RESEARCH ARTICLE

Elicitation of Anxiety Without Time Pressure and Its Detection Using Physiological Signals and Artificial Intelligence: A Proof of Concept

ANTONIO DI TECCO [1,2], **FRANCESCO PISTOLESI** [2], (Member, IEEE), **AND BEATRICE LAZZERINI** [2], (Member, IEEE)
[1]Department of Information Engineering, University of Florence, 50121 Florence, Italy
[2]Department of Information Engineering, University of Pisa, 56122 Pisa, Italy

Corresponding author: Antonio Di Tecco (antonio.ditecco@unifi.it)

ABSTRACT Stress can be defined as a state of anxiety (or mental tension) caused by a particular situation. Everybody experiences stress to some level, but how we respond to stress significantly affects our well-being. Various events generate anxiety that leads to stress. For example, not having enough time to complete a task or being late are situations where anxiety (and stress) depends on a temporal factor: the scarcity of time. But people also slide into anxiety as they live in a condition that causes them to be tense, independently of time. The studies eliciting anxiety in laboratory settings have less widely considered this variant. This paper presents a proof of concept (PoC) that investigated the possibility of stimulating anxiety without time pressure through a purposely edited horror movie trailer, giving new insights into the emotional experiences evoked by controlled audiovisual stimuli. The PoC comprised an AI-based classifier to detect a person's emotion among *anxiety*, *relaxation*, and *none of the two* based on the galvanic skin response (GSR), photoplethysmogram (PPG), and heart rate (HR), achieving an accuracy higher than 95%. Key application areas include media and marketing, and psychology. Media producers could improve their content to capture the audience better; psychologists could create tailored exposure experiences to promote gradual desensitization to stress triggers.

INDEX TERMS Affective computing, anxiety, artificial intelligence, galvanic skin response, heart rate, pattern recognition, photoplethysmogram, physiological signal, stress, wearable sensors.

I. INTRODUCTION

Modern stress never stops and can impact well-being and health. People are continuously bombarded by news updates and notifications that can induce a feeling of uncertainty and apprehension. Also, modern work environments are rapidly changing, leading to uncertainty and a need for quick decisions. This creates a "perfect storm" that decreases mental well-being and spoils the quality of life. According to an article cited on the *American Institute of Stress*, 83%

of people felt stressed at work in 2022 [1]. The *American Psychological Association* defines stress as *the physiological or psychological response to internal or external stressful events (stressors)*. Stress can be acute or chronic based on the duration of the stressor [2]. It causes changes involving nearly all the body systems, impacting how people feel and behave. Acute stress arises on a single occasion, typically within a short timeframe, whereas chronic stress develops when stressors are recurrent or prolonged. Chronic stress can manifest either episodically or continuously. Episodic stress arises at regular intervals and is of limited duration; continuous stress occurs over an extended period. The number of stressors is

The associate editor coordinating the review of this manuscript and approving it for publication was Byung-Gyu Kim.

a key predictor for an individual's health and psychological well-being [3]. Continuous exposure to stressors leads to adverse health problems, including cardiovascular diseases and cognitive decline [4], along with physiological changes, such as increased heart rate, blood pressure, and cortisol levels [5].

Various studies have used physiological changes to detect stress, such as the galvanic skin response (GSR), heart rate (HR), heart rate variability (HRV), electroencephalography (EEG), photoplethysmography (PPG), and respiration rate (RESP) [6], [7], [8]. The EEG and GSR signals led to levels of accuracy of 60% and 92.9% using artificial neural networks [9] and support vector machines (SVMs), respectively [10]. Multiple signals were also combined to increase precision using SVMs and k-nearest neighbor (k-NN) classifiers. For example, simultaneously using the ECG, GSR, RESP, blood pressure, and blood oximeter, and the combination of EEG, GSR, and PPG led to levels of accuracy of 95.8% [11] and 96.25% [12], respectively. Another study involved construction workers, detecting stress with an 80.32% precision via Gaussian SVMs based on EEG and salivary cortisol [13]. High levels of accuracy (up to 94.78%) were also reached using inertial sensors to predict the stress levels of drivers based on steering wheel movements [14]. Stress detection was also carried out in Industry 4.0 environments using HR, GSR, and RGB videos via supervised and unsupervised techniques, whose respective precisions were 94.9% and 77.4% [15]. Recent studies have also started focusing on detecting and potentially tracking a person's mood in daily life. For example, detecting happiness from 14-channel EEGs via deep learning [16].

These studies highlighted that physiological signals help design accurate machine-learning algorithms for stress detection. Accurate stress detection can enhance mental health monitoring and help people improve the way they deal with stress. However, most studies in the literature focused on specific contexts, thus limiting their generalization capabilities. Moreover, most approaches mainly elicited stress, forcing people to deal with tight time constraints, for example, asking them to carry out tasks under time pressure [17]. But being under time pressure is not the only cause of stress. Stress also arises when exposing people to *stressors* that do not necessarily put a person under time pressure. This variant of stress (common in daily life) needs to be better explored in the literature. A deeper understanding of how people respond to different types of stress could be key in various application fields. For example, it could assist in designing multimedia content that evokes stress at specific moments to capture the viewers' attention or help people cope with stress, thereby enhancing psychological well-being.

This paper describes a proof of concept (PoC) that presents a way of eliciting anxiety without putting people under time pressure using a purposely edited horror movie trailer. The PoC also shows that this type of anxiety—which is similar to long-lasting tension—can be detected only using the GSR and PPG signals. A total of 34 participants took
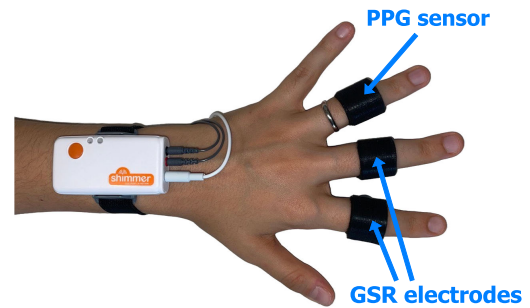
FIGURE 1. The *Shimmer3 GSR+* on the wrist. GSR electrodes are around the index and middle fingers; the PPG sensor is around the ring finger.

part in an experiment where they were shown the edited horror movie trailer, preceded and followed by a relaxing video clip, while their GSR and PPG signals were being recorded. Participants' data were used to design and develop an intelligent system made up of two 2-class k-NN classifiers distinguishing between two pairs of classes: *Anxiety* and *Another Emotion*, and *Relaxation* and *Another Emotion*. The system output is *Anxiety*, *Relaxation*, or *Another Emotion*, i.e., a set of negative affective states, such as *anger* and *sadness*. The system achieved a mean accuracy of 95.22%.

A system implemented based on our PoC could work alongside the existing methods to enrich the types of anxiety elicited in lab settings in years to come. This could help stressed people suffering from various forms of anxiety to improve their quality of life via personalized interventions tailored to specific forms of anxiety. The paper is structured as follows: Section II gives a background on the physiological signals used by the PoC; Section III describes the data collection and editing procedure; Section IV describes the anxiety detection method; Section V presents the results; Section VI discusses the results and presents some application areas that could benefit from implementing the PoC; Section VII draws the conclusions.

II. PHYSIOLOGICAL SIGNALS
This section gives a background of the physiological signals used in this study.

A. GALVANIC SKIN RESPONSE (GSR)
GSR measures the skin electrical conductance in micro-Siemens (μS), which increases in proportion to the micro-variations in sweat levels generated by the autonomic nervous system as a reaction to external stimuli. Sweat is saline, and sodium chloride increases skin conductance: a micro-increase (-decrease) in sweat generates an increase (decrease) in electrical conductivity.

The GSR consists of two main components: phasic and tonic. The *phasic component* is associated with quick variations in the skin conductance in response to emotional stimuli; the *tonic component* is the "baseline", i.e., the skin conductance resting level. GSR is measured in body areas with micro sweat levels controlled by the autonomic nervous system, e.g., palms of the hands [18].
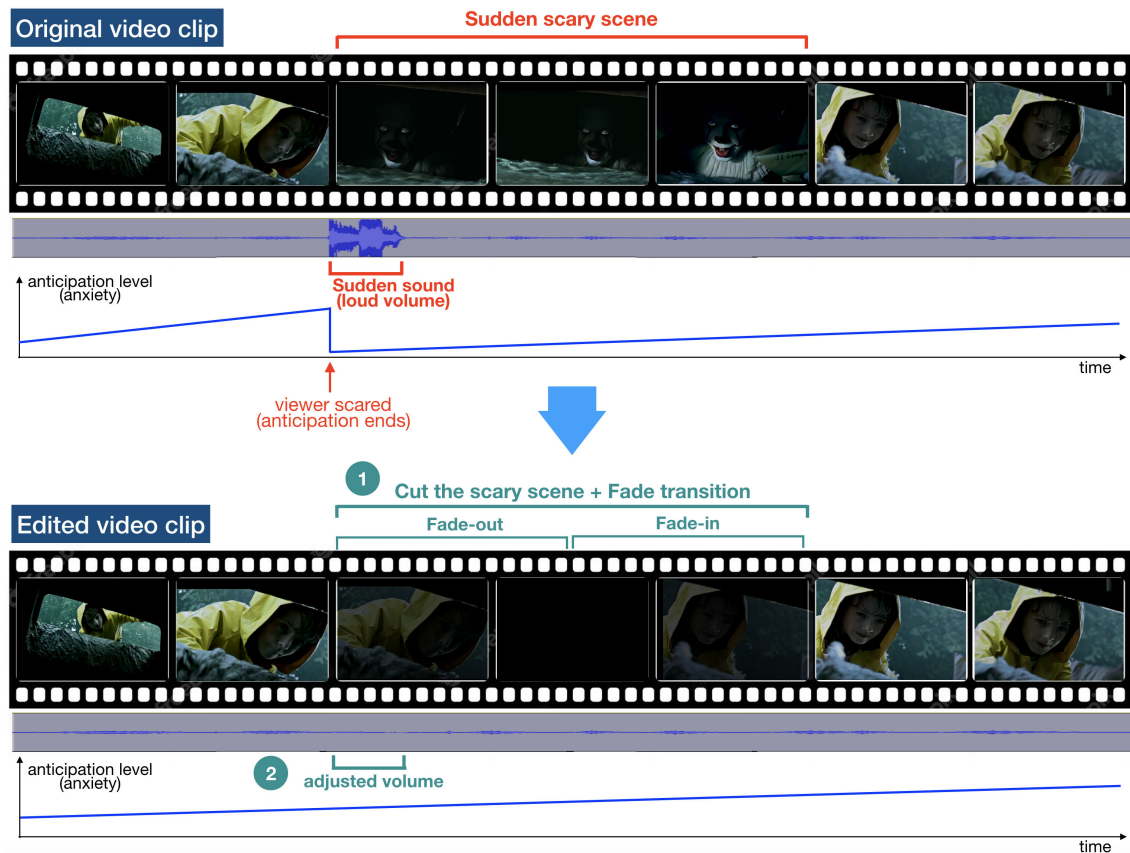
FIGURE 2. Overview of the editing procedure. Sudden scary scenes are cut out of the video clip and replaced with a fade-out and fade-in transition. The sudden loud sounds accompanying each scary scene were removed. This editing prevents the viewer from being suddenly scared and the accumulated tension from drifting away, thereby constantly increasing the anticipation level. The result is a continuous sense of anticipation that stresses the viewer, evoking long-lasting anxiety.

GSR levels highly correlate with anxiety. In particular, when experiencing stress, tension, or anxiety, there is an increase in micro sweat secretion that increases skin conductance. Conversely, in moments of relaxation, sweat secretion decreases, thereby diminishing skin conductance [19].

B. PHOTOPLETHYSMOGRAM (PPG)

PPG detects the blood oxygenation level ($SpO_2\%$). It involves shining a light onto the skin and measuring the percentage of light absorbed/reflected by the underlying blood vessels. PPG is commonly measured in millivolt (mV) [20] and is widely used in wearables to estimate a person's heart rate.

