Portable Emotion Recognition System Based on EMG Signals and Deep Learning

Paola A. Sanpatín-Días, Paul D. Rosero-Montalvo, Pınar Tözün, and Wilmar Hernandez, *Senior Member, IEEE*

Abstract—One of the biggest challenges of computers is interacting with humans by recognizing and interpreting human emotions. Traditionally, this process is carried out by computer vision or multichannel electroencephalogram emotion recognition techniques. However, they comprise heavy computational resources allocated in the cloud, far from final users or where the dataset was made off. Therefore, portable applications in harsh environments are challenging. Rather, the sensors can capture the reactions of the muscles and respond on the spot, preserving the information locally without constant messages to the cloud. Therefore, this research aims to recognize the six primary human emotions using electromyography sensors. They are placed on specific facial muscles to detect happiness, anger, surprise, fear, sadness, and disgust.The experimental results showed that working with the CortexM0 microcontroller, enough computational capabilities were achieved to store a deep learning model with a classification store of 92%. Furthermore, we demonstrate the necessity to collect data from natural environments and how they need to be processed by a machine learning pipeline.

Index Terms—EMG analysis, sensors, human emotions, emotion recognition, machine learning, and deep learning.

1 Introduction

In recent years, humans and machines have worked closely to understand the human brain and its functionalities to control body movements [1]. Specifically, machines try to recognize human behavior through muscle activity to interact with them [2]. Therefore, facial muscles play a fundamental role in human communication, even with machines, since facial expressions represent a very high percentage of non-verbal communication, where 55% of the message is transmitted through facial expressions, 38% by intonation, and only 7% of the message by linguistic language [3]. Hence, smooth communication between people and machines requires interfaces, which commonly are sensors or cameras. However, cameras use remote sensing techniques to collect data from humans that, in some scenarios, these techniques cannot describe muscle movements when people react to a specific emotion. Indeed, cameras can recognize emotions by applying deep learning techniques comparing thousands of images. These techniques require heavy computational resources, mainly when the deep model applies filters to find patterns that remove unnecessary picture information. Therefore, the Human-machine interaction systems (HMI) by cameras are commonly built in robust computational infrastructures installed on servers [4]. Furthermore, when cameras are deployed in uncontrolled environments, they insert luminosity variations and vibrations that affect the quality of the image and the ability of the machine to recognize human emotions. As a result, cameras struggle to detect the voluntary/involuntary messages produced by the nervous system in response to human emotions [5].

Conversely, modern data acquisition systems collect data from human behavior through electrical measurements of muscle activity. Furthermore, emotion recognition methods (ERM) attempt to recognize facial expressions in different situations to determine the internal emotional state of a person by subjective experiences (i.e., how the person feels). ERM techniques are focused on gathering information from physiological signals, such as: electroencephalogram (EEG), electromyogram (EMG), and electrocardiogram (ECG) [6]. Previous work, such as Ekman et al. [7], proposes the existence of six basic facial emotions: anger, disgust, fear, happiness, sadness, and surprise. These emotions follow a specific physiological response pattern and can be demonstrated by ERM techniques, such as EMG analysis. Therefore, collecting EMG signals need small electrodes (i.e., sensors) to detect light voltage variations in the muscles of the face [8]. Besides, it is important to mention that facial muscles have a single motor unit associated with movements of different facial regions. which contribute to the representation of facial expressions and that is located under the subcutaneous tissue of the face and neck [9].

Therefore, machine learning (ML) methods allow recognizing those light muscle movements to classify them using different mathematical approaches to detect patterns. In this scenario, emerging microcontrollers have enough computational capabilities to collect data from electrodes and run the inference of the ML model deployed in an external server [10]. Therefore, the data is processed close to the final user (tinyML), reducing data travel to the server and power consumption. However, developing small electronic devices

- *Paola A. Sanipatín-Días is with SDAS Research Group, Ben Guerir, Morocco. (e-mail:*
- *Paul D. Rosero-Montalvo and Pınar Tözün are with the Computer Science Department at the IT University of Copenhagen, Denmark. (e-mail: paur@itu.dk and pito@itu.dk)*
- *Wilmar Hernandez is with Facultad de Ingenieria y Ciencias Aplicadas, Universidad de Las Americas, Quito, Ecuador. (e-mail: wilmar.hernandez@udla.edu.ec)*

with sensors also brings open challenges, such as collecting data by humans in an invasive way by putting sensors on the face. It is because importing an external dataset to train ML learning models is challenging, since sensors, sample rates, and pre-processing techniques needed to process EMG signals are usually unknown [11]. Consequently, sensors face replicating similar signals from an external dataset. Therefore, even when building a dataset is also expensive (in terms of time, materials, and volunteers), it is a suitable solution to increase the accuracy of the ML model with less system time-response than cameras, which can use external datasets easily by just resizing images.

Taking into account the above-mentioned statements, this research aims to present an EMG electronic device that can classify the most relevant human emotions locally. Therefore, the system uses three sensors placed on specific muscles of the face to collect data and create a dataset. Then, filters in hardware and software are developed to reduce noise and unpredictable light muscle contractions of the face. Next, three ML approaches are designed to define the suitable ML method with the highest classification accuracy and low computational consumption. In short, the main contributions are as follows:

- An extensive literature review is carried out to select the minimum number of sensors and samples to classify six human emotions, proving that EMG analysis is an adequate alternative in harsh environments where cameras struggle to take high-quality images of humans.
- A proper ML analysis is performed by three approaches to determine the best one with a light workload in the electronic device. Therefore, analog signals are converted into different data structures to fit ML algorithms' training phase.
- An adequate electronic system design combining hardware and software to reach an ML application, keeping a high classification score and less power consumption than cameras.

