Cognitive Computation (2025) 17:109
https://doi.org/10.1007/s12559-025-10458-6

RESEARCH

Emotions for Everyone: A Low-Cost, High-Accuracy Method for Emotion Classification

Nabil I. Ajali-Hernández[1] · Carlos M. Travieso-González[1]

Received: 22 January 2025 / Accepted: 26 April 2025

Abstract
The classification of emotions is of vital importance in health care, particularly in the context of early detection of cognitive disorders. Given the critical role of emotions as early indicators of cognitive health, this study addresses the need to develop effective and accessible classification methods. In this research, we present an innovative approach to emotion classification using a proprietary dataset and harnessing the power of deep learning. In particular, we use a specific, innovative combination of attentional layers and Long-Short Term Memory (LSTM) algorithms to achieve an emotion classification. A key differentiator of our methodology is the use of a compact and low-cost array of biometric sensors. This approach provides a cost-effective alternative to traditional systems, which often rely on more complex and expensive sensor arrays, such as those using electroencephalography (EEG). Despite the affordability of our sensor configuration, our classification model achieves an outstanding accuracy rate of 93.75%. This performance not only demonstrates the effectiveness of our method but also positions it at the forefront of emotion classification using these sensors. By significantly reducing cost while increasing classification accuracy, our method helps to push the boundaries of current state-of-the-art approaches and provides a novel and cost-effective solution for emotion classification and cognitive health monitoring.

Keywords Emotion recognition · Cognitive health · Emotional classification · Biometric sensors · Deep learning

Introduction

In the ever-evolving landscape of human–machine interaction, emotion recognition has emerged as a critical frontier that transcends disciplinary boundaries and has profound implications across multiple domains. The historical trajectory of emotion recognition has been and continues to be, deeply rooted in societal advances, technological innovations, and our evolving understanding of human cognition, since we are emotions [1]. From its earliest stages, when psychological theories laid the groundwork for understanding emotions, to the present era of unprecedented technological advances, the pursuit of deciphering human emotions has become a compelling imperative.

The beginnings of emotion recognition can be traced back to early psychological paradigms, such as Paul Ekman's work on facial expressions and their universality in the 1970 s. In subsequent decades, the integration of sensor technologies, particularly electroencephalography (EEG), has provided unprecedented insights into the neurophysiological underpinnings of emotions [2, 3]. This integration has not only refined our understanding of emotional states but also expanded the horizons of potential applications, ranging from mental health diagnostics to human–computer interaction [4].

The advent of machine learning, especially deep learning techniques, has exponentially increased the accuracy and scope of emotion recognition. The interplay of sophisticated algorithms ranging from Support Vector Machines (SVMs), k-neighbors, or perceptrons to Neural Networks (NNs), Convolutional Neural Networks (CNNs), attention layers, or Long Short-Term Memory (LSTM) networks has redefined the landscape [5–9]. This has enabled a nuanced analysis of emotional states from multiple data modalities.

Some examples of the use of these algorithms in emotion classification are speech recognition, which has ushered in a new era of multimodal emotion recognition, where the fusion of acoustic features and linguistic patterns contributes

✉ Nabil I. Ajali-Hernández
nabil.ajali101@alu.ulpgc.com

[1] Signals and Communications Department (DSC), University of Las Palmas de Gran Canaria, Campus Universitario de Tafira, Las Palmas de Gran Canaria 35017, Spain

Published online: 26 May 2025

to a more holistic understanding of emotional expressions [6]. EEG, where brain signals combined with algorithms have enabled a better understanding of brain patterns and improved emotion recognition and classification [2–4].

The fusion of image recognition methods with artificial intelligence has also received great recognition in recent years, such as the one carried out by the University of Santa Monica together with the company Snapchat [10]. Or the more economical search for the application of a set of IoT sensors that provide the information to understand and configure a set of tools capable of providing valuable information to the algorithms that classify emotions [11].

Related Works

This article aims to explore the current state of the art in emotion recognition, as we consider it a key aspect in the detection of cognitive disorders and also a useful tool in working with all kinds of people with social difficulties (autism, Asperger's). This approach focuses on the use of sensors and deep learning algorithms to predict emotions with high accuracy.

There are many approaches in the literature, ranging from speech recognition analysis to image analysis, using multimodal approaches where several of these techniques are mixed with others such as EEG.

Here, thanks to work such as that developed by Khare et al. [12] and others, we try to highlight the transformative potential of these technologies, highlighting their applications in different domains, see Fig. 1, and predicting future trajectories of emotion recognition research, as the development of computational algorithms and artificial intelligence has allowed work on emotion recognition and classification to multiply in recent decades.

Using the speech recognition approach, Zielonka et al. [13] used voice sensors as an input method to perform emotion recognition (anger, disgust, fear, happiness, sadness, and neutral) and achieved a 76% success rate in their classification. Subsequently, Anvarjon et al. [14] improved this approach to emotion recognition in speech and achieved a range of 77.01–92.02%. The proposed CNN model was trained on the frequency features extracted from the speech data and then tested to predict the emotions (anger, happiness, sadness, and neutral).

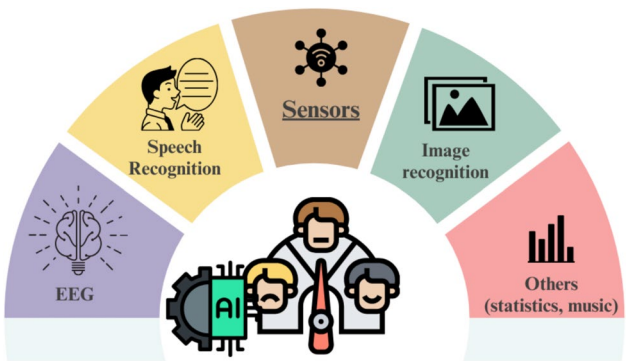

Fig. 1 Paradigms used in emotion classification and recognition

As mentioned above, another common approach is to extract features from images. This technique can be used alone or in combination with sensor-based or EEG-based techniques. For example, Yadav et al. [15] use facial images to detect different emotions (sadness, surprise, happiness, anger, disgust, fear, and neutral). This is done by extracting covariance matrices and feature vectors from the facial images and then applying Principal Component Analysis (PCA), Gaussian Mixture Models (GMM), and Grey Level Co-occurrence Matrices (GLCM). Finally, they used an SVM to classify the emotions. The results vary between 87.7 and 93%.

In their study, Hassouneh et al. [16] also focus on the line of feature extraction from facial images. However, in this study, they rely on the use of EEG to complement the results and try to add value using the brainwaves. They aim to classify the emotional expressions of children with autism and physically disabled people (mute, deaf, and bedridden) using facial landmarks and EEG signals. CNN and LSTM networks are used and implemented in a real-time emotion recognition algorithm. This algorithm uses virtual markers through an optical flow algorithm that works effectively in non-uniform lighting and subject head rotation (up to 25°), different skin tones, and different backgrounds. After carrying out the experiments, the results show an average accuracy of 87.25%.

The confluence of emotion recognition and electroencephalography (EEG) has led to promising advances in joint applications. Recent studies such as those by Iacoviello et al. [17] and Yang et al. [18] exemplify this synergy, applying machine learning and deep learning to brain signals to classify emotions such as fear, happiness, and disgust with success rates ranging from 84.2 to 90.2%. All this work from around the world has led to major advances in the field of emotion recognition. This is illustrated in Fig. 2.

However, these approaches are often limited by cost, requiring complex setup and significant computational resources. In this work, we present a contrasting approach that prioritizes low-cost, high-value systems. We achieve this by using inexpensive and readily available sensors, coupled with the power of deep learning to uncover predictive patterns, ultimately aiming for a favorable cost–benefit ratio.

