Real-time Multimodal Emotion Classification System in E-learning Context

Arijit Nandi[1,2], Fatos Xhafa[1], Laia Subirats[2,3], and Santi Fort[2]

[1] Department of Computer Science, Universitat Politècnica de Catalunya (BarcelonaTech), 08034 Barcelona, Spain
[2] Eurecat, Centre Tecnològic de Catalunya, 08005 Barcelona, Spain
[3] ADaS Lab, Universitat Oberta de Catalunya, 08018 Barcelona, Spain
arijit.nandi@eurecat.org, fatos@cs.upc.edu, laia.subirats@eurecat.org, santi.fort@eurecat.org

Abstract. Emotions of learners are crucial and important in e-learning as they promote learning. To investigate the effects of emotions on improving and optimizing the outcomes of e-learning, machine learning models have been proposed in the literature. However, proposed models so far are suitable for offline mode, where data for emotion classification is stored and can be accessed boundlessly. In contrast, when data arrives in a stream, the model can see the data once and real-time response is required for real-time emotion classification. Additionally, researchers have identified that single data modality is incapable of capturing the complete insight of the learning experience and emotions. So, multimodal data streams such as electroencephalogram (EEG), Respiratory Belt (RB), electrodermal activity data (EDA), etc., are utilized to improve the accuracy and provide deeper insights in learners' emotion and learning experience. In this paper, we propose a Real-time Multimodal Emotion Classification System (ReMECS) based on Feed-Forward Neural Network, trained in an online fashion using the Incremental Stochastic Gradient Descent algorithm. To validate the performance of ReMECS, we have used the popular multimodal benchmark emotion classification dataset called DEAP. The results (accuracy and *F1-score*) show that the ReMECS can adequately classify emotions in real-time from the multimodal data stream in comparison to the state-of-the-art approaches.

Keywords: Affective computing · e-learning · Real-time Multimodal Emotion Classification System · Feed Forward Neural Network.

1 Introduction

Emotion, human intelligence and learning are interlinked. Emotions affect the learner's focus, exert their learning desire and influence self-regulated learning. Emotions, particularly positive emotions, have more impact on academic excellence through self-regulated learning and engagement. In e-learning, it is usually observed that the same lectures or even courses become boisterous to students

Nandi, A. [et al.]. Real-time multimodal emotion classification system in E-Learning context. A: International Conference on Engineering Applications of Neural Networks. "Proceedings of the 22nd Engineering Applications of Neural Networks Conference, EANN 2021". Berlín: Springer, 2021, p. 423-435. ISBN 978-3-030-80568-5. DOI 10.1007/978-3-030-80568-5_35.
The final authenticated version is available online at https://doi.org/10.1007/978-3-030-80568-5_35

due to negative emotions. Also, emotion stimulates the activation in the long-term memory of the associative learning material. As a result, positive emotions can improve learners' skills to learn more and perform well in evaluations, accumulate extensive expertise. This relationship between emotion and learning led many scientists to study the recognition of emotion in e-learning.

Emotion is a fundamental component of individuals, influencing their behavior, decision-making, ability to think, adaptability, well-being, and interpersonal interactions [11]. Emotions have a large effect on human actions and they must be included in human practices such as e-learning [12]. The impact of experimentally induced positive and negative emotions on multimedia learning studied in [18] showed that students with the greatest understanding of prior education or working capacity had offset the emotional effect on learning outcomes. According to [3,13], it is not only learning but also the interdependence between learning and feeling that is mediated in the e-learning. With the expansion of Learning Management Systems (LMS), conventional face-to-face learning is adapting each time more e-learning. While in conventional classroom instruction, a teacher may alter his or her teaching approach by observing students' facial expressions and body movements, in e-Learning environments, this becomes difficult.