The rest of the manuscript is structured as follows: Section 2 shows related works and a background of facial muscles. Then, Section 3 presents the electronic design focused on the location of sensors. Afterward, Section 4 describes the machine learning pipeline. Next, the results are discussed in Section 5. Finally, Section 6 presents the conclusions of the paper and future work.

2 Background

This section summarizes earlier surveys covering the intersection between the data acquisition and deep learning techniques used to achieve emotion recognition (Section 2.1). Following, Section 2.2 shows a facial muscle description for a better understanding the location of sensors.

2.1 Early works on EMG in the field of emotion recognition

In detection based computer vision, Thiam et al. [12] have developed their own data classification pipeline, including three cameras to acquire a human emotion database, with results that yielded an accuracy of about 94%. In addition, Perusquia et al. [13] introduced a device for micro-smile recognition using signal processing and neural networks. These works demonstrate that cameras need controlled environments and an external server to deploy deep learning models. On the other hand, early studies such as [6], [14], [15], [16], [17] present emotion recognition approaches with classical classifiers such as Support Vector Machines (SVM) and convolutional neural networks (CNN). They mainly use ECG and EMG signals. Furthermore, recent works such as [18], [19], [20] have presented extensive methods to classify emotion recognition by means of EEG signals, achieving 97% accuracy. However, they use existing data sets to develop their neural network models [21].

In EMG signals space, works such as [8], [22], [23] presented electronic devices to acquire EMG signals from some facial muscles. The studies included SVM and neural networks, with an accuracy of about 91% up to 99% by taking thousand of samples and running experiments on computers. Therefore, the above-mentioned electronic devices are developed just for data collection. Furthermore, in [24], presented a multichannel EMG analysis and a Long and Short Term Memory classifier (LSTM), as a classifier to recognize eight emotions with 95% average accuracy.

The above-metioned works have complex systems, and their application in natural conditions would be challenging since they depend on the number of sensors and human emotions collected. In addition, they might need high computational resources and specific hardware in controlled environments. Therefore, few applications have worked on exporting the ML model to the device, demonstrating that the presented works have low portability. Furthermore, a small number of works use hardware and software filters to reduce the noise and the complexity of the ML model.

2.2 Facial Muscles

The facial muscles are the fibers (elastic tissue) that support sensory organs to perform their activities. For example, chewing, smelling, hearing, or listening. Motor neurons carry out this process by giving instructions to muscles by electrical pulses to maintain muscle contraction. This electrical activity is called motor unit action potential (MUAP), representing the duration, electrical amplitude, and phases of muscle contraction. Typical MUAP duration is between 5 and 15 ms [25]. In this scenario, the facial muscles are divided into three: (1) the top group, which generates eyebrow movement, frowning mainly to allow eye-opening; (2) the middle group, closely involved with mouth movements, helps phonation, and some very particular muscles allow chewing; and (3) the bottom group is responsible for generating facial movements due to the lips' soft tissues [26].

The recognition of facial expressions from physiological signals is generated from the central nervous system through the MUAP analysis. Thus, there are low electrical levels when the muscle is at rest. In contrast, muscle contraction is generated when a nerve impulse raises the electrical levels [27]. Therefore, the facial actions coding system (FACS) indexes the facial expressions in actions units (AU) to match with the biomechanic information obtained in MUAP analysis of EMG signals. Furthermore, this approach aims at splitting the face into individual movements

of their specific muscle activation to describe emotions [28]. Therefore, FACS has intensity scoring by appending letters A–E (for minimal-maximal intensity) to the action unit number. Consequently, the target emotions that FACS helps to determine are as follows: surprise, sadness, happiness, anger, disgust, and fear. Lastly, Table 1 shows the summary of all the muscles people commonly use to express emotions according to FACS.

TABLE 1: Muscles involved in the generation of emotions by FACS [1]

Emotion	Muscular basis	FACS name
Happiness	Orbicularis oculi Zygomaticus major	Cheek raiser Lip corner puller
Anger	Depressor glabellae Depressor supercilii Corrugator supercilii Orbicularis oculi Levator palpebrae superioris	Brow lowerer Upper lid raiser Lid tightener
Surprise	Frontalis Levator palpebrae superioris Masseter Temporal	Inner brow raiser Outer brow raiser Upper lid raiser Jaw drop
Fear	Frontalis Orbicularis oculi Corrugator supercilii Depresor supercilii Levator superioris Risorius Masseter	Inner brow raiser Outer brow raiser Brow lowerer Upper lid raiser Lid tightener Lip stretcher Jaw drop
Disgust	Levator labii superioris Depresor Anguli oris Levator labii Inferioris Mentails	Nose wrinkler Lip corner depressor Chin raiser
Sadness	Frontalis Depressor Anguli oris Corrugator supercilii Depresor superciliar	Inner brow raiser Brow lowerer Lip corner depressor

3 Electronic Design

Section 3.1 mentions setting sensors for the data acquisition process. Then, the electronic system is described in Section 3.2.

3.1 Sensors' Location

EMG sensors are composed of electrodes that can recognize electricity fluctuations in muscles when humans express emotions. In addition, EMG sensors are commonly built with three electrodes; two electrodes are in charge of detecting the electricity variations; they can be placed on the top or in the middle of the muscle [29]. In this research, one electrode is placed on the top of the muscle, other electrode is used as the ground [30]. In the human face, voltage reference are bones, since they do not receive electricity from neurons.

