A Wavelet-based Approach for Multimodal Prediction of Alexithymia from Physiological Signals

Valeria Filippou*
The Cyprus Institute
Nicosia, Cyprus
v.filippou@cyi.ac.cy

Nikolas Theodosiou*
The Cyprus Institute
Nicosia, Cyprus
n.theodosiou@cyi.ac.cy

Mihalis A. Nicolaou
The Cyprus Institute
Nicosia, Cyprus
m.nicolaou@cyi.ac.cy

Elena Constantinou
The University of Cyprus
Nicosia, Cyprus

Georgia Panayiotou
The University of Cyprus
Nicosia, Cyprus

Marios Theodorou
The University of Cyprus
Nicosia, Cyprus

ABSTRACT

Alexithymia is a trait reflecting a person's difficulty in identifying and expressing their emotions that has been linked to various forms of psychopathology. The identification of alexithymia might have therapeutic, preventive and diagnostic benefits. However, not much research has been done in proposing predictive models for alexithymia, while literature on multimodal approaches is virtually non-existent. In this light, we present, to the best of our knowledge, the first predictive framework that leverages multimodal physiological signals (heart rate, skin conductance level, facial electromyograms) to detect alexithymia. In particular, we develop a set of features that primarily capture spectral-information that is also localized in the time domain via wavelets. Subsequently, simple classifiers are utilized that can learn correlations between features extracted from all modalities. Via several experiments on a novel dataset collected via an emotion processing imagery experiment, we further show that (i) one can detect alexithymia in patients using *only one* stage of the experiment (elicitation of joy), and (ii) that our simpler framework outperforms compared methods, including deep networks, on the task of alexithymia detection. Our proposed method achieves an accuracy of up to 92% when using simple classifiers on specific imagery tasks. The simplicity and efficiency of our approach makes it suitable for low-powered embedded devices.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning approaches**.

KEYWORDS

affective computing, multimodal machine learning, alexithymia

1 INTRODUCTION

Alexithymia was firstly defined as the inability to describe emotions with words by Sifneos in 1972 [1]. Alexithymic individuals experience difficulties in recognising, identifying and describing emotions, and have a tendency towards focusing their thoughts on external stimuli rather than inner ones [2]. However, it is not considered a psychiatric disorder but rather as a personality trait which is associated with, and strongly predicts such disorders [3], but also physical and psychosomatic problems. In general, clinically relevant alexithymia affects approximately 10% of the general population [4].

Alexithymia is often assessed subjectively using self-reporting tools, using the the Toronto Alexithymia Scale (TAS-20). It is the most widely used self-reporting tool, hence considered as the gold standard in the field [5]. In the search to identify more objective markers of alexithymia, however, it has been noted that it is also characterized by differences in physiological reactivity towards emotional experiences, although research in the field has often resulted in inconsistent findings about the nature of these differences [6]. Even more, these physiological differences are often considered as central features contributing to the phenomenology of alexithymia, and perhaps also its etiology and maintenance, an assumption [7] that has not been empirically validated yet. To be able to identify the relative contribution of physiological markers during emotion processing to the core difficulties of alexithymia, research needs to employ advanced technological and statistical methods, that allow the simultaneous assessment of multiple signals, as they unfold dynamically over time, as can be accomplished with Machine Learning (ML) [8]. Furthermore, efficient predictive models can also lead to implementations in embedded systems and particularly wearables. This would have technical benefits such as reducing experimental time, monitoring patients remotely in real-time, measuring physiological signals that are less obtrusive, and working efficiently in comparison to more costly ML algorithms [9]. This line of research has also theoretical benefits such as increasing our understanding of alexithymic deficits, e.g. are these arousal-related (i.e. depend on emotional intensity reflected in autonomic nervous system (ANS) signals, like skin conductance) or valence-related (i.e. depend on the unpleasantness of a situation reflected in facial expression signals). Importantly, it can also have clinical interventions, as the identification and treatment of alexithymic difficulties may prevent the future development of disorders that are strongly correlated with it (i.e. depression).

In this light, we present, to the best of our knowledge, the first attempt to use multi-modal signals as predictors for alexithymia using ML, and one of the few that do this with any method. In particular, we present an approach for detecting alexithymia from five physiological signals, sampled from three different modalities as seen in Table 1 (electrocardiogram (ECG), electrodermal activity (EDA) and electromyography (EMG) [10]). In particular, we develop a novel set of features that capture spectral and temporal information from multimodal signals, whereas simple classification algorithms are subsequently used to capture cross-modal correlations in the feature space. We evaluate our framework on a study

*Both authors contributed equally to this research.

that used affective imagery to elicit emotions, aiming to identify differences between control and alexithymic volunteers using the TAS-20 [11]. We show that the proposed framework can lead to an accuracy of up to 92%, using data from only one small portion of the experiment - an insight that can lead to reduction of experimental time and cost. Most interestingly, we show that our approach consistently outperforms deep neural networks on the task of alexithymia prediction albeit being much more efficient, reaching an accuracy of up to 92%. The efficiency of the proposed approach can make it suitable for embedded systems and low-power devices, and can lead to improvements in the utilization of self-reporting tools.

2 RELATED WORK

Emotion recognition and the study of human emotional processes using physiological signals received increased attention recently [12]. The study of alexithymia through the lens of ML using physiological signals to characterize it is quite new, hence not many studies have been conducted for this matter. For this reason, this section presents studies that used physiological signals for emotion, depression, and anxiety detection.

ML approaches follow two main paths, a) classical ML algorithms with manual feature engineering approaches (i.e. hand-crafting) and b) Deep Learning (DL) models, where hierarchical, compositional representations are learned in a task dependent manner. Different feature engineering approaches exist to extract and select useful features. Studies used sequential forward and backward feature approaches [13] and manual extraction of features [14, 9]. Additionally, studies showed that the number of features needed to achieve high accuracy scores ranges between five and 14 features [13, 14, 9] for emotion recognition.

