

CAFS: Cost-Aware Features Selection Method for Multimodal Stress Monitoring on Wearable Devices

Niloofer Momeni, *Member, IEEE*, Adriana Arza Valdés, *Member, IEEE*, João Rodrigues, Carmen Sandi, and David Atienza, *Fellow, IEEE*

Abstract—Objective: Today, stress monitoring on wearable devices is challenged by the tension between high-detection accuracy and battery lifetime driven by multimodal data acquisition and processing. Limited research has addressed the classification cost on multimodal wearable sensors, particularly when the features are cost-dependent. Thus, we design a Cost-Aware Feature Selection (CAFS) methodology that trades-off between prediction-power and energy-cost for multimodal stress monitoring. **Methods:** CAFS selects the most important features under different energy-constraints, which allows us to obtain energy-scalable stress monitoring models. We further propose a self-aware stress monitoring method that intelligently switches among the energy-scalable models, reducing energy consumption. **Results:** Using CAFS methodology on experimental data and simulation, we reduce the energy-cost of the stress model designed without energy constraints up to 94.37%. We obtain 90.98% and 95.74% as the best accuracy and confidence values, respectively, on unseen data, outperforming state-of-the-art studies. Analyzing our interpretable and energy-scalable models, we showed that simple models using only heart rate (HR) or skin conductance level (SCL), confidently predict acute stress for $HR > 93.30BPM$ and non-stress for $SCL < 6.42\mu S$, but, outside these values, a multimodal model using respiration and pulse wave's features is needed for confident classification. Our self-aware acute stress monitoring proposal saves 10x energy and provides 88.72% of accuracy on unseen data. **Conclusion:** We propose a comprehensive solution for the cost-aware acute stress monitoring design addressing the problem of selecting an optimized feature subset considering their cost-dependency and cost-constraints. **Significant:** Our design framework enables long-term and confident acute stress monitoring on wearable devices.

Index Terms—Stress monitoring, cost-aware machine learning, cost-constraints feature selection, low-power wearable devices.

I. INTRODUCTION

IN today's society, we are constantly experiencing acute stress from our way of living and working. The "acute stress" term is often referred to the non-specific response of the organism to adapt to external demands for change that challenges its capabilities and resources [1]. On the one hand, being exposed to high levels of acute stress decreases humans' performance and the impaired cognitive workload can even lead to accidents in critical tasks [2]. On the other hand, in the long term, frequently high levels of stress become 'toxic'

for both physical and mental health [2], [3]. That is why every person could benefit from continuous monitoring of their stress reactivity in daily life [3]. It could allow the identification of acute stress episodes, then triggering preventive or corrective action. In the case of safety-critical tasks (e.g., pilots, fire-fighters, and rescuers), identifying high levels of acute stress can prevent catastrophic accidents. However, a reliable stress detection still represents a common major challenge since it is not directly observable [4] nor a monolithic concept [3].

The analysis of stress reactivity by its physiological response (i.e., sympathetic arousal) that considers multiple signals, i.e., multimodal, can be used for reliable, non-intrusive, and continuous acute stress monitoring [4]–[11]. Wearable and low-power edge-computing technologies, together with machine learning techniques, facilitate real-time, multimodal, and continuous stress reactivity monitoring on wearable devices. However, several challenges must be overcome to fully implement a stress detection method on wearable devices, including safety, security and privacy, memory usage, battery lifetime, reliability, etc. Today, battery lifetime is a major constraining factor for wearable technology due to the restrictive battery factors of size and shape that ensure portability and wearability. Besides, it is especially challenging on a multimodal wearable platform since a significant part of the energy is spent by the power-hungry bio-sensors. It is within this context that, in this study, we focus on designing and developing a low-power stress monitoring framework suitable for accurate long-term monitoring of stress reactivity.

State-of-the-art studies on multimodal monitoring systems do not address the battery lifetime issue directly on the machine learning detection models design [9]–[11]. To the date, such studies do not differentiate between the physiological features that feed the models in terms of energy consumption. Their multimodal machine learning models are trained to accurately predict an output and traditional feature selection algorithms are used to select the most informative features without considering the cost of individual features. Thus, they assign equal weights to features of different costs and priorities. However, in reality, sensors and biosignal processing algorithms bear their costs and increase the complexity and energy consumption of the edge device.