The analysis of PPG waveforms has proved a valuable approach to evaluating stress because of their significant changes during mental stress and relaxation [21]. Also, a person's heart rate increases during physical activities or because of discomfort and a feeling of danger: it is an innate reaction that prepares them to flee [22].

III. DATA COLLECTION

This section describes how data were collected. The *University of Pisa Bioethical Committee* approved the protocol (authorization no. 22/2022, protocol no. 77857/2022).

A. PARTICIPANTS

A total of 34 participants (14 women and 20 men) aged between 19 and 28 took part in the experiment. Their details are in Table 1. Before collecting data, participants read and signed an informed consent explaining the experiment's aim and procedure and that data would be anonymized. After the experiment, participants were asked if they had previously watched *Andy Muschietti's IT: Chapter One* (2017) (or its trailer) to consider potential bias from prior exposure. Only two participants confirmed having seen that content before, leading to the exclusion of their data.

B. VIDEO CLIP

The first step to selecting the video clip was analyzing various candidate film genres. We first eliminated those less likely to generate continuous and profound anxiety stimuli, such as comedy, drama, science fiction, romance, western, adventure, biographical, etc. Thriller and horror genres resulted as potential candidates. Then, by watching various trailers of both genres, we realized that thriller scenes could elicit anxiety but often relied on diverse emotional arcs as they contained action, mystery, or surprise elements, leading to frequent peaks in emotional responses. Thriller trailers
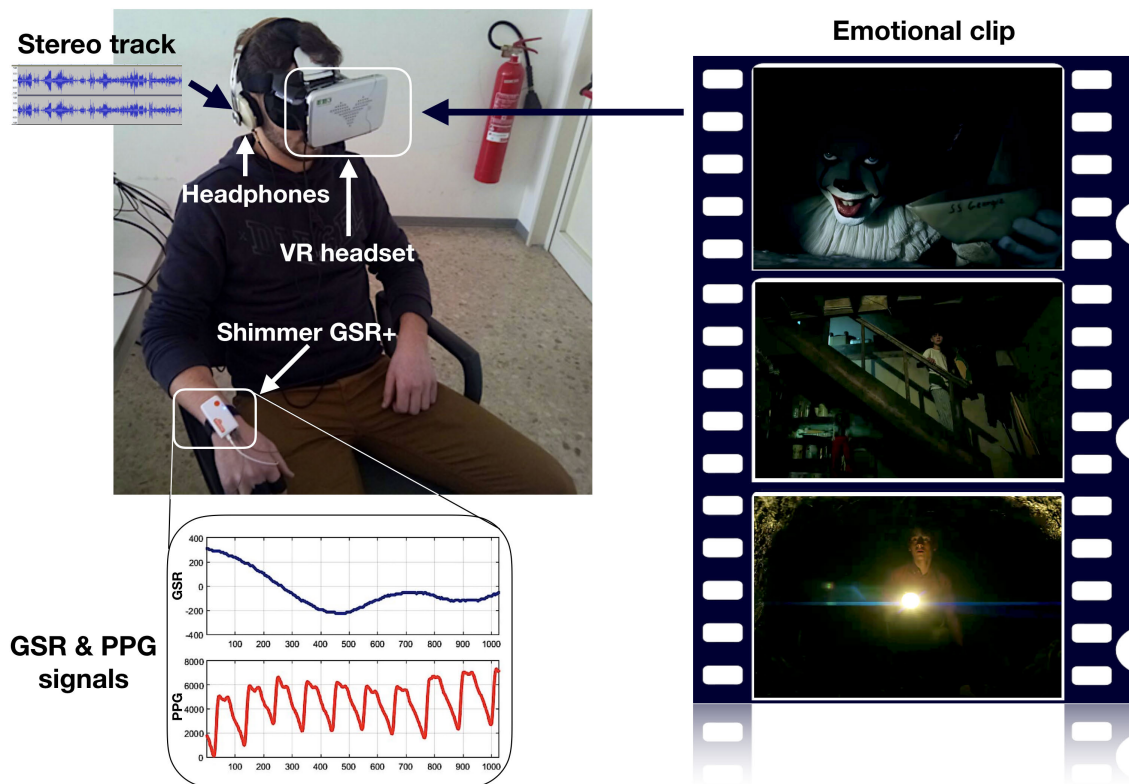
FIGURE 3. The experiment setup. The participant wears the mobile VR headset and the headphones for an immersive experience. A wrist-worn *Shimmer3 GSR+* records the galvanic skin response (GSR) and photoplethysmogram (PPG). The heart rate (HR) is derived from the PPG using the internal *PPG to HR* algorithm of the *Shimmer3 GSR+*.

generally needed to be edited with frequent cuts to remove the scenes causing peaks of emotion, thereby generating continuous anxiety. Although we carefully performed the editing, trying to make the content as smooth as possible, these cuts were too frequent and challenging to make unnoticeable. The viewer frequently became aware that the video had been edited, thus damaging the attempt to generate continuous anxiety.

Contrarily, we found that some horror trailers could easily be edited with fewer cuts that were unnoticeable to the viewer. Watching the edited horror trailer resulted in a uniform and protracted state of anxiety. This characteristic made the horror movie trailer particularly suitable for our research objectives. Other film genres or scenes may be appropriate after applying our editing procedure to remove the variability of emotions characterizing those scenes. In this paper, we wanted to show that horror movie trailers are one of the possible genres that can generate long-lasting anxiety after being edited with the proposed technique. And that anxiety can be elicited using alternatives to time pressure based on edited movie trailers.

Table 2 summarizes the considered horror movie trailers. Those characterized by excessive violence or paranormal themes were discarded because their scenes could not be edited only to elicit anxiety. The trailer selected for the experiment was *Andy Muschietti's IT: Chapter One (2017)*,

which featured calm and anxiety-inducing scenes. To create a long-lasting scene where anxiety is prevalent in the viewer, each scary scene was removed from the trailer and the parts preceding and following that scene were joined through a fade transition. Also, the editing procedure removed the sudden loud sounds accompanying each scary scene, and smoothed the soundtrack volume to make the changes imperceptible to the ear. Fig. 2 shows the editing procedure. In particular, the top of the figure shows a part of the original video clip with a scary scene. Under the video clip, the figure shows its soundtrack and a plot highlighting that a sudden scary scene causes a drop in anticipation level that leads to decreased anxiety. Instead, the bottom of Fig. 2 shows no chance to reduce anxiety (anticipation level) when watching the edited video clip. Viewers constantly expect something scary but do not know it will never occur: this results in a movie trailer only characterized by suspense and anticipation.

Before and after the edited trailer, participants watched a Meditation Relax Music's relaxing video clip characterized by natural content and peaceful soundtracks. The relaxing video clip and trailer were adapted to be played using a mobile Virtual Reality (VR) headset. Fig. 3 shows a participant during the experiment while watching the edited trailer (emotional clip).
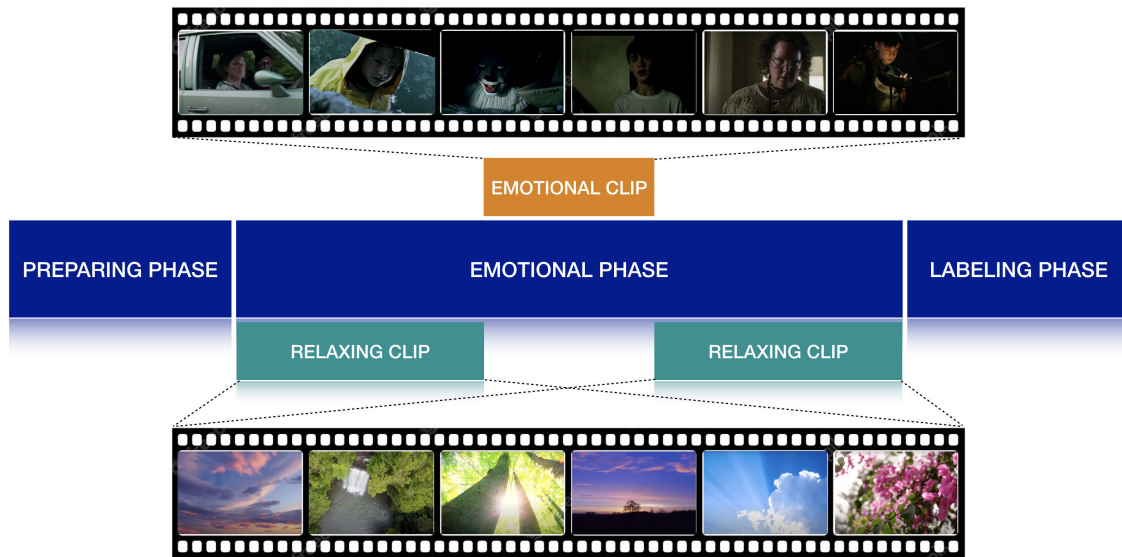
FIGURE 4. The three phases of the experiment.

C. PHYSIOLOGICAL SIGNALS RECORDED

Two physiological signals were collected: GSR and PPG. Signals were measured using a *Shimmer3 GSR+* worn as shown in Fig. 1, with a sampling rate of 128 Hz. The HR was derived from the PPG using the internal *PPG to HR* algorithm of the *Shimmer3 GSR+*. These signals were chosen as they are reliable for detecting stress [18], [23]. Also, they can be collected using affordable and non-invasive sensors embedded in most current smartwatches and smart bands.

D. PROTOCOL

The experiment consisted of three phases (blue rectangles in Fig. 4): a preparing phase, an emotional phase, and a labeling phase, whose details are as follows:

1) *Preparing phase*: the participant was asked to comfortably sit in a chair while positioning the *Shimmer3 GSR+* electrodes on the index and middle fingers (GSR) and ring finger (PPG). The participant then put on the mobile VR headset (with a smartphone inside as a display) and headphones (see Table 3 for specs). Then, the participant was induced into a resting state using dim lights and relaxing background music. The participant's physiological signals were observed in real-time on a PC monitor using *Consensys* software. When the HR and GSR values were low and stable [24], [25], a participant was considered relaxed, and their signal values were considered as a baseline;

2) *Emotional phase*: this phase comprised a *relaxing clip*, an *emotional clip*, and another *relaxing clip*. Both *relaxing clips* lasted 4 minutes to rest the participant with stress relief music and video [26], whereas the *emotional clip* (i.e., the edited version of *Andy Muschietti's IT: Chapter One*) lasted 2 minutes;

3) *Labeling phase*: the participant filled out a questionnaire asking for the emotion felt during the *emotional clip* and whether they felt relaxed after watching the second *relaxing clip*.

IV. METHOD

This section describes how the datasets were generated to train a classifier to distinguish *anxiety*, *relaxation*, or *another emotion*, based on physiological signals recorded while watching the video clips. The procedure is illustrated in Fig. 5. The section then presents the system architecture and how the classifiers were trained and tested.

A. SIGNAL LABELING

Participants reported they experienced a single predominant emotion during the *emotional clip*. Specifically, 20 participants reported feeling *anxiety*, 8 participants reported feeling *sadness*, and 4 participants reported feeling *anger*. All the participants reported feeling *relaxed* when approaching the end of the *post-stimulus relaxing clip*. Labeling was performed as follows:

- the signals of those who reported feeling *anxiety* while watching the *emotional clip* were labeled as *Anxiety*;
- the signals of those who reported feeling *sadness* or *anger* while watching the *emotional clip* were labeled as *Another emotion*;
- the signals collected while watching the *relaxing clip* by those who said they felt *relaxed* were labeled as *Relaxation*.