It should be noted that data sources used for emotion classification are of paramount importance. Researchers have found that single data modality might come short to capture a complete insight of the learning experience. So, multiple data streams, such as EEG, EDA, eye tracking, audio, video, RB, ECG etc.) are envisioned [7,25]) to support higher accuracy in emotion classification [33]. In [10], authors have also demonstrated the necessity of building robust user models and learning through integration of information with fusion technologies. In fact, Learning Analytics and Knowledge (LAK) has recognized the need to take such dynamic behavioral data into account in addition to traditional e-learning data (e.g., MOOCs, LMS data, etc.) [22]. According to [33], combining physiological data (such as electroencephalogram (EEG) or electrocardiogram (ECG), etc.) with external behaviors (such as eye movement or facial expressions, etc.) is a promising approach to capture learner's emotions and experience. In [4], authors have introduced Multimodal Machine Learning (MML), as an approach to deal with multimodal data sources. Learning from multimodal sources (heterogeneous sources) offers the possibility to catch the interaction between modalities and to obtain a detailed understanding of natural phenomenon. A recent study [20] has shown that multimodal data combination improves accuracy and provides greater insights into learner's emotions and experience.

We recently proposed in [23], a real-time emotional classification processing methods for a single data stream and used physiological data (EEG) stream. In this paper, we propose real-time emotional classification processing methods for multimodal data streams based on decision fusion approach aiming to improve accuracy and robustness of online classifiers.

The contributions in this paper are as follows:

(1) A real-time multimodal emotion classification system that uses Feed Forward Neural Network trained in an online fashion using the Incremental Stochas-

tic Gradient Descent for processing each data modality and classify the emotional states. Then, the emotion class decisions from each modality are fused in a decision level by Weighted Majority Voting to classify the final emotional state. A three-modal physiological data stream (EEG, EDA, Respiratory Belt) is used.

(2) The experimental results show that our proposed ReMECS classifier outperforms state-of-the-art emotion classifiers for multi-modal stream data; namely, Random Forest, Stacked Auto-encoder, Convolutional NN, and others.

The rest of the paper is organized as follows: in Sect. 2, preliminaries of various concepts are introduced. In Sect. 3, materials and methods used in our proposed approach are presented. Analysis of results and evaluations are presented in Sect. 4. Finally, Sect. 5 draws the conclusion and future work.

2 Preliminaries

Herewith, a brief introduction of concepts related to the learning from the data streams and relevant emotional models are presented.

2.1 Learning from Multi-modal Data Streams

In multimodal data stream learning, the corresponding model of each modality learns progressively from data tuples as they arrive, with a single pass through them [15]. Here, the time dimension in Fig. 1 is important to note, where different data tuples arrive in a stream mode at different time (t_1, t_2, ...) from different modality. The models (one per each modality) will be tested using the earlier model as soon as the data arrives. The training will then be completed based on the error, and a new model will be available for the next data set; this process will continue as long as the multimodal data stream is arriving. The model will perform poorly at first because it lacks sufficient understanding of the data, but it will learn gradually and upgrade itself from the data stream and improve its classification performance ultimately. Furthermore, the model can be analyzed at any point in time, and no part of the data set will be looped back over.

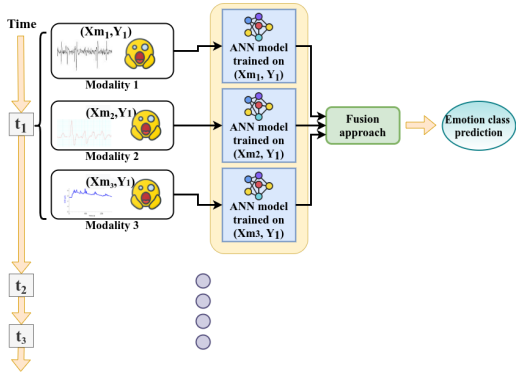


Fig. 1. Online emotion classification from a multimodal data stream.

2.2 Emotion Representation

Studies about emotion have faced a constant challenge in defining multiple emotions in a meaningful way. The number of emotion representation categories has long been a source of debate in psychology [31]. Researchers have focused extensively on two emotion representation models: categorical emotion model (CEM) and dimensional emotion model (DEM). DEM contains human emotions in a dimensional structure –each dimension reflects an emotional characteristic– and can be put into 3D or 2D as a continuous and coordinated point in multi-dimensional space. Rather than selecting discreet emotional labels, an individual's emotion are expressed at different continuous or distinct levels, e.g., pleasant—unpleasant—attention—rejection or valence—dominance (VAD) [27]. The most common model under DEMs is VAD, providing valence from positive to negative, arousal expresses the strength of emotions from calm to excited, and dominance varies from controlled to in control. It is impossible to quantify the dominance (mostly omitted) that leads to the two-dimensional space-arousal (VA) [14]. Russell's 2D emotion model is most commonly used model in DEM.