More recently, researchers applied DL algorithms to predict emotions, and components of emotions e.g. arousal, valence, dominance. Some of the most common DL algorithms used are: Deep Neural Network (DNN) [15, 16], Convolutional Neural Network (CNN) [15, 17, 16] and recurrent neural network (RNN) such as Long Short-Term Memory (LSTM) [18, 19, 16]. In [15], DNN and CNN models were developed, and in [18], a bimodal LSTM was developed for emotion recognition. Both studies achieved mean accuracy greater than 75% for the classification of valence and arousal. Other studies, developed models for anxiety detection by using different physiological signals. In [17], a 1D CNN was trained on ECG-based features to detect anxiety of arachnophobic individuals, reaching accuracy of 83.29%. In general, DL models outperformed the simpler, traditional algorithms as they are more likely to optimize extracted features [19, 16].

Furthermore, other studies extracted spectral features from the physiological signal, converted those features into images and then used pre-trained models for classification tasks. For example, both [20] and [21] used deep transfer learning for emotion recognition using physiological signals to classify arousal and valence. Nima et al. [8] was the only paper identified to predict alexithymia using ML but by recording a video with the facial expressions of the volunteers. Therefore, our model is the first model developed to predict alexithymia using physiological signals that include ANS indices of arousal, as well as facial responses that describe emotional valence.

3 DATASET

In this section, we describe the experimental conditions and settings under which the dataset has been collected. In more detail, we describe details of participants (Sec. 3.1), the settings (Sec. 3.2), and the data acquisition process (Sec. 3.3).

3.1 Subjects

Participants (N=52) were young, healthy adults, recruited from two universities in Cyprus. They were screened with the TAS-20, which was translated into Greek and validated by Anagnostopoulou and Kossieoglou [22]. The scale assesses on a 5-point scale difficulty identifying feelings, difficulty describing feelings, and externally oriented thinking. The total score was used to identify clinical levels of alexithymia using established standards of ≥ 60 for high alexithymia (M = 65.44, SD = 4.29) and ≤ 51 for low alexithymia (M = 40.37, SD = 6.59) [23]. Exclusion criteria for the study were any medical or mental health conditions that can interfere with ANS reactivity, e.g. cardiac problems, depression etc. and regular medication use. Of those eligible for the study, data from 52 participants was used (27 high alexithymia, 25 low alexithymia) due to data availability. The mean age was 21.31 (SD: 2.96). Of the 52 participants, eight identified as male and 44 as female. Half of the male participants were low alexithymic and the other half were high alexithymic. In terms of the female participants, 21 were low alexithymic and 23 were high alexithymic.

3.2 Experiment design

The experiment was conducted as a $2 \times 3 \times 2$ mixed design. That is, the population was split into *two* groups (high / low alexithymia), where each group imagined *three* emotions (fear / joy / neutral) in *two depths of processing* (shallow / deep) [22]. In total each participant completed ten imagery trials, four fear, four joy, two neutral, where half of them were under shallow processing task and the other half were under deep processing task. Dependent variables included: two signals that are consider to reflect levels of arousal (heart rate (HR) and skin conductance level (SCL)), and three signals, which reflects changes in experienced valence i.e. responses differentiate based on pleasantness vs. unpleasantnes feelings (orbicularis (ORB), corrugator (COR), zygomaticus (ZYG)) assessed during baseline and emotional imagery. Participants also provided self-reports oh how they felt (emotional labelling, valence, arousal and dominance ratings) at then end of each imagery trial. Details about the filtering of each physiological signal can be found in Constantinou et al. [11].

3.3 Data acquisition

3.3.1 Data sources. Physiological data were collected using BIOPAC MP150 for Windows and AcqKnowledge 3.9.0 data acquisition software (Biopac Systems Inc., Santa Barbara, CA). Electrodes were placed on the face and arms of the participants following standard procedures. Figure 1 demonstrates the placements of sensors for each signal.

3.3.2 Imagery materials. Ten standardized scripts, describing everyday fear, joy and neutral situations were selected from a larger pool of emotional scripts validated in the specific population [24].

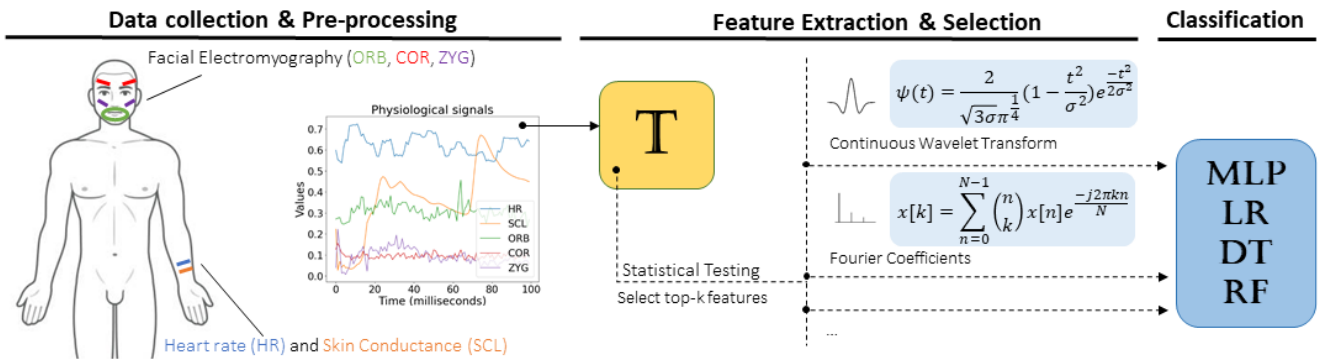

Figure 1: Overview of the proposed framework. Data collection of the five multi-modal physiological signals from several body locations. The data is extracted and pre-processed, while feature selection is done via statistical testing using the tsfresh package. The most commonly selected features are visualized - namely wavelet and Fourier coefficients. Finally, parametric (MLP, LR), and non-parametric models (DT, RF) are used for classification