Table 1 shows that emotional expressions share the use of some muscles. With this information, three EMG sensors are deployed to develop the electronic device. One sensor is located between the *Orbicularis oculi* muscle and the *Corrugator supercilii* muscle. These muscles are in the forehead's upper part of the eyebrow. It is the primary muscle over the negative emotions of sadness, anger, and disgust (sensor 1). Another sensor is placed on the *Zygomaticus major* muscle, which is located in the cheek near the mouth.

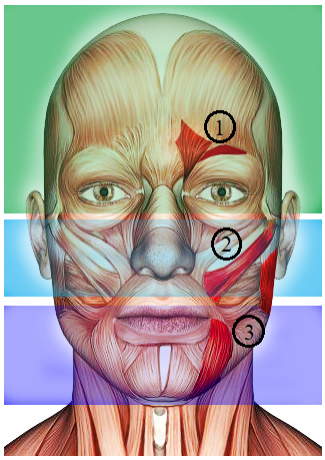


Fig. 1: Location of three EMG sensors to collect data. Sensor 1: Orbicularis oculi and the Corrugator supercilii muscles. Sensor 2: Zygomaticus major. Sensor 3: Depressor Anguli Oris and Masseter muscles. Green section: top face's muscles. Blue section: middle face's muscles, and Purple section: bottom face's muscles

This muscle has a high activity on emotions of happiness, fear, and surprise (sensor 2). Finally, the last sensor collects data from the *Depressor Anguli Oris* and *Masseter* muscles. Both muscles are in the lower part of the mouth towards the jaw and act mainly on the emotions of surprise, fear, happiness, and sadness(sensor 3) [28]. In short, Fig 1 shows the graphical description for sensors' location[i].

3.2 Electronic system description

Once the location of each sensor is defined, we select new-brand sensors to be deployed into the portable electronic device. Therefore, according to hardware, functionalities, accuracy, stability, and replicability requirements, MyoWare muscle biosensors were selected to collect EMG signals. These sensors are placed in a bipolar configuration with at least 2 cm of separation between them to avoid noise by electrical interference. Besides, they have an instrumentation amplifier internally with a gain of 200x, a bandpass filter with lower 3 dB cutoff frequency 20Hz and upper 3 dB cutoff frequency 498Hz, and a rectification circuit. Then, to avoid the sensor sending a raw signal directly to the microcontroller, an extra amplifier is used to couple the signal to the analog-digital converter and reduce noise for unpredictable facial movements. This task is done using a voltage follower amplifier. However, even when the signal is rectified, some peaks represent DC components in the signal. Therefore, filters in software are needed. Consequently, Arduino nano IoT is defined as a microcontroller to receive EMG signals from sensors. Next, the Arduino sends those signals to the computer to train ML models and export the inference to the microcontroller to make decisions locally. Fig. 2 shows the above-mentioned electronic system design.

i. *Figure 1 was purchased in: depositphotos and modified according to the interests of the authors*

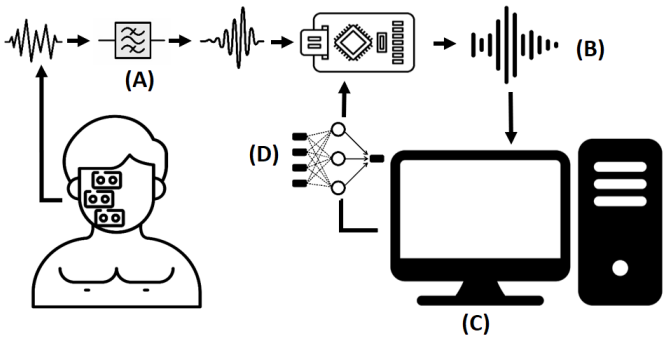


Fig. 2: Electronic system design. Step A: sensors gather data by employing analog filters. Step B: The microcontroller receives the EMG signals and converts them into digital signals to send to the computer. Step C: The computer stores the EMG signals to train ML models, and step D: The inference is allocated to the microcontroller.

4 Machine Learning Pipeline

Sensors usually exhibit noise and drift, which is often infeasible to remove or model statistically. Therefore, ML models should be robust enough that this noise or drift does not adversely impact them during regular operations. In a similar vein, sensors that are not adequately isolated electrically can have the potential to distort other measurements. Thus, the development of a machine learning model needs to follow a multi-step design methodology, commonly referred to as the machine learning lifecycle or pipeline. In this section, we covered the data collection, data pre-processing, and model design. Then, the following section mentioned how to evaluate the model, optimize and, finally, export the model to the electronic device.

4.1 Data Collection

For conducting the data collection stage, 30 people (i.e., 18 men and 12 women) used cleaning/shaving products to remove the beard, long sideburns, and make-up. Then, new electrodes were put on the face of each person. Next, EMG signals were collected at a sampling frequency of 25 Hz for 4 seconds, and then people rested for 10 seconds between experiments to eliminate the effect of muscle fatigue. This process was performed ten times where people intentionally made each emotion by their criteria, representing 300 instances per expression. Afterward, they were exposed to images to express emotions naturally under the Then, the following section mentioned how to evaluate the model, optimize and, finally, export the model to the electronic device. We use a Lenovo Thinkpad P50s laptop system to store data whose specifications are listed in Table 2.