B. SIGNAL PRE-PROCESSING

First, the raw GSR, PPG, and HR signals were preprocessed by filling in missing values through linear interpolation.
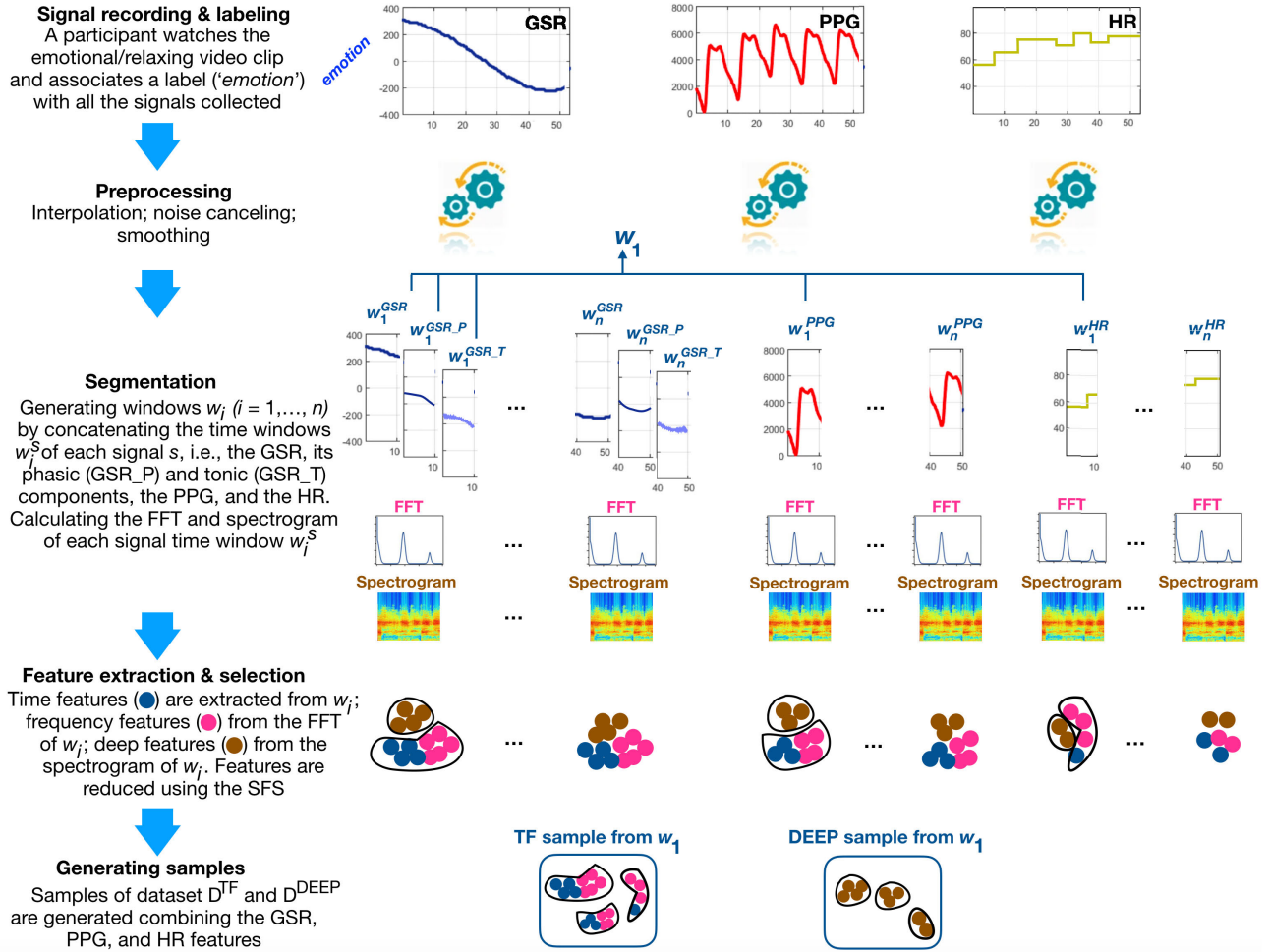
FIGURE 5. Overview of the procedure that generates the samples of all datasets D^X with $X \in \{TF, DEEP\}$. From top to bottom, the signals (GSR, PPG, and HR) recorded while watching the emotional/relaxing video clip are labeled, preprocessed, and divided in windows w_i. Then, each window w_i generates a sample by extracting features from the time, frequency, and deep domains (i.e., from w_i, its FFT, and spectrogram, respectively). The figure shows how the process generates a sample from an example window (w_1). In particular, a sample of a D^{TF} dataset contains the features extracted from all signals' windows (w_1^{GSR}, $w_1^{GSR\_P}$, $w_1^{GSR\_T}$, w_1^{PPG}, and w_1^{HR}) in the time and frequency domains; a sample of a D^{DEEP} dataset combines the features extracted from the spectrograms of all signals' windows.

The raw GSR signal underwent a noise-canceling procedure based on a first-order low-pass Butterworth filter with a cut-off frequency of 5 Hz [27]. Then, it was decomposed into the phasic and tonic components [28]. This led to three GSR signals per participant: preprocessed, phasic, and tonic.

The PPG and HR signals were smoothed using a third-order Savitzky-Golay filter with a span of 11 [29], [30].

Each participant's emotional and post-stimulus signals were divided into 10-second non-overlapping time windows. Shorter windows resulted in higher noise and lower classification accuracy, whereas longer windows failed to capture the rapid changes in the signal. The emotional and post-stimulus signals generated 12 and 24 consecutive time windows, respectively. Then:

1) each time window coming from the *emotional clip* was associated with the emotion the participant declared after watching that clip;

2) each time window coming from the *relaxing clip* of those who declared relaxation after watching that clip was associated with relaxation.

The first time window of each clip (emotional or post-stimulus) was discarded as potentially influenced by the participant's emotion during the previous clip [31]. Also, the first eight and last eight post-stimulus time windows were neglected to reduce the dataset imbalance, as the relaxing and emotional clips had different lengths.

C. FEATURE EXTRACTION

Three feature sets were extracted from each window in the time, frequency, and deep domains. Their features were as follows:

1) *Time domain*: the features extracted from the GSR, PPG, and HR signals were the maximum, minimum, the ratio between maximum and minimum, root mean square, mode, crest factor, mean, median,

TABLE 1. Participant profiles.

ID	Gender	Age	Prior exposure
1	M	22	No
2	M	22	No
3	M	22	No
4	M	22	No
5	M	22	No
6	F	24	No
7	M	22	No
8	F	22	No
9	M	20	No
10	M	20	No
11	M	19	Yes
12	F	19	No
13	M	19	No
14	F	19	No
15	F	25	No
16	M	24	No
17	M	25	No
18	M	28	No
19	M	26	No
20	F	22	No
21	M	23	No
22	F	22	No
23	F	23	No
24	F	23	No
25	F	23	No
26	M	24	No
27	M	22	No
28	M	21	No
29	M	24	No
30	F	19	Yes
31	F	22	No
32	F	21	No
33	M	19	No
34	F	22	No

TABLE 2. Film names and their director(s).

Film name	Director(s)
M3GAN	Gerard Johnstone
Pahanhautoja	Hanna Bergholm
Nope	Jordan Peele
Scream V	Matt Bettinelli-Olpin, Tyler Gillett
IT: Chapter Two	Andy Muschietti
Rabid	Jen Soska, Sylvia Soska
Hereditary	Ari Aster
A Quiet Place	John Krasinski
IT: Chapter One	Andy Muschietti
Annabelle: Creation	David F. Sandberg
The Witch	Robert Eggers

TABLE 3. Equipment name and producer.

Equipment name	Producer
Shimmer3 GSR+	IMotions
RIEM 3 Mobile VR Headset	Ritech
Galaxy S4	Samsung
MOMENTUM On-Ear	Sennheiser

median absolute deviation, mean absolute deviation, geometric mean, 10% trimmed mean, 2nd-, 3rd-, and 4th-order moments, 33rd-percentile, 1st-, and 3rd-quantile, difference between maximum and minimum value (range), harmonic mean, skewness, kurtosis,
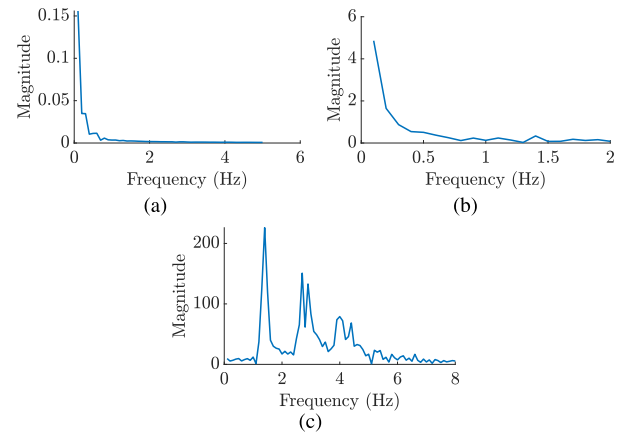
FIGURE 6. GSR (a), HR (b), and PPG (c) single-sided amplitude spectra.
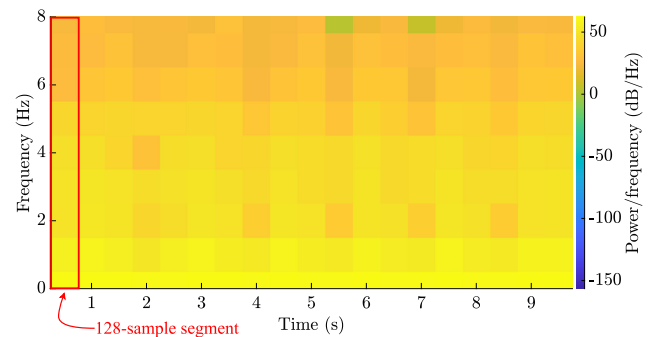
FIGURE 7. Spectrogram of a 10-second time window of the PPG signal from 0 to 8 Hz, using 128-sample segments.

interquartile, mean of absolute values, mean square, standard deviation, and variance [32]. Additional GSR features [33] were the number of peaks, rising time, mean height of the peaks, standard deviation of the heights of the peaks, mean of the time intervals between each pair of consecutive peaks, the standard deviation of the time intervals between each pair of successive peaks, the difference between last and first value, the normalized difference between the maximum and minimum values, and the arc tangent of normalized difference between the maximum and minimum values (angle). PPG features [34], [35], [36] were the mean pulse rate, the standard deviation of pulse rate, the mean root square of pulse rate, mean RR intervals, the standard deviation of RR intervals, the standard deviation of successive RR interval differences, mean root square of successive differences, Poincaré standard deviation perpendicular to the line of identity (SD1), Poincaré standard deviation along the line of identity (SD2), the ratio of SD1 to SD2, number of adjacent NN intervals that differ from each other by more than 50 ms, percentage of the number of adjacent NN intervals, and Baevsky's stress index;

2) *Frequency domain*: the discrete Fourier transform (DFT) of each GSR, HR, and PPG window highlighted that the spectral power distribution concentrated at

TABLE 4. Frequency bands and sub-bands for feature extraction.

Signal	Band	Sub-band(s)
GSR	[0, 5]	[0, 0.3]
HR	[0, 2]	[0, 0.2], [0.5, 2]
PPG	[0, 8]	[0, 2], [2, 3.5], [3.5, 8]

(a)

(b)

FIGURE 8. Block diagrams of the system versions S_1 (a) and S_2 (b). Each output of classifiers C_1, C_{21}, and C_{22} is 0 or 1. The right-hand side of S_2 shows three AND logic gates that generate the output of S_2. Little white circles preceding some inputs to the AND gates denote NOT logic gates.

frequencies lower than 8 Hz (see Fig. 6). The frequency resolution of the signals was 0.1 Hz, which was sufficient to detect physiological changes associated with emotions [37], [38]. Regarding the GSR, the highest power density was in the band ranging from 0 to 0.3 Hz (associated with sympathetic tone variations [39]). Additional features were extracted from the band [0, 5] Hz. Regarding HR, features were extracted from frequencies between 0 and 0.2 Hz, widely used to assess heart rate variations related to respiration [40]. Features were also extracted from 0.5 to 2 Hz because the heartbeat range is typically between 30 bpm (i.e., 0.5 Hz) and 120 bpm (i.e., 2 Hz) [41]. Finally, as highlighted in [42] and [43], the frequency band that showed the highest PPG power density distribution was [0.5, 8] Hz: PPG features were thus extracted from this band. Additional PPG features were extracted in the frequency ranges of [0, 2] Hz, [2, 3.5] Hz, and [3.5, 8] Hz, trying to capture more detailed information. Table 4 summarizes the frequency intervals used for feature extraction.