Specifically, joy and fear scripts were selected so that they differed significantly on valence, but not on arousal, whereas neutral scripts differed from both joy and fear scripts on both dimensions. Affective imagery has been used effectively to induce emotions of varying valence and arousal levels [25], and these three categories of emotions allow us to examine valence and arousal effects on physiological reactivity independently. All participants received all scripts in three semi-counterbalanced orders. Scripts were in the first person, two-sentence long and contained references to physical reactions. Further details and examples of scripts can be found in Constantinou et al. [11].

3.3.3 Experimental protocol and set-up. Participants arrived at the lab, and sat in a reclining chair in a dark, sound-attenuated room. Following informed consent, they completed a brief set of questionnaires, and were given instructions and fitted with physiological monitors. Prior to the experiment, a five-minute adjustment period was executed to stabilise physiological recordings and familiarise participants with the equipment. Then, participants were firstly instructed with which depth of processing they will execute the trial, and secondly they were given an index card with the imagery script to memorise. They started to imagine the scripts at the tone cue. Shallow and deep instructions differed in the elements of imagery. For the shallow condition, participants were asked to imagine the scene as vividly as possible, and were told that at the end they had to recall and write down elements of the imagined environments (objects, people, animals). For deep processing, instructions guided participants to emphasise affective reactions and other subjective experiences. They were asked to imagine the scene as vividly as possible, as if they were actively participating, and later write down a summary of their reaction, including thoughts, behaviours, and/or bodily changes, whatever they considered necessary for an actor to exactly impersonate them. The information provided by each participant at the end of each trial (number of objects/people/animals for shallow processing and thoughts/behaviours/bodily changes for deep processing) were used as a manipulation check that participants engaged in the specific type of task. No participants were removed for not engaging in the tasks, and the two groups (alexithymic, control) did not differ in task performance (see Constantinou [11]).

Ten imagery trials followed, one for each of ten scripts [22]. Each trial consisted of three phases: 1) a 20s resting baseline, during participants were instructed to clear their mind and relax, 2) a 60s imagery phase-1, during participants were instructed to imagine the script that they read and 3) a 40s imagery phase-2, during participants were instructed to re-imagine the script. At the end of each phase-1, the participants were asked to report what they imagined depending on the depth of processing for the particular script.

4 METHODOLOGY

Nowadays, Deep Learning (DL) is mostly used because the developed network is able to learn and make intelligent decisions on its own. However, due to the high dimensionality of the signals, which is usually higher than the number of subjects (considering the same experimental conditions), DL results may be relatively poor. This was confirmed in our case by our results (Sec. 5). In this light, by means of statistical hypothesis time-series on multiple time-series features, we propose a set of spectral features for alexithymia detection (Sec. 4.2). In this way, we reduce the dimensionality of the problem to a few scalars per time-series signal. As we show by our experiments, the devised features are discriminative for alexithymia, reaching accuracy of up to 92%, using only *one* stage of the experiment e.g. Joy-Shallow-Phase-1. In this section, we discuss data pre-processing (Sec. 4.1), feature extraction and selection (4.2), the classification models employed (Sec. 4.3), as well as the validation method used for the algorithms (Sec. 4.4). Finally, in Figure 1 we provide an overview of the steps followed for this study.

4.1 Data pre-processing

The files from the BIOPAC software were imported into Python™ (v.3.8.8) for analysis. The extracted files contained several columns representing: time, HR, SCL, ORB, COR, ZYG, and digital channels indicating the stage of the experiment. The digital channels (i.e. stages of the experiment) were: phase-1, phase-2, arousal, valence,

tone-1 (shallow processing) and tone-2 (deep processing), and they were presented by binary representations. The phases were used to indicate the imagery period of the experiment which were: baseline (20sec), first imagery period (60sec) and second imagery period (40sec). The arousal and valence were used to indicate the emotion of the script used. Lastly, the tones were used to indicate the depth of processing of the experiment which was either shallow or deep. As aforementioned, this describes the way that subjects responded to what they imagined at the end of the experiment.

Table 1 defines the signals and their modalities.

Table 1: Modalities used to measure physiological signals

Modality	Definition	Physiological signal
ECG	Measures the potential differences identified at the skin surface due to electrical activity of the heart	HR
EDA	Measures the electrical conductivity of the skin	SCL
EMG	Measures the skeletal muscle electric activity at the skin surface. It is used for both facial or body expressions	ZYG, COR, ORB

All the signals were recorded with a sampling frequency of 1000 Hz, however phase-2 in some cases had sampling frequency of 125 Hz. Therefore, the first step was to resample phase-2 signals to 1000 Hz where necessary. The next step was to select only the data of interest that included the data recorded from the 10 trials. This means that any inter-trial intervals data was removed. This was indicated by the values of the digital signals and based on these values the data was broken down into different cases.

As shown in Figure 1, the raw data was then downsampled for each participant independently to improve memory complexity. The down-sampling was achieved by using the polyphase filter resampling [26] with down-sampling factor N = 300 [27]. This number is equal to three seconds, which is considered sufficient for emotional changes to be identified (verified empirically and with trials). The data was scaled according to the requirements of each algorithm. Additionally, the mean and standard deviation of each of the five signals for each participant were calculated to be used as a baseline in the binary classifiers.