On multimodal monitoring systems, the energy cost of computing a single physiological feature is composed of three actions: i) signal acquisition by the sensor, ii) bio-parameter calculation that includes signal filtering and delineation, and iii) feature extraction algorithm. Note that from one signal, different bio-parameters can be extracted and, subsequently, different features. Thus, if two features from one specific signal are selected, then the energy cost of the sensor needs to

This work was supported by the NCCR Robotics through the Symbiotic Drone project and Synapsy [51NF40-125759, -158776 and -185897], by the Swiss National Science Foundation (NSF) through the ML-Edge research (GA No. 200020182009/1) projects and the grants [31003AB-135710, 176206], by the ONR-G Grant (N62909-20-1-2063) and by the Oak Foundation to CS.

N. Momeni, A. Arza and D. Atienza are with the Embedded Systems Laboratory (ESL) and J. Rodriguez and C. Sandi are with the Laboratory of Behavioral Genetics (LGC) from Swiss Federal Institute of Technology in Lausanne (EPFL), 1015 Lausanne, Switzerland, e-mail: niloofer.momeni@epfl.ch. N. Momeni is also with the Department of Mathematical Statistics, Lund University, Sweden. e-mail: niloofer.momeni@matstat.lu.se

be counted only once, similarly if they come from the same bio-parameter. The cost of each feature is not a fixed value and depends on what other features have been selected. Thus, we propose a Cost-Aware Feature Selection (CAFS) algorithm that automatically considers the cost-dependencies variation in this work. It selects a low-energy feature set that is sufficient to provide an accurate and reliable acute stress prediction such that a cost-constraint is satisfied.

Our CAFS methodology allows exploiting the trade-off between energy-cost and prediction-power of individual features. Changing the cost-constraint provides different sets of selected features, hence different stress detection machine learning models with different confidence levels. Thus, we can have a complex model that makes confident predictions under no/weak cost-constraint at the expense of high energy consumption, and a simpler low-power model that makes less accurate predictions under strong cost-constraint (e.g., very limited battery lifetime). We further propose a self-aware stress monitoring framework that, when needed, toggles between different stress detection models to reduce energy consumption without sacrificing the system quality. The methodology is evaluated with experimental data and simulation.

Our main contributions are summarized as follows:

- We propose CAFS, a novel cost-aware feature selection methodology applicable to any supervised classification problem using cost-dependent features, which aims to reduce any form of classification cost like energy consumption, computation time, complexity, memory footprint, etc.
- We develop reliable and energy-scalable machine learning models for acute stress monitoring based on multimodal physiological signals that are interpretable and generalizable for our experimental condition, obtaining the best accuracy of 90.98% and prediction confidence of 95.74% on the unseen dataset, which is higher than the latest state-of-the-art studies.
- We show how to minimize resource usage and complexity of the used machine learning model applying our CAFS resulting in up to 94.37% energy saving with 6.76% and 34.44% loss in accuracy and prediction confidence, respectively, compared with a non-energy-constrained model.
- We introduce a self-aware machine learning framework for dynamic energy management on wearable devices that allows 89.25% of energy reduction with similar accuracy and 23.54% drop in the prediction confidence concerning the non-energy-constrained model in our study case.

II. RELATED WORK

A. Multimodal Acute Stress Monitoring on Wearable Devices

Acute stress situations provoke a body physiological stress response that triggers several reactions orchestrated by the autonomic nervous system, including sweating skin, increased heart rate, and increased respiratory frequency. Those reactions can be measured on several physiological signals with wearable sensors, including Respiration (RSP), Electrocardiogram (ECG), Photoplethysmogram (PPG), Electrodermal Activity (EDA), and skin temperature (SKT) [8]. It has been shown that a robust acute stress prediction needs fused information from multiple modalities (signals) [3], [4], [7], [12], [13].

Multimodal acute stress monitoring based on physiological signals has been previously studied by adopting different

machine learning algorithms such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), XGBoost (eXtreme Gradient Boosting), and Naïve Bayes trained on different combinations of biosignals [4]–[7]. Most of the reported works assess their machine learning framework on training or cross-validation datasets [5], [14], without proving their generalization power [15]. In contrast, few other studies evaluate their stress and cognitive workload detection on new unseen data reporting an accuracy range of 60 – 86% [9], [10], [16], [17].

