Neural network for automatic emotion classification using virtual environments

J. Adrian Lopez[0009-0004-2523-9680], Alma Y. Alanis[0000-0001-9600-779X], and Ricardo E. Garcia[0000-0002-1210-3459]

University Center for Exact and Engineering Sciences,
University of Guadalajara,
1421 Marcelino Garcia Barragan, Guadalajara, Jalisco 44430, Mexico
jose.lopez2863@alumnos.udg.mx
alma.alanis@academicos.udg.mx
emmanuel.garcia@academicos.udg.mx

Abstract. Given the inherent variability of human emotional responses, their objective measurement remains a complex challenge. Emotions are typically manifested through physiological changes, such as sweating, body temperature, and heart rate variations. This study utilized these three parameters to map participants' emotional states within the emotional space proposed by Kołakowska (2015). To elicit controlled emotional responses (happiness, anger, sadness, and fear) virtual environments were created using stimuli from the International Affective Picture System (IAPS) and the International Affective Digitized Sounds (IADS). Following data acquisition, signal processing techniques were applied to project the responses onto the emotional space, thereby enhancing the accuracy of emotional state monitoring. A neural network was subsequently trained for multiclass classification, achieving a 10% improvement in accuracy compared to previous models.
Keywords: UAV, PID Control, Neuronal Control, Extended Kalman Filter, Hardware-in-the-Loop.

1 Introduction

1.1 Emotions and affective computing

"Emotion is an organism's response to a particular stimulus (person, situation, or event). Typically, it is an intense, short-lived experience, and the person is usually highly aware of it" [1].

In the last decade, technological advancements have enabled an increasingly deep integration of human emotions into the computational domain, giving rise to the field of affective computing. This interdisciplinary field focuses on the interaction between humans and machines, aiming not only to process logical information, but also to understand and respond to user emotions [2]. Affective computing has applications in various areas, from education to mental health, and draws on knowledge from psychology, neuroscience, and artificial intelligence to create systems that simulate empathy and enhance human-computer interaction [2].

https://doi.org/10.61728/AE20255190

One of the most promising approaches within affective computing is automatic emotion detection (AED), which refers to the ability of machines to identify and process emotions expressed by users through different modalities, such as voice, facial expressions and more recently body movements in virtual environments [4].

This recognition becomes particularly relevant in virtual environments, where users can interact with digital representations of themselves and others in three-dimensional spaces. Virtual environments provide a rich context for emotional expression and allow researchers and developers to experiment with new forms of interaction and communication [4].

Many of the studies related to emotion processing and detection involve the generation of biosignals for analysis. In this regard, a significant portion of state-of-the-art research uses resources such as videos, sounds, images, or questionnaires as emotional stimuli, which are typically displayed on a monitor or screen [8].

The convergence of AED and virtual environments presents a wide array of innovative opportunities. These environments can be systematically designed to foster immersive experiences in which emotions play a pivotal role, facilitating natural and intuitive interactions. For instance, in virtual reality applications for psychological therapy, automatic emotion recognition can assist therapists in dynamically adjusting their approaches based on patients' emotional responses. Likewise, in educational contexts, tailoring content to students' emotional states has the potential to enhance learning outcomes and improve information retention [9].

However, despite significant advances in this field, automatic emotion recognition in virtual environments faces several challenges. Cultural variability in emotional expression, individual differences in the perception and interpretation of emotions, and the complexity of human behavior are factors that complicate design of accurate and effective systems. In addition, ethics and privacy in the handling of emotional data represent crucial considerations that must be addressed.

1.2 Previous work

Previous research has already been conducted with this same focus, using virtual environments as a primary method of emotional stimulation. In these studies, classification relied not only on bio signals but also incorporated voice and text acquisition. Although this study yielded good results, classification method implemented in [4] is exclusively binary.

Another related investigation focused on brain signals through electroencephalograms (EEG). For this study, virtual environments are implemented to subject patients to emotional thresholds, which are categorized through activation and valence. For classification, signals are first characterized by partitioning their power, resulting in two characterization methods called bipartition and tripartition. Support vector machines (SVM) and a decision tree model are used for classification [5].