4.2 Feature extraction and selection

In order to efficiently extract and select relevant features from the multivariate signals, the tsfresh python library was used [28]. Tsfresh extracts a large number of time-series features (794), in order to describe a time series dataset with respect to a target variable. Statistical hypothesis testing is employed in order to evaluate the discriminative performance and importance of extracted features for a specific task. In our case, the ten most relevant features were selected to be used as input into the classifier.

For each individual experiment, various tsfresh features were extracted, resulting in 3915 attributes for each subject. Using the target variable, the relevance of the aforementioned features was calculated in order to keep the top ten for each subject. Therefore, the resulting data structure was a matrix $X \in \mathbb{R}^{52\times10}$. The top features for the best stage of the experiment, which were Joy-Shallow-Phase-1 and Neutral-Shallow-Phase-1, are presented in Table 2.

Table 2: Top k=10 features extracted via tsfresh

Stage: Joy-Shallow- Phase-1
SCL regression[a], attribute: intercept, chunk length: 50, function: variance
SCL Continuous Wavelet Transform, coefficient: 3, width: 20
SCL Continuous Wavelet Transform, coefficient: 4, width: 20
ORB Fast Fourier Transform, attribute: real, coefficient: 49
SCL Continuous Wavelet Transform, coefficient: 2, width: 20
SCL Continuous Wavelet Transform, coefficient: 1, width: 20
SCL Continuous Wavelet Transform, coefficient: 5, width: 20
SCL Continuous Wavelet Transform, coefficient: 0, width: 20
SCL Continuous Wavelet Transform, coefficient: 6, width: 20
SCL Continuous Wavelet Transform, coefficient: 7, width: 20
Stage: Neutral-Shallow-Phase-1
SCL Continuous Wavelet Transform, coefficient: 0, width: 2
SCL Continuous Wavelet Transform, coefficient: 9, width: 5
SCL Continuous Wavelet Transform, coefficient: 8, width: 5
SCL Continuous Wavelet Transform, coefficient: 5, width: 2
SCL Continuous Wavelet Transform, coefficient: 6, width: 2
SCL Fast Fourier Transform, attribute: real, coefficient: 23
SCL Continuous Wavelet Transform, coefficient: 7, width: 2
SCL Fast Fourier Transform, attribute: angle, coefficient: 10
SCL Fast Fourier Transform, attribute: real, coefficient: 14
SCL regression[a], attribute: slope, chunk length: 5, function: variance

[a]Linear least-squares regression aggregated over chunks

Considering the top features from Table 2, `fft_coefficient` and `cwt_coefficients` were the most common features that appeared. Namely, the former calculates the discrete Fourier Transform coefficients [29], given by:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{i2\pi}{N}kn} \tag{1}$$

using the Fast Fourier Transform algorithm [30]. The coefficient for either the real, imaginary, magnitude or angle in degrees components of the expansion are extracted. On the other hand `cwt_coefficients`, performs a Continuous Wavelet Transform for the “Mexican hat wavelet" [31] given by:

$$\psi(t) = \frac{2}{\sqrt{3\sigma}\pi^{\frac{1}{4}}}\left(1-\frac{t^2}{\sigma^2}\right)e^{\frac{-t^2}{2\sigma^2}} \tag{2}$$

where σ is the scale factor. Empirical evidence suggests that the specific wavelet can describe a signal using a relatively small number of parameters, when compared to other wavelets [32]. The aforementioned features are extracted via a parallel feature selection algorithm based on statistical hypothesis tests such as the Mann-Whitney U [33] or Kolmogorov Smirnov [34]. These tests are

configured based on the label type (categorical or continuous) and the supervised ML problem (regression or classification) in hand[1].

4.3 Classification

Classification models were used in order to identify whether the volunteer is alexithymic or control. Specifically, deep neural networks (DNNs), decision trees (DT), random forests (RF), multilayer percpetrons (MLP) and logistic regression (LR) models with several hyper-parameter configurations.

Deep Neural Networks. With the exponential increase of computational power and availability of data, Deep Learning (DL) has become a vital tool for researchers, especially in medical applications and computer vision. As CWTs were amongst the most descriptive features, we decided to apply CNN architectures in order to classify the CWT features for both classes of subjects. This was also motivated by poor performance on raw features, especially with recurrent nets (e.g. LSTMs). Also, spectral features are commonly used for similar tasks, as discussed in Sec. 2.

Training DL models from scratch is a demanding process in terms of data volume. To alleviate poor performance, we utilised *transfer learning* [35]. That is, we used well established computer-vision DL models, already trained on large image datasets, by only modifying their feature extraction layer weights. The latter was achieved by training the specific layer on our dataset, for a small number of epochs, while keeping the rest of the weights fixed. Specifically, we used our CWT images to fine-tune the ResNet [36], DenseNet [37] and AlexNet [38], all of which have been pre-trained on the ImageNet dataset [39], and have performed extremely well in computer vision tasks.

Logistic Regression. LR is a classification algorithm. It is used to calculate (or predict) a binary (yes/no) event based on a set of independent variables. The model builds a regression model to predict the probability that a given data entry belongs to the category numbered as "1". LR models the data using the sigmoid function ($g(z) = \frac{1}{1+e^{-z}}$).

Decision Tree. This is an algorithm represented by a tree where nodes represent the features, leaves represent the outcomes and branches represent the decisions. The idea behind the DT algorithm is that the dataset is divided into smaller subsets based on the features until all the sample points get a final label. The particular algorithm uses gini impurity to decide the best split starting from the root node and further to the other subsequent splits, $Gini = 1 - \sum_{i=1}^{c} p_i^2$), where p_i is the probability of the class i in a node. Gini impurity selects the best possible split by measuring the quality of the split. Value zero is the lowest and best possible impurity. This is achieved when all the samples have the same label.