Our Proposal

This study presents a novel method for recognizing fear and disgust by integrating biometric sensors with a deep

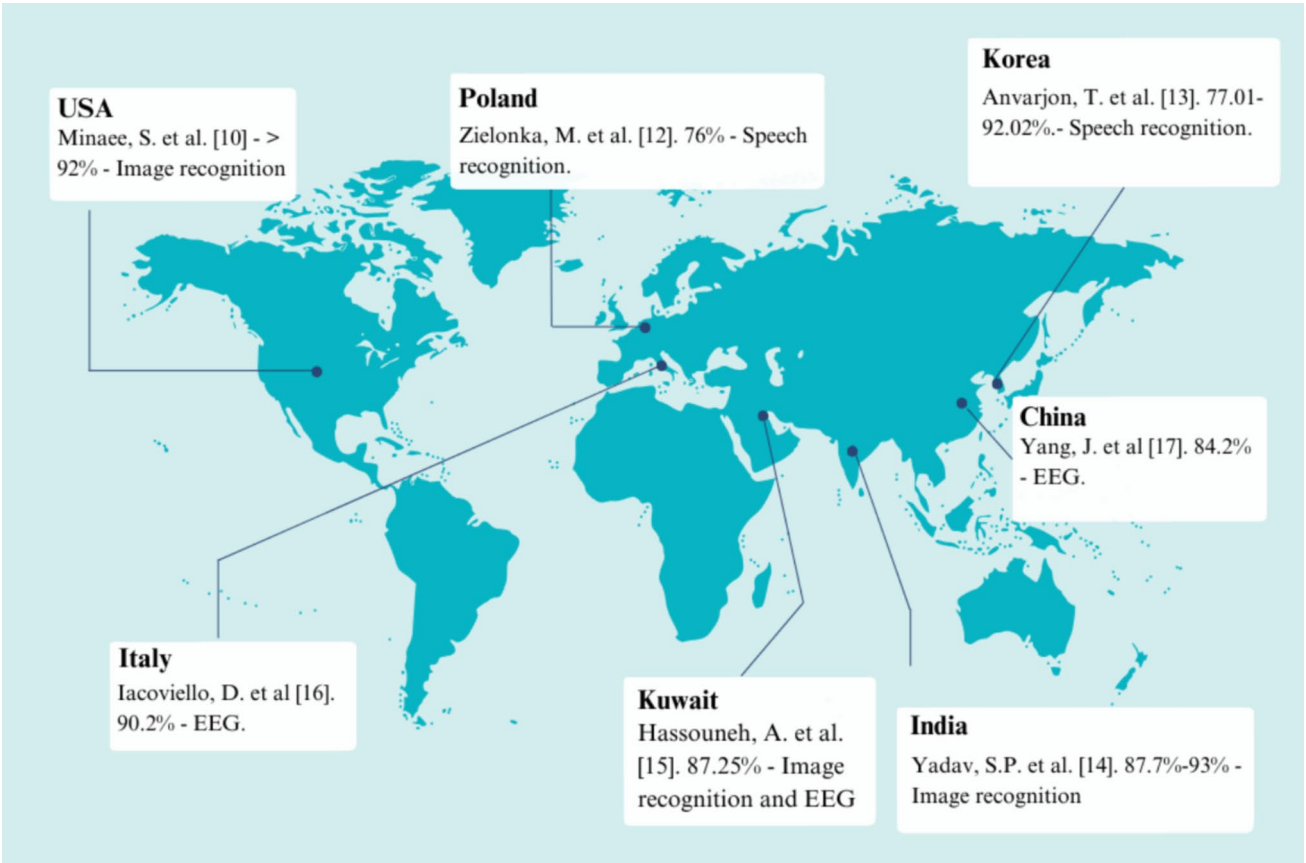

Fig. 2 Emotion recognition research around the world

learning classifier. The proposed approach offers a cost-effective solution with low computational and material requirements, achieving performance that is comparable or superior to existing methods. A set of biometric sensors—including pulse, oxygen, temperature, conductance, and airflow sensors—captures physiological signals associated with emotional states while subjects view video stimuli. These signals are then pre-processed and filtered to enhance data quality before being input into a deep learning classifier based on a hybrid CNN-LSTM architecture augmented with an attention layer. This classifier categorizes the signals into fear, disgust, or neutral states.

The combination of low-cost biometric sensors with advanced deep learning techniques provides an accessible solution that achieves accuracies exceeding 92% while significantly reducing both computational cost and overall expense. Given the limited number of studies exploring emotion detection through low-cost sensors, this work represents a promising new paradigm in the field.

In summary, the proposed method offers an innovative, cost-effective, and highly scalable solution for emotion recognition. Its development is expected to advance knowledge in the field and positively impact various application areas, including psychology, security, marketing, and education.

Material and Methods

Sensors and Controllers

Libelium e-Health Board

The e-Health board, made by Libelium, Fig. 3, is used to add the different types of sensors and capture the signals [19]. This board supports all the communication modules produced by Cooking Hacks, a brand that belongs to Libelium and oversees extending electronics to creators of all audiences in an accessible and educational way.

If our board is too small for the dimensions of the project, layers can be added. The layers complement the functionality of the board model used, adding circuits, sensors, and external communication modules to the original board.

Fig. 3 Libellium e-Health board

Fig. 4 Plan view of an Arduino UNO-R3 chip. Its components and its motherboard can be seen

Microcontroller: Arduino UNO-R3

The Arduino UNO-R3 board, a versatile microcontroller powered by the ATmega328 chip, serves as the central hub for data acquisition and initial processing in this study, see Fig. 4. The choice of Arduino stems from its open-source nature, ease of programming, and vast community of support, making it an ideal platform for rapid prototyping and experimentation in the context of emotion recognition research. It has 14 digital input/output pins, 6 analog inputs, a ceramic resonator at 16 MHz, can be powered by battery or USB cable, and is programmed by computer. Communication between the two is through the serial port.

Sensor Array

To capture nuanced physiological signals indicative of emotional states, a carefully curated array of non-invasive, cost-effective sensors forms the foundation of our data collection:

Pulse and Blood Oxygenation (SpO_2): This sensor illuminates the skin with red and infrared light, measuring the differential absorption of oxygenated and deoxygenated hemoglobin. Variations in blood oxygenation levels reveal changes in heart rate and respiration, which are closely linked to emotional states.
Temperature: Body temperature fluctuations offer valuable insights into autonomic nervous system activity, which is highly responsive to emotional experiences. A dedicated temperature sensor continuously monitors the subject's skin temperature.
Galvanic Skin Response (GSR): By measuring changes in the electrical conductivity of the skin, the GSR sensor provides a window into sweat gland activity regulated by the sympathetic nervous system. This offers a sensitive indicator of emotional arousal.
Airflow: This sensor measures breathing patterns, which can be used to detect changes in breathing rate and depth that are associated with emotional arousal.

The e-health board has several additional sensors, such as ECG, accelerometer, blood glucose, electromyogram (EMG), and blood pressure. However, these are not considered for several reasons. On the one hand, it has been observed that their value is not correlated with emotions in many cases, or they add too much noise, as in the case of the electromyogram. For this reason, it has been decided to adopt the simplest possible approach, seeking to increase the cost/benefit ratio.

The deliberate choice of this sensor suite prioritizes the following principles:

- *Non-invasiveness*: User comfort and ease of deployment are prioritized by employing sensors that do not require complex procedures or direct penetration of the skin.
- *Cost-effectiveness*: The selected sensors align with the study's aim to develop an accessible emotion recognition system, making it feasible for broader applications.
- *Established Physiological Correlates*: Each sensor targets well-understood physiological processes with documented links to the autonomic nervous system activity and emotional experiences.
- Sensors are pre-calibrated according to the manufacturer's specifications (see documentation [19])

Dataset

The database consists of 14 subjects aged between 25 and 67 years. The 5 physiological variables mentioned in the previous section (Sensor Array), i.e., pulse, oxygen, temperature, conductance, and airflow, were recorded during the measurement, all of which together with the date make up the database, which is anonymized to meet ethical and privacy criteria.

This measurement is recorded from a moment before the generation of the emotion, so first, we check that the sensors are giving normal values, then we introduce the stimulus that generates the emotion, in this case, a video of disgust or fear, and finally, we end the video and wait for the person to be stable again.