The features extracted from the frequency domain of the GSR, PPG, and HR windows were as follows: the 99% occupied bandwidth, the mean frequency of the power spectrum, the median frequency of the power spectrum, average power, minimum amplitude frequency, maximum amplitude frequency, the ratio between the maximum and minimum amplitude frequencies (if minimum amplitude frequency was zero, it was assigned a value of 10^{-6}), the ratio between minimum amplitude and maximum amplitude, and the ratio between maximum amplitude-frequency and its amplitude. Additionally, HR features were extracted using a resolution frequency of 0.1 Hz and adapting frequency bands as follows [44]: the average power in [0, 0.2] Hz, average power in [0.1, 0.4] Hz, the sum of average powers in [0, 0.2] Hz and [0.1, 0.4] Hz, and the ratio of average powers between [0, 0.2] Hz and [0.1, 0.4] Hz;

3) *Deep domain*: the short-time Fourier transform (STFT) generated a spectrogram from each signal's 10-second time window. In particular, spectrograms were computed considering the frequency ranges of [0, 5] Hz for GSR, [0, 2] Hz for HR, and [0, 8] Hz for PPG signals. Segments of 128 samples, with a 50% overlap, were extracted from the time window. The *Hamming window* was applied to each segment to create the spectrogram. These choices were a trade-off between temporal and spectral resolution to capture short-term
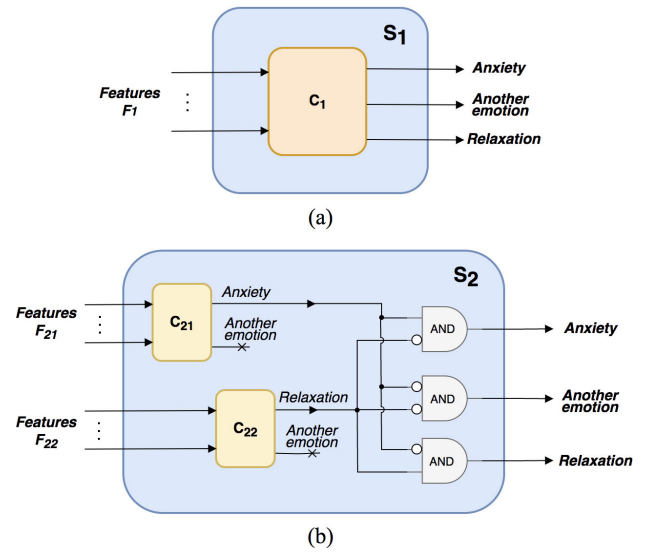
variations and spectral information. Fig. 7 shows an example spectrogram.

Deep features were extracted from spectrograms using the *relu3* layer of *Alexnet* [45] for its fast feature extraction and high classification accuracy compared to other deep models [46]. We selected the *relu3* layer because it led to higher levels of accuracy than other layers in pattern recognition problems [46], [47], [48], [49]. This layer extracted 64,896 features from each spectrogram's signal (324,840 in total).

D. SYSTEM ARCHITECTURE

Two system versions were developed, S_1 and S_2, based on a three-class classifier and two binary classifiers, respectively. Fig. 8 shows their block diagrams. Version S_1 takes a sample in input represented using features F_1 extracted from the GSR, PPG, and HR signals, considering the time, frequency, and deep domains. S_1 uses a classifier (C_1) that classifies a sample as Anxiety, Relaxation, or Another emotion. Version S_2 uses two classifiers, C_{21} and C_{22}, that take a sample represented using features F_{21} and F_{22}, respectively, extracted from the time, frequency, and deep domains. The output classes of C_{21} are Anxiety or Another emotion; those of C_{22} are Relaxation or Another emotion. S_2 classifies a sample as:

- *Anxiety*: if C_{21} detects anxiety;
- *Relaxation*: if C_{22} detects relaxation;
- *Another emotion*: if C_{21} and C_{22} detect *Another emotion*.

Fig. 9 shows a flowchart of S_2.

E. DATASETS

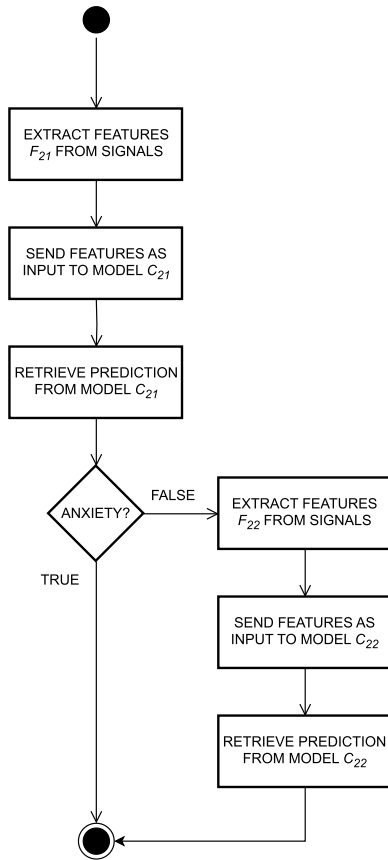As Fig. 5 shows, the union of the features extracted from five contemporary 10-second windows (three from the GSR,

FIGURE 9. Flowchart of system S_2.

TABLE 5. Number of samples in the training and test sets of D_1^{TF} and D_1^{DEEP}.

Class	Test Set	Before SMOTE	After SMOTE
Anxiety	18	162	162
Another emotion	11	97	162
Relaxation	17	156	162
Total	**46**	**415**	**486**

TABLE 6. Number of samples in the training and test sets of D_{21}^{TF} and D_{21}^{DEEP}.

Class	Test Set	Before SMOTE	After SMOTE
Anxiety	18	162	253
Another emotion	28	253	253
Total	**46**	**415**	**506**

TABLE 7. Number of samples in the training and test sets of D_{22}^{TF} and D_{22}^{DEEP}.

Class	Test Set	Before SMOTE	After SMOTE
Another emotion	11	97	156
Relaxation	17	156	156
Total	**28**	**253**	**312**

The χ^2 coefficient was calculated for each feature to assess this hypothesis as follows:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{N} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{2}$$

where O_{ij} is the observed frequency in the i-th category of the feature and the j-th category of the output class, E_{ij} represents the expected frequency in the i-th category of the feature and the j-th category of the output class, assuming independence. We chose a significance level (α) of 0.05, representing the acceptable level of Type I error. The features that obtained a chi-squared coefficient higher than the critical value at the 0.05 significance level and with degrees of freedom corresponding to the number of classes ($N-1$) were considered to be significantly associated with the output, and were maintained. This statistical process ensured that the features selected had a statistically significant relationship with the output.

The Synthetic Minority Oversampling Technique (SMOTE) [50] was applied to fix the class imbalance in the training set. Tables 5, 6, and 7 summarize the number of samples in the datasets before and after the balancing.

F. TRAINING AND EVALUATION

Each classifier (C_1, C_{21}, and C_{22}) was designed, trained, and tested in four steps by using each of its datasets, as follows:

1) *Determining the best feature set*: the forward sequential feature selection (SFS) [51] found the optimal feature set considering, as a criterion, the minimization of the cross-entropy of an ANN. The ANN was chosen for its ability to learn complex linear and nonlinear

one from the PPG, and one from the HR) generated a sample. Outlier samples caused by artifacts from the sensors or participant movement were identified and removed from the dataset using the Isolation Forest algorithm with a contamination ratio of 10% for each class.

Three datasets were created, each in two versions: one version (denoted as *TF*) contained time and frequency features, whereas the other version (denoted as *DEEP*) comprised deep features. The datasets were as follows:

- D_1^{TF} and D_1^{DEEP} to design and train C_1;
- D_{21}^{TF} and D_{21}^{DEEP} to design and train classifier C_{21};
- D_{22}^{TF} and D_{22}^{DEEP} to design and train classifier C_{22}.

Each dataset was split into a training set (90%) and a test set (10%). Both sets contained the same proportion of samples per class.

Features were preliminarily filtered using the chi-squared (χ^2) test, a statistical technique designed to evaluate the association between individual features (f_i) and the output class (c) in a classification problem involving N classes. The null hypothesis (H_0) was that no significant relationship existed between a given feature and the output class, indicating that the two are statistically independent:

$$H_0 : F_i \text{ and } C \text{ are independent.} \tag{1}$$

input-output relationships [52]. We chose an ANN with a single hidden layer, with as many neurons in the output layer as the number of classes (three for C_1, and two for C_{21} and C_{22}). The number of hidden neurons was set according to the following equation [53]:

$$H \approx \sqrt{N_i N_o}, \qquad (3)$$

where H is the number of hidden neurons, and N_i and N_o are the number of input and output, respectively. N_i was set to 20 as we considered selecting a maximum of 20 features, whereas N_o was equal to 3 for C_1, and 2 for C_{21} and C_{22}. Then, $H = 7$ for C_1, and $H = 6$ for C_{21} and C_{22}.

A total of 100 executions of the SFS were performed using stratified 10-fold cross-validation on the training set, thereby obtaining 100 candidate feature sets. Each fold contained randomly selected samples; folds did not share any sample. A statistical evaluation led to the most predictive feature set among the 100 candidates. In particular, each candidate feature set (F_i) was the input to train an ANN using a 30-run 10-fold cross-validation. At each iteration, the performance of the ANN was assessed by measuring the cross-validation error (E_{ik}), representing the model's predictive accuracy, i.e., the average of the mean cross entropies obtained for each of the 10 test folds at iteration k using feature set F_i as input. The Student's t-test with a significance level (α) of 0.05 led to finding the optimal feature set. This significance level represented the statistical significance threshold. In particular, Student's t-test was used to compare the means of the cross-validation errors generated by the ANN models using distinct candidate feature sets F_i and F_j. The null hypothesis was

H_0 : the average mean cross-entropies of F_i and F_j

do not show a statistically significant difference,

(4)

suggesting that preferring one of the two feature sets over the other would not impact the model performance. The t-test produced a statistic (t) that measured the difference between the means, determined as:

$$t = \frac{\overline{E}_i - \overline{E}_j}{s_p \sqrt{2/K}}, \qquad (5)$$

where $\overline{E}_i$ and $\overline{E}_j$ are the means of the cross-validation errors for feature sets F_i and F_j, respectively, K is the number of iterations (i.e., 30), and s_p denotes the pooled standard deviation of errors, calculated as:

$$s_p = \sqrt{\frac{(n-1)\sigma_i^2 + (m-1)\sigma_j^2}{n + m - 2}}, \qquad (6)$$

where σ_i and σ_j are the standard deviations of errors for feature sets F_i and F_j, respectively, n and m are the

sample sizes of the cross-validation errors for F_i and F_j, both equal to 30.

The t-test also generated a p-value measuring the probability of observing a t-statistic as extreme as the one computed if the null hypothesis (H_0) were true. If the p-value is lower than α, then the observed difference in the means of cross-validation errors is highly unlikely to have occurred by random chance. This indicated that the results were statistically significant, and the null hypothesis (H_0) was rejected as there was a statistically significant difference in the performance of candidate feature sets F_i and F_j. The feature set that resulted in the highest number of rejections of the null hypothesis—indicating statistically significant differences in cross-validation errors compared to the others—was considered the most powerful for the classification task. This selection process statistically proved that the chosen feature set had the best predictive power, thus being an optimal choice for the classification task.