4.2 Data Pre-processing

Once EMG signals were collected and stored in the computer, the dataset was built, showing the six types of emotional expressions: happiness, anger, surprise, fear, disgust, and sadness. Each sample is a three-dimensional array (i.e., 100 instances per expression with 3 EMG sensors [3][100]), representing 3600 samples with their respective

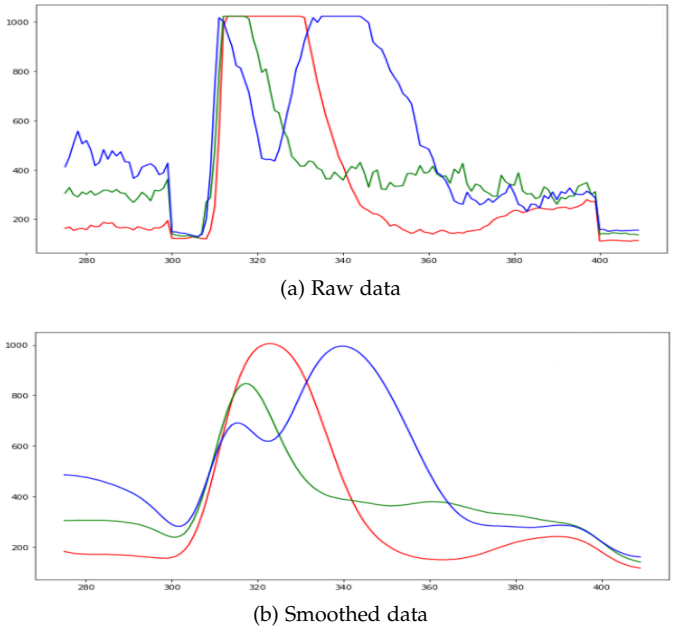
TABLE 2: Lenovo Thinkpad P50s System Specifications

Feature	Description
Processor	Model: **Intel Core i7-6500U (9th Gen)** Speed: **2.5 GHz** Cache: **4 MB Intel® Smart Cache** Instruction Set: **64-bit**
Memory	Type: **DDR3L-1600** Speed: **1600 MHz** Capacity: **32 GB**
Storage	type: **SATA HDD** Speed: **5400 RPM** Capacity: **1 TB**
GPU	NVIDIA Quadro M500M Memory: **2 GB**
Operating System	**64-bit** Windows 10 Professional Edition

labels. However, when these samples are plotted, it is evident that some noise components were injected into the analog signal. Therefore, smoothing filters in software are deployed to improve the signal. The signal-to-noise (SNR) metric is used to determine the suitable smooth algorithm, the results are represented in Table 3 [31], [32]. Also, this Table shows the statistical description of the signal, such as mean, standard deviations, and variability coefficient. Thus, a Gaussian filter is selected to smooth EMG signals. Fig. 3 demonstrates how the Gaussian filter smooths the signal and reduces the noise. Finally, Fig 4 shows each emotion with its corresponding EMG smoothed signals.

TABLE 3: SNR analysis applied smoothing algorithms into the analog signals

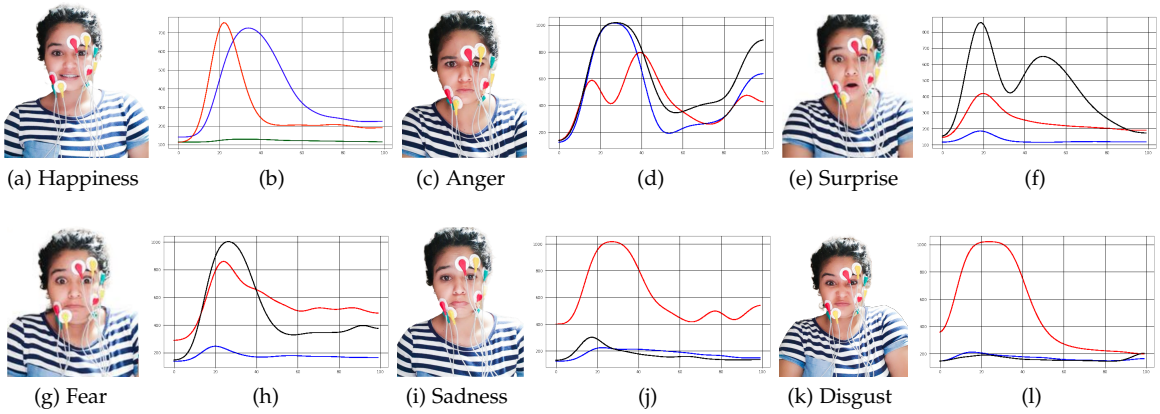
Parameters	Original Signal	FILTERS			
		Media Mobile	Moving Average	Savitzky	Gaussian
Mean	145.50	221.04	218.08	220.86	220.86
Standard deviation	78.31	93.62	99.28	98.19	89.30
SNR	1.85	2.36	2.19	2.24	2.47
Variability coefficient	0.53	0.42	0.45	0.44	0.40



(a) Raw data

(b) Smoothed data

Fig. 3: Anger representation by EMG signals, Y-axis: Analog-digital converter resolution, and X-axis: sample size. Sensor 1: (–), Sensor 2 (–), and sensor 3: (–)



(a) Happiness (b) (c) Anger (d) (e) Surprise (f)

(g) Fear (h) (i) Sadness (j) (k) Disgust (l)

Fig. 4: Emotion analysis by EMG signals. Y-axis: Analog-digital converter resolution, and X-axis: sample size. Sensor 1: (–), Sensor 2 (–), and sensor 3: (–)

4.3 Model Design

Since the dataset is a multidimensional array, several ML approaches could be trained to classify human emotions. Therefore, we design an ML pipeline to test different ML models with their corresponding data representation. Then, according to the computational limitations in microcontrollers, such as memory size and power consumption, the suitable ML model is selected.