3 Materials and Methods

3.1 Data Set Description

DEAP [19] (Database for Emotion Analysis using Physiological Signals) is a widely used multimodal dataset in emotion classification. DEAP contains EEG, peripheral (such as EDA, RB, etc.) and video signals. The DEAP experiment was conducted on 32 participants in which 16 were male and 16 female. Each participant watched 40 different music videos of 60 s in length. A total of 48 channels including 32 EEG channels, 12 peripheral channels, 3 unused channels and 1 status channels were used to record the raw data. Each data file is stored in a 3D matrix representation ($40 \times 40 \times 8064$), which represents video/trial $\times$ channel $\times$ data. The emotion labels are stored in a 2D matrix (40×4) in the same file. In the dataset the channels from 1-32 are for EEG signals, channel no. 37 is for EDA signal and channel no. 38 is for RB signal.

3.2 Feature Extraction

Extraction of features is important for retrieving EEG, EDA and RB information, which effectively represents the emotional state. The extracted features are then used to train emotion classification algorithms.

Wavelet decomposition (WD) is a time-frequency analysis procedure that is common, practical, and widely used. Because of its localized analysis approach that uses time as well as a frequency window, multi-rate filtering, multi-scale zooming, and is better suited for non-stationary signals, it is the most commonly used feature extraction technique applied to EEG, EDA and RB signals [17]. The multi-scale analysis of EEG signals provides both details and approximations of the EEG signal at different wavelet scales [31]. It also provides a series of wavelet

coefficients at different scales. All these coefficients are capable of describing the original signal's complete characteristics; that is why these are considered features of the signal. Most frequently, Meyer WD, Morlet Mother WD, Haar Mother WD and Daubechies WD are the wavelet base functions [29]. The most frequently used features extracted from each sub-bands of EEG, EDA and RB are entropy, median, mean, standard deviation, variance, 5th percentile value, 25th percentile value, 75th percentile value, 95th percentile value, root means square value, zero crossing rate, mean crossing rate [8,2,1]. In our experiment we have extracted and used these features for emotion classification.

3.3 Feed-Forward Neural Network (FFNN)

FFNN is one of the forms of multi-layer perceptron (MLP) [26], used in many applications due to its high forecasting and classification capabilities. It consists of three layers (input layer, hidden layer and output layer). Briefly, *Neurons* are the basic processing element of an ANN and there are no direct connections within the neurons of the same layers. Training of an FFNN aims at minimizing the error, where mean square error (MSE) is most popularly used error function in classification context. In our proposed approach, we have used 3-layers FFNN for processing the data streams of each modality. So, there is a total of 3 FFNNs (1 for EEG stream, 1 for EDA stream, and 1 for RB stream) used to develop the ReMECS system. *Sigmoid* activation function is used in all the FFNNs throughout the layers. The reason for choosing 3-layers FFNN for our experiment is as follows: (1) It has ability to learn and perform classification based on the data given for training; (2) We do not make any assumption about the pattern classes underlying probability density functions or other probabilistic information.

3.4 Incremental Stochastic Gradient Descent (ISGD)

In practical scenario, a proper supervised training of a model needs multiple passes (multiple epochs) through the training data. However, in a streaming environment, batch mode gradient descent is inefficient and creates system overhead. Because in streaming scenario data comes in continuous rate and the number of observations increases batched mode gradient descent operations are expensive to perform it in online scenario [5]. In online scenario the Incremental Stochastic Gradient Descent, a version of Stochastic Gradient Descent, is more suitable to train the model sequentially based on the data stream arrival i.e., weights are updated sequentially. The weight update is as follows: $w_i = w_{i-1} - \gamma_i \nabla V(\langle w_{i-1}, x_{t_i} \rangle, y_{t_i})$. The main difference with the stochastic gradient method is that here a sequence t_i is chosen to decide which training point is visited in the i^{th} step. Such a sequence can be stochastic or deterministic.

