Innovations in Systems and Software Engineering
https://doi.org/10.1007/s11334-022-00470-6

S.I. : COUPLING DATA AND SOFTWARE ENGINEERING TOWARDS SMART SYSTEMS

A lightweight deep neural network for detection of mental states from physiological signals

Debatri Chatterjee[1,2] · Souvik Dutta[2] · Rahul Shaikh[2] · Sanjoy Kumar Saha[2]

Received: 16 June 2022 / Accepted: 3 July 2022

Abstract
Detection of mental states like stress/anxiety, mediation is a widely researched topic and is important for ensuring overall well-being of an individual. Several approaches have been reported in the literature for prediction or assessment of mental states. Recently, with advances in sensor technology, various physiological signals are being used by researchers for detecting mental states. In the present study, we have used a light weight deep convolutional neural network (CNN) for creating a mental state prediction model. The proposed detection model is created using publicly available WESAD dataset. The dataset contains electrocardiogram (ECG), galvanic skin response (GSR), skin temperature and electromyogram (EMG) signals recorded using a wearable device. Results show that for binary classification of *stress vs no-stress* condition our results are comparable with that reported in state-of-the-art machine learning/deep learning-based approaches. However, for three class classification of *baseline vs stress vs amusement* states, our model gives an accuracy of 90% which is much higher compared to that reported in the literature. In addition, we have also tried to classify various binary states like *stress vs baseline*,*stress vs amusement* and *stress vs meditation* conditions. The *f1 score* obtained for these classes are 0.96, 0.87 and 0.91, respectively, which are much higher than that reported in state-of-the-art literature using same dataset. Proposed light weight CNN-based mental state classification model is computationally less complex compared to other deep networks used by the researchers. Thus, it can be used for monitoring mental state successfully in real-life scenarios.

Keywords Mental stress · Detection of mental states · Anxiety and stress assessment · Deep learning · Deep convolutional layer

1 Introduction

In psychology, affective state of mind is an umbrella term that refers to the experience of feeling the underlying emotional state. The long-term experience is called *mood* and more focused short-term experience is called *emotions* [1]. Affective computing is an interdisciplinary branch that deals with assessing the affective state of an human user. It includes the hardware, software and underlying theoretical models that helps in development of affect sensitive computer systems. Such systems are more intuitive and provide natural computer interfaces by using the emotional state of its users [1]. Positive affective states include enthusiasm, amusement, confidence etc. and these result into a positive impact on overall mental health of an individual. On the other hand, negative affective states like anger, stress, fear etc. negatively impact our well being. Literature suggests that all affective states have some physiological manifestation that may be

Souvik Dutta, Rahul Shaikh and Sanjoy Kumar Saha have contributed equally to this work.

✉ Debatri Chatterjee
debatri.chatterjee@tcs.com

Souvik Dutta
souvik.dutta.1399@gmail.com

Rahul Shaikh
rahulshaikh5720.ac@gmail.com

Sanjoy Kumar Saha
sanjoykumar.saha@jadavpuruniversity.in

[1] TCS Research, Tata Consultancy Services, Kolkata, West Bengal, India

[2] Department of Computer Science and Engineering, Jadavpur University, Kolkata, West Bengal, India

Published online: 23 July 2022

subtle, but potentially observable. Psychophysiology bridges the domains of psychology and physiology of human affects.

Psychologists usually use self-rating-based questionnaires like, State-Trait Anxiety Inventory (STAI) [2], Mini-mental state exam (MMSE) [3], Montreal cognitive assessment (MoCA) [4] etc. for assessment of mental state. However, this approach requires manual intervention for interpretation and is often not applicable for continuous assessment in real-life scenarios. Recent advancements of sensor technology enable the researchers to measure various physiological changes of an individual in a non-invasive way. For studying human affective states, most commonly used technologies are galvanic skin response (GSR), electrocardiogram (ECG), electromyography (EMG), respiration rate, electroencephalogram (EEG) etc. Such biofeedback-based devices are portable, usable, affordable, and applicable for measuring stress in our everyday lives. Diverse machine learning and deep learning-based approaches are applied on the captured signal for noise cleaning, feature selection and classification of mental states thereof. A number of publicly available affect datasets are also being used by researchers. Few such datasets are DEAP [5], ASCERTAIN [6], WESAD [8], AMIGOS [7] etc. Researchers have also reported their findings using their own collected data. However, the accuracy of such approaches varies a lot across participants and across datasets.