2) *Optimizing models*: classifiers based on Decision Trees (Tree), Discriminant Analysis, Naive Bayes, Support Vector Machines (SVM), ANNs, k-Nearest Neighbors (k-NN), Kernels, and Ensembles were optimized using the Bayesian optimizer. The optimizer used a 30-run stratified 10-fold cross-validation with repartitioning to optimize each classifier. Each model's optimal hyperparameters were identified using the dataset's training set and the corresponding optimal feature set. This common practice corresponds to performing a hyperparameter fine-tuning to best exploit the feature sets found by the SFS algorithm. In fact, a given feature set may have special characteristics whose best predictive power can be unveiled by tailoring the hyperparameters of a model to maximize the synergy between the selected features and the model. Also, this approach can improve performance by efficiently fine-tuning the models, avoiding the computational burden of running the SFS method for all models independently.

The optimized models were finally used for internal and external evaluations;

3) *Internal evaluation*: classifiers were trained and evaluated using a 30-run stratified 10-fold cross-validation;

4) *External evaluation*: classifiers were trained 30 times using the dataset's training set and evaluated on the test set (see Section IV-E). The model that achieved the highest classification accuracy on the test set was selected as the best.

System S_1 consisted of the C_1 classifier that obtained the highest accuracy after the external evaluation (i.e., using the test set) among those trained using D_1^{TF} and D_1^{DEEP}. System S_2 comprised the C_{21} and C_{22} classifiers that obtained the highest accuracy in the external evaluation among those trained using D_{21}^{TF} and D_{21}^{DEEP}, and D_{22}^{TF} and D_{22}^{DEEP},

TABLE 8. Number of features of each signal for each dataset.

Dataset	Feature Set	GSR	HR	PPG	Total
D_1^{TF}	F_1	4	1	8	13
D_{21}^{TF}	F_{21}	2	2	6	10
D_{22}^{TF}	F_{22}	3	3	6	12
	Total	9	6	20	
D_1^{DEEP}	F_1	12	3	0	15
D_{21}^{DEEP}	F_{21}	13	2	0	15
D_{22}^{DEEP}	F_{22}	6	3	0	9
	Total	31	8	0	

respectively. System version S_2 named hybrid (HYB) was implemented combining C_{21} based on D_{21}^{TF} and C_{22} based on D_{22}^{DEEP}.

Analyses were performed using MATLAB 2023b running on a personal computer equipped with an Intel CPU i7-12700K, 32 GB of DDR4 DRAM, and NVIDIA GPU 3060 Ti.

V. RESULTS

This section describes the results of the two system versions regarding classification accuracy.

A. FEATURES SELECTED

The numbers of features per signal in each dataset are summarized in Table 8.

As the table shows, 20 features were in the TF domain of the PPG, about triple the number of those of the other signals. Conversely, the GSR features were prevalent in the DEEP domain, where none of the PPG features was considered. The feature sets of TF datasets were as follows:

- F_1: The range and crest factors were the GSR features, along with the percentile and angle degree from the phasic component. The HR feature was the 3rd-quantile. The PPG features consisted of median absolute deviation, crest factor, 10% trimmed mean in the frequency range [2, 3.5] Hz, standard deviation in the frequency range [3.5, 8] Hz, occupied bandwidth in the frequency range [0, 2] Hz, the mean frequency of the power spectrum in the frequency range of [0, 2] Hz, [2, 3.5] Hz, and [0, 8] Hz;
- F_{21}: The GSR tonic features were the ratio between the maximum and minimum values and the geometric mean. The HR features consisted of the mean of absolute values and the 3rd-quantile. The PPG features were the median absolute deviation, the standard deviation in the frequency range [3.5, 8] Hz, the occupied bandwidth in the frequency range [0, 2] Hz, the mean frequency of the power spectrum in the frequency range [0, 2] Hz, and the mean frequency of the power spectrum in the frequency range [2, 3.5] Hz and [0, 8] Hz;
- F_{22}: The GSR tonic features were the ratio between the maximum and minimum values, the difference between the last and first values, and the variance in the frequency range [0, 0.3] Hz. The HR features were the mean of absolute values, the maximum value, and the range in the

TABLE 9. Summary of internal evaluation results of optimal machine learning algorithms trained with TF and DEEP datasets.

Type	Classifier	Model	Avg	Std	F1 score
TF	C_1	Ensemble	0.8393	0.0129	0.8386
	C_{21}	Ensemble	0.9155	0.0062	0.9154
	C_{22}	SVM	0.9036	0.0068	0.9036
DEEP	C_1	k-NN	0.8106	0.0067	0.8098
	C_{21}	k-NN	0.8879	0.0073	0.8879
	C_{22}	ANN	0.9176	0.0038	0.9175

TABLE 10. Hyperparameters of the TF models trained during internal evaluation.

Classifier	Model	Hyperparameter
C_1	Ensemble	Method: AdaBoostM2; NumLearningCycles: 422; LearnRate: 0.9267; MinLeafSize: 3; MaxNumSplits: 147; SplitCriterion: Gdi
C_{21}	Ensemble	Method: AdaBoostM1; NumLearningCycles: 251; LearnRate: 0.9874; MinLeafSize: 1; MaxNumSplits: 201; SplitCriterion: Gdi
C_{22}	SVM	BoxConstraint: 10.078; KernelScale: 4.9718; KernelFunction: Gaussian

frequency range [0, 2] Hz. The PPG features comprised the median absolute deviation, the mean root square of the pulse rate, the occupied bandwidth in the frequency range [0, 2] Hz, the maximum amplitude frequency in the frequency range [0, 2] Hz, the mean frequency of the power spectrum in the frequency range [3.5, 8] Hz, and the maximum amplitude frequency in the frequency range [3.5, 8] Hz.

B. INTERNAL AND EXTERNAL EVALUATIONS

Performance was assessed by calculating the average (Avg), standard deviation (Std), and weighted F1 score of accuracy for both internal and external evaluations.

The highest levels of accuracy achieved during the internal evaluation of each dataset are summarized in Table 9. The hyperparameters selected during the internal evaluation by the Bayesian optimizer for the optimized models are described in Table 10 and Table 11, respectively.

The average accuracy of the best models used in internal evaluation exceeded 80%, reaching accuracy peaks above 90% with both binary classifiers in TF and C_{22} in DEEP domains. Specifically, the C_{21} model trained with the TF dataset accurately distinguished Anxiety from Another emotion. In contrast, the C_{22} model trained with the DEEP dataset effectively distinguished Relaxation from Another emotion.

The best results obtained during the external evaluation for systems S_1 and S_2 are summarized in Table 12. As the table shows, the hybrid system S_2 (HYB) achieved the highest accuracy. This system was developed by selecting the models that reached the highest accuracy throughout the external evaluation, i.e., Subspace k-NN for C_{21} and Weighted k-NN for C_{22}, trained on the TF and DEEP datasets, respectively (see Tables 14 and 15 in the Appendix).

TABLE 11. Hyperparameters of the DEEP models trained during internal evaluation.

Classifier	Model	Hyperparameter
C_1	k-NN	NumNeighbors: 2; Distance: Mahalanobis; DistanceWeight: SquaredInverse
C_{21}	k-NN	NumNeighbors: 1; Distance: Euclidean; DistanceWeight: Equal; Standardize: True
C_{22}	ANN	Activations: Tanh; Lambda: 0.0025648; LayerSizes: 76

TABLE 12. Summary of external evaluation results of systems S_1 and S_2.

Type	System	Avg	Std	F1 score
TF	S_1	0.8913	0	0.8905
DEEP	S_1	0.7826	0	0.7830
HYB	**S_2**	**0.9522**	**0.0028**	**0.9521**
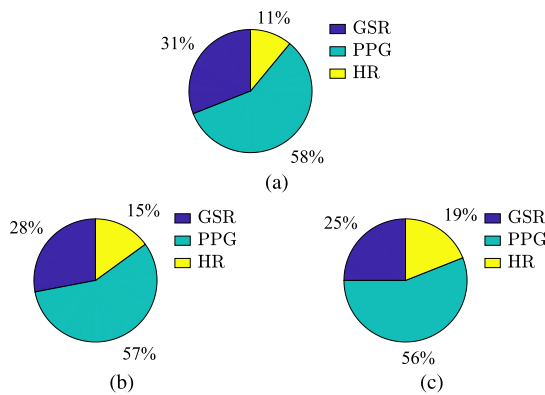
FIGURE 10. Cumulative importance scores of TF feature sets: F_1 (a), F_{21} (b), and F_{22} (c).

FIGURE 11. Cumulative importance score of DEEP features sets: F_1 (a), F_{21} (b), and F_{22} (c).

FIGURE 12. Confusion matrix on system S_2. The matrix shows the number of correctly and incorrectly classified samples for each class.
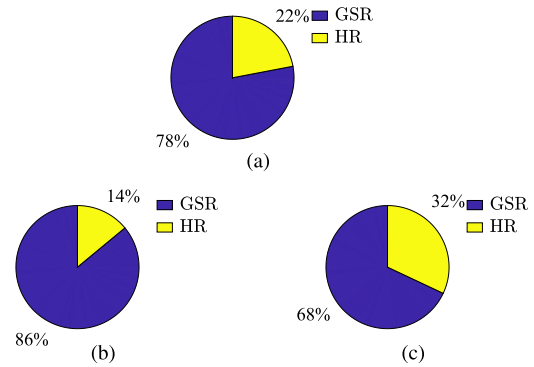
VI. DISCUSSION

This section discusses the results and presents some application areas that could benefit from implementing a system based on the proof of concept presented in the paper.

A. ROLE OF THE PHYSIOLOGICAL SIGNALS

The predictive power of the GSR, PPG, and HR signals in each optimal feature set (F_1, F_{21}, and F_{22}) was assessed using the feature permutation importance (FPI) technique based on the out-of-bag random forest [54]. The FPI involves measuring the impact of randomly shuffling feature values on the model's performance across multiple decision trees. The features causing a more significant decline in performance are considered more important for the model's predictions. In detail, each feature was assigned an importance score by changing its value 30 times. The cumulative importance score of each signal (GSR, PPG, and HR) was calculated by summing up the importance scores of all features extracted from that signal. Fig. 10 and Fig. 11 show three pie charts highlighting the cumulative importance score of each signal in the three optimal feature sets: F_1, F_{21}, and F_{22}.

In particular, Fig. 10 shows that the PPG signal obtained the highest cumulative importance score in the TF domain for all three feature sets (57%, on average). Also, there is

no marginal signal in the feature sets used by the most accurate system version (F_{21} and F_{22}). The least important signal, i.e., the HR in F_{21}, has a score percentage of 15%, suggesting that this signal is not negligible when determining a person's emotion. The pie charts in Fig. 11 highlight that the GSR signal is crucial when classifying emotions in the DEEP domain, for all the feature sets: its mean cumulative importance score is 77.3%. Moreover, the PPG is useless in the DEEP domain. Finally, comparing the pie charts in Fig. 10 and Fig. 11 shows that the GSR and HR signals provide higher predictive contributions in the DEEP domain than in the TF domain.

Regarding performance, Table 12 shows that system S_2 outperformed S_1 during the external evaluation. System S_2 distinguished between *Anxiety*, *Relaxation*, and none of the two (i.e., emotions like *anger* and *sadness*). Fig. 12 shows the confusion matrix of 30 systems S_2 evaluated using dataset D_{21}, whereas the system micro-average and weighted recall, precision, and F1 score are summarized in Table 13. The weighted average considers weights in proportion to the number of samples in each class.

The table shows that all the measures exceeded 0.88: system S_2 thus achieved a high performance. In particular, a precision of 0.9529 highlights that when the classifier predicted that an instance belonged to a particular class, it was

TABLE 13. Micro-average and weighted-average scores of system S_2.

Metric	Class	Precision	Recall	F1 score
Micro	Anxiety	0.9783	1	0.9890
	Another emotion	0.8837	0.9212	0.9021
	Relaxation	0.9711	0.9216	0.9457
Weighted		0.9529	0.9522	**0.9521**

FIGURE 13. TF, DEEP, and TF+DEEP classifiers' performance.