3.5 Decision Level Fusion

Techniques of data fusion incorporate data from various sources. The fusion of data is classified as *feature-level fusion* and *decision-level fusion*. In our study,

we have used decision-level fusion. To fuse the decisions from each classifiers at decision level, we have used a popular decision ensemble method called Weighted Majority Voting (WMV) [6] (see Fig. 2). The algorithm is fed a stream of objects that need to be classified, each one followed by the correct label.

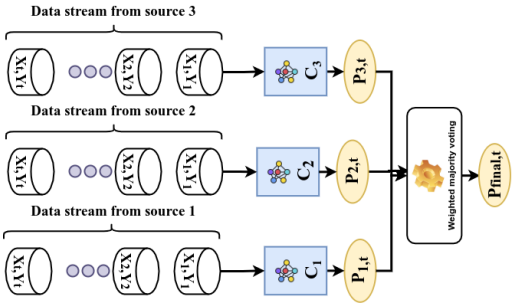


Fig. 2. Decision level fusion using weighted majority voting.

The pseudo code of WMV is as follows:

Algorithm: Pseudocode for Weighted Majority Voting

Input: a stream of pairs (x, y), parameter $\beta \in (0, 1)$
Output: a stream of predictions $\hat{y}$ for each x.
`Weighted Majority Voting`($multimodal\ stream$, β)
  initialize stream classifiers $C_1, C_2 \ldots C_N$ with weights $w_i = \frac{1}{N}$ for each data stream modality
  for *each* x *in stream* **do**
    collect predictions $C_1(x) \ldots C_N(x)$
    $p \leftarrow \sum_{i=1}^{N} w_i C_i(x)$ `// decision fusion`
    $v \leftarrow (p - \frac{1}{2})$
    if $v > 0$ **then**
      $\hat{y} = 1$ `// producing the predicted class`
    else
      $\hat{y} = 0$
    for $i \leftarrow 1$ *to* N **do**
      if $C_i(x) \neq y$ **then**
        $w_i \leftarrow \beta \cdot w_i$ `// penalizing the weights by` β
    $S_w \leftarrow \sum_{i=1}^{N} w_i$ `// adding all the weights`
    for $i \leftarrow 1$ *to* N **do**
      $w_i \leftarrow \frac{w_i}{S_w}$ `// weight scaling`
  return $\hat{y}$

3.6 Experimental Study

The working principles of real-time multimodal emotion classification for an EEG, EDA and RB data streams are presented here. The illustrative view of our proposed ReMECS is shown in Fig. 3, according to the followings steps:

(1) Data set consideration and data rearrangement: The pre-processed multimodal DEAP data [19] is used for a multimodal data stream using EEG,

EDA and RB signal stream. As the DEAP data is stored in 3D matrix format, we have rearranged the EEG signals into 1D matrix as follows:

$[participant, video, channel\ no, channel\ data, valence\ class, arousal\ class]$.

Similarly, for EDA and RB data the 1D matrix looks as follows:

$$[participant, video, data, valence\ class, arousal\ class].$$

In the experiment, the EDA, RB and EEG multimodal data streams are used to classify high/low valence and arousal emotions. Thus, while streaming, the valence and arousal scores are automatically scaled. So, valence score greater than 5 is considered as high valence (i.e. 1); otherwise, it is considered as low valence (i.e. 0). For arousal class labels similar scaling is done.

(2) Stream reading: In the multimodal data (EEG, EDA and RB) stream simulation, we have used a sliding window protocol to stream the data for every participant. The sliding window size is set to 10s because it has already been used in previous emotion literature [1,9] and its effectiveness has been validated. The multimodal data stream rate is approximately 3 Mb/10s. With the help of WebSockets, the multimodal streaming system is simulated. In the ReMECS system, continuous multimodal physiological data (EEG, EDA and RB) streams are coming to the server from the client side, and the server processes the multimodal data streams for classifing emotions in real-time.