Recently, deep learning-based approaches are being used that consider nonlinear transformations of physiological signals for the detection of features of human emotional behavior. Majority of these models are based on a single sensor and are computationally heavy. In this work, we have proposed a multi-modal, physiological data fusion framework using deep convolutional neural network (CNN) to classify physiological signals corresponding to various mental states collected from wearable devices. Toward that aim, we have used the WESAD [8] dataset that includes multi-modal data recorded during different affective states, especially during stress. The classification results obtained using the proposed approach are compared with state of the results reported in the literature.

The paper is organized as follow: Sect. 2 describes the prior arts related to affect recognition. Section 3 describes the dataset and explains our proposed approach. Section 4 provides the detail results and discussion. Finally, the paper is concluded in Sect. 5.

2 Related works

Affective states are reflected physiologically on different parts of the body, such as the brain, heart, face, and skin [9]. Such signals are used in a machine leaning algorithm to detect human emotions such as stress, anxiety etc [9,10]. These signals are reliable markers as those are controlled by the autonomic nervous system and are less likely to be faked [9,10]. Researchers have extensively studied the detection of human emotions using physiological signals and machine learning models. For this purpose, a machine learning model is trained to classify emotion into discrete categories (e.g., happy, sad, disgust etc). These models are mostly based on traditional machine learning with handcrafted features [11,12]. For ECG signals, there are large number of researches focusing on different types of feature extraction methods. Few such methods include heart rate variability (HRV), empirical mode decomposition (EMD) with-in beat analysis (WIB), FFT analysis, and various methods of wavelet transformations [23]. For GSR, as the skin conductance is mainly related with arousal level, thus useful information related with its amplitude and frequency is analyzed in time and frequency domains by applying various techniques and extracting some statistical parameters as: median, mean, standard deviation, minimum, maximum, as well as ratio of minimum and maximum [24]. In [13], authors adopted multi-modal signal-based approach in order to improve the overall performance. In [25] authors fused EEG, GSR and PPG signals and achieved an accuracy of 79% for 4-class emotion classification. They have used four time-domain features including entropy (E), variance (V), kurtosis (K), and skewness (S) from GSR signal, heart rate (HR), and heart rate variability (HRV) from the PPG data and signal asymmetry of EEG data.

Recently, researchers have tried to detect emotion using deep learning models [10]. Majority of these models were created using single type of sensor, such as, EEG [14], respiration [15]. In [10], authors fused respiration and heart rate variability (HRV). In [14], authors used differential entropy of various EEG sub-bands as the feature to train deep belief network (DBN). In [10], HRV parameters (heart rate, HRV amplitude, LF, HF, and LF/HF) and RSP parameters (RSP value and RSP rate) were applied to CNN as parameter to classify emotion. In [16], authors used CNNs for the extraction of SCR and BVP features and they achieved around 70 to 75% accuracy for prediction of emotion (relaxation, anxiety, excitement, and fun). For both the signals, they extracted some statistical features like standard deviation, minimum, maximum, difference between minimum and maximum. In addition, they used HRV parameters like RR interval, standard deviation of RR intervals and GSR features like initial skin conductance value, final skin conductance values, and their differences to train the classification model. In another work, the accuracy of affection models with deep learning on multimodal DEAP database [17] is 0.83 and 0.84 for valence and arousal, respectively. In [18], authors proposed a hybrid model composed of a CNN and a recurrent neural network (RNN). Here, the prediction was made in the long short-term memory (LSTM) unit of the RNN and they obtained

an accuracy of 74.1% for arousal and 72.1% for valence. Another models based on CCN and DNN [19] showed better results in the affective classification while using the EEG signals [20]. In [20], authors have transformed the raw EEG signal and five EEG sub-bands into fixed size gray images. They have also used peripheral physiological signals (EOG, GSR, EMG, skin temperature, blood volume pressure, respiration, ECG). Features like mean, variance, zero crossing rate, mean of derivatives, number of local minima etc were computed from these signals. Finally, the generated images and features obtained from peripheral sensors are fed into four pre-trained AlexNet models for emotion recognition.