To facilitate the analysis of different emotional states, each measurement will be segmented into five stages that reflect the temporal dynamics of the emotional response:

- Stage 1 (Baseline): Represents the initial stabilization period before stimulus presentation, capturing the subject's resting physiological state.
- Stage 2 (Pre-Emotion): Encompasses the period from the start of the stimulus until the onset of the target emotion (fear or disgust).
- Stage 3 (Emotion): Marks the active experience of the target emotion, triggered by a specific event within the stimulus. It begins when an event of disgust or fear occurs.
- Stage 4 (Post-Emotion): Captures the physiological changes as the subject transitions from the emotional state to its baseline.
- Stage 5 (Recovery): Represents the return to the subject's physiological baseline after the emotional experience.

This segmentation strategy allows for a nuanced analysis of the physiological patterns associated with each emotional stage. It will help identify which stages are most accurately classified by the artificial intelligence model and where potential challenges lie. For example, the similarity between Stage 1 (Baseline) and Stage 5 (Recovery) may pose difficulties for the system, but the crucial focus remains on the accurate identification of Stage 3 (Emotion). It has therefore been decided that, as they do not make a significant contribution to the results, stages 1 and 5 should be excluded to facilitate the reading and presentation of these results. Figure 5 depicts the stages of the data acquisition process.

Selection of Subjects

Participants were selected randomly among volunteers to ensure that all subjects fell within a range of "normality." All volunteers were screened by a psychologist to confirm that they were mentally healthy, as the study is not clinical, but aims to capture emotional responses only.

AI Algorithms

To construct the architecture of the emotion recognition model, a combination of CNNs, LSTMs, and attention layer was chosen among all the algorithms because it was the one that gave the best results when compared with other models.

CNN

A CNN is a deep learning architecture that is adept at image and video recognition, pattern analysis, and pixel-based data processing [20–22]. CNN's use the convolutions to extract features from the input, see Fig. 6. Filters

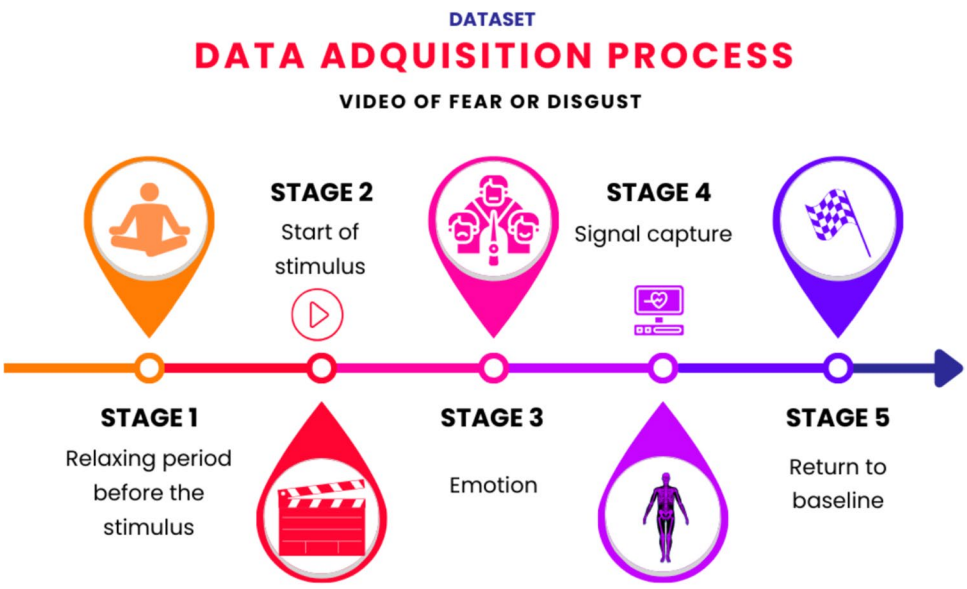

Fig. 5 General data acquisition scheme

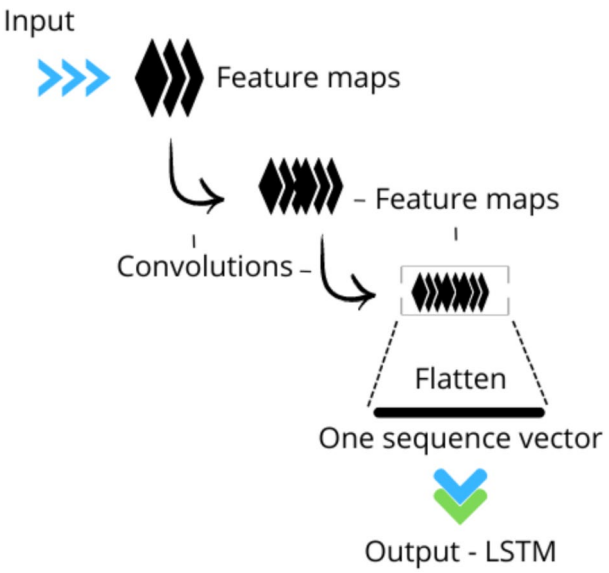

Fig. 6 Schematic of a CNN architecture with several convolutional layers similar to the one used in the model

slide across the input image, performing element-wise multiplications and summations to generate feature maps (or activation maps). Through multiple layers of convolution and pooling, followed by fully connected layers, a CNN learns hierarchical representations—from simple edges and textures to complex objects and spatial patterns [23].

The basic mathematical expression of a convolution operation in a CNN is as follows:

$$(f * g)(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt \quad (1)$$

where f is the input function, g is the kernel or filter that is applied to the input image, and x is the position in the image where the convolution operation is being performed. The above expression refers to convolution in one dimension, but in CNNs convolution operations are used in multiple dimensions, particularly in images that are two-dimensional matrices. This mathematical expression of a 2D layer in CNN is [24]:

$$O_{i,j,k} = f\left(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\sum_{c=0}^{C-1} I_{i+m,j+n,c} * K_{m,n,c,k} + b_k\right) \quad (2)$$

where:

- $\boldsymbol{O}$ is the output tensor of dimensions ($\boldsymbol{I_o}$, $\boldsymbol{J_o}$, $\boldsymbol{K}$)
- I is the input tensor of dimensions ($\boldsymbol{I_i}$, $\boldsymbol{J_i}$, $\boldsymbol{C}$)
- $\boldsymbol{K}$ is the kernel, or filter, tensor of convolution of dimensions ($\boldsymbol{M}$, $\boldsymbol{N}$, $\boldsymbol{C}$, $\boldsymbol{K}$)
- $\boldsymbol{f}$ is the activation function. We chose ***ReLU*** because it introduces non-linearity into the network, which is essential for learning complex patterns in the data. Moreover, ***ReLU*** is computationally efficient and helps mitigate the vanishing gradient problem, thereby facilitating faster and more stable convergence during training.
- $\boldsymbol{b}$ is the bias tensor.
- The summation is over the indices $\boldsymbol{m}$, $\boldsymbol{n}$, and $\boldsymbol{c}$ to traverse the input tensor and the kernel tensor.

LSTM

Long-Short Term Memory Networks are a specialized variant of Recurrent Neural Networks (RNNs) designed to address the issue of vanishing gradients when handling long-term dependencies in sequential data. LSTMs achieve this through internal memory cells and gates that selectively regulate information flow, allowing them to retain and utilize contextually relevant information across extended time steps. This architecture has found success in domains such as speech recognition, natural language processing, and time series forecasting [25]. Variants like Bidirectional LSTMs (BiLSTMs) and Gated Recurrent Units (GRUs) offer additional refinements and capabilities.

The LSTM cell comprises a memory cell and three primary gates, which can be expressed mathematically as follows [26].

For our LSTM, assume that the hidden state $\boldsymbol{h_{t-1}}$ has dimension $\boldsymbol{d}$ and the input $\boldsymbol{x_t}$ has dimension $\boldsymbol{k}$. Then, the weight matrices $\boldsymbol{W_i}$, $\boldsymbol{W_f}$, and $\boldsymbol{W_o}$ are of dimension $\mathbb{R}^{d\times(d+k)}$ and the corresponding bias vectors $\boldsymbol{b_i}$, $\boldsymbol{b_f}$, and $\boldsymbol{b_o}$ are in $\mathbb{R}^{d}$.