4.3.1 Supervised classification algorithms

Classification algorithms can learn using different mathematical approaches [33]. In consequence, reviewed works mention the most usual classification algorithm methods based on:

- Distance-based: K-Nearest Neighbors (KNN)
- Model-based: Support Vector Machine (SVM)
- Density-based: Bayesian classifier (BC)
- Heuristic: Decision Tree (DT)

Afterward, the dataset is reshaped since the classification above-mentioned algorithms have as inputs one-dimensional arrays. Therefore, Fig 5 shows how the samples were reshaped into one-dimensional arrays.

4.3.2 Auto encoders

This approach is focused on generating synthetic data from known samples. This is achieved by learning the data distribution and drawing random samples from the latent space of the generative model. Therefore, autoencoders have two stages; first, the encoder compresses data into a low representation extracting the most representative features. Second, the decoder reconstructs data, denoising information. In this scenario, long short-term memory networks (LTSM) are a type of recurrent neural network. They are helpful in processing longer sequences while capturing the temporal structure of the data, like the time-series data. This model was applied in the dataset to obtain cleaned-synthetic data to determine the distribution of each label to classify incoming data. However, even when the number of layers were increased, the LTSM could not replicate samples correctly. It happens since even when the LTSM needs few samples to train the model, the complexity of the dataset requires more samples than the one the dataset already has. Therefore, this approach is inadequate for classifying emotions. Lastly, Fig. 6 shows the synthetic samples obtained by the LTSM network.

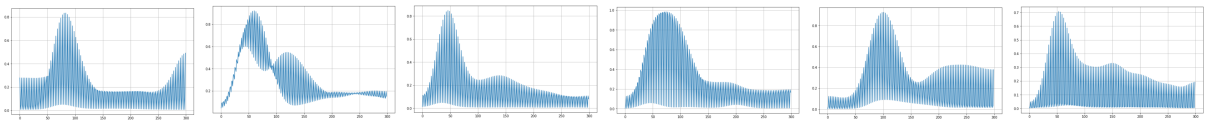
4.3.3 Neural networks

Neural networks learn to extract relevant features when data go through their architecture. Then each neuron updates its weight to create rules. Therefore, they are capable of learning from time series, sounds, or images. In this case, given that the dataset is multidimensional and in the previous step was reshaped, two neural network models were deployed with similar structures but different input shapes. The first model is shown in Table 4 and it is called DP_1. The DP_1 architecture uses a sequential model and dense layers, which means each layer is deeply connected with its preceding layer. In addition, "Rectifier Linear Unit" (ReLU) is used for sparse activation of neurons, efficient computation, and better gradient propagation. Furthermore, a flatten layer is used to obtain a flattened output. Finally, a dense layer with six neurons and softmax as activation function is added. Applying softmax makes the output vector in the interval $I = [0, 1]$ such that the components will add up to 1. Therefore, they can be interpreted as probabilities.

TABLE 4: Proposed multidimensional architecture of the neural network (DP_1)

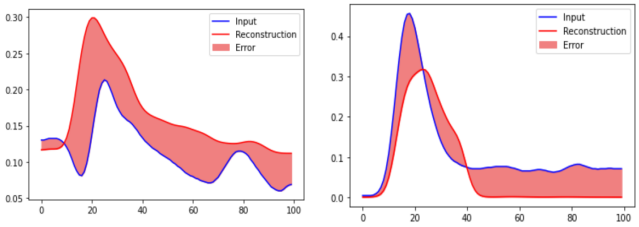
Layer (Type)	Output Shape	# Param
Input (Dense)	(None,100,80)	320
Layer 1 (Dense)	(None,100,40)	3240
Layer 2 (Dense)	(None,100,20)	820
Layer 3 (Flatten)	(None,2000)	0
Output (Dense)	(None,6)	12006

Consequently, the second neural network model called DP_2 uses a similar architecture to the first model except for the flatten layer since the input is a one-dimensional array (Table 5). It is worth mentioning that both neural network models do not add a drop layer because we carefully make some tests to avoid over-fitting and reduce the model complexity.



(a) Happiness (b) Anger (c) Surprise (d) Fear (e) Sadness (f) Disgust

Fig. 5: Samples reshaped into one-dimensional array



(a) Raw data (b) Smoothed data

Fig. 6: synthetic samples obtained by the LTSM network. Input: (–), Reconstruction: (–)

TABLE 5: Proposed one-dimensional architecture of the neural network (DP_2)

Layer (Type)	Output Shape	# Param
Input (Dense)	(None,100)	30100
Layer 1 (Dense)	(None,50)	5050
Layer 2 (Dense)	(None,25)	1275
Output (Dense)	(None,6)	156

5 Results

This section shows both the evaluation and optimization of the ML and the electronic device that was designed.

5.1 Evaluation of ML models

Evaluating their results is the next step, given that all the ML models were trained. Therefore, we used the confusion matrix to obtain classification metrics, such as: precision, recall, F1-score, and accuracy. Furthermore, it is important to remember that the dataset is described in different forms. First, the dataset is represented as a multidimensional array to train deep learning models. Second, the dataset was reshaped into a one-dimensional array to train supervised classification algorithms. Consequently, Table 1 shows the metrics mentioned above. Indeed, it is clear that KNN and SVM are not suitable techniques to classify emotions since their performance is the lowest. Furthermore, both techniques struggled with especially recognizing surprise and disgust. Conversely, Naive Bayes and Decision Tree have high scores in classifying emotions even when the precision in detecting disgust is below 90%.