In short, in the previous works, researchers have used various features, machine learning and deep learning models for classification of affective states. However, the accuracy of such models are still low for practical applications. Deep learning-based approaches provide comparatively better accuracy but are computationally complex. Thus, there is a need to come up with a multi-modal classification model having better accuracy and should be computationally light weight so that it can be implemented in real-life scenarios.

We found few state-of-the-art literature that used WESAD dataset for detection of mental states. In [8], the authors of the dataset extracted variety of features from different sensor signals. On the raw ECG/BVP signal, heart rate peaks, heart rate and corresponding statistical features (mean, standard deviation) were extracted. Similarly, on EDA signal, statistical features like mean, standard deviation, dynamic range, etc. were extracted. Furthermore, the tonic level and a phasic components and number of peaks were used. For EMG signal, they mostly used number of peaks, power spectral density at various frequency bands. They reported an accuracy of 92.83% for stress vs no-stress classification and 74.74% for 3 class (baseline vs stress vs amusement) classification. In another state-of-the-art work by Bobade et al. [26], authors used the same WESAD dataset for binary and 3 class classification of mental states. Here they have used similar statistical features like the mean, standard deviation, minimum, and maximum signal values. Additionally they calculated tonic, phasic components and number of peaks from EDA signal. The accuracy and *f-score* were reported using leave-one-subject-out approach. They reported, deep learning artificial neural network (ANN) gives the overall best performance with an accuracy of 84.3% for 3 class and 95.2% for binary classification of mental state. In another work by Maciej et al. [27], authors proposed various end-to-end DL learning using raw sensor signals for 3 class classification and reported that best performance was achieved using fully convolutional network (FCN) architecture with an accuracy of **78.9%** (***f-score of 0.73***). In another recent work by Md Taufeeq Uddin et al. [28], authors used the same WESAD dataset for prediction of stress and meditation states. They synthesized the physiological signals of the dataset and tried various binary classifications like *stress vs baseline*, *stress vs amusement* and *stress vs meditations*. They have used raw signal values to train the 1DCNN structure and tried to differentiate stress condition from other mental states like baseline, mediation and amusement. Precision, recall and *f1 score* reported for each of these classifications.

We have compared the performance of our proposed approach with these state-of-the-art approaches, the details of which are presented in results section.

3 Proposed approach

3.1 Dataset

Wearable stress and affect detection (WESAD) dataset [8] contains physiological signals (GSR, ECG, EMG, respiratory signal and skin temperature) and motion data recorded from chest and wrist using RespiBAN and Empatica E4 device. Sensor data were collected from 15 participants, of which 12 were male and 3 female, under four different conditions namely, baseline, amusement, stress and meditation condition. During baseline condition, subjects were sitting and reading neutral reading materials. This phase was 20 min long. During amusement condition, participants watched affective movie clips that induced a happy state. This phase was 6.5 min long. Authors of the dataset used trier social stress test (TSST) [21] protocol for inducing stress. Participants were asked to give a 5 min speech on their personal strengths and weaknesses in presence of an audience. Next, they performed a mental count down from 2023 in steps of 17. If they made a mistake, then they were requested to start over again. The ground truths of mental states were collected using state-trait anxiety inventory (STAI). The stress phase was 10 min long. Finally during meditation state, which was 7 min long, participants performed breathing exercise to induce neutral/relaxed state. It is to be noted that, in the present work, we have used the signals captured via the chest-worn device only.

3.2 Signal processing and feature extraction

The overall approach adopted in our work is shown in Figure 1. The WESAD dataset contains five types of sensor signals namely, GSR, ECG, EMG, respiratory signal and skin temperature. For each sensor, the rest period signal and task period (*i.e.,* amusement, stress and meditation phase) signal are first extracted based on the time stamp denoting the rest and task periods/intervals.