Input Gate This gate is responsible for updating the cell state using a sigmoidal activation function. It is defined as follows:

$$i_t[= \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (3)$$

- $\boldsymbol{W_i}$ is the weight matrix associated with the input gate of the LSTM. It is used to map the concatenated vector $[\boldsymbol{h_{t-1}}, \boldsymbol{x_t}]$ (which combines the previous hidden state and the current input) into the cell state space.
- $b_i \in \mathbb{R}^d$ is the corresponding bias vector.
- $\boldsymbol{x_t}$ is the input vector at the current time step. It contains the features or measurements from the system at time t and has a dimension $\boldsymbol{k}$.
- $\boldsymbol{h_{t-1}}$ represents the hidden state from the previous time step in the LSTM. It contains the information that was computed up to time $\boldsymbol{t-1}$ and is used, together with the current input $\boldsymbol{x_t}$, to determine the operations of the LSTM's gates and to update the cell state.

- **σ** denotes the sigmoid function, which outputs values in the interval [0,1] to determine the proportion of new information to be retained [25, 26].

Forget Gate This gate decides which information should be discarded from the cell state. It is expressed as:

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{4}$$

where $\boldsymbol{W_f} \in \mathbb{R}^{d\times(d+k)}$ *and* $\boldsymbol{b_f} \in \mathbb{R}^{d}$ serve similar roles for the forget gate (weight representation and bias).

Output Gate This gate determines the output of the LSTM cell based on a filtered version of the cell state [26]. It is computed as follows:

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t.tanh(C_t) \tag{6}$$

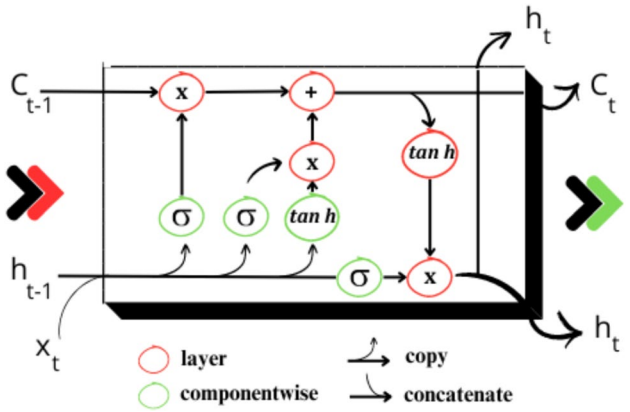

Fig. 7 Scheme of the components of a LSTM cell

Here, $\boldsymbol{W_o} \in \mathbb{R}^{d\times(d+k)}$ *and* $\boldsymbol{b}_0 \in \mathbb{R}^{d}$ are the weight matrix and bias for the output gate. The *tanh* function compresses the cell state C_t into the range [− 1,1] for output generation.

Cell State Update Finally, the previous cell state C_{t-1} must be updated as follows. The complete process can be seen in Fig. 7.

$$C_t = f_t.C_{t-1} + i_t.g_t \tag{7}$$

where $g_t = tanh(W_g \cdot [h_{t-1}, x_t] + b_g)g\_t$ represents the candidate cell state computed using the hyperbolic tangent function (*tanh*), with $\boldsymbol{W_g} \in \mathbb{R}^{d\times(d+k)}$ and $\boldsymbol{b_g} \in \mathbb{R}^{d}$.

Attention Layer

The attention layer is a fundamental component in deep neural network architectures, especially for tasks involving sequence processing such as machine translation, text summarization, and speech recognition. Its main function is to focus the model's attention on the most relevant parts of the input in order to produce an accurate output.

An attention layer takes as input a sequence of vectors $\boldsymbol{h_1}, \boldsymbol{h_2}, \ldots, \boldsymbol{h_n}$ and produces a sequence of attention vectors $\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots, \boldsymbol{a_n}$, where each a_i represents the relative importance of h_i in the final output, see Fig. 8.

The attention vectors are computed using a scoring mechanism that assigns a score to each input vector. The scoring function can be as simple as a dot product or as complex as a deep neural network [27].

Where $\boldsymbol{X}$ is the data matrix, $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ are the subnetworks of the attention head, $\boldsymbol{w}$ indicates that the subnetworks are weighted and, finally, after the dot products, a concatenation is performed using softmax to output a linear vector.

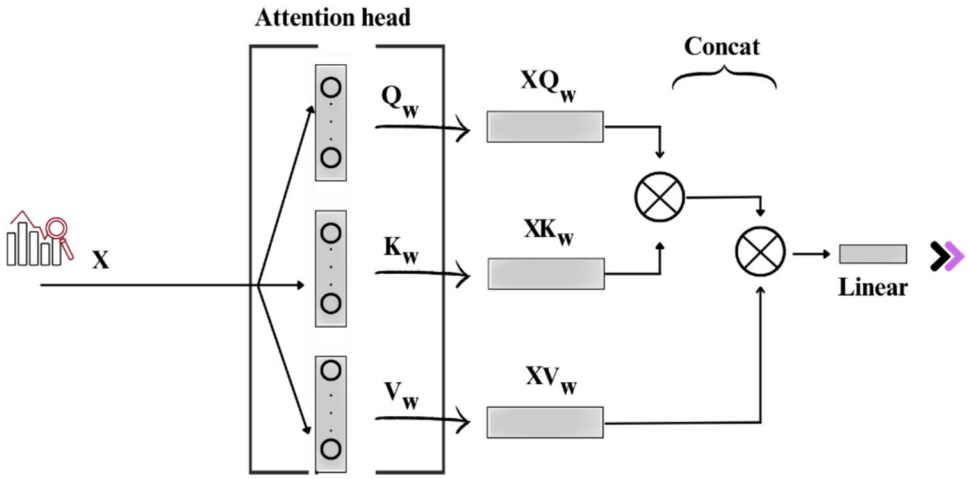

Fig. 8 Scheme of a multi-head attention layer

Important Additional Layers

It is necessary to add other types of layers to optimize the procedure and also for the model to work correctly, the most important ones are as follows:

Dense Layer A dense or fully connected layer establishes connections between every neuron in that layer and every neuron in the subsequent layer. This structure is essential for learning complex relationships within data. The mathematical representation for a simple fully connected network with one hidden layer and an output layer is [28]:

$$y_{fc} = f\left(\sum_{i=1}^{n}(W_i * x_i) + b\right) \tag{8}$$

where:

- $\boldsymbol{x_i}$ is the input vector to the network.
- $\boldsymbol{W_i}$ are the weight matrices for the connections between layers.
- $\boldsymbol{b}$ is the bias.
- $\boldsymbol{f}$ is the activation function applied to the output of each layer to compute the attention weights. The election was the softmax function. This choice ensures that the weights are normalized into a probability distribution across the input sequence, allowing the model to focus more on the most relevant features.

Note that deep neural networks often contain multiple hidden layers, leading to more intricate formulas with additional weight matrices and bias vectors.

Dropout layer Dropout is a regularization technique designed to mitigate overfitting in deep learning models. During training, it randomly deactivates (sets to zero) a specified percentage of neurons at each iteration. Mathematically, a dropout layer multiplies the input vector by a randomly generated binary mask with the exact dimensions. The probability that a mask element is 1 is termed the "dropout rate" (commonly between 0.2 and 0.5).

Note that dropout is typically deactivated during testing (dropout rate set to 0), and dropout rates depend on the model's complexity and the dataset [29].

Hyperparameters

Fine-tuning the hyperparameters is essential to optimize performance and prevent overfitting or underfitting. Here are shown the most important used in the model:

Optimizer: Adam is a popular optimization algorithm that combines the strengths of RMSprop and Momentum. It computes adaptive learning rates for each parameter based on estimates of first and second moments of the gradients. The following Eqs. (9–11) describe the Adam optimizer [30, 31]:

$$m_t = \beta_1 m_{t-1} + \left(1 - \beta_1\right) g_t \tag{9}$$

$$v_t = \beta_2 v_{t-1} + \left(1 - \beta_2\right) g_t^2 \tag{10}$$

$$\theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{v_t} + \epsilon} m_t \tag{11}$$

where:

- $\boldsymbol{m_t}$ is the first updated moment (mean) estimate.
- $\boldsymbol{v_t}$ is the second updated moment (variance) estimate.
- $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the moment decay parameters.
- $\boldsymbol{g_t}$ is the gradient at the current step.
- $\boldsymbol{\alpha}$ is the learning rate.
- $\boldsymbol{\epsilon}$ (epsilon) is a small numerical constant to avoid division by zero.
- $\boldsymbol{\theta_t}$ is the current value of the parameter being updated. This is the parameter that the algorithm optimizes.