However, multimodal machine-learning approaches are complex and their deployment on wearable systems is limited due to the constraints of memory, energy consumption, and duty cycles [7]. Even though, for such multimodal applications, processing on the edge (i.e., edge-computing) when comparing to cloud computing is still advantageous in terms of latency, bandwidth cost, data safety and privacy, and battery lifetime [6], [18], [19]. To overcome those constraints, efforts have mainly focused on improvements for the hardware platform (sensors and micro-controllers) [6] but little on designing a cost-aware machine learning model. Previous studies have mainly focused on using simpler models and reduced features set where all the features' costs are assumed to be equal without considering their cost-dependencies [9]–[11].

Therefore, cost-aware machine learning design methods for multimodal applications on the edge are needed to reduce cost while maximizing the application classification performance. Furthermore, such design should consider the cost of selected features on the targeted hardware platform and count for their energy-cost dependencies.

B. Cost-Aware Machine Learning Design Techniques for Multimodal Applications on the Edge

The cost-aware design of machine learning models has been extensively studied [20]–[23]. This optimization problem is divided in techniques for reducing time and models' cost on training (i.e., cost-sensitive learning) and prediction (i.e., cost-constrained test). In the context of the internet of things (IoT) and wearable devices that target edge-computing, most proposals aim to reduce the classification cost by employing feature selection techniques to reduce the features number [16], [24], [25]. However, they select features based on their computational complexity and chosen thresholds defined by classification performance or specific application constraints. Such approaches may result in a sub-optimal feature set of lower cost and do not count for features' cost-dependencies. In other domains, several works discuss the trade-off between features' cost and classification accuracy by using a lost function that integrates features cost and classification error [24], [26]–[28]. Nevertheless, methods employing a lost function are not designed for multimodal classification but rather for making a prediction based on data where the cost of the selected features is calculated by adding an individual feature's cost, e.g. data of only one sensor. Similarly, other heuristic methods that try to solve the optimization problem of accuracy vs cost [23], [29] do not consider the cost-dependency of the selected features.

On the other hand, [30] proposes a mixed-integer linear programming model based on SVM to incorporate the variable acquisition costs in the feature selection procedure. Recently, a study in [31] extends the work in [30] by incorporating

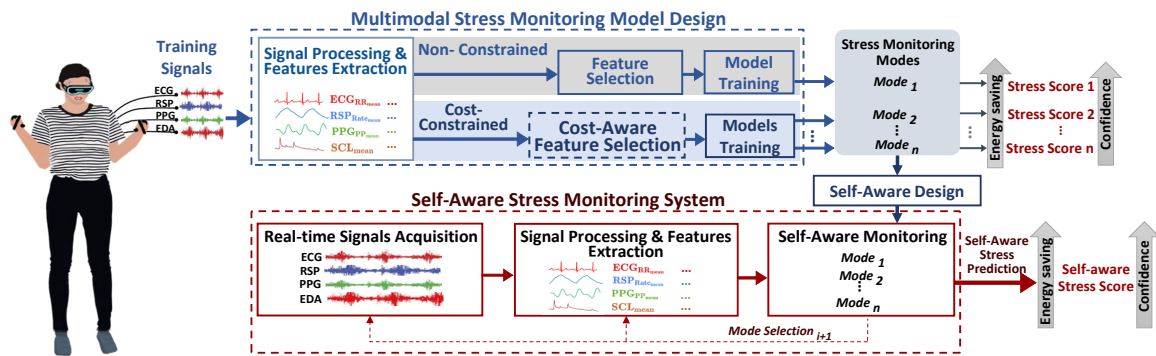


Figure 1: Multimodal acute stress monitoring model design (top) for energy-scalable modes starting from a more confident mode but demanding more energy to a less confident mode but with lower energy cost. Self-aware acute stress monitoring system (bottom) bringing energy savings while keeping high the prediction confidence.

a budget constraint to preserve classification accuracy with the least expensive features. In [23] it is proposed a greedy forward feature selection to also satisfy a budget constraint. Although these previous proposals present valuable ideas for a cost-aware feature selection of multimodal data, they do not consider the cost-dependency variation of different features. Still, there is a lack of machine learning techniques that account for the energy consumption of wearable devices