The signal corresponding to the task intervals is first subdivided into a number of windows. Window size should be sufficient to capture the signature of the mental state and signal frequency also has an impact on it. Keeping all these in

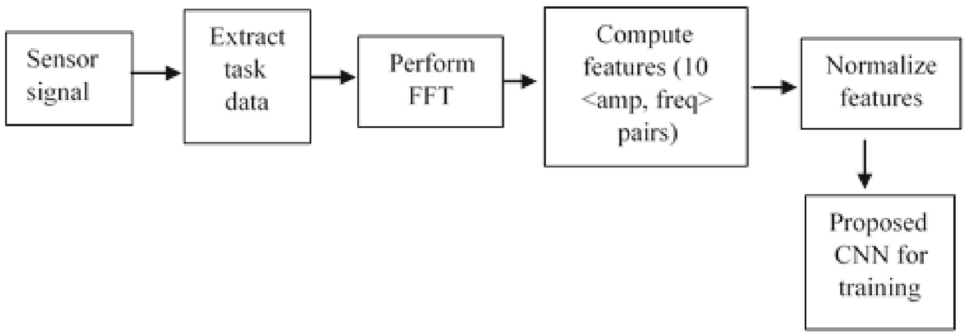

Fig. 1 Overall approach adopted in the study

mind, in our work, a window of duration 30 seconds has been considered. For each window, the time domain sensor signals are converted into frequency domain by applying fast Fourier transform (FFT). Prior to transform, the signal window is smoothened using hamming window function. Frequency domain components obtained after FFT are then arranged in descending order of their amplitudes and corresponding frequency values are also noted. Top ten components *(amplitude, frequency)* pairs are considered as the feature for the particular window of the signal. We varied the number of features from 5 to 15 and checked the model performance. However, best performance was achieved using 10 *(amplitude, frequency)* pairs and hence we used 10 features in the final model. Mental state is likely to be reflected in the variations in the sensor signal. The information (frequency and amplitude) related to the major components at different instances is likely to have an inherent pattern enabling the discrimination between different mental conditions. This encouraged us to consider top-order components as the features. It is to be noted that for almost all the participants such top ten pairs were found to be significant. Finally, these features are normalized using maximum and minimum feature values for a particular subject and labeled using the ground truth provided in the dataset. Normalization helps to reduce the inter-subject variation and suppresses the noise also. The process is repeated for each of the five sensor signals. These top 10 *(amplitude, frequency)* pairs are then fed to CNN for training purpose.

3.3 Deep CNN-based architecture

Deep learning based on convolutional neural network is being widely used in various image processing activities. It has motivated us to look for deep learning-based solution to deal with the classification task involving physiological sensor signal. In this work, our main objective is to identify various mental states like amusement, stress, meditation from the physiological signals. To achieve this goal, a convolutional neural network has been proposed to obtain a general model that can handle the variety of bio-signals.

Figure 2 shows the schematic diagram of the proposed architecture. In contrast to the common practice of feeding the raw data to the deep network, in the proposed model an initial set of features is provided in the low-dimensional matrix form. Thus, the network is free from the burden of learning the signature from the scratch. This makes the network lightweight and reduces the number of network parameters drastically. Moreover, the capturing process of physiological signals is error prone. As a result, learning from raw data is likely to get affected. In the proposed work, dominating signal components in the form of *(amplitude, frequency)* pairs are taken as input. It mitigates the issue of noise. As it is non-trivial to extract the meaningful relationships between these feature pairs within and across channels to discriminate between the mental states, we rely on a deep network. Input to the architecture is fed in the form of a 3D matrix. Let (X, Y, Z) denote the dimension of the matrix. Here, Z stands for the depth and it corresponds to the number of channel/sensor signal. As we have worked with five bio-signals, so Z is 5. $(X \times Y)$ is the dimension of feature matrix for each channel. As for each channel, top 10 *(amplitude, frequency)* pairs are taken as feature values, so X and Y are 10 and 2, respectively. Layer-wise brief description of the proposed network is as follows.