Adam was selected due to its robustness in handling sparse gradients and its proven performance across various deep learning tasks [30, 31].

Learning Rate: 1×10^{-4}

The learning rate controls the magnitude of parameter updates during training. The choice of a learning rate of 1×10^{-4} is based on preliminary experiments and supported by prior literature, ensuring stable convergence without overshooting minima.

Batch Size: 5, 10 and 15

The batch size defines the number of examples processed per iteration. We evaluated different batch sizes and found that a batch size 15 offered a good balance between computational efficiency and convergence behavior. Larger batch sizes generally improved gradient stability, but we observed diminishing returns beyond a batch size of 15 in our experiments.

Epochs: 1000

The number of epochs represents the number of complete passes through the training dataset. An epoch count of 1000 was determined to be sufficient to allow the model to learn the underlying patterns in the data, with early stopping applied based on validation performance.

Momentum: 0.99 and 0.999

Momentum values were employed to control the influence of previous updates on new ones. These values help

accelerate gradient descent in the relevant direction and dampen oscillations, contributing to faster convergence.

Shuffling:

Shuffling the training data at each epoch is applied to ensure a random distribution of examples, which helps prevent the model from learning the order of the data.

Activation Function: ReLU

As it was explained in section 2.3.1. *ReLu* was used in the CNN layers due to its ability to introduce non-linearity, facilitate efficient gradient propagation, and mitigate the vanishing gradient problem.

Regularization: (rate 0.2 to 0.5)

Dropout is a regularization technique where a fraction of neurons is randomly deactivated during training. This helps prevent overfitting by ensuring that the model does not rely too heavily on any single feature

The selection of these values was based on prior literature and experimentation with different configurations. Finding the optimal hyperparameter configuration is an iterative process that requires careful testing and adjustments.

Architecture of the Network

The diagram that is shown in Fig. 9 shows a system that uses sensors to capture the emotional responses during video viewing. The captured signal is vectorized and normalized using the *Z*-score function, resulting in 1D vectors of dimension (19,1,1). These overlapped vectors, encompassing multiple events, are fed into a 3-layer 1D CNN with feature maps of decreasing size (128, 64, 32) for efficient feature extraction.

The output then passes through a fully connected layer with a single output neuron, followed by a dropout layer to reduce overfitting. The resulting vector is flattened and then processed through sequential LSTMs, each containing 64 neurons, to capture temporal dependencies within the emotional response. Finally, the output of the LSTM layer is passed through an attention layer to focus on the most relevant features before final classification through a softmax layer.

Performance Indices This study employs a set of widely accepted metrics in the literature to validate the obtained results [32–34]:

Accuracy: It measures the overall correctness of the model's predictions, calculated as the ratio of true positives and true negatives to the total number of samples.

Precision: It reflects the proportion of positive predictions that are correct, calculated as the ratio of true positives to the sum of true positives and false positives.

$$P=TP / (TP + FP) \tag{12}$$

Recall (Sensitivity): It quantifies the model's ability to identify all positive instances, calculated as the ratio of true positives to the sum of true positives and false negatives.

F1 Score: It provides a weighted average of precision and recall, offering a more balanced assessment of the model's performance, calculated as 2 * (precision * recall)/(precision + recall).

$$F1 = 2.(Precision * Recall)/(Precision + Recall) \tag{13}$$

Specificity: It measures the model's ability to correctly identify negative instances, calculated as the ratio of true negatives to the sum of true negatives and false positives.

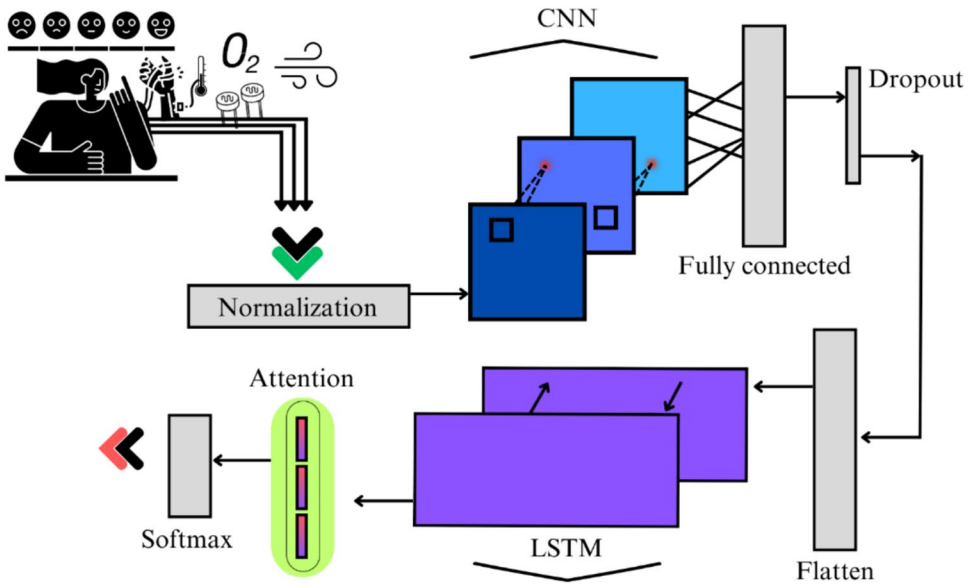

Fig. 9 Architecture of the network

$$S=TN / (TN + FP) \tag{14}$$

where *TP* are true positives, *FP* are false positives, *TN* are true negatives, and *FN* are false negatives.

Confidence interval (CI): is an estimated range of values, derived from sample data, that is likely to contain the true value of a parameter (such as the model's accuracy) 95% of the time. The CI provides insight into the precision and reliability of the estimate. We use the binomial proportion approach [35].

$$CI = \hat{p} \pm Z\alpha/2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{15}$$

where $\hat{\boldsymbol{p}}$ is the observed accuracy, $\boldsymbol{n}$ is the total number of samples, and $\mathcal{Z}_{\alpha/2}$ is the critical value from the standard normal distribution (e.g., 1.96 for a 95% CI). Although the Clopper–Pearson exact interval [36] can also be used for smaller sample sizes, the normal approximation typically suffices for moderately large $\boldsymbol{n}$. This confidence interval helps assess the robustness of our accuracy estimate by providing a plausible range of values.

p-value: represents the probability of observing the given results, or something more extreme, under the assumption that the null hypothesis is true. In the context of our model, it quantifies the likelihood that the observed accuracy (e.g., 93.75%) could have occurred by random chance if the true accuracy were at a baseline level (typically 50% for binary classification). A small *p*-value (commonly < 0.05) indicates that such an outcome is highly unlikely under the null hypothesis, thus suggesting that our model's performance is statistically significant [37].

To determine whether the observed accuracy significantly exceeds chance (50% for random binary classification), a one-sided binomial test is performed [38, 39]. Under the null hypothesis $\boldsymbol{H}_0 : \boldsymbol{p} = 0.5$, the probability of obtaining $\boldsymbol{k}$ or more correct predictions out of $\boldsymbol{n}$ is:

$$\boldsymbol{p - value} = \sum_{i=k}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n \tag{16}$$

where $\boldsymbol{n}$ is the total number of predictions (or observations), $\boldsymbol{k}$ is the number of successes observed, and $\binom{n}{i}$ is the binomial coefficient which represents the number of ways to choose $\boldsymbol{i}$ successes out of $\boldsymbol{n}$ trials.

These definitions help ensure that our reported performance metrics are both precise (through the CI) and statistically validated (through the *p*-value).

Hardware and Time Computing.