Related to deep learning models, the DP_1 model had a similar performance to classification algorithms (i.e., accuracy is about 93%). However, DP_2 proved to be the model with the highest classification performance. Also, this model has high scores in recall and precision. As a result, Naive Bayes, Decision Tree, and Neural network models (i.e., DP_2) are quantized and exported to the electronic device to test them in a real environment, which means predicting the label in incoming data.

TABLE 6: Evaluation of ML classifation models. Precision (Prec.), Recall (Rec.), F1-score (F1-sco.), and Accuracy (Acc.)

ML model	Classification metrics				
	Prec.	Rec.	F1-sco.	Error	Acc.
SVM	0.9154	0.9166	0.9154	0.1	0.92
kNN	0.8296	0.8166	0.8153	0.25	0.82
Desicion Tree	0.9384	0.9333	0.9334	0.13	0.93
Naive Bayes	0.9718	0.9666	0.9667	0.06	0.97
DP1	1.00	1.00	1.00	0.0	1.0
DP2	0.9436	0.9333	0.9319	0.06	0.93

5.2 Model optimization and deployment

The electronic device has been tested to prove its decision capability. The system collected data from five women and ten men who were not part of the training set. Thus, the system does not have previous information about their facial movements. Each participant expressed different emotions reacting to different stimuli. First, the participant decides the emotion to express by himself/herself, which we call artificial emotions. Second, the participant sees pictures to express emotions naturally. Then, the decision tree and the DP_1 models were exported into the electronic device since Naive Bayes does not have proper support in microcontrollers for large arrays. On the one hand, a decision tree is a lightweight algorithm that consumes only 4 Kbytes of Flash and 2 Kbytes of RAM and takes the decision in 2 seconds. However, given that its rules are stringent, this model depends on the specific location of sensors on the face, and light movements affect the EMG signal. Furthermore, whether one electrode is inadequately connected to the sensor or is worn, it injects noise into the signal and affects decision-making. On the another hand, even when the neural network model was quantized before being exported into the microcontroller, it consumes 150 Kbytes of Flash, needs 10 Mbytes of RAM, and takes the decision in 8 seconds. Nevertheless, the model can stand the noise and drift.

Following the planned test, in the first scenario (i.e., people emulate emotions), the system recognized emotions with an accuracy of 82%, considering that the system had trouble recognizing sadness, disgust, and fear. In the second scenario (i.e., people react to stimuli), the accuracy increased to 92%. The neural network model obtained those results due to its robustness. Pitifully, the decision tree model did not obtain high classification scores.

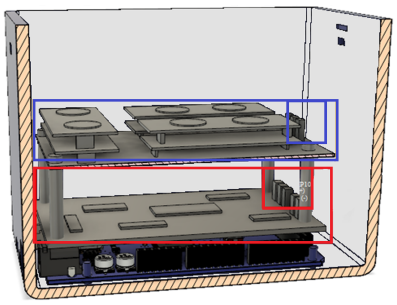
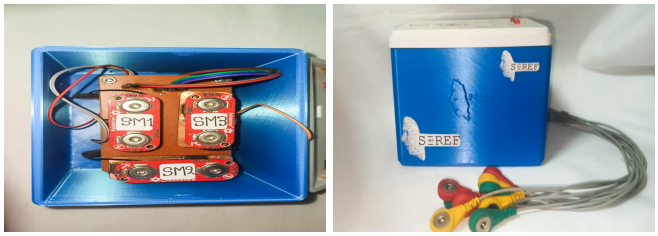


Fig. 7: Digitalized electronic system. Blue rectangle: sensors' board, small blue and red boxes: Jumpers, and Red rectangle: noise filter board allocated over the microcontroller

5.3 Electronic device

Once the neural network model is selected, its inference is exported into the microcontroller. The next step is designing the PCB board and case. Therefore, the electronic system was digitalized to simulate all the connections between the microcontroller, noise filter board, and sensors. Based on building a robust electronic case, we decided that design the PCB board in levels (i.e., racks) where the first level is the microcontroller and power supply, and the second level comprises the analog amplifiers to reduce the noise of the EMG signal. On the third level stands the sensor boards. Lastly, jumpers in the second and third levels allow for transferring the EMG signals to the microcontroller and the voltage needed for sensors. Fig. 7 shows the digitalized electronic system. In addition, Fig. 8 shows the electronic case made in a 3D printer. Furthermore, on top of the case, there are LEDs to inform when the electronic system is collecting data and making decisions. A USB connection powers the system and can send data to the computer.



(a) Sensors' PCB board (b) Electronic case with electrodes

Fig. 8: Emotion recognition electronic case

6 Conclusions and Future Works

In this paper, the proposed emotion recognition system by sensors and ML methods, which integrates an ML pipeline and electronic system design, achieves excellent performance in classifying the six principal human emotions by using EMG sensors. In addition, it was shown that a deep learning model with a multidimensional array as input is a suitable solution compared to classification methods. Additionally, this model was quantized to be compiled into the microcontroller to reduce sending information to the cloud. Therefore, the electronic system is portable and can send information to computers with low computational resources since the model does not need to be updated. Furthermore, The ML pipeline shows an adequate data collection step with the correct allocation of EMG sensors on the face using the facial actions coding system, which describes the specific facial muscle activation. Therefore, the electronic system was built with its case with a noise cancellation board between sensors and the microcontroller.