FIGURE 14. Cumulative importance score of TF+DEEP features sets: F_1 (a), F_{21} (b), and F_{22} (c).

correct over 95% of the time. The recall of 0.9522 means that the system correctly identified over 95% of the samples belonging to the corresponding class. Finally, the system achieved an F1 score of 0.9521, demonstrating its high performance in terms of both precision and recall. The results highlighted that using two physiological signals along with features from different domains led to an accurate assessment of emotion.

To further assess performance, the classifiers of the system were trained using a dataset obtained by first combining time-frequency and deep features and then reducing the resulting feature set as explained in Sections IV-E and IV-F. This dataset is referred to as "TF+DEEP". Fig. 13 shows a histogram highlighting the mean levels of accuracy of each classifier trained using the TF, DEEP, and TF+DEEP datasets. As the histogram shows, each classifier's mean level of accuracy obtained using the TF+DEEP dataset (yellow column) was not better than the corresponding classifiers trained only using TF or DEEP features (blue and orange columns).

In conclusion, analyzing the cumulative importance scores of the features from the TF+DEEP dataset highlighted that DEEP features achieved higher cumulative importance scores in predicting emotional states, as shown in Fig. 14.

B. ADVANTAGES OF ANXIETY INDUCTION VIA PURPOSELY EDITED VIDEO CONTENT

Generating anxiety through purposely edited horror movie trailers is an alternative way of exploring emotional responses. The precision and control this method offers allow for a finely tuned emotional experience. Manipulating narrative elements, visuals, and auditory cues generates specific emotional reactions, offering a more nuanced approach than conventional time-constrained tasks. Also, inducing anxiety within controlled environments (such as those using video content) is promising for therapeutic and training applications. The structured nature of the experience can provide a safe platform for exposure therapy or emotional management interventions, contributing to enhanced emotional regulation.

Using purposely edited videos also enables a detailed analysis of emotional responses. Continuous tracking of physiological signals such as GSR, PPG, and heart rate offers a comprehensive understanding of the emotional trip, revealing the intricate interplay between psychological and physiological dimensions. By investigating the emotional patterns in response to video-induced anxiety, we gain insights into the complex landscape of human emotions. This approach can thus enhance our understanding of how anxiety
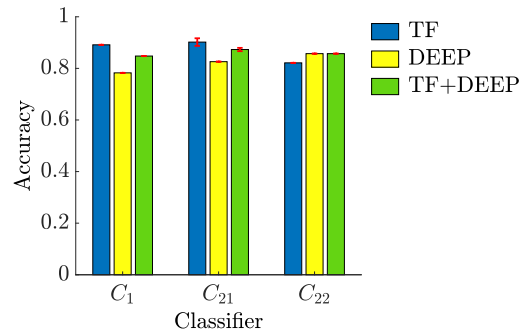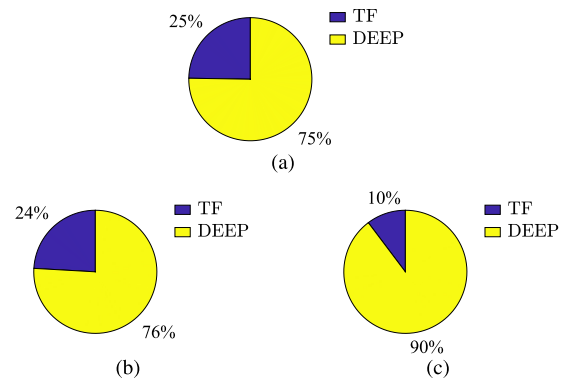
manifests and evolves, potentially impacting various fields, including psychology, neuroscience, and media studies.

An essential distinction lies between anxiety induced by video content and that arising from time-constrained tasks. The former may elicit subjective emotional responses and imagined scenarios, whereas the latter stems from objective pressure imposed by limited timeframes. The intensity and duration of these anxiety forms also differ. Video-induced anxiety may persist over a prolonged period, providing a sustained emotional experience. Time-constrained tasks tend to elicit a brief but intense wave of anxiety. The proof of concept presented in this paper can help improve the understanding of these variations in anxiety experiences, thereby impacting various application fields.

C. POTENTIAL IMPLICATIONS OF THE RESEARCH

The proof of concept described in this paper has several potential applications and implications that are described in the next sections.

1) MEDIA AND MARKETING

Understanding the interplay between emotional experiences and audience engagement is crucial in media and marketing. Emotional experience can influence how audiences perceive, interact with, and respond to media content. A system implementing the PoC can analyze the anxiety types induced by purposely edited video content using various video clips.

TABLE 14. External evaluation results of machine learning algorithms trained using data from the TF dataset.

Model	C_1			C_{21}			C_{22}		
	Avg	Std	F1 score	Avg	Std	F1 score	Avg	Std	F1 score
Fine Tree	0.7608	0	0.7586	0.8043	0	0.8032	0.6071	0	0.5745
Medium Tree	0.6739	0	0.6668	0.7391	0	0.7391	0.6071	0	0.5745
Coarse Tree	0.5000	0	0.4385	0.5217	0	0.5253	0.6785	0	0.6310
Linear Discriminant	0.6739	0	0.6546	0.6739	0	0.6771	0.6071	0	0.6097
Quadratic Discriminant	0.6304	0	0.6330	0.5652	0	0.5652	0.5714	0	0.5232
Gaussian Naive Bayes	0.5434	0	0.5372	0.4130	0	0.3919	0.5714	0	0.5232
Kernel Naive Bayes	0.4130	0	0.4145	0.5434	0	0.5271	0.6785	0	0.6806
Linear SVM	0.6479	0.0087	0.6387	0.6316	0.0050	0.6332	0.5882	0.0180	0.5918
Quadratic SVM	0.8248	0.0050	0.8233	0.7167	0.0036	0.7195	0.6349	0.0150	0.6383
Cubic SVM	0.7801	0.0162	0.7800	0.7898	0.0103	0.7919	0.6894	0.0166	0.6911
Fine Gaussian SVM	0.8695	0	0.8668	0.7391	0	0.7421	0.7857	0	0.7857
Medium Gaussian SVM	0.7608	0	0.7611	0.6956	0	0.6979	0.7142	0	0.6975
Coarse Gaussian SVM	0.5434	0	0.5031	0.5869	0	0.5887	0.6071	0	0.5490
Fine k-NN	0.7608	0	0.7589	0.8043	0	0.8063	0.6071	0	0.6097
Medium k-NN	0.7391	0	0.7332	0.7173	0	0.7202	0.7142	0	0.7172
Coarse k-NN	0.4782	0	0.4856	0.5434	0	0.5350	0.6785	0	0.6662
Cosine k-NN	0.7391	0	0.7332	0.7173	0	0.7202	0.7500	0	0.7476
Cubic k-NN	0.7608	0	0.7589	0.8043	0	0.8063	0.6071	0	0.6097
Weighted k-NN	0.7826	0	0.7826	0.8043	0	0.8066	0.6785	0	0.6806
SVM Kernel	0.6793	0.0435	0.6799	0.7506	0.0257	0.7533	0.7807	0.0373	0.7828
Logistic Regression Kernel	0.6594	0.0412	0.6616	0.7512	0.0223	0.7539	0.7807	0.0410	0.7831
Boosted Trees	**0.8913**	**0**	**0.8920**	0.8043	0	0.8066	0.6785	0	0.6662
Bagged Trees	0.8043	0.034	0.8001	0.8146	0.0262	0.8162	0.7202	0.0439	0.7039
Subspace Discriminant	0.6425	0.0175	0.6444	0.6250	0.0197	0.6262	0.6646	0.0176	0.6397
Subspace k-NN	0.8568	0.0150	0.8548	**0.9019**	**0.0148**	**0.8746**	0.7301	0.0275	0.7313
RUSBoosted Trees	0.6739	0	0.6668	0.7391	0	0.7391	0.6071	0	0.5745
Narrow ANN	0.7300	0.0518	0.7249	0.7318	0.0324	0.7343	0.7093	0.0484	0.7118
Medium ANN	0.7971	0.0422	0.7948	0.7560	0.0280	0.7587	0.7073	0.0481	0.7100
Wide ANN	0.8333	0.0363	0.8319	0.7844	0.0261	0.7866	0.6964	0.0233	0.6989
Bilayered ANN	0.7451	0.0502	0.7400	0.7379	0.0284	0.7407	0.6944	0.0493	0.6963
Trilayered ANN	0.7010	0.0518	0.6899	0.7282	0.0417	0.7307	0.7142	0.0579	0.7162
Optimized Tree	0.5217	0	0.5113	0.5652	0	0.5220	0.6428	0	0.6026
Optimized Discriminant	0.6739	0	0.6546	0.5652	0	0.5652	0.6071	0	0.6097
Optimized Naive Bayes	0.6304	0	0.6278	0.7173	0	0.7202	**0.8214**	**0**	**0.8226**
Optimized SVM	0.8260	0	0.8251	0.8695	0	0.8695	0.7142	0	0.7172
Optimized k-NN	0.8260	0	0.8236	0.8478	0	0.8496	0.6785	0	0.6755
Optimized Kernel	0.7475	0.0209	0.7489	0.7856	0.0181	0.7856	0.7658	0.0275	0.7679
Optimized Ensemble	0.8912	0	0.8905	0.8478	0	0.8496	0.8210	0	0.8145
Optimized ANN	0.8321	0.0310	0.8316	0.7524	0.0209	0.7551	0.7589	0.0178	0.7599

This lets the system investigate the emotional nuances causing audience engagement.

When viewers engage with media content that elicits anxiety triggered by suspenseful elements, their emotional responses contribute to a more immersive experience. The exalted emotional arousal can result in increased attention, cognitive processing, and enhanced memory retention [55]. Manipulating the emotional content allows the presence or absence of specific anxiety-inducing elements to vary. Media creators (but also marketers) could thus tailor their strategies by assessing the reactions to different emotional triggers and how these reactions impact engagement metrics, such as viewer retention. Implications include:

- *Improved content planning:* content creators and marketing professionals could use our research to understand emotional responses to various narrative elements. For example, an online streaming platform that aims to optimize viewer engagement can make data-driven decisions by analyzing viewer reactions to different emotional triggers within the content. If a specific twist in a series elicits a particular type of anxiety, the platform can use this information to recommend similar content to viewers. This keeps viewers engaged and provides a more satisfying viewing experience. People who enjoy content with specific types of suspenseful elements can receive tailored recommendations that align with their preferences. This approach expands beyond streaming services to traditional television.

- *Personalized advertising:* marketing professionals can use our research to measure the emotional nuances of advertisements and promotional campaigns. In particular, they can refine content and strategies to elicit specific emotional responses that drive consumer actions by assessing viewer reactions and engagement metrics. For example, an e-commerce platform may use this information to develop emotionally resonant advertisements that increase sales.

2) THERAPEUTIC CONTEXTS

In therapeutic contexts, people experience anxiety in diverse and complex ways. Anxiety can manifest in response to triggers from traumatic experiences, phobias, generalized

TABLE 15. External evaluation results of machine learning algorithms trained using data from the DEEP dataset.