(3) Feature extraction: Wavelet feature extraction technique is used to extract features from multimodal signal streams (EDA, EEG and RB signals) in this experiment. The base function for the feature extraction is wavelet Daubechies 4 (Db4). In our experiment, decomposition of EEG signal streams into five levels, EDA into three levels and RB into three levels are performed.

(4) Emotion classifier: For emotion classification from the multimodal data streams we have used a Feed Forward Neural Network of 3-layers (input, hidden and output layer) model for high/low valence and arousal classification. Here, one FFNN for EEG signal stream, a second FFNN for EDA stream and a third FFNN for RB signal stream procession for emotion classification. So a total of 3 FFNNs are used in parallel to process the streams from each modality.

(5) Model test and training: FFNN models are trained with ISGD in online fashion using the interleaved test-then-train method. The interleaved test-then-train approach is chosen as it uses the same memory; an individual tuple is used to test the model before it is trained and then the accuracy and F1-score metrics are updated. Thus, the FFNN is always tested on data tuples never seen before. Initially, the model performs poorly but gradually it will be more stable and perform better as it sees more tuples from multimodal data streams.

3.7 Experimental Setup

Machine setup, software environment, parameter setup and performance metrics are as follows.

(a) Machine configuration: Ubuntu 18.04 64 bit OS, processor core-i7-7700HQ with RAM 16 Gb–2400 MHz and 4Gb-Nvidia GTX-1050 graphics.

(b) Software development: ReMECS is implemented from scratch in Python 3.7.

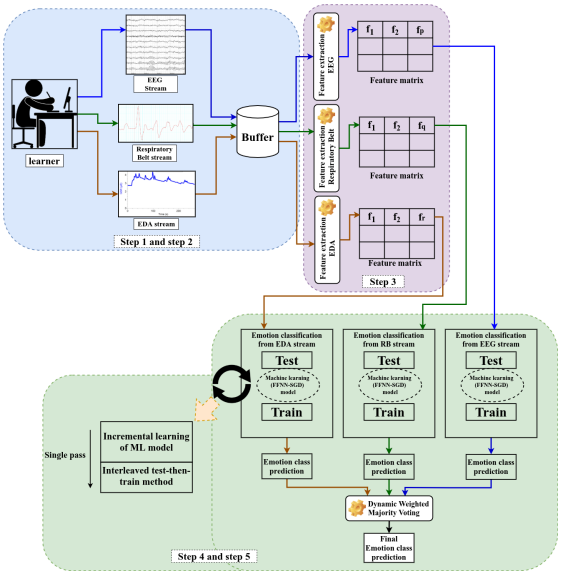


Fig. 3. Diagram of real-time multimodal emotion classification system (ReMECS).

(c) Parameter setup: In our ReMECS, the learning rate for ISGD is 0.05, and it is fixed throughout the emotion classification. To set the learning rate, we have done a *Cyclical Learning Rates* (CLR) [28] on one subject using ReMECS, where the learning rate for ISGD is varied from 0 to 1 with a step size increment of 0.01. The learning rate for which the ReMECS approach has less error, is selected for the experiment.

(d)Performance Metrics: For evaluating the classifiers performance, F-measure ($F1\text{-}score$) and balanced accuracy (Acc) [2] are used. These metrics are calculated as $Acc = \frac{sensitivity+specificity}{2}$ and $F1 - score = 2 \times (\frac{precision \times recall}{precision+recall})$. Where $sensitivity = \frac{TP}{TP+FN}$, $specificity = \frac{TN}{FP+TN}$, $Precision(Pre) = \frac{TP}{TP+FP}$ and $Recall(Rec) = \frac{TP}{TP+FN}$, with the usual meaning of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

4 Results, Evaluation and Discussion

Here, we report a three-fold comparison: (1) single modal emotion recognition approaches vs multimodal emotion recognition approach (ReMECS); (2) ReMECS with our previous Real-time Emotion Classification System (RECS [24]) and, (3) our ReMECS with state-of-the-art offline emotion classifiers from literature.