Moreover, the deep learning model can correctly recognize over 92% of human emotions in women and men. Besides, it demonstrated that artificial emotions without stimuli could be detected with sensors and shows that natural facial expressions are standardized in people. Finally, tests in real scenarios prove that in electronic systems with sensors, data collection is necessary to fine-tune models even when it is expensive (i.e., in terms of cost of the materials and time).

This research provides a new perspective on recognizing human emotions with sensors in harsh environments where cameras are considered invasive to humans or have trouble with vibrations. In future work, the proposed electronic system will be further expanded and optimized to work with a few samples of people who have low face mobility or facial muscle limitations in medical treatments.

Acknowledgment

This research was supported in part by Novo Nordisk Fonden (award ref: NNF20OC0064411), in part by the Corporacion Ecuatoriana para el Desarrollo de la Investigacion y la Academia (CEDIA), Ecuador (Research project: CEPRA XII-2018-13), and in part by the Universidad de Las Americas (UDLA), Quito, Ecuador (Research project: IEA.WHP.21.02).

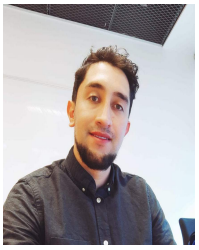
References

[1] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions," vol. 21, p. 974–989, Oct. 1999.
[2] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using eeg signals: A survey," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 374–393, 2019.
[3] H. I. Dino and M. B. Abdulrazzaq, "Facial expression classification based on svm, knn and mlp classifiers," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, pp. 70–75, 2019.
[4] A. Ghods and D. J. Cook, "A survey of deep network techniques all classifiers can adopt," *Data mining and knowledge discovery*, vol. 35, no. 1, pp. 46–87, 2021.
[5] M. Degirmenci, M. A. Ozdemir, R. Sadighzadeh, and A. Akan, "Emotion recognition from eeg signals by using empirical mode decomposition," in *2018 Medical Technologies National Congress (TIPTEKNO)*, pp. 1–4, 2018.
[6] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, 2020.
[7] P. Ekman, "Universal facial expresions of emotion,"
[8] Y. Cai, Y. Guo, H. Jiang, and M.-C. Huang, "Machine-learning approaches for recognizing muscle activities involved in facial expressions captured by multi-channels surface electromyogram," *Smart Health*, vol. 5-6, pp. 15–25, 2018. WEARABLE SENSOR SIGNAL PROCESSING FOR SMART HEALTH.
[9] S. Chen, Z. Gao, and S. Wang, "Emotion recognition from peripheral physiological signals enhanced by eeg," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2827–2831, 2016.

[10] K. Fukano, K. Iiazawa, T. Soeda, A. Shirai, and G. Capi, "Deep learning for gesture recognition based on surface emg data," in *2021 International Conference on Advanced Mechatronic Systems (ICAMechS)*, pp. 41–45, 2021.
[11] A. J. Young, L. H. Smith, E. J. Rouse, and L. J. Hargrove, "Classification of simultaneous movements using surface emg pattern recognition," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 5, pp. 1250–1258, 2013.
[12] P. Thiam, V. Kessler, M. Amirian, P. Bellmann, G. Layher, Y. Zhang, M. Velana, S. Gruss, S. Walter, H. C. Traue, D. Schork, J. Kim, E. André, H. Neumann, and F. Schwenker, "Multi-modal pain intensity recognition based on the senseemotion database," *IEEE Transactions on Affective Computing*, vol. 12, no. 3, pp. 743–760, 2021.
[13] M. Perusquía-Hernández, M. Hirokawa, and K. Suzuki, "A wearable device for fast and subtle spontaneous smile recognition," *IEEE Transactions on Affective Computing*, vol. 8, no. 4, pp. 522–533, 2017.
[14] Z. Guendil, Z. Lachiri, C. Maaoui, and A. Pruski, "Multiresolution framework for emotion sensing in physiological signals," in *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 793–797, 2016.
[15] P. S. Ghare and A. Paithane, "Human emotion recognition using non linear and non stationary eeg signal," in *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*, pp. 1013–1016, 2016.
[16] J. Shin, J. Maeng, and D.-H. Kim, "Inner emotion recognition using multi bio-signals," in *2018 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia)*, pp. 206–212, 2018.
[17] X.-h. Wang, T. Zhang, X.-m. Xu, L. Chen, X.-f. Xing, and C. L. P. Chen, "Eeg emotion recognition using dynamical graph convolutional neural networks and broad learning system," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1240–1244, 2018.
[18] D. Kollias and S. Zafeiriou, "Exploiting multi-cnn features in cnn-rnn based dimensional emotion recognition on the omg in-the-wild dataset," *IEEE Transactions on Affective Computing*, vol. 12, no. 3, pp. 595–606, 2021.
[19] Y. Zhao, J. Yang, J. Lin, D. Yu, and X. Cao, "A 3d convolutional neural network for emotion recognition based on eeg signals," in *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6, 2020.
[20] T. Song, S. Liu, W. Zheng, Y. Zong, Z. Cui, Y. Li, and X. Zhou, "Variational instance-adaptive graph for eeg emotion recognition," *IEEE Transactions on Affective Computing*, pp. 1–1, 2021.
[21] G. Li, D. Ouyang, Y. Yuan, W. Li, Z. Guo, X. Qu, and P. Green, "An eeg data processing approach for emotion recognition," *IEEE Sensors Journal*, vol. 22, no. 11, pp. 10751–10763, 2022.
[22] S. A. Mithbavkar and M. S. Shah, "Recognition of emotion through facial expressions using emg signal," in *2019 International Conference on Nascent Technologies in Engineering (ICNTE)*, pp. 1–6, 2019.
[23] P. Das, J. Bhattacharyya, K. Sen, and S. Pal, "Assessment of pain using optimized feature set from corrugator emg," in *2020 IEEE Applied Signal Processing Conference (ASPCON)*, pp. 349–353, 2020.
[24] S. A. Mithbavkar and M. S. Shah, "Analysis of emg based emotion recognition for multiple people and emotions," in *2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*, pp. 1–4, 2021.
[25] D. C. Preston and B. E. Shapiro, "15 - basic electromyography: Analysis of motor unit action potentials," in *Electromyography and Neuromuscular Disorders (Third Edition)* (D. C. Preston and B. E. Shapiro, eds.), pp. 235–248, London: W.B. Saunders, third edition ed., 2013.
[26] H. Parsaei and D. W. Stashuk, "Emg signal decomposition using motor unit potential train validity," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 2, pp. 265–274, 2013.
[27] P. Ekman and E. L. Rosenberg, "What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)," *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, pp. 1–672, mar 2012.
[28] P. Lewinski, T. M. Den Uyl, and C. Butler, "Automated facial coding: Validation of basic emotions and FACS AUs in facereader," *Journal of Neuroscience, Psychology, and Economics*, vol. 7, pp. 227–236, dec 2014.
[29] A. Turgunov, K. Zohirov, R. Nasimov, and S. Mirzakhalilov, "Comparative analysis of the results of emg signal classification based on machine learning algorithms," in *2021 International Conference on Information Science and Communications Technologies (ICISCT)*, pp. 1–4, 2021.
[30] O. Ahmed and A. Brifcani, "Gene expression classification based on deep learning," in *2019 4th Scientific International Conference Najaf (SICN)*, pp. 145–149, 2019.
[31] P. D. Rosero-Montalvo, V. F. López-Batista, and D. H. Peluffo-Ordóñez, "A new data-preprocessing-related taxonomy of sensors for iot applications," *Information*, vol. 13, no. 5, 2022.
[32] P. Kowalski and R. Smyk, "Review and comparison of smoothing algorithms for one-dimensional data noise reduction," in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, pp. 277–281, 2018.
[33] P. D. Rosero-Montalvo, E. A. Fuentes-Hernández, M. E. Morocho-Cayamcela, L. M. Sierra-Martínez, and D. H. Peluffo-Ordóñez, "Addressing the data acquisition paradigm in the early detection of pediatric foot deformities," *Sensors*, vol. 21, no. 13, 2021.