The first layer of the proposed network is a depth-wise convolutional layer. In this layer, each input channel (i.e., 10×2 feature matrix of corresponding bio-signal window) is convolved with two kernels of size 2×1. It ensures a column-wise convolution of each channel of the input data and avoids mixing of the amplitude and the frequency columns. Using two kernels per channel ensures that the kernels are able to prioritize either amplitude or frequency or both, during the convolution step, while extracting information from the input data. Moreover, depth increases to 10 as two kernels result in two feature maps for two kernels. A dropout layer follows this. It randomly sets drops for the input units. A dropout rate of 0.1 is chosen empirically. It ensures that the channel is robust to minor errors in the input and thus it effectively acts as a regularization layer. It is followed by four more depth-wise convolutional layers of which the first two layers are having the kernel of size 2×1 and the last two are with the

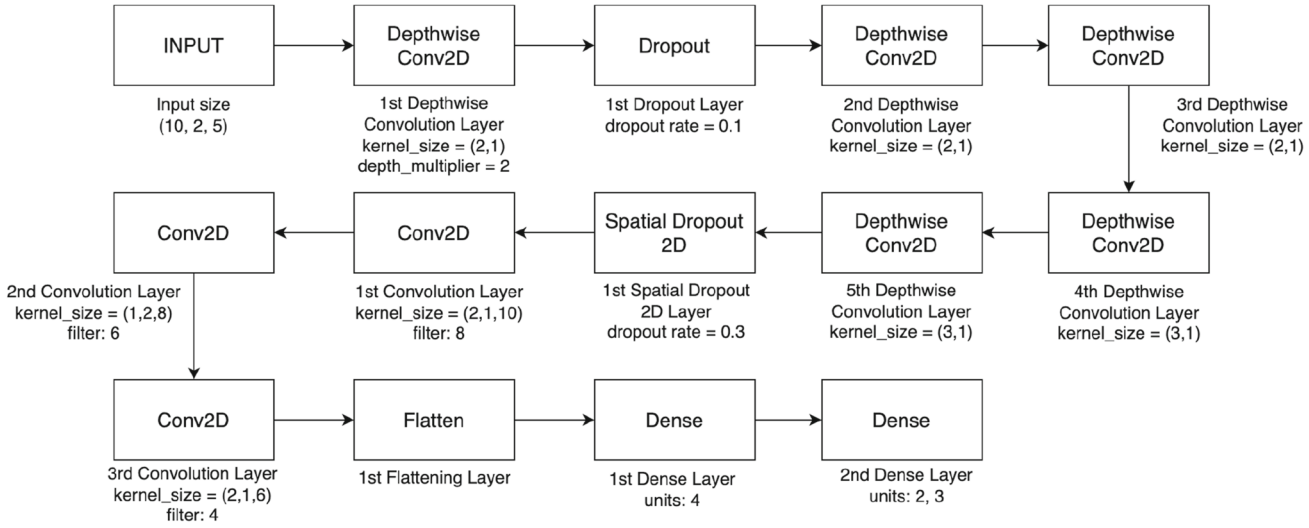

Fig. 2 Proposed deep learning architecture

kernel of size 3×1. These layers are used just to summarize the information in the columns of matrices of each channel. It is to be noted that information between the columns in the matrix or across the channels are yet to be combined/mixed. Thereafter, a spatial dropout layer is used to drop entire input channel. The dropout rate is taken as 0.3. This dropout layer promotes the independence between the channels.

The feature maps obtained so far retains the meaningful information in each dimension of 3D input matrix. Next, the cross-relationships across the dimensions are explored. The feature maps pass through three convolutional layers. The first convolutional layer facilitates mixing of amplitude information and frequency across the channels. It has eight kernels of size $2 \times 1 \times 10$. The second convolutional layer facilitates pairwise mixing of amplitude and frequency across the channels. It has six kernels of size $1 \times 2 \times 8$ and is the first layer which handles both the amplitude and the frequency simultaneously. The last convolutional layer acts a summarization layer across the channels and has four kernels of size $2 \times 1 \times 6$. In general, these convolutional layers allow mixing of the data across the channels and hence facilitates in learning complex, spatial relationships between the signal features. It is then subjected to a flattening layer.