Random Forest. The RF algorithm is a type of parallel ensemble method. Ensemble methods are a set of techniques that combine multiple ML algorithms into one predictive model. This is done either to decrease bias (boosting), variance (bagging), or improve predictions (stacking). RF is an ensemble of decision trees, which means that RF builds several decision trees and combines them by a voting process to get a more stable and accurate prediction. RF falls in the family of bagging algorithms.

Multilayer Perceptron. This algorithm consists of multiple single perceptrons, where each consists of weights and bias, a combination function and an activation function. It deals with non-linearly separable data. The algorithm accepts n input variables, x, which are the multiplied with the weights w. Then, a sum of all these calculations is determined.

$$y = \sum_{i=1}^{n} w_i x_i + b \tag{3}$$

The activation function ϕ is then used to check the sum if it exceeded a certain threshold.

$$y = \begin{cases} 1, if \phi(w \cdot x + b) > 0 \\ 0, if \phi(w \cdot x + b) \leq 0 \end{cases} \tag{4}$$

4.4 Evaluation

A leave-one-subject-out cross-validation was used to evaluate the models. This approach utilizes each individual subject as a "test" set. It is a specific type of k-fold cross validation, where the number of folds, k, is equal to the number of participants in the dataset. We choose this approach to ensure that our results are *subject-independent* while exploiting the majority of the dataset, and to reduce the possibility of over-fitting.

5 EXPERIMENTS AND RESULTS

Two approaches were followed to classify alexithymic and control subjects ($n = 52$). The first approach used pre-trained deep neural networks and the second approach used extracted features and machine learning models. These two approaches showed that spectral features are informative for the specific classification.

The feature extraction process results showed that spectral representations constitute appropriate descriptors for alexithymia in physiological signals, demonstrating the most statistical significance. Additionally, the most informative physiological signal was SCL as shown in Table 2. Therefore, for the pre-trained DNN, only the SCL signal was used. As aforementioned, we fine-tuned three CNN architectures, namely DenseNet, AlexNet and ResNet. Table 3 demonstrates the average accuracy of the stages of the experiment using the networks. The majority of the results are around 50%, however Neutral-Shallow-Phase-1 achieved the highest score that was above 72.9 % in all three networks.

Regarding the results of the simpler ML algorithms, an accuracy of over 92% was achieved using both LR and MLP as shown in Table 4 (experiments Fear-Shallow-Phase-1 and Neutral-Shallow-Phase-1). In the majority of experiments, MLP and LR had comparable results - the best overall in terms of accuracy, followed by RF and subsequently DT with the worst performance. As shown in Table 4, Joy-Shallow-Phase-1 experiment had the best performance using MLP and LR in comparison to all the tests performed. Figure 2 shows the confusion matrix of both control and alexithymic groups, where they achieved a precision and a recall scores of 0.92. That is, 92% of the predicted alexithymic subjects were indeed alexithymic (precision). Further, the classifier correctly classified 92% of the alexithymic subjects as alexithymic (recall).

[1]https://tsfresh.readthedocs.io/en/latest/

Table 3: Accuracy of DNN models fine-tuned on 2D CWT spectograms derived from SCL signal

Depth	Phase	Emotion	AlexNet	DenseNet	ResNet
		Universal	0.579	0.496	0.438
		Universal[a]	0.579	0.496	0.438
		Fear[a]	0.546	0.529	0.558
Deep	1	Joy[a]	0.608	0.588	0.625
		Neutral[a]	0.533	0.529	0.554
		Fear[a]	0.546	0.562	0.617
Shallow	1	Joy[a]	0.588	0.571	0.600
		Neutral[a]	**0.738**	**0.788**	**0.729**
		Fear[a]	0.558	0.629	0.612
Deep	2	Joy[a]	0.550	0.462	0.612
		Neutral[a]	0.533	0.600	0.500
		Fear[a]	0.583	0.629	0.700
Shallow	2	Joy[a]	0.521	0.629	0.562
		Neutral[a]	0.629	0.646	0.642

[a]Scaling prior to feature extraction

Table 4: Accuracy results using LOSO cross-validation

Depth	Phase	Emotion	MLP	LR	DT	RF
		Universal	0.830	0.800	0.520	0.772
		Universal[a]	0.712	0.731	0.712	0.788
		Mean-Std Universal	0.588	0.500	0.692	0.538
		Fear[a]	0.788	0.769	0.750	0.750
Deep	1	Joy[a]	0.784	0.784	0.804	0.804
		Neutral[a]	0.735	0.592	0.653	0.694
		Fear[a]	0.827	0.846	0.635	0.846
Shallow	1	Joy[a]	**0.920**	**0.920**	0.820	0.840
		Neutral[a]	0.891	0.913	0.826	0.848
		Fear[a]	0.731	0.769	0.731	0.788
Deep	2	Joy[a]	0.780	0.760	0.800	0.800
		Neutral[a]	0.694	0.776	0.653	0.776
		Fear[a]	0.788	0.731	0.750	0.846
Shallow	2	Joy[a]	0.880	0.820	0.860	0.860
		Neutral[a]	0.870	0.826	0.761	0.848
		Mean	0.788	0.769	0.731	0.787
		Std	0.088	0.108	0.090	0.083

Figure 3 presents the receiver operating curve (ROC). The best performing model is MLP, achieving 0.96 AUC score.

As already mentioned, Joy-Shallow-Phase-1 was the experiment that achieved the best performance for classifying alexithymic and control participants. Figure 4 demonstrates the accuracy scores while using different number of features. MLP and LR achieved 92% accuracy score by only using the first four features. These features are the ones mentioned in Table 2 with ascending order.

