, pp. 1385–1394, Sep. 2018.

[28] G. Udovic, J. Derek, M. Russo, and M. Sikora, "Wearable emotion recognition system based on GSR and PPG signals," in *Proc. 2nd Int. Workshop Multimedia Pers. Health Health Care*, 2017, pp. 53–59.

[29] D. Singh and B. Singh, "Feature wise normalization: An effective way of normalizing data," *Pattern Recognit.*, vol. 122, 2022, Art. no. 108307.

[30] J. A. Pandian, G. Geetharamani, and B. Annette, "Data augmentation on plant leaf disease image dataset using image manipulation and deep learning techniques," in *Proc. IEEE 9th Int. Conf. Adv. Comput.*, 2019, pp. 199–204.

[31] C. Summers and M. J. Dinneen, "Improved mixed-example data augmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2019, pp. 1262–1270.

[32] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, "Random erasing data augmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 13001–13008.

[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[35] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[36] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[37] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[38] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[39] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1492–1500.

[40] X. W. Chen and J. C. Jeong, "Enhanced recursive feature elimination," in *Proc. IEEE 6th Int. Conf. Mach. Learn. Appl.*, 2007, pp. 429–435.

[41] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, 2010.

[42] "Google coral TPU Dev board," [Online]. Available: https://coral.ai/products/dev-board