2 Methodology

For this study, a database collected in another work titled "Automatic emotion detection through biosignals and Virtual Environments" is used. This database compiles three different physiological signals: heart rate, temperature, and skin impedance. The database contains information from 30 patients subjected to four different emotions (happiness, anger, sadness, and fear) through virtual environments generated using images from IAPS [7] and sounds from IADS [6].

Physiological signals are measured in voltage and then processed to convert them into interpretable units: beats per minute (bpm), degrees Celsius (° C) and impedance (Ω). Once the database information is preprocessed, a final transformation is performed, where these interpretable units are converted to values ranging from -1 to 1, corresponding to the three-dimensional space proposed by Mehrabian and Russell [3]. In this space, each signal corresponds to one of the axes (as shown in Figure 1):

- Valence: Degree of pleasure or displeasure associated with the emotion.
- Activation: Level of arousal or physiological excitement.
- Dominance: Sensation of control or submission in response to emotion.

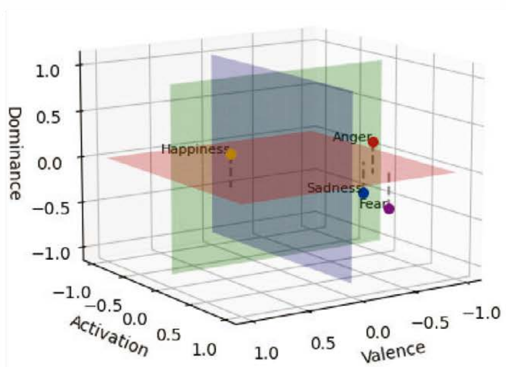

Fig. 1. Representation of the three-dimensional space defined by dominance, valence and activation.

2.1 Signal Characterization

To label emotions, an algorithm based on Euclidean distance is used as:

$$d(P, Q) = \sqrt[2]{(x_2 - x_1) + (y_2 - y_1)(z_2 - z_1)} \tag{1}$$

Emotion	X	Y	Z
Happiness	0.76	0.48	0.35
Anger	-0.43	0.67	0.34
Fear	-0.64	0.60	-0.43
Sadness	-0.63	0.27	-0.33

Table 1. Coordinates for each emotion.

Each signal is compared to predefined reference points in the three-dimensional space (Table 1), representing emotions of happiness, sadness, fear, and anger. Signal is classified with emotion corresponding to the closest point in terms of Euclidean distance, thus obtaining usable data for emotional classification.

2.2 Neural Network Design

Two neural network models are designed to observe their behavior with different activation functions.

First neural network is designed with 3 inputs, one hidden layer containing 20 neurons with a hyperbolic tangent ($tanh$) activation function, and finally an output layer with 4 neurons and a linear activation function. This design aims to generate 4 hyperplanes capable of correctly separating emotions.

$$f(x) = x \tag{2}$$

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^x} \tag{3}$$

The second model followed the same pattern as the first, with 3 inputs, a hidden layer of 5 neurons with a tanh activation function, and a final output layer with a Softmax activation function and 4 outputs to classify the 4 different emotions present in this study to predict both neural networks are trained using data transformed into dominance-activation-valence space, with the goal of predicting the corresponding emotion.

$$Softmax(x_i) = \frac{e^{x_i}}{\sum_j e^{e_j}} \tag{4}$$

Inputs to use are the three values obtained in characterization phase. Values are dominance for temperature, valence for heart rate and activation for skin impedance.

Selection of these activation functions is based on the type of data that is being used. Function selected for the hidden layer (tanh) is chosen because of data being handled is normalized between -1 and 1 and this function can work with both negative and positive values. As for activation functions for output layer (linear and Softmax), this choice is made to compare results in classification