Furthermore, literature suggested that spectral features in general (including CWT) are informative. Additionally, our feature extraction results showed that CWT is one of the most informative features. Therefore, CWTs of the required signals were plotted in order to visualise the cross correlation between the signal and the

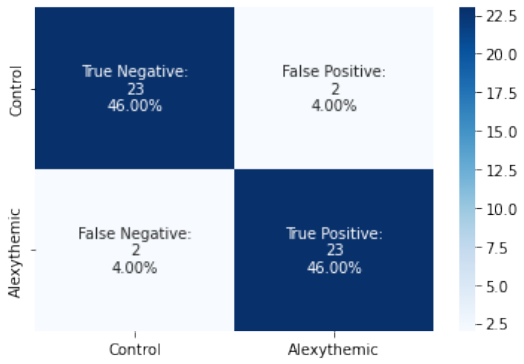

Figure 2: Confusion matrix of Joy-Shallow-Phase-1 experiment using LR and MLP, which had exactly the same performance.

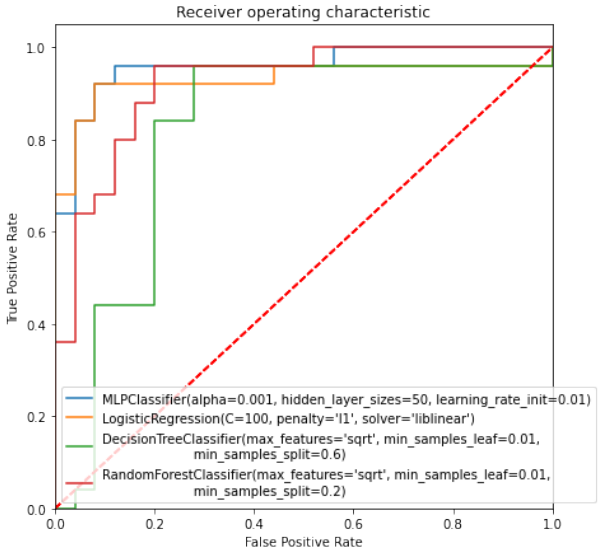

Figure 3: ROC of elicited joy (JSP1)

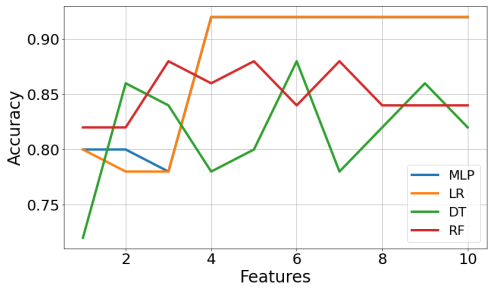

Figure 4: Accuracy for Joy-Shallow-Phase1 with 1:10 features

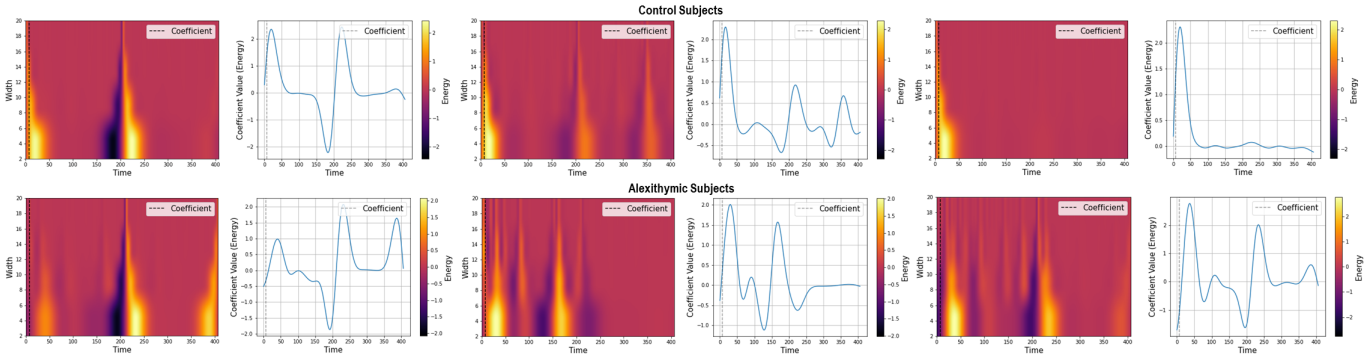

Figure 5: Visualization of the CWT using the Mexican hat wavelet for three Control (top) and three Alexithymic Subjects (bottom)

Mexican hat wavelet, at different widths of the transform. As demonstrated in Figure 5 brighter regions correspond to a stronger cross-correlation whereas darker regions indicate low cross correlation. Here, the most relevant features are the coefficients that correspond to the brightest regions, in combinations with the width of the transform where the best cross-correlations occur. For example, in the case of Joy-Shallow-Phase-1, these coefficients lie early in the signal and show a high cross-correlation with the Mexican hat wavelet of width 20. Additionally, the left part of the Figure 5 was used as the input into the pre-trained networks. DNNs were fine-tuned on data that had the same format as the left-hand side of the figures.

It is worth noting that the wavelet width and CWT coefficients identified by the feature extractors are *the same for all subjects* as the data transformation applied is global. By examining Figure 5 we notice that the coefficient (dashed line) is near the region of the first peak for the majority of the subjects. This suggests that the specific region contains valuable information regarding the presence of alexithymia. What is more, we also notice that the aforementioned peak is slightly shifted to the right for alexithymic subjects, as seen in Figure 5. This also implies that the specific region might be quite expressive.