The generated feature map is then fed to a flattening layer. It converts the feature matrix from a two-dimensional matrix to a one-dimensional array which is required to pass it through the dense layer. We have used a dense layer which is fully connected with 4 units. The dense layer allows the model to perform a linear operation on the feature matrix generated by the convolution layer. Moreover, as the convolution layers work locally for the spatial set of defined filters that traverses along with the data matrix, the dense layer acts as a global layer where all the nodes of the layer participate and are connected to all other nodes of the following layers. Therefore, the usage of dense layers in this work allows the model to establish a global relationship between the features and also accounts for the abstraction of more complex patterns in the data. The final output layer yields the prediction probabilities of each sample for the classes using the *Softmax* activation function.

Normally, sub-sampling layers make the model computationally less expensive. It is worth where spatial arrangement of data (as in case of an image) carries a meaning and a data value has relation/coherence with the neighboring values. So losing certain data is not costly. However, in our case, spatial proximity of the values in the input matrix does not have any influence. So, the usage of such layers for our scenario might result in information loss. Therefore, in the proposed architecture, pooling or sub-sampling layers has been avoided. In total, there are 527 network parameters that are to be trained. This is drastically small in comparison with any other deep network architectures used in various applications.

3.4 Training of the proposed model and testing

A cross-validation approach has been considered to train the proposed CNN model. A 15-fold cross-validation approach was used where the data of 14 subjects are used to train the model and the data of one subject were kept aside as the test set. The data of 14 training subjects are further divided into the training and validation set based on a 75–25 randomized split. Model callbacks like "early stopping" and "model checkpointing" are used while training. The model is trained multiple times before reporting the results, to ensure that the model is robust to hyper-parameters like model-weights initialization and the order of inputs used to update the weights. The model is implemented using the Tensorflow and Keras framework. The proposed model has been optimized by tun-

ing hyper-parameters such as batch size, model loss function and model optimizer.

The batch size is one of the most important parameters used to tune the networks. Larger batch sizes are computationally efficient, while smaller batch sizes lead to better generalization. For the proposed model, we performed experimentation with various batch size and finalized a batch size of 48, which is computationally efficient and allows the model to generalize as well.

This model performs classification between the four classes namely, baseline, amusement, stress and meditation. Thus, a weighted categorical cross-entropy loss function is used to train the model. The categorical cross-entropy function is generally used as a loss function for multi-class classification. Moreover, in the original dataset, the duration of four phases was different. Hence, the number of windows obtained for each of these phases was also different. Maximum number of instances were obtained for baseline phase and minimum number of windows were obtained for amusement phase. In order to handle this class imbalance, the loss function is weighted to automatically adjust weights inversely proportional to class frequencies in the input data. This enables the model to "pay more attention" to under-represented classes. In the context of deep feature representation learning using CNNs, re-sampling may either introduce large amounts of duplicate samples, which slows down the training and makes the model susceptible to overfitting [22]. Thus, we avoided re-sampling approach and used inverse class weight-based approach to handle class imbalance. The optimizer functions play a critical role in optimizing the internal parameters, i.e., the weights and biases, of a model. In this model, Adam [29] is used as an optimizer with a learning rate of 0.001.

4 Results and discussion

4.1 Affective state classification

In order to examine the performance of our proposed model, we opted for binary and 3 class classification as done by authors of WESAD dataset[8]. For binary classification, authors of [8] presented classification accuracy and f-score for *stress vs no-stress* condition. For this, they combined the states baseline and amusement into a non-stress class. For 3 class classification, they classified baseline vs stress vs amusement. All models were evaluated using the leave-one-subject-out (LOSO) cross-validation approach. We followed the same approach and generated classification accuracy and *f-score* for both binary and 3 class classification. The results are presented in Table 1. It is observed that for binary classification our results are marginally better than that reported in [8]. However, for 3 class classification we achieved an accuracy of 90.3%(with *f-score* = 0.90) which is much higher than that reported in [8].