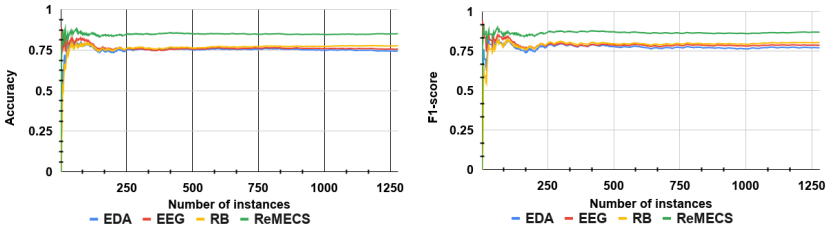
First comparison: In Table 1, the average accuracy and F1-score are provided for valence and arousal classification using single modal data stream classifiers and multimodal data stream classifier ReMECS. The accuracy comparison for valence and arousal emotion classification of single modal approach vs mul-

timodal ReMECS are presented in Fig. 4(a) and 5(a), resp. and in Fig. 4(b) and 5(b) valence and F1-score comparison of single modality vs ReMECS are shown.

Second comparison: In our previous work RECS, where we have developed a realtime emotion classification from EEG data stream using Logistic Regression trained in online fashion using SGD. So from the comparison, ReMECS has achieved better average accuracy and F1-score than RECS and outperformed RECS for valence and arousal emotion classification. Thus, from the comparison, it can be concluded that ReMECS performed better in terms of real-time valence and arousal emotion classification from multimodal physiological data (EEG, EDA and RB) stream because the average accuracy and F1-score of ReMECS is superior than RECS and those of the considered single modal approaches.

Table 1. Avg. accuracy and F1-score comparison of single modal approaches vs multimodal ReMECS

		Valence accuracy	Valence F1-score	Arousal accuracy	Arousal F1-score
Single modal approach	EEG	0.7635 (±0.01)	0.7902 (±0.15)	0.7196 (±0.02)	0.7424 (±0.03)
	EDA	0.7513 (±0.03)	0.7730 (±0.03)	0.7348 (±0.02)	0.7540 (±0.06)
	RB	0.7644 (±0.02)	0.7902 (±0.05)	0.8255 (±0.03)	0.8413 (±0.03)
	RECS	0.6796 (±0.004)	0.71 (±0.003)	0.6483 (±0.002)	0.71 (±0.002)
Multimodal approach	ReMECS	**0.8477 (±0.02)**	**0.8649 (±0.02)**	**0.9551 (±0.04)**	**0.9589 (±0.04)**



(a) (b)

Fig. 4. Accuracy and F1-score comparison of single modal approaches with multimodal ReMECS for Valence classification.

Selected works from literature for comparison: In [30], authors have proposed a deep learning based model called multiple-fusion-layer ensemble classifier of stacked auto-encoder (MESAE) for emotion classification form multimodal physiological signals. They have utilised the DEAP data set for performance validation –six data modality (EEG, EDA, EMG, RB, Blood volume pressure and skin temperature features). Researchers in [21] have shown a multimodal emotion recognition approach using bimodal deep auto encoder (BDAE) where they have utilized fusion of EEG data with other features. A multimodal emotion approach using an ensemble of convolutional neural network (ECNN) can be found in [16]. A plurality voting approach is adopted to make the ensemble model, fusing four data modalities (EEG, EDA, RB and EOG). Another CNN multimodal emotion recognition can be found in [32], using a Hierarchical Fusion Convolutional Neural Network (HFCNN) to develop the multimodal emotion

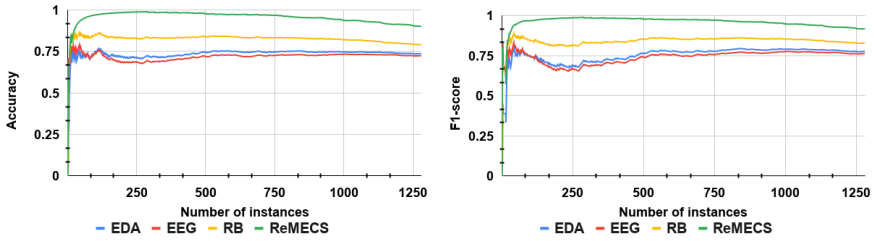


Fig. 5. Accuracy and F1-score comparison of single modal approaches with multimodal ReMECS for Arousal classification.

recognition system. They have used EEG, galvanic skin response (GSR), respiration belt (RESP), skin temperature (TEMP), and plethysmograph (PLET) data. Another novel emotion recognition approach for emotionally sensitive health systems based on multimodal physiological signals can be found in [2] based on three data modalities (respiratory belt (RB), photoplethysmography (PPG) and fingertip temperature (FTT)); a decision level fusion is performed to produce the final emotional state for their Random Forest classifier. The comparison of our approach with all these approaches is shown in Table 2.