6 DISCUSSION

In general, our results demonstrated high levels of accuracy for the binary classification for the simpler ML approaches, namely MLP and LR, especially when using data from the Joy-Shallow-Phase-1 and Neutral-Shallow-Phase-1. This can possibly be attributed to greater differences in emotion variation between control and alexithymic groups during this experiment. Parametric approaches (MLP & LR) have outperformed the non-parametric approaches (tree-based approaches) for Joy-Shallow-Phase-1, although there is no clear pattern of a best-performing model, as these methods vary from phase to phase. When comparing the mean accuracy we notice that MLPs and RFs slightly outperform the two other approaches, although the mean accuracy of all four methods is similar. It can be argued that MLP and LR performed better in the cases of interest due to the fact that only continuous features were identified as relevant during the feature extraction stage. Although tree based models tend to perform well in binary classification tasks, they usually require a larger sample size, compared to LR models. This might also be a reason behind the relatively good performance of LR in Joy-Shallow-Phase-1 and Neutral-Shallow-Phase-1. As expected, RF outperform DT, since RF constitute a more generalised solution. Considering the LR and MLP algorithms, in the majority of the experiments MLP showed greater performance than LR, which is again expected as it is a more expressive model.

Interestingly, the proposed approach outperforms deep networks on this task. In particular, three pre-trained models were fine-tuned to classify alexithymia (ResNet, AlexNet and DenseNet). Results suggest that CWT features extracted from the physiological signals can perform well in the specific transfer learning context. Highest accuracy was achieved in experiment Neutral-Shallow-Phase-1. Similarly, the specific experiment achieved the second highest score while using simple ML models. This indicates that the corresponding stage of the experiment can provide insight to a person's level of alexithymic traits. We note that experiments with LSTMs have been conducted using raw physiological signals, however performance was only slightly better than random.

7 CONCLUSION

We presented a predictive framework that leverages multimodal physiological signals to detect alexithymia in subjects. Namely, we develop a set of spectral-based features using statistical hypothesis testing, that are shown to be discriminative in terms of the alexithymia score of a given subject across a variety of methods. The proposed framework outperforms complex Deep Networks that have been pre-trained and utilized via transfer learning, achieving an accuracy of up to 92% when using simple classifiers. Furthermore, we find insights that can inform experiment design - such as two particular emotion elicitation experiments, namely joy and neutral, that are consistently easier to classify - indicating that the difference in emotion variation might be higher in these cases. The proposed approach is efficient, which is a desirable property for implementations in low-power embedded systems and sensors, while its simplicity can further lead to interpretable results.

REFERENCES

[1] Peter Emanuel Sifneos. 1972. *Short-term psychotherapy and emotional crisis.* Harvard University Press.

[2] Zohreh Khayyam Nekouei, Hamid Taher Neshat Doost, Alireza Yousefy, Gholamreza Manshaee, and Masoumeh Sadeghei. 2014. The relationship of alexithymia with anxiety-depression-stress, quality of life, and social support in coronary heart disease (a psychological model). en. *J. Educ. Health Promot.*, 3, (June 2014), 68.

[3] Mustafa Kemal Yontem and Kemal Adem. 2019. Prediction of the level of alexithymia through machine learning methods applied to automatic thoughts. en. *Psikiyatr. Guncel Yaklasimlar - Curr. Approaches Psychiatry*, 11, S1, (January 2019), 64–79.

[4] Katharina S. Goerlich. 2018. The multifaceted nature of alexithymia – a neuroscientific perspective. *Frontiers in Psychology*, 9. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2018.01614. https://www.frontiersin.org/article/10.3389/fpsyg.2018.01614.

[5] Danyelle Greene, Penelope Hasking, Mark Boyes, and David Preece. 2020. Measurement invariance of two measures of alexithymia in students who do and who do not engage in non-suicidal self-injury and risky drinking. *Journal of Psychopathology and Behavioral Assessment*, 42, 4, 808–825. ISSN: 1573-3505. DOI: 10.1007/s10862-020-09806-7. https://doi.org/10.1007/s10862-020-09806-7.

[6] Olivier Luminet, R. Michael Bagby, and Graeme J.Editors Taylor, editors. 2018. *Processing emotions in alexithymia: a systematic review of physiological markers. Alexithymia: Advances in Research, Theory, and Clinical Practice*. Cambridge University Press, 291–320. DOI: 10.1017/9781108241595.018.

[7] Christos Charis and Georgia Panayiotou, editors. 2018. *Alexithymia as a core trait in psychosomatic and other psychological disorders. Somatoform and Other Psychosomatic Disorders: A Dialogue Between Contemporary Psychodynamic Psychotherapy and Cognitive Behavioral Therapy Perspectives*. Springer International Publishing, Cham, 89–106. ISBN: 978-3-319-89360-0. DOI: 10.1007/978-3-319-89360-0_5. https://doi.org/10.1007/978-3-319-89360-0_5.

[8] Nima Farhoumandi, Sadegh Mollaey, Soomaayeh Heysieattalab, Mostafa Zarean, and Reza Eyvazpour. 2021. Facial emotion recognition predicts alexithymia using machine learning. *Computational Intelligence and Neuroscience*, 2021, 2053795. ISSN: 1687-5265. DOI: 10.1155/2021/2053795. https://doi.org/10.1155/2021/2053795.

[9] Bohdan Myroniv, Cheng-Wei Wu, Yi Ren, Albert Christian, Ensa Bajo, and Yu-chee Tseng. 2017. Analyzing user emotions via physiology signals. *Data Science and Pattern Recognition*, 2, (December 2017).

[10] Thomas Christy, Ludmila I Kuncheva, and Kerry W Williams. 2012. Selection of physiological input modalities for emotion recognition. *UK: Bangor University*.

[11] Elena Constantinou, Georgia Panayiotou, and Marios Theodorou. 2014. Emotion processing deficits in alexithymia and response to a depth of processing intervention. en. *Biol. Psychol.*, 103, (December 2014), 212–222.