In another state-of-the-art work by Bobade et al. [26], authors used WESAD dataset for binary and 3 class classification of mental states using various classification algorithms. The reported accuracy for binary and 3 class classification is shown in Table 1. Our achieved accuracy for 3 class is much higher than that reported in [26], and for binary class, our results are comparable with that reported in [26].

In [27], authors used various DL approaches on WESAD dataset and the accuracy reported are **78.9**% (***f-score of 0.73***) which is much less compared to our results.

In another state-of-the-art work by Md Taufeeq Uddin et al. [28], authors reported precision, recall and *f1 score* for various binary classifications which are presented in Table 2. We followed the same approach and generated results using our classification model and also presented in Table 2. It is observed that for *stress vs baseline* class, our model performs better than that reported in [28]. Similarly, for both *stress vs amusement* and *stress vs meditation*, our model outperforms the state-of-art approach.

Thus, we can conclude that our proposed lightweight deep neural network can effectively classify various mental states and performs better than other approaches reported in the literature.

5 Conclusion and future scope

We have presented a method for analyzing physiological signals for predicting various mental states. For this purpose, we have used publicly available WESAD dataset, containing sensor data from multiple physiological sensors like respiration, electrodermal activity, electrocardiogram, body temperature and electromyogram. We have proposed a light weight deep convolutional neural network (CNN) for creating a mental state prediction model. Instead of using raw sensor data to train the deep convolution network, we have used top 10 *<amplitude, frequency>* pair as the feature for each sensor data. Our proposed model has achieved the accuracy of 90.3% and 94.2% for three-class and a binary classification problems. We have shown the efficacy of our method by comparing the results with the original ground truth data. In addition, we compared our proposed approach with other state-of-the-art approaches reported in the literature. Results show that for three class classification, our proposed approach outperforms other machine learning or deep learning-based approaches reported in the literature. For binary classification problem, our results are marginally better than the results reported by the authors of WESAD dataset [8]. Our proposed approach is lightweight as it does not require learning the signature from the scratch thereby reduces the number of network parameters drastically. Thus,

Table 1 Classification results using our approach and comparison with state of the art

Classification	Philip et al[8]		Bobade et al. [26]		Proposed approach	
	Acc (%)	*f-score*	Acc (%)	*f-score*	Acc (%)	*f-score*
Stress vs no-stress	92.83	0.91	95.2	0.94	94.2	0.90
Baseline vs stress vs amusement	74.74	0.65	84.3	0.84	90.3	0.90

Table showing the performance of our proposed approach in terms of accuracy and f-score. It also shows the results reported in two state-of-the-art papers using same dataset

Table 2 Comparison of binary classification results with state-of-the-art

Classification	Precision		Recall		F1-score	
	Proposed approach	Taufeeq et al. [28]	Proposed approach	Taufeeq et al. [28]	Proposed approach	Taufeeq et al. [28]
Stress vs baseline	0.97	0.71	0.96	0.68	0.96	0.69
Stress vs amusement	0.93	0.54	0.84	0.61	0.87	0.57
Stress vs meditation	0.90	0.72	0.93	0.71	0.91	0.71

Comparison of performance of our proposed approach with state-of-the-art approach for various binary classifications. State-of-the-art approach also used same dataset as ours

our proposed CNN-based mental state classification model is computationally less complex compared to other deep networks used by the researchers and hence can be deployed in real-life scenarios.

In future, we would like to validate our proposed model by using other publicly available datasets. Moreover, since the proposed approach is generic in nature, hence we would like to apply the same for assessment of emotion, attention, cognitive load etc.

Data Availability The datasets used and analyzed during the current study are available in the UC Irvine Machine Learning Repository repository, https://archive.ics.uci.edu/ml/datasets/WESAD+%28Wearable+Stress+and+Affect+Detection\%29.