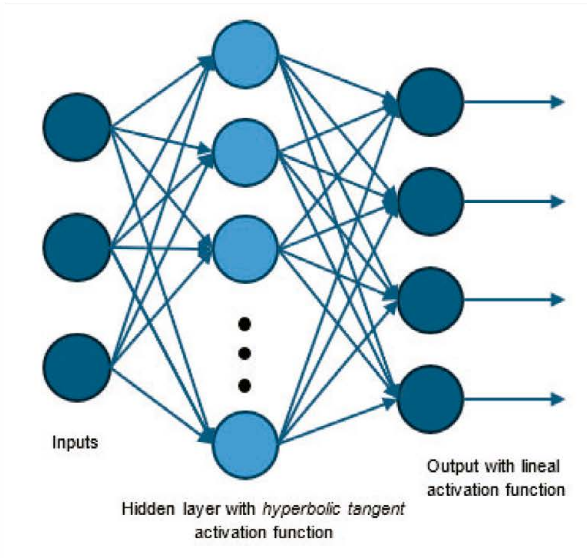

Fig. 2. Architecture design with linear output.

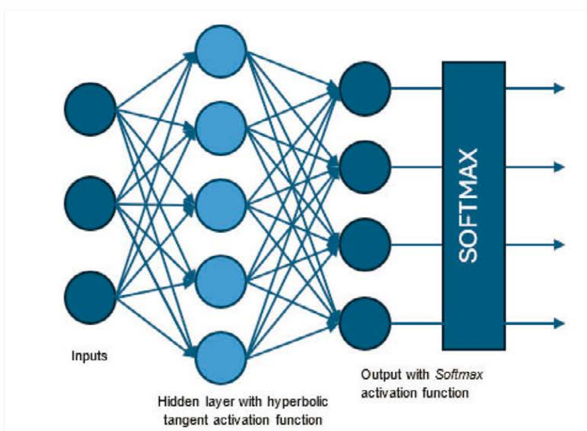

Fig. 3. Architecture design with Softmax output.

using hyperplanes in case of linear function and by probability in the case of Softmax function.

Simulation is carried out in a virtual Python environment on the Google Colab server, using the TensorFlow library to generate neural network models. By implementing this library, the Adam optimizer model is used and loss (Mean square error for lineal out and categorical cross entropy for Softmax) and accuracy (f1-score) metrics are assigned. The training process is planned to start with a low number of epochs and gradually increase them based on the behavior of the neural network.

2.3 Implementation of Backpropagation

Training neural networks is carried out using backpropagation algorithm, which adjusts the weights of neural network to minimize loss function. This process consists of two main phases: forward pass and backward pass.

Forward Pass: In this step, input data is propagated through neural network, calculating outputs of each layer until final prediction is obtained. For each neuron, corresponding activation function is applied. Backward Pass:

In this phase, error between neural network's prediction and actual value is calculated. This error is propagated backward through network, updating weights of each neuron using chain rule to compute gradients. General updating weight is:

$$w_{ij} = w_{ij} - n\frac{\partial L}{\partial w_{ij}} \tag{5}$$

1. w_{ij} represents the weight between neuron i and neuron j.
2. n Is the learning rate, which controls the size of the weight updates.
3. ∂L Is the gradient of the error L for the weight w_{ij}.

3 Results

3.1 Analysis of Physiological Signals

Analysis of the physiological signals in dominance-activation-valence space revealed clear patterns associated with each emotion:

Happiness: High valence, moderate activation, and high dominance.

Sadness: Low valence, low activation, and low dominance.

Fear: Low valence, high activation, and low dominance.

Anger: Low valence, high activation, and dominance.

These patterns align with findings reported in literature, validating effectiveness of proposed approach.

3.2 Performance of the Neural Networks

Both implemented neural networks performed well in classification tasks. Main difference between them is number of iterations required to achieve obtained result.

As shown in Table 2, differences F1-score result are minimal, suggesting that both networks have high classification potential. Now, it is possible to conclude about both models:

Model	F1-score
Linear Neural Network	0.96
Softmax Neural Network	0.99

Table 2. Results of Neural Network Training.