Table 2. Comparison with state-of-the-art works for multimodal emotion classification

Research	Valence		Arousal	
	Accuracy	F1-score	Accuracy	F1-score
MESAE [30]	0.7719	0.6901	0.7617	0.7243
BDAE [21]	0.852	-	0.805	-
ECNN [16]	0.829	-	0.829	-
HFCNN [32]	0.8328	-	0.8471	-
RF [2]	0.7308	-	0.7218	-
ReMECS (our proposal)	**0.8477**	**0.8649**	**0.9551**	**0.9589**

5 Conclusions

In this paper, a framework for emotion recognition (ReMECS) using multimodal physiological signals stream (EEG, EDA and RB) is proposed. Our system is based on Feed-Forward Neural Network, trained in an online fashion with the Incremental Stochastic Gradient Descent. To validate the performance of ReMECS, we have used the DEAP multimodal emotion dataset. It is shown that decision level fusion (by Weighted Majority Voting) from multiple classifiers (one per signal sensor source) has improved the emotion classification in terms of average accuracy and F1-score in both valence and arousal dimensions. The comparison among single modal emotion classifiers and state-of-the-art multimodal emotion classification approaches with proposed system shows that it has outperformed the considered approaches.

As future work, we plan to apply our system in an application scenario in e-Learning called augmented workspace in Eurecat's[4] materials laboratory, where students perform learning tasks (materials characterization, flexibility measurement, etc.) ReMECS will classify students' emotions in real-time during the learning activities. The classified students' emotions will be shown to the teachers in a dashboard to undertake appropriate action.

Acknowledgement

Work partially funded by ACCIÓ under the project TutorIA.

References

1. Ayata, D., Yaslan, Y., Kamaşak, M.: Emotion recognition via random forest and galvanic skin response: Comparison of time based feature sets, window sizes and wavelet approaches. In: Medical Technologies National Congress. pp. 1–4 (2016)
2. Ayata, D., Yaslan, Y., Kamasak, Mustafa, E.: Emotion recognition from multimodal physiological signals for emotion aware healthcare systems. J. of Medical and Biological Eng. pp. 149–157 (2020)
3. Bahreini, K., Nadolski, R., Westera, W.: Towards multimodal emotion recognition in e-learning environments. Interactive Learning Env. **24**(3), 590–605 (2016)
4. Baltrušaitis, T., Ahuja, C., Morency, L.: Multimodal machine learning: A survey and taxonomy. IEEE TPAMI **41**(2), 423–443 (Feb 2019)
5. Bertsekas, D.P.: Incremental gradient, subgradient, and proximal methods for convex optimization: A survey. Optimization for Machine Learning **2010**(1-38), 3
6. Bifet, A., Holmes, G., Kirkby, R., Pfahringer, B.: Moa: Massive online analysis. J. Mach. Learn. Res. **11**, 1601–1604 (2010)
7. Blikstein, P., Worsley, M.: Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. J. Learn. Anal. **3**, 220–238 (09 2016)
8. Bota, P., Wang, C., Fred, A., Silva, H.: Emotion assessment using feature fusion and decision fusion classification based on physiological data: Are we there yet? Sensors **20**(17) (2020)
9. Candra, H., Yuwono, M., Chai, R., Handojoseno, A., Elamvazuthi, I., Nguyen, H.T., Su, S.: Investigation of window size in classification of eeg-emotion signal with wavelet entropy and support vector machine. In: 37th Annual Int'l Conference of the IEEE Engineering in Medicine and Biology Society. pp. 7250–7253 (2015)
10. Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., Specht, M.: Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data. p. 188–197. ACM (2017)
11. Ekman, P.: An argument for basic emotions. Cognition and Emotion **6**(3-4), 169–200 (1992)
12. Faria, A.R., Almeida, A., Martins, C., Gonçalves, R., Martins, J., Branco, F.: A global perspective on an emotional learning model proposal. Telematics and Informatics **34**(6), 824 – 837 (2017)