Declarations

Conflict of interest The authors did not receive support from any organization for the submitted work. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Picard RW (2000) Affective computing. MIT Press, Cambridge
2. Bauer G, et al (2012) Can smartphones detect stress-related changes in the behaviour of individuals? In: 2012 IEEE international conference on pervasive computing and communications workshops, pp 423–426. IEEE
3. Lenore K et al (1999) The mini-mental state examination (MMSE). J Gerontol Nursing 25(5):8–9
4. Smith Tasha et al (2007) The Montreal Cognitive Assessment: validity and utility in a memory clinic setting. Canadian J Psych 52(5):329–332
5. Sander Koelstra et al (2012) Deap: A database for emotion analysis; using physiological signals. IEEE Trans Affect Comput 3(1):18–31
6. Subramanian Ramanathan et al (2016) ASCERTAIN: Emotion and personality recognition using commercial sensors. IEEE Trans Affect Comput 9(2):147–160
7. Miranda Correa et al (2018) Amigos: A dataset for affect, personality and mood research on individuals and groups. IEEE Trans Affect Comput 12(2):479–493
8. Schmidt P, et al (2018) Introducing wesad, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM international conference on multimodal interaction, pp 400–408
9. Shu L et al (2018) A review of emotion recognition using physiological signals. Sensors. 18(7):2074
10. Oh S et al (2020) The design of CNN architectures for optimal six basic emotion classification using multiple physiological signals. Sensors. 20(3):866
11. Dzedzickis A et al (2020) Human emotion recognition: Review of sensors and methods. Sensors. 20(3):592
12. Wijsman J, et al (2011) Towards mental stress detection using wearable physiological sensors, In: Proceedings of annual international conference of the IEEE engineering in medicine and biology society, pp 526-532
13. Bota P et al (2020) Emotion assessment using feature fusion and decision fusion classification based on physiological data: Are we there yet? Sensors. 20(17):4723
14. Zheng WL, et al (2014) EEG-based emotion classification using deep belief networks, In: Proceedings of IEEE international conference on multimedia and expo (ICME), pp 1-6
15. Zhang Q et al (2017) Respiration-based emotion recognition with deep learning. Comput Ind 92–93:84–90
16. Martinez HP et al (2013) Learning deep physiological models of affect. IEEE Comput Intell Mag 8(2):20–33
17. Yin Z et al (2017) Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. Comput Methods Programs Biomed 140:93–110
18. Li X, et al (2016) Emotion recognition from multi-channel EEG data through convolutional recurrent neural network, In: Proceedings of IEEE international conference on bioinformatics and biomedicine (BIBM), pp 352–359
19. Tripathi S, et al (2017) Using deep and convolutional neural networks for accurate emotion classification on deap dataset, In: Deployed application case studies
20. Wenqian L et al (2017) Deep convolutional neural network for emotion recognition using EEG and peripheral physiological signal. Proc ICIG 12:385–394

21. Clemens Kirschbaum et al (1993) The 'trier social stress test'a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 28(1–2):76–81
22. Cui Yin, et al (2019) Class-balanced loss based on effective number of samples. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition
23. Dissanayake T, et al (2019) An ensemble learning approach for electrocardiogram sensor based human emotion recognition. Sensors, p 4495
24. Udovici'c, et al (2017) Wearable emotion recognition system based on GSR and PPG signals. In: Proceedings of the 2nd international workshop on multimedia for personal health and health care, Mountain View, pp 53–59
25. Aasim R et al (2020) Physiological sensors based emotion recognition while experiencing tactile enhanced multimedia. Sensors 20(14):4037
26. Pramod B, et al (2020) Stress detection with machine learning and deep learning using multimodal physiological data. In: Second international conference on inventive research in computing applications (ICIRCA). IEEE
27. Dzieżyc M et al (2020) Can we ditch feature engineering? end-to-end deep learning for affect recognition from physiological sensor data. Sensors. 20(22):6535
28. Taufeeq U Md, et al (2019) Synthesizing physiological and motion data for stress and meditation detection. In: 8th international conference on affective computing and intelligent interaction workshops and demos (ACIIW). IEEE
29. Kingma DP, et al (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980