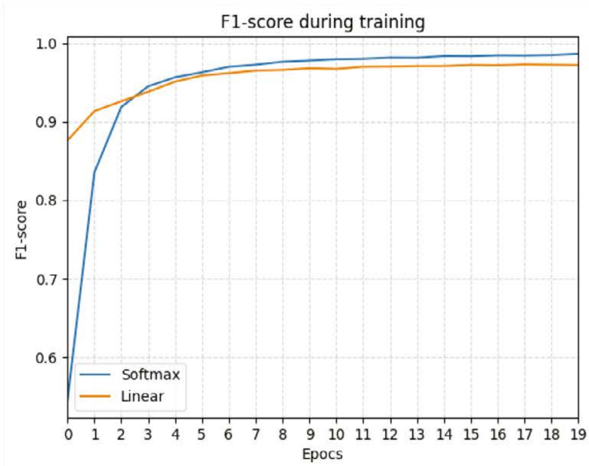

Fig. 4. Evolution of the Softmax neural network and Linear neural network training for validation dataset.

Linear Neural Network: Despite having a larger hidden layer (20 neurons), this model showed similar performance to Softmax Neural Network, suggesting that additional complexity is not necessary for this dataset.

Softmax Neural Network: With a smaller hidden layer (5 neurons) and a Softmax activation function in output layer, this model also achieved exceptional performance, highlighting effectiveness of Softmax function for multiclass classification problems.

Tests are conducted using a wide variety of configurations to determine which setup provided the best performance and optimal learning time. The best training results for both networks were achieved with a batch size of 10, leading to fast training and good performance, as observed in Fig. 4.

To ensure that these results are not merely due to overfitting to the training data, tests were conducted by adding a small perturbation to test data, ranging from -0.1 to 0.1. Such a small noise level is chosen because emotions to be classified are located at short distances in the three-dimensional space. If perturbation exceeded ± 0.2, points belonging to one emotion would cross threshold of the hyperplanes classifying the emotions.

4 Discussion

Experimental results showed high performance for both neural networks, suggesting that three-dimensional space governed by dominance, activation, and valence has a strong capacity to effectively represent emotions based on physiological signals. Additionally, it demonstrates that this representation is effective for training neural networks. Ability of neural networks to achieve a high F1-score value indicates that implemented physiological signals contain sufficient information to discriminate among four emotions studied.

Similarity in the performance of both models suggests that, for the dataset used, simple architecture can achieve good and accurate classification. This is highly relevant for future research, as using simpler architecture makes their implementation in real-time systems more straightforward, where computational efficiency is crucial.

Conclusion

This study shows feasibility of using neural networks to classify emotions based on physiological signals transformed into a dominance-activation-valence space. Results highlight the importance of representing emotions in a multidimensional space and effectiveness of neural networks in learning complex patterns from physiological data. Future research could explore generalization of these models to other datasets or incorporation of additional signals, such as brain activity (EEG) or facial expressions, to further improve classification accuracy.

References

1. Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. Human emotion recognition: Review of sensors and methods. *Sensor*, **20**(3), 592 (2020)
2. Poria, S., Cambria, E., Bajpai, R., Hussain, A.: A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion* **37**, 98–125 (2017).
3. Mehrabian, A., Russell, J. A.: An Approach to Environmental Psychology. *MIT Press*, Cambridge, MA (1974).
4. García Manzo, R. E.: Detección automática de emociones a través de bioseñales y ambientes virtuales. *Tesis de Maestría*, Maestría en Ciencias en Bioingeniería y Cómputo Inteligente, Universidad de Guadalajara (2022).
5. Menezes, M. L. R., Samara, A., Galway, L., Sant'Anna, A., Verikas, A., Alonso-Fernandez, F., Wang, H., Bond, R.: Towards emotion recognition for virtual environments: an evaluation of EEG features on benchmark dataset. *Personal and Ubiquitous Computing* **2017**.
6. Bradley, M. M., Lang, P. J.: The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. *Technical Report B-3*, University of Florida, Gainesville, FL (2007).
7. Lang, P. J., Bradley, M. M., Cuthbert, B. N.: International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual. *Technical Report*, University of Florida, Gainesville, FL (2008).

8. Katsigiannis, S., Ramzan, N.: DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices. *IEEE Journal of Biomedical and Health Informatics* **22**(1), 98–107 (2018).
9. Marín-Morales, J., Llinares, C., Guixeres, J., Alcañiz, M.: Emotion recognition in immersive virtual reality: From statistics to affective computing. *Sensors* **20**(18), 5163 (2020).