[4] Eurecat Technology Centre of Catalonia, Spain: https://www.eurecat.org/

13. Finch, D., Peacock, M., Lazdowski, D., Hwang, M.: Managing emotions: A case study exploring the relationship between experiential learning, emotions, and student performance. Int'l Journal of Management Education **13**(1), 23–36 (2015)
14. Hanjalic, A.: Extracting moods from pictures and sounds: towards truly personalized tv. IEEE Signal Processing Magazine **23**(2), 90–100 (2006)
15. Hayes, T.L., Kanan, C.: Lifelong machine learning with deep streaming linear discriminant analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (June 2020)
16. Huang, H., Hu, Z., Wang, W., Wu, M.: Multimodal emotion recognition based on ensemble convolutional neural network. IEEE Access **8**, 3265–3271 (2020)
17. Islam, M.R., Ahmad, M.: Wavelet analysis based classification of emotion from eeg signal. In: Int'l Conf. on Electrical, Computer and Comm. Eng. pp. 1–6 (2019)
18. Knörzer, L., Brünken, R., Park, B.: Emotions and multimedia learning: The moderating role of learner characteristics. J. Comp. Assist. Learn. **32**(6), 618–631 (2016)
19. Koelstra, S., Muhl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: Deap: A database for emotion analysis;using physiological signals. IEEE Trans. Affect. Comput. **3**(1), 18–31 (2012)
20. Lee, D.H., Anderson, A.K.: Reading what the mind thinks from how the eye sees. Psychological Science **28**(4), 494–503 (2017)
21. Liu, W., Zheng, W., Lu, B.: Multimodal emotion recognition using multimodal deep learning. CoRR **abs/1602.08225** (2016)
22. Mitri, D.D., Schneider, J., Specht, M., Drachsler, H.: The big five: Addressing recurrent multimodal learning data challenges. vol. 2163. CrossMML (2018)
23. Nandi, A., Xhafa, F., Subirats, L., Fort, S.: A survey on multimodal data stream mining for e-learner's emotion recognition. In: 2020 International Conference on Omni-layer Intelligent Systems (COINS). pp. 1–6 (2020)
24. Nandi, A., Xhafa, F., Subirats, L., Fort, S.: Real-time emotion classification using eeg data stream in e-learning contexts. Sensors **21**(5) (2021)
25. Prieto, L., Sharma, K., Kidzinski, L., Rodríguez-Triana, M., Dillenbourg, P.: Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. J. Comput. Assist. Learn. **34**(2), 193–203 (2018)
26. Savran, A.: Multifeedback-layer neural network. IEEE Trans. on Neural Networks **18**(2), 373–384 (2007)
27. Schlosberg, H.: Three dimensions of emotion. Psych. rev. **61(2)**, 81–88 (1954)
28. Smith, L.N.: Cyclical learning rates for training neural networks. In: IEEE Winter Conference on Applications of Computer Vision. pp. 464–472 (2017)
29. Subasi, A.: Eeg signal classification using wavelet feature extraction and a mixture of expert model. Expert Systems with Applications **32**(4), 1084 – 1093 (2007)
30. Yin, Z., Zhao, M., Wang, Y., Yang, J., Zhang, J.: Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. Computer Methods and Programs in Biomedicine **140**, 93–110 (2017)
31. Zhang, J., Yin, Z., Chen, P., Nichele, S.: Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review. Information Fusion **59**, 103 – 126 (2020)
32. Zhang, Y., Cheng, C., Zhang, Y.: Multimodal emotion recognition using a hierarchical fusion convolutional neural network. IEEE Access **9**, 7943–7951 (2021)
33. Zheng, W., Liu, W., Lu, Y., Lu, B., Cichocki, A.: Emotionmeter: A multimodal framework for recognizing human emotions. IEEE Transactions on Cybernetics **49**(3), 1110–1122 (March 2019)