Model	C_1			C_{21}			C_{22}		
	Avg	Std	F1 score	Avg	Std	F1 score	Avg	Std	F1 score
Fine Tree	**0.7826**	**0**	**0.7830**	0.6086	0	0.6132	0.8214	0	0.8235
Medium Tree	0.7173	0	0.7165	0.5434	0	0.5454	0.8214	0	0.8235
Coarse Tree	0.5652	0	0.4897	0.5217	0	0.4996	0.6785	0	0.6806
Linear Discriminant	0.5000	0	0.5006	0.5000	0	0.5050	0.6428	0	0.6428
Quadratic Discriminant	0.6739	0	0.6662	0.6956	0	0.6991	0.7142	0	0.7172
Gaussian Naive Bayes	0.5869	0	0.5790	0.5652	0	0.5652	0.6071	0	0.6097
Kernel Naive Bayes	0.6521	0	0.6437	0.6521	0	0.6521	0.6428	0	0.6428
Linear SVM	0.5869	0	0.5817	0.4957	0.0186	0.5015	0.6785	0	0.6806
Quadratic SVM	0.6630	0.0176	0.6641	0.7391	0.0090	0.7421	0.7857	0	0.7879
Cubic SVM	0.6382	0.0157	0.6383	0.6189	0.0151	0.6233	0.8214	0	0.8235
Fine Gaussian SVM	0.6739	0	0.6716	0.7173	0	0.7202	0.8214	0	0.8235
Medium Gaussian SVM	0.6304	0	0.6241	0.6956	0	0.6980	0.6428	0	0.6465
Coarse Gaussian SVM	0.5434	0	0.5287	0.5434	0	0.5489	0.5714	0	0.5714
Fine k-NN	0.6956	0	0.6951	0.6956	0	0.6980	0.8214	0	0.8226
Medium k-NN	0.6521	0	0.6444	**0.8260**	**0**	**0.8280**	0.7500	0	0.7477
Coarse k-NN	0.4347	0	0.4221	0.5217	0	0.5272	0.6428	0	0.6465
Cosine k-NN	0.6086	0	0.5908	0.7826	0	0.785	0.7500	0	0.7477
Cubic k-NN	0.6956	0	0.6951	0.6956	0	0.6980	0.8214	0	0.8226
Weighted k-NN	0.7173	0	0.7119	0.8043	0	0.8052	**0.8571**	**0**	**0.8586**
SVM Kernel	0.6962	0.0358	0.6936	0.7119	0.0342	0.7147	0.7857	0	0.7857
Logistic Regression Kernel	0.6431	0.0451	0.6396	0.7324	0.0327	0.7352	0.7976	0.0225	0.7981
Boosted Trees	0.7826	0	0.7775	0.6956	0	0.6991	0.7500	0	0.7516
Bagged Trees	0.7379	0.0293	0.7318	0.6859	0.0423	0.6894	0.7242	0.0235	0.7251
Subspace Discriminant	0.5579	0.0194	0.5558	0.5217	0.0207	0.5264	0.6408	0.0169	0.641
Subspace k-NN	0.7246	0.0249	0.7237	0.7409	0.0240	0.7426	0.7946	0.0156	0.7968
RUSBoosted Trees	0.7391	0	0.7391	0.5652	0	0.5652	0.7857	0	0.7879
Narrow ANN	0.6479	0.0340	0.6471	0.6437	0.0403	0.6475	0.7678	0.0248	0.7699
Medium ANN	0.6557	0.0286	0.6556	0.6425	0.0404	0.6463	0.7767	0.0355	0.7791
Wide ANN	0.6606	0.0216	0.6610	0.6225	0.0341	0.6267	0.7946	0.0231	0.7968
Bilayered ANN	0.6515	0.0488	0.6495	0.6431	0.0496	0.6468	0.7757	0.0337	0.7781
Trilayered ANN	0.6443	0.0493	0.6388	0.6527	0.0523	0.6559	0.7807	0.0343	0.7829
Optimized Tree	0.6086	0	0.5956	0.4347	0	0.3787	0.7142	0	0.7142
Optimized Discriminant	0.6739	0	0.6662	0.6956	0	0.6991	0.7142	0	0.7172
Optimized Naive Bayes	0.5869	0	0.5758	0.6086	0	0.6132	0.7142	0	0.7172
Optimized SVM	0.6086	0	0.6095	0.7173	0	0.7186	0.8214	0	0.8235
Optimized k-NN	0.6956	0	0.6916	0.7391	0	0.7411	0.8570	0	0.8580
Optimized Kernel	0.7089	0.0107	0.7083	0.7614	0.0142	0.7602	0.7857	0	0.7857
Optimized Ensemble	0.7826	0	0.7811	0.7173	0	0.7202	0.7857	0	0.7879
Optimized ANN	0.6521	0.0201	0.6525	0.6159	0.0328	0.6201	0.7817	0.0113	0.7840

worries, or social interactions. Each person's emotional landscape is unique.

The proposed PoC could help tailor interventions to customize therapeutic strategies to align with specific anxiety triggers and emotional responses. Purposely edited video content generates a controlled and targeted anxiety-provoking stimulus. This stimulus accentuates anticipatory tension and suspense, enabling therapists to develop exposure-based interventions tailored to each individual's triggers and sensitivities. Implications of our research include:

- *Personalized exposure therapy:* an exposure therapy that gradually introduces social scenarios within a controlled audiovisual environment may be helpful for those experiencing anxiety in social situations. Therapists could benefit from our research to adjust the narrative elements and emotional scenes. This leads to customizing the exposure experience to align with each individual's specific triggers and sensitivities, thereby

allowing people to face and control their emotional responses.
- *Continuous monitoring and desensitization:* therapists could closely monitor the progress and comfort levels of patients. By controlling video-induced anxiety's emotional content and intensity, a system based on our PoC could help therapists gradually desensitize patients to anxiety triggers, improving emotional regulation and resilience and promoting better emotional well-being [56]. In cases of post-traumatic stress or anxiety disorders, the methodology could also be adapted to simulate controlled environments for therapeutic desensitization. Mitigating anxiety disorders and their consequences is crucial in modern society, especially among adolescents. Alarming statistics reportedly estimated an anxiety prevalence of up to 12%, considering 82 countries with wide geographic variety and cultural backgrounds, with these numbers dramatically increasing in the last years [57].

D. FUTURE RESEARCH DIRECTIONS AND POTENTIAL IMPROVEMENTS

There are various research directions for future work. Some of the most attractive are as follows:

1) *Refinement of emotional classifiers:* further research can focus on improving the accuracy and specificity of emotional classification. This may involve using other machine learning techniques, larger datasets, and more diverse emotional classes to create a model to determine more emotional states.

2) *Comparison of anxiety induction methods:* future studies could compare the efficacy of eliciting anxiety via purposely edited video content to other methods, such as time-constrained tasks, real-life stressors, or different audiovisual stimuli. Comparative research could provide insights into the strengths and weaknesses of each approach and our understanding of emotional experiences.

3) *Long-term effects of exposure therapy:* exploring the long-term effects of exposure-based intervention using controlled audiovisual stimuli is essential in psychological therapy. In particular, analyzing whether individuals who undergo such treatment maintain reduced anxiety levels over time could provide insights into the sustainability of this approach in improving people's emotional management.

4) *Neurological investigations:* analyzing the neurological aspects of anxiety induction through audiovisual stimuli is another interesting future direction. For example, using neuroimaging techniques (e.g., functional magnetic resonance imaging), researchers could examine the neural mechanisms underlying the emotional responses and the impact of purposely edited audiovisual content on brain activity during anxiety induction.

Regarding potential improvements to our methodology, future studies will consider investigating a more exhaustive variety of audiovisual stimuli, including different genres and narrative structures. This will help enrich the study by determining specific elements that induce anxiety. This diversification could lead to greater flexibility for various applications.

Another future direction will be focused on investigating participants' differences. In particular, an in-depth exploration of the influence of individual differences (age, gender, and prior anxiety experiences) on emotional responses to audiovisual stimuli could lead to a more comprehensive understanding of anxiety induction. Tailoring interventions based on individual profiles could also enhance the effectiveness of therapeutic applications.

Finally, improvements in data collection methods (e.g., using less intrusive devices), more advanced techniques for signal processing and data analysis could increase precision and reliability.

Some of these activities are already ongoing to improve the proof of concept presented in this paper.

VII. CONCLUSION

This paper has presented a proof of concept (PoC) exploring the possibility of eliciting anxiety through purposely edited horror movie trailers, offering new insights into the emotional experiences evoked by controlled audiovisual stimuli. The editing was based on accentuating anticipatory tension and suspense while excluding explicit frightening scenes. The result was a form of anxiety similar to long-lasting tension, which is highly subjective and different from that elicited by time pressure, considered by most papers in the literature.

An AI-based classifier was also designed and developed to determine a person's emotion among *anxiety*, *relaxation*, and *none of the two* based on the GSR, PPG, and HR physiological signals, achieving an accuracy higher than 95%.

Combining anxiety induction via purposely edited video content with that based on time-constrained tasks could shed light on the nuances of emotional experiences, having various practical applications spanning content creation in media and marketing, and therapeutic interventions in psychology.

In media and marketing, implementing the PoC could optimize content creation strategies, enhance audience connections, and achieve communication goals.

Implications in psychological therapy could lead to designing tailored interventions. For example, generating a controlled and targeted anxiety-provoking stimulus could enable therapists to develop tailored exposure-based interventions that gradually introduce social scenarios within a controlled environment. This could help achieve gradual desensitization to anxiety triggers, improving emotional management and psychological well-being.

APPENDIX
EXTERNAL EVALUATION TABLES
See Tables 14 and 15.

REFERENCES

[1] C. Mazur. (2023). *Workplace Stress*. Zippia. [Online]. Available: https://www.zippia.com/advice/workplace-stress-statistics/

[2] C. Sandi, "Stress and cognition," *Wiley Interdiscipl. Rev., Cogn. Sci.*, vol. 4, no. 3, pp. 245–261, 2013.

[3] A. Baum, J. P. Garofalo, and A. M. Yali, "Socioeconomic status and chronic stress: Does stress account for SES effects on health?" *Ann. New York Acad. Sci.*, vol. 896, no. 1, pp. 131–144, Dec. 1999.

[4] B. S. McEwen, "Physiology and neurobiology of stress and adaptation: Central role of the brain," *Physiol. Rev.*, vol. 87, no. 3, pp. 873–904, Jul. 2007.

[5] G. P. Chrousos, "Stress and disorders of the stress system," *Nature Rev. Endocrinol.*, vol. 5, no. 7, pp. 374–381, Jul. 2009.

[6] A. M. Brouwer and M. A. Hogervorst, "A new paradigm to induce mental stress: The sing-a-song stress test (SSST)," *Frontiers Neurosci.*, vol. 8, p. 224, Jul. 2014.

[7] A. de Santos Sierra, C. S. Ávila, J. G. Casanova, and G. B. D. Pozo, "A stress-detection system based on physiological signals and fuzzy logic," *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4857–4865, Oct. 2011.

[8] S. S. Panicker and P. Gayathri, "A survey of machine learning techniques in physiology based mental stress detection systems," *Biocybern. Biomed. Eng.*, vol. 39, no. 2, pp. 444–469, Apr. 2019.

[9] R. Mahajan, "Emotion recognition via EEG using neural network classifier," in *Proc. Soft Comput., Theories Appl. (SoCTA)*, vol. 1, 2018, pp. 429–438.

[10] L. Zhu, P. Spachos, P. C. Ng, Y. Yu, Y. Wang, K. Platanitis, and D. Hatzinakos, "Stress detection through wrist-based electrodermal activity monitoring and machine learning," *IEEE J. Biomed. Health Informat.*, vol. 27, no. 5, pp. 2155–2165, May 2023.

[11] A. O. Akmandor and N. K. Jha, "Keep the stress away with SoDA: Stress detection and alleviation system," *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 3, no. 4, pp. 269–282, Oct. 2017.

[12] A. Arsalan and M. Majid, "Human stress classification during public speaking using physiological signals," *Comput. Biol. Med.*, vol. 133, Jun. 2021, Art. no. 104377.

[13] H. Jebelli, S. Hwang, and S. H. Lee, "EEG-based workers' stress recognition at construction sites," *Automat. Construct.*, vol. 93, pp. 315–324, Sep. 2018.

[14] B.-G. Lee and W.-Y. Chung, "Wearable glove-type driver stress detection using a motion sensor," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1835–1844, Jul. 2017.

[15] G. Rescio, A. Manni, A. Caroppo, M. Ciccarelli, A. Papetti, and A. Leone, "Ambient and wearable system for workers' stress evaluation," *Comput. Ind.*, vol. 148, Jun. 2023, Art. no. 103905.