Paola A. Sanipatín-Días is a research engineer at Smart Data Analysis System Group (SDAS), Ben Guerir, Morocco. She received her engineering degree in electronics from the Technical University of the North, Ecuador. Her research focuses on embedded systems, machine learning, deep learning, and electronic design.



Paul D. Rosero-Montalvo He is a Post. Doc at IT University of Copenhagen (ITU) in Denmark since June 2021. His research focuses on emerging microcontrollers to run machine learning models in decentralized networks. Earlier, He was a research assistant professor in the Applied Science Department at the Universidad Tecnica del Norte (UTN), Ecuador, for seven years. At the same time, he was a part-time lecturer in the TI Department at Instituto Tecnologico 17 de Julio, Ecuador. He received his Ph.D. from the University of Salamanca in Spain in November 2020, where Prof. Vivian Lopez-Batista advised him. He has a master's degree in Data Management systems from Universidad de las Fuerzas Armadas ESPE, Ecuador (2018), and an Engineering degree in Electronics and Telecommunications from UTN (2013).



Pınar Tözün is an Associate Professor at IT University of Copenhagen (ITU) in Denmark since 2018. Her research focuses on hardware-conscious machine learning, performance characterization of data-intensive systems, and scalability and efficiency of data-intensive systems on modern hardware. Before joining ITU, she was a research staff member at IBM Almaden Research Center (San Jose, CA, USA) for three years, where she was part of the team that led the development of IBM Db2 Event Store. Before joining IBM, she received her PhD from École polytechnique fédérale de Lausanne (EPFL) in Switzerland in 2014. Her PhD thesis received ACM SIGMOD Jim Gray Doctoral Dissertation Award Honorable Mention in 2016.



Wilmar Hernandez received the degree in electronics engineering and the specialist degree in microelectronics from the Instituto Superior Politecnico Jose Antonio Echeverria (ISPJAE), Havana, Cuba, in 1992 and 1994, respectively, and the M.S. degree in signal treatment and the Ph.D. degree in electronic engineering from Enginyeria La Salle, Universitat Ramon Llull, Barcelona, Spain, in 1997 and 1999, respectively. From 1992 to 1995, he was a Lecturer with the Electrical Engineering Faculty, ISPJAE, and a Researcher with the Microelectronics Research Center, ISPJAE. From 1999 to 2003, he was with the Department of Electronics and Instrumentation, University Institute for Automobile Research, Universidad Politecnica de Madrid (UPM), Spain, where he was the Technical Director of the department, from January 2003 to January 2004. From January 2004 to March 2013, he was an Associate Professor of circuits and systems with the Department of Circuits and Systems, EUIT de Telecomunicacion, UPM. From September 2014 to September 2015, he was a Researcher with SENESCYT, Ecuador, under the Prometeo Fellowship Program. From December 2015 to November 2017, he was a Professor with the Universidad Tecnica Particular de Loja, Ecuador. Since January 2018, he has been a Professor with the Universidad de Las Americas, Ecuador.