The present work was developed on a hardware with the following specifications; CPU i7-9700 K, 16 Gb RAM-DDR4, Nvidia RTX-2060 Super Windforce OC (8 VRAM). The computation time for training and test was 85.16 ± 2.065 s and 0.01165 ± 0.004 s (11 ms approx.).

Methodology

In this section, we describe the methodology we used to evaluate the performance of our proposed tool and the experiments that were conducted. The results of these experiments are presented in the "Results" section. We believe that the results of our experiments will provide valuable information for the design and implementation of emotion recognition tools.

First Tests

To get a starting point, evaluations were started using a simple neural network with 5 neurons.

Ensembles Tests

A series of ensembles were then used to test different parameters. A summary of these and the different variations is shown below:

- Variation of the number of networks in the ensemble (5, 10, 50, 100).
- Variation of the number of neurons in the hidden layer (15–25).
- Tests with different number of states (5, 3, 2). These stages were explained in the "Dataset" section.

Machine and Deep Learning Algorithms

A set of algorithms famous for pattern recognition with spatiotemporal character was used; these include:

- LSTM
- Bi-LSTM
- GRU
- CNN-LSTM
- CNN-LSTM-Attention (The proposed method)

Table 1 presents a summary of the tests conducted.

Results and Discussion

The following section presents the experimental results. For clarity, the accuracy metrics for the complete set of tests are first reported, followed by a detailed focus on the performance of the proposed method. Finally, the results are discussed in order to evaluate the effectiveness of the proposed

Table 1 Summary of all the tests performed

Type of architecture	Number of tests	Characteristics
Simple neural network	Tests 1 and 2	5 neurons in the hidden layer
Ensemble of networks (5 stages)	Tests 3–6	5, 10, 50, and 100 ensembled networks
Ensemble of networks (3 stages) *	Tests 7–10	5, 20, 50, and 100 ensembled networks
Ensemble of networks (2 stages) **	Tests 11 and 12	Ensemble of 100 and 200 networks
LSTM	Test 13	Double LSTM. 64 neurons in the hidden layer
Bi-LSTM	Test 14	Double Bi-LSTM. 64 neurons in the hidden layer
GRU	Test 15	64 neurons in the hidden layer
CNN-LSTM	Test 16	Triple CNN – Double LSTM
CNN-LSTM-Attention	Test 17	Defined in Sect. 2.3.6

*3 stages exclude the baseline stage and the recovery stage

**2 stages exclude the baseline stage, the post emotion stage, and the recovery stage

Table 2 Global accuracy of all the tests performed

Type of architecture	Number of tests	Neurons	Global accuracy
Simple neural network	Tests 1 and 2	15	54%
Ensemble of networks (5 stages)	Test 3 (5 ensembled networks)	23	51%
	Test 4 (10 ensembled networks)	21	44%
	Test 5 (50 ensembled networks)	19	44%
	Test 6 (100 ensembled networks)	21	36%
Ensemble of networks (3 stages)	Test 7 (5 ensembled networks)	45	62%
	Test 8 (20 ensembled networks)	55	62%
	Test 9 (50 ensembled networks)	85	60%
	Test 10 (100 ensembled networks)	110	60%
Ensemble of networks (2 stages)	Test 11 (100 ensembled networks)	5	89%
	Test 12 (200 ensembled networks)	5	89%
LSTM	Test 13	64	93%
Bi-LSTM	Test 14	64	93%
GRU	Test 15	64	93%
CNN-LSTM	Test 16	-	92.5%
CNN-LSTM-Attention	**Test 17**	**-**	**93.75%**

model. Table 2 summarizes the global accuracy achieved in the experiments.

As can be seen, the simple neural network achieved a success rate of 54%. This gave us a starting point to understand that a much more complex network should be used.

Therefore, various ensemble configurations and different state selections were evaluated to assess their impact on overall classification performance. The results varied considerably across the different ensemble tests. Notably, the best outcomes were observed when only stages 2 and 3 were considered, followed by those including three stages (2, 3, and 4). This observation likely stems from the fact that stages 1 and 5—representing baseline and recovery—provide limited discriminatory information and may introduce noise into the classification process. For instance, an ensemble of 20 networks using three states yielded an accuracy of 62%, whereas an ensemble of 100 networks using two states achieved a surprising accuracy of 89%.

Subsequently, when employing more specialized algorithms for processing these types of signals, similar performance levels of around 93% were attained across LSTM, Bidirectional-LSTM, and GRU models, likely due to their similar operational characteristics. However, the addition of convolutional layers to the classifier resulted in a slight decrease in accuracy to 92.5%, possibly because CNNs primarily focus on extracting spatial features rather than temporal ones. Remarkably, by integrating 1D CNNs with an attention mechanism, the accuracy increased to 93.75%—the best performance observed in this study. This improvement is attributed to the synergistic effect of combining these three types of layers, where the attention layer enhances the focus on the most relevant parts of the signal.

Figure 10 illustrates the accuracies obtained across the different tests and classifiers.

The confusion matrix obtained for the best case, i.e., that of the proposed model, is shown in Fig. 11, where it