[16] M. Baldassini, F. Pistolesi, and B. Lazzerini, "Detecting happiness from 14-channel binary-valued EEG charts via deep learning," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2022, pp. 66–73.

[17] J. A. Mortensen, M. E. Mollov, A. Chatterjee, D. Ghose, and F. Y. Li, "Multi-class stress detection through heart rate variability: A deep neural network based study," *IEEE Access*, vol. 11, pp. 57470–57480, 2023.

[18] M. Memar and A. Mokaribolhassan, "Stress level classification using statistical analysis of skin conductance signal while driving," *SN Appl. Sci.*, vol. 3, no. 1, p. 64, 2021.

[19] R. Markiewicz, A. Markiewicz-Gospodarek, and B. Dobrowolska, "Galvanic skin response features in psychiatry and mental disorders: A narrative review," *Int. J. Environ. Res. Public Health*, vol. 19, no. 20, p. 13428, Oct. 2022.

[20] A. A. Alian and K. H. Shelley, "Photoplethysmography: Analysis of the pulse oximeter waveform," in *Monitoring Technologies in Acute Care Environments: A Comprehensive Guide to Patient Monitoring Technology*. Springer, 2013, pp. 165–178.

[21] P. Celka, P. H. Charlton, B. Farukh, P. Chowienczyk, and J. Alastruey, "Influence of mental stress on the pulse wave features of photoplethysmograms," *Healthcare Technol. Lett.*, vol. 7, no. 1, pp. 7–12, Nov. 2019.

[22] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: A review," *Med. Biol. Eng. Comput.*, vol. 44, no. 12, pp. 1031–1051, Dec. 2006.

[23] P. H. Charlton, P. Celka, B. Farukh, P. Chowienczyk, and J. Alastruey, "Assessing mental stress from the photoplethysmogram: A numerical study," *Physiol. Meas.*, vol. 39, no. 5, 2018, Art. no. 054001.

[24] D. Nanchen, "Resting heart rate: What is normal?" *Heart*, vol. 104, no. 13, pp. 1048–1049, 2018.

[25] S. Valenti, G. Volpes, A. Parisi, D. Peri, J. Lee, L. Faes, A. Busacca, and R. Pernice, "Wearable multisensor ring-shaped probe for assessing stress and blood oxygenation: Design and preliminary measurements," *Biosensors*, vol. 13, no. 4, p. 460, 2023.

[26] K. Adiasto, D. G. J. Beckers, M. L. M. van Hooff, K. Roelofs, and S. A. E. Geurts, "Music listening and stress recovery in healthy individuals: A systematic review with meta-analysis of experimental studies," *PLoS ONE*, vol. 17, no. 6, 2022, Art. no. e0270031.

[27] E. Piciucco, E. Di Lascio, E. Maiorana, S. Santini, and P. Campisi, "Biometric recognition using wearable devices in real-life settings," *Pattern Recognit. Lett.*, vol. 146, pp. 260–266, Jun. 2021.

[28] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "CvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 797–804, Apr. 2016.

[29] R. W. Schafer, "What is a Savitzky-Golay filter? [Lecture Notes]," *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 111–117, Jul. 2011.

[30] A. Raheel, M. Majid, and S. M. Anwar, "DEAR-MULSEMEDIA: Dataset for emotion analysis and recognition in response to multiple sensorial media," *Inf. Fusion*, vol. 65, pp. 37–49, Jan. 2021.

[31] J. K. Olofsson, S. Nordin, H. Sequeira, and J. Polich, "Affective picture processing: An integrative review of ERP findings," *Biol. Psychol.*, vol. 77, no. 3, pp. 247–265, Mar. 2008.

[32] D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion based music recommendation system using wearable physiological sensors," *IEEE Trans. Consum. Electron.*, vol. 64, no. 2, pp. 196–203, May 2018.

[33] Y. S. Can, B. Arnrich, and C. Ersoy, "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey," *J. Biomed. Informat.*, vol. 92, Apr. 2019, Art. no. 103139.

[34] A. Bánhalmi, J. Borbás, M. Fidrich, V. Bilicki, Z. Gingl, and L. Rudas, "Analysis of a pulse rate variability measurement using a smartphone camera," *J. Healthcare Eng.*, vol. 2018, Feb. 2018, Art. no. 4038034.

[35] M. Brennan, M. Palaniswami, and P. Kamen, "Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability?" *IEEE Trans. Biomed. Eng.*, vol. 48, no. 11, pp. 1342–1347, Nov. 2001.

[36] S. Carrasco, M. J. Gaitán, R. González, and O. Yánez, "Correlation among Poincaré plot indexes and time and frequency domain measures of heart rate variability," *J. Med. Eng. Technol.*, vol. 25, no. 6, pp. 240–248, 2001.

[37] Y. Shu, P. Yang, N. Liu, S. Zhang, G. Zhao, and Y. J. Liu, "Emotion distribution learning based on peripheral physiological signals," *IEEE Trans. Affect. Comput.*, vol. 14, no. 3, pp. 2470–2483, Jul./Sep. 2023.

[38] D. Sadhukhan, S. Pal, and M. Mitra, "Adaptive band limit estimation based PPG data compression for portable home monitors," *Measurement*, vol. 134, pp. 153–165, Feb. 2019.

[39] S. Vieluf, M. Amengual-Gual, B. Zhang, R. El Atrache, C. Ufongene, M. C. Jackson, S. Branch, C. Reinsberger, and T. Loddenkemper, "Twenty-four-hour patterns in electrodermal activity recordings of patients with and without epileptic seizures," *Epilepsia*, vol. 62, no. 4, pp. 960–972, Apr. 2021.

[40] R. Balocchi, D. Menicucci, E. Santarcangelo, L. Sebastiani, A. Gemignani, B. Ghelarducci, and M. Varanini, "Deriving the respiratory sinus arrhythmia from the heartbeat time series using empirical mode decomposition," *Chaos, Solitons Fractals*, vol. 20, no. 1, pp. 171–177, Apr. 2004.

[41] K. Yamamoto and T. Ohtsuki, "Non-contact heartbeat detection by heartbeat signal reconstruction based on spectrogram analysis with convolutional LSTM," *IEEE Access*, vol. 8, pp. 123603–123613, 2020.

[42] J. Park, H. S. Seok, S.-S. Kim, and H. Shin, "Photoplethysmogram analysis and applications: An integrative review," *Frontiers Physiol.*, vol. 12, Mar. 2022, Art. no. 808451.

[43] J. de Pedro-Carracedo, D. Fuentes-Jimenez, A. M. Ugena, and A. P. Gonzalez-Marcos, "Transcending conventional biometry frontiers: Diffusive dynamics PPG biometry," *Sensors*, vol. 21, no. 16, p. 5661, Aug. 2021.

[44] N. P. Novani, L. Arief, R. Anjasmara, and A. S. Prihatmanto, "Heart rate variability frequency domain for detection of mental stress using support vector machine," in *Proc. Int. Conf. Inf. Technol. Syst. Innov. (ICITSI)*, Oct. 2018, pp. 520–525.

[45] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[46] S. Barburiceanu, S. Meza, B. Orza, R. Malutan, and R. Terebes, "Convolutional neural networks for texture feature extraction. Applications to leaf disease classification in precision agriculture," *IEEE Access*, vol. 9, pp. 160085–160103, 2021.

[47] S. L. Sahu and R. Chintala Bhargavi, "Prediction of diseases in potato plant using pre-trained and traditional machine learning models," in *Proc. 4th Int. Conf. for Emerg. Technol. (INCET)*, May 2023, pp. 1–8.

[48] J. Chen, H. Wang, S. Wang, E. He, T. Zhang, and L. Wang, "Convolutional neural network with transfer learning approach for detection of unfavorable driving state using phase coherence image," *Expert Syst. Appl.*, vol. 187, Jan. 2022, Art. no. 116016.

[49] P. Xi, A. Law, R. Goubran, and C. Shu, "Pilot workload prediction from ECG using deep convolutional neural networks," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2019, pp. 1–6.

[50] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[51] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.

[52] O. A. M. López, A. M. López, and J. Crossa, "Fundamentals of artificial neural networks and deep learning," in *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer, 2022, pp. 379–425.

[53] T. Masters, *Practical Neural Network Recipes in C++*. San Mateo, CA, USA: Morgan Kaufmann, 1993.

[54] W. Lin, W. Fan, H. Liu, Y. Xu, and J. Wu, "Classification of handheld laser scanning tree point cloud based on different KNN algorithms and random forest algorithm," *Forests*, vol. 12, no. 3, p. 292, Mar. 2021.

[55] T.-H. Lee, M. Sakaki, R. Cheng, R. Velasco, and M. Mather, "Emotional arousal amplifies the effects of biased competition in the brain," *Social Cognit. Affect. Neurosci.*, vol. 9, no. 12, pp. 2067–2077, Dec. 2014.

[56] E. Landkroon, K. Meyerbröker, E. Salemink, and I. M. Engelhard, "Future-oriented imagery rescripting facilitates conducting behavioral experiments in social anxiety," *Behav. Res. Therapy*, vol. 155, Aug. 2022, Art. no. 104130.

[57] T. Biswas, J. G. Scott, K. Munir, A. M. N. Renzaho, L. B. Rawal, J. Baxter, and A. A. Mamun, "Global variation in the prevalence of suicidal ideation, anxiety and their correlates among adolescents: A population based study of 82 countries," *eClinicalMedicine*, vol. 24, Jul. 2020, Art. no. 100395.

FRANCESCO PISTOLESI (Member, IEEE) received the master's degree (summa cum laude) in computer engineering (enterprise curriculum) from the University of Pisa, Pisa, Italy, and the Ph.D. degree in information engineering (computer system architectures curriculum) from the Leonardo da Vinci Doctoral School of Engineering, University of Pisa, in 2015. He is currently an Assistant Professor with the Department of Information Engineering, University of Pisa. He is also the Scientific Coordinator of two EU projects with the University of Pisa Research Unit, aiming to design innovative AI algorithms to perform a multidimensional risk assessment in case of an earthquake and prevent landslides and floods, including the psychological consequences for those involved. He has coauthored papers on hybrid evolutionary algorithms and sensor-based AI solutions for OSH published in international journals and conference proceedings. His research interests include artificial intelligence (AI) and industrial informatics, particularly multiobjective optimization of sustainable industrial processes and occupational safety and health (OSH). He is an Associate Editor of *Journal of Intelligent Manufacturing* and *Industrial Artificial Intelligence* journal (Springer).

BEATRICE LAZZERINI (Member, IEEE) received the master's degree in computer science from the University of Pisa, Pisa, Italy, and the master's degree in computer engineering from the University of Florence, Florence, Italy. She is currently a Full Professor of computer engineering with the Department of Information Engineering, University of Pisa. She has coauthored seven books and has contributed to more than 230 papers in international journals and conferences. She is the co-editor of two books. She was involved and had roles of responsibility in several national and international research projects and scientific events. She has recently been the scientific coordinator of two regional research projects and an organizer of two special sessions at two IEEE international conferences. Her research interests include computational intelligence, with a particular emphasis on fuzzy systems, neural networks, evolutionary computation, and their applications to pattern classification, image processing, data fusion, context awareness, risk analysis and management, diagnosis, forecasting, multiobjective optimization, multicriteria decision making, affective computing and sentiment analysis, occupational safety and health, and optimization of industrial processes.

ANTONIO DI TECCO received the master's degree in computer engineering (computer systems and networks curriculum) from the University of Pisa, Pisa, Italy, and the M.B.A. degree from Valencian International University, Valencia, Spain. He is currently pursuing the Ph.D. degree in smart computing with the Department of Information Engineering, University of Florence. He spent six months as a Visiting Fellow with the University of Hertfordshire, Hatfield, U.K., conducting research on recommendation systems for domotics. His research interests include artificial intelligence, embedded systems, and the Internet of Things, with applications to affective computing and smart environments for safety and health and usability and quality.

• • •