[12] Fabien Ringeval, Björn Schuller, Michel Valstar, Nicholas Cummins, Roddy Cowie, Leili Tavabi, Maximilian Schmitt, Sina Alisamir, Shahin Amiriparian, Eva-Maria Messner, Siyang Song, Shuo Liu, Ziping Zhao, Adria Mallol-Ragolta, Zhao Ren, Mohammad Soleymani, and Maja Pantic. 2019. Avec 2019 workshop and challenge: state-of-mind, detecting depression with ai, and cross-cultural affect recognition. In (AVEC '19). Association for Computing Machinery, Nice, France, 3–12. ISBN: 9781450369138. DOI: 10.1145/3347320.3357688. https://doi.org/10.1145/3347320.3357688.

[13] Vitaliy Kolodyazhniy, Sylvia D Kreibig, James J Gross, Walton T Roth, and Frank H Wilhelm. 2011. An affective computing approach to physiological emotion specificity: toward subject-independent and stimulus-independent classification of film-induced emotions. en. *Psychophysiology*, 48, 7, (July 2011), 908–922.

[14] Cheng He, Yun-jin Yao, and Xue-song Ye. 2017. An emotion recognition system based on physiological signals obtained by wearable sensors. In *Wearable Sensors and Robots*. Canjun Yang, G. S. Virk, and Huayong Yang, editors. Springer Singapore, Singapore, 15–25. ISBN: 978-981-10-2404-7.

[15] Soorena Salari, Amin Ansarian, and Hajar Atrianfar. 2018. Robust emotion classification using neural network models. In *2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, 190–194. DOI: 10.1109/CFIS.2018.8336626.

[16] Patrícia J. Bota, Chen Wang, Ana L. N. Fred, and Hugo Plácido Da Silva. 2019. A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. *IEEE Access*, 7, 140990–141020. DOI: 10.1109/ACCESS.2019.2944001.

[17] Adrian Vulpe-Grigoraşi and Ovidiu Grigore. 2021. A neural network approach for anxiety detection based on ecg. In *2021 International Conference on e-Health and Bioengineering (EHB)*, 1–4. DOI: 10.1109/EHB52898.2021.9657544.

[18] Hao Tang, Wei Liu, Wei-Long Zheng, and Bao-Liang Lu. 2017. Multimodal emotion recognition using deep neural networks. In (October 2017), 811–819. ISBN: 978-3-319-70092-2. DOI: 10.1007/978-3-319-70093-9_86.

[19] Yuyang Tan, Qinwei Zeng, and Hongchang Zhang. 2021. Research on anxiety detection based on personalized data markers. *Journal of Physics: Conference Series*, 1948, 1, 012035. DOI: 10.1088/1742-6596/1948/1/012035. https://doi.org/10.1088/1742-6596/1948/1/012035.

[20] Fatih Demir, Nebras Sobahi, Siuly Siuly, and Abdulkadir Sengur. 2021. Exploring deep learning features for automatic classification of human emotion using eeg rhythms. *IEEE Sensors Journal*, 21, 13, 14923–14930. DOI: 10.1109/JSEN.2021.3070373.

[21] Rayan Elalamy, Marios Fanourakis, and Guillaume Chanel. 2021. Multi-modal emotion recognition using recurrence plots and transfer learning on physiological signals. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 1–7. DOI: 10.1109/ACII52823.2021.9597442.

[22] Georgia Panayiotou and Elena Constantinou. 2017. Emotion dysregulation in alexithymia: startle reactivity to fearful affective imagery and its relation to heart rate variability. en. *Psychophysiology*, 54, 9, (September 2017), 1323–1334.

[23] James D A Parker, Graeme J Taylor, and R Michael Bagby. 2003. The 20-item toronto alexithymia scale. en. *J. Psychosom. Res.*, 55, 3, (September 2003), 269–275.

[24] Georgia Panayiotou. 2008. Emotional dimensions reflected in ratings of affective scripts. *Personality and Individual Differences*, 44, 8, 1795–1806. ISSN: 0191-8869. DOI: https://doi.org/10.1016/j.paid.2008.02.006. https://www.sciencedirect.com/science/article/pii/S0191886908000652.

[25] C V Witvliet and S R Vrana. 1995. Psychophysiological responses as indices of affective dimensions. en. *Psychophysiology*, 32, 5, (September 1995), 436–443.

[26] Chia-Chuan Hsiao. 1987. Polyphase filter matrix for rational sampling rate conversions. In *ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Volume 12, 2173–2176. DOI: 10.1109/ICASSP.1987.1169404.

[27] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261–272. DOI: 10.1038/s41592-019-0686-2.

[28] Maximilian Christ, Nils Braun, Julius Neuffer, and Andreas W Kempa-Liehr. 2018. Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing*, 307, 72–77.

[29] Gilbert Strang. 1994. Wavelets. *American Scientist*, 82, 3, 250–255.

[30] E. O. Brigham and R. E. Morrow. 1967. The fast fourier transform. *IEEE Spectrum*, 4, 12, 63–70. DOI: 10.1109/MSPEC.1967.5217220.

[31] Norman Ricker. 1951. The form and laws of propagation of seismic wavelets. In *3rd World Petroleum Congress*. OnePetro.

[32] Afshin Gholamy and Vladik Kreinovich. 2014. Why ricker wavelets are successful in processing seismic data: towards a theoretical explanation. In *2014 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES)*. IEEE, 11–16.

[33] Patrick E McKnight and Julius Najab. 2010. Mann-whitney u test. *The Corsini encyclopedia of psychology*, 1–1.

[34] Frank J Massey Jr. 1951. The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46, 253, 68–78.

[35] Sinno Jialin Pan and Qiang Yang. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22, 10, 1345–1359.

[36] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

[37] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

[38] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

[39] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: a large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.