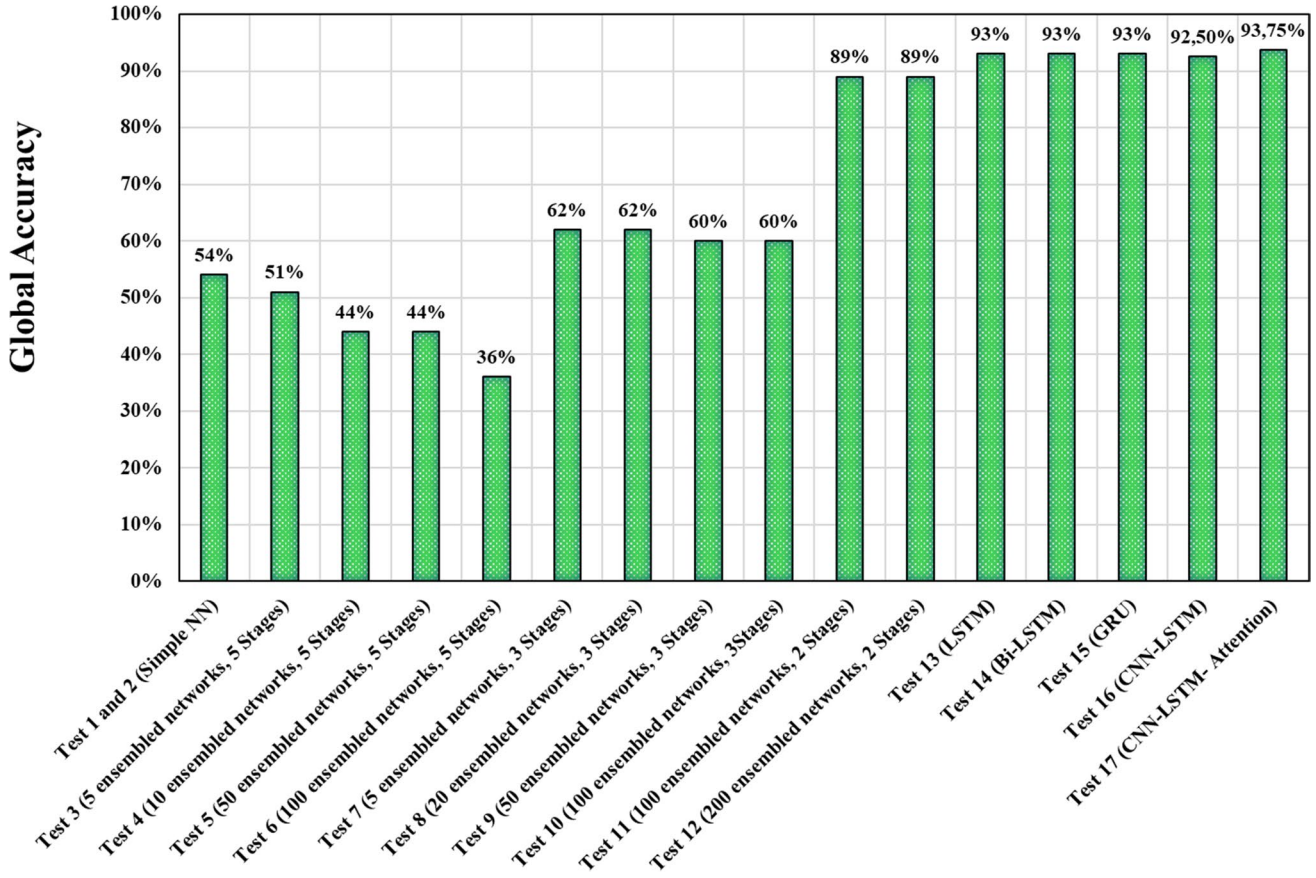

Fig. 10 Global accuracy of the proposed experiments

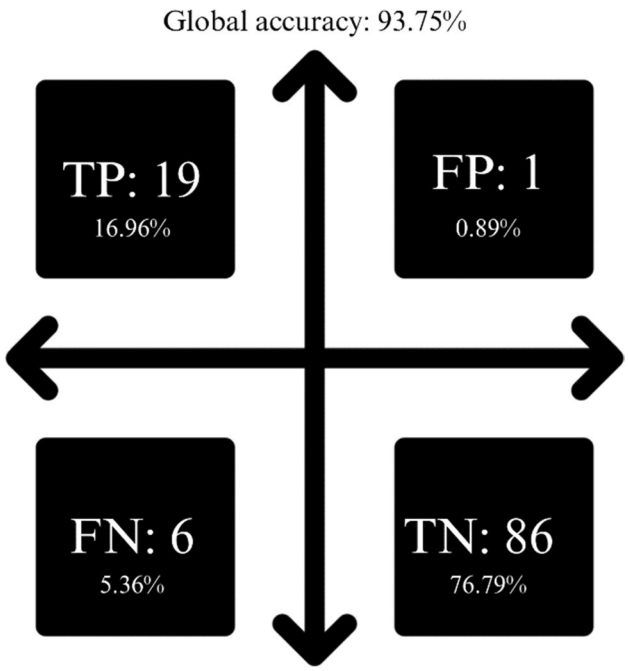

Fig. 11 Matrix confusion of the classification

Table 3 Classification performance indices with the proposed model

Metric	Value
Accuracy	**0.9375**
Precision	0.95
Recall	0.76
*F*1-score	0.848
Specificity	0.989
CI	(0.8927, 0.9823)
p-value	≈ 0

can be seen that the model matches the TPs and TNs quite accurately.

Table 3 shows the classification statistics parameters of the proposed model derived from the results presented in the confusion matrices.

The high accuracy (0.9375) indicates that the model correctly classifies a large proportion of samples, while the precision (0.95) and specificity (0.989) demonstrate that it maintains low rates of false positives and false negatives, respectively. Although the recall value (0.76) shows that most positive cases are correctly identified, this lower recall also suggests that some instances of the target emotions may

have been missed. This may be attributable to overlapping physiological responses among different emotions and the inherent limitations of the low-cost sensor setup. Future refinements in sensor data acquisition and feature extraction are anticipated to improve recall further.

Moreover, while the overall $F1$ score (0.848) reflects a balanced performance, there is a possibility that class imbalance in the dataset has influenced this measure. It is acknowledged that certain emotions might be underrepresented, potentially affecting the $F1$ score. To mitigate this issue, future work will explore strategies such as re-weighting samples or employing oversampling techniques.

In addition to these performance metrics, statistical analyses reinforce the robustness of the model's performance. The 95% confidence interval for accuracy, calculated using a binomial approach, ranges approximately from 89.3 to 98.2%, indicating that the true accuracy is likely within this high-performance region. Furthermore, a one-sided binomial test yields a p-value on the order of 10^{-20}, providing overwhelming evidence that the observed accuracy is not due to chance. Together, these findings confirm the reliability of the classifier and underscore its potential for cost-effective cognitive health monitoring applications.

Context vs. Other Studies

As can be seen in Table 4, the percentages of accuracy in classifying emotions are compared. The purpose of this comparison is to show the general context and to situate the accuracy achieved with our highly efficient model, as there are no similar projects using this sensor technology that achieve these success rates. Therefore, the accuracy rate and the type of sensors and methodology used are shown to take into account the cost/accuracy ratio.

To better contextualize the performance of our proposed emotion classification system, a comparative analysis of its cost-effectiveness relative to more advanced EEG-based systems was conducted. Although EEG systems can achieve marginally higher accuracy, they typically necessitate expensive equipment and complex configurations, thereby incurring significantly higher overall costs. In contrast, our method employs a compact set of low-cost, readily available biometric sensors, rendering it substantially more affordable.

For instance, while typical EEG systems may cost several times more than our sensor configuration [42], our approach achieves competitive accuracy (93.75%) at only a fraction of the cost. This results in a favorable price-performance ratio, measured as the success per unit cost. In addition, recent studies further corroborate the advantages of low-cost approaches in cognitive health monitoring and emphasize their significant impact [43].

Moreover, previous work such as that of Francese et al. [40] demonstrated a user-centric approach to emotion detection with low-cost sensors, achieving promising classification performance on a relatively small dataset. However, our study attains a superior accuracy of 93.75% while incurring a significantly lower cost and yielding a more favorable cost-performance ratio. This improvement can be primarily attributed to the integration of advanced deep learning techniques—including an attention mechanism—which enhances the extraction of discriminative features from biometric signals. Collectively, these findings underscore the superiority of our approach for practical and cost-effective emotion detection in cognitive health monitoring applications.

Nonetheless, several limitations remain to be addressed in future work, including the need to increase both the number of subjects and the diversity of databases, as well as to expand the range of emotions analyzed to gain a more comprehensive understanding of the system's response. It is also important to note that this study is based on healthy subjects and focuses solely on the detection of fear and disgust. Future investigations will aim to compare these results with those obtained from subjects suffering from

Table 4 Comparison of the model with other similar studies

Study name	Emotions	Sensors	Accuracy (%)
Zielonka et al. [13]	Anger, disgust, fear, happiness, sadness and neutral	Voice	76%
Anvarjon et al. [14]	Anger, happiness, sadness and neutral	Voice	77.01–92.02%
Yadav et al. [15]	Anger, disgust, fear, happiness, sadness, surprise and neutral	Facial image	87.7–93%
Hassouneh et al. 16]	Anger, disgust, fear, happiness, sadness, and surprise	Facial image and EEG	87.25%
Iacoviello et al. [17]	Anger, disgust, fear, and happiness	EEG	90.2%
Yang et al. [18]	Fear, happiness, sadness and neutral	EEG	84.21%
Francese, R et al. [40]	Anger, fear, contentment, and sadness	Heart rate, movement, and audio	84.41–91.47%
Moon, E.et al. [41]	Arousal, valence	Hearth rate and speech	84.22%
This study	**Fear, disgust, and neutral**	**Blood oxygenation, temperature, galvanic skin response, and airflow**	**93.75%**

neurodegenerative diseases, such as Alzheimer's disease, to assess differences in emotional outputs.

Conclusions

This study presents an innovative emotion detection and classification system that uses a unique combination of low-cost biometric sensors, deep learning, attentional layers, and LSTM algorithms to achieve exceptional accuracy (93.75%) while significantly reducing costs compared to traditional EEG-based approaches. These results represent a breakthrough in accessible and accurate emotion classification with potential applications in cognitive health monitoring.

Novelty and Impact

This study represents a truly groundbreaking approach to emotion classification. The exceptional accuracy achieved using low-cost sensors outperforms many traditional facial image and EEG-based systems, highlighting the potential to revolutionize emotion recognition and related healthcare applications.

Cost-effectiveness

The affordability of the system is a significant advantage. It addresses a critical need for cost-effective solutions in cognitive health monitoring, making it accessible to a wider range of individuals and healthcare settings.

High Efficiency

The study's high accuracy (93.75%), precision (0.95), recall (0.76), $F1$ score (0.848) and specificity (0.989) indicate excellent efficiency. The deep learning models, attention layers, and LSTM algorithms show optimal performance in extracting and classifying emotional markers from the sensor data.

Future Directions

On the one hand, larger scale deployment and further validation of the tool with different populations and age groups to ensure wider applicability. Secondly, to explore the classification of a wider range of emotions for more nuanced cognitive health applications.

This study sets a new benchmark in emotion classification by combining innovation, affordability, and excellence. It holds great promise for transforming cognitive health assessment and intervention and makes a significant contribution to the field.

Acknowledgements This work was funded by"Agencia Canaria de Investigación, Innovación y Sociedad de la Información de la Consejería de Economía Conocimiento y Empleo y por el Fondo Social Europeo (FSE) Programa Operativo Integrado de Canarias 2014-2020, Eje 3 Tema Prioritario 74 (85%)" from "Gobierno de Canarias" in Spain, under the reference"TESIS2020010118".

Author Contribution N.I. wrote the main manuscript text, prepared the figures and made experimentation. C.M. supervised the experiments and made the methodology. All authors reviewed the manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work was funded by "Agencia Canaria de Investigación, Innovación y Sociedad de la Información de la Consejería de Economía Conocimiento y Empleo y por el Fondo Social Europeo (FSE) Programa Operativo Integrado de Canarias 2014–2020, Eje 3 Tema Prioritario 74 (85%)" from "Gobierno de Canarias" in Spain, under the reference "TESIS2020010118."

Data Availability No datasets were generated or analysed during the current study.

Declarations

Ethics Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Competing Interests The authors declare no competing interests.

References

1. Ekman P. Universals and cultural differences in facial expressions of emotion. In Nebraska symposium on motivation: University of Nebraska Press; 1971.
2. Liu, Y., Sourina, O., & Nguyen, M. K. (2011). Real-time EEG-based emotion recognition and its applications. *Transactions on Computational Science XII: Special Issue on Cyberworlds*, 256–277.
3. Alarcao SM, Fonseca MJ. Emotions recognition using EEG signals: a survey. IEEE Trans Affect Comput. 2017;10(3):374–93.
4. Rahman, M. M., Sarkar, A. K., Hossain, M. A., Hossain, M. S., Islam, M. R., Hossain, M. B., ... & Moni, M. A. (2021). Recognition of human emotions using EEG signals: a review. *Computers in biology and medicine*, *136*, 104696.
5. R. M. Mehmood and H. J. Lee, "Emotion classification of EEG brain signal using SVM and KNN," 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy, 2015, pp. 1–5, https://doi.org/10.1109/ICMEW.2015.7169786.

6. Bhatti, M. W., Wang, Y., & Guan, L. (2004, May). A neural network approach for human emotion recognition in speech. In *2004 IEEE International Symposium on Circuits and Systems (ISCAS)* (Vol. 2, pp. II-181). IEEE.
7. Huang, C. (2017, November). Combining convolutional neural networks for emotion recognition. In *2017 IEEE MIT undergraduate research technology conference (URTC)* (pp. 1–4). IEEE.
8. Wöllmer M, Kaiser M, Eyben F, Schuller B, Rigoll G. LSTM-modeling of continuous emotions in an audiovisual affect recognition framework. Image Vis Comput. 2013;31(2):153–63.
9. Alhagry, S., Fahmy, A. A., & El-Khoribi, R. A. (2017). Emotion recognition based on EEG using LSTM recurrent neural network. *International Journal of Advanced Computer Science and Applications*, *8*(10).
10. Minaee S, Minaei M, Abdolrashidi A. Deep-emotion: facial expression recognition using attentional convolutional network. Sensors. 2021;21(9):3046.
11. Tiwari, S., Agarwal, S., Syafrullah, M., & Adiyarta, K. (2019, September). Classification of physiological signals for emotion recognition using IoT. In *2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)* (pp. 106–111). IEEE.
12. Khare SK, Blanes-Vidal V, Nadimi ES, Acharya UR. Emotion recognition and artificial intelligence: a systematic review (2014–2023) and research recommendations. Information fusion. 2024;102: 102019.
13. Zielonka M, Piastowski A, Czyżewski A, Nadachowski P, Operlejn M, Kaczor K. Recognition of emotions in speech using convolutional neural networks on different datasets. Electronics. 2022;11(22):3831.
14. Anvarjon T, Mustaqeem, & Kwon, S. Deep-net: a lightweight CNN-based speech emotion recognition system using deep frequency features. Sensors. 2020;20(18):5212.
15. Yadav SP. Emotion recognition model based on facial expressions. Multimedia Tools and Applications. 2021;80(17):26357–79.
16. Hassouneh A, Mutawa AM, Murugappan M. Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods. Informatics in Medicine Unlocked. 2020;20: 100372.
17. Iacoviello D, Petracca A, Spezialetti M, Placidi G. A real-time classification algorithm for EEG-based BCI driven by self-induced emotions. Comput Methods Programs Biomed. 2015;122(3):293–303.
18. Yang J, Huang X, Wu H, Yang X. EEG-based emotion classification based on bidirectional long short-term memory network. Procedia Computer Science. 2020;174:491–504.
19. Libelium IoT. (2023, 22 february). *E-health: low cost sensors for early detection of childhood disease - libelium*. Libelium. https://www.libelium.com/libeliumworld/success-stories/e-health-low-cost-sensors-for-early-detection-of-childhood-disease-inspire-project-hope/
20. S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 international conference on engineering and technology (ICET)*, 2017, pp. 1–6.
21. R. Chauhan, K. K. Ghanshala, and R. C. Joshi, "Convolutional neural network (CNN) for image detection and recognition," in *2018 first international conference on secure cyber computing and communication (ICSCCC)*, 2018, pp. 278–282.
22. Valueva MV, Nagornov NN, Lyakhov PA, Valuev GV, Chervyakov NI. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. Math Comput Simul. 2020;177:232–43.
23. R. Venkatesan and B. Li, Convolutional neural networks in visual computing: a concise guide. CRC Press, 2017.
24. Goodfellow Y. Bengio, and A. Courville: Deep learning. MIT press; 2016.
25. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80.
26. Van Houdt G, Mosquera C, Nápoles G. A review on the long short-term memory model. Artif Intell Rev. 2020;53:5929–55.
27. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
28. Heaton J. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: the MIT Press, 2016, 800 pp, ISBN: 0262035618. Genet Program Evolvable Mach. 2018;19(1–2):305–7.
29. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research. 2014;15(1):1929–58.
30. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014;
31. Jesús. ¿Qué es un optimizador y para qué se usa en deep learning? DataSmarts Español [Internet]. 2020 Jul; Available from: https://datasmarts.net/es/que-es-un-optimizador-y-para-que-se-usa-en-deep-learning/
32. Fawcett T. An introduction to ROC analysis. Pattern Recogn Lett. 2006;27(8):861–74.
33. Tharwat A. Classification assessment methods. Applied computing and informatics. 2020;17(1):168–92.
34. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020;21(1):1–13.
35. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. Stat Sci. 2001;16(2):101–33.
36. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 1934;26(4):404–13.
37. Goodman, S. (2008, July). A dirty dozen: twelve p-value misconceptions. In *Seminars in hematology* (Vol. 45, No. 3, pp. 135–140). WB Saunders.
38. Ugarte MD, Militino AF, Arnholt AT. Probability and statistics with R. CRC Press; 2008.
39. Rohatgi VK, Saleh AME. An introduction to probability and statistics. John Wiley & Sons; 2015.
40. Francese R, Risi M, Tortora G. A user-centered approach for detecting emotions with low-cost sensors. Multimedia Tools and Applications. 2020;79(47):35885–907.
41. Moon, E., Sagar, A. S., & Kim, H. S. (2024). Multimodal daily-life emotional recognition using heart rate and speech data from wearables. *IEEE Access*.
42. Farnsworth, B. (2024, 16 diciembre). EEG headset prices – an overview of 15+ EEG devices. *iMotions*. https://imotions.com/blog/learning/product-guides/eeg-headset-prices/
43. Zhu X, Liu Z, Cambria E, Yu X, Fan X, Chen H, Wang R. A client–server based recognition system: non-contact single/multiple emotional and behavioral state assessment methods. Comput Methods Programs Biomed. 2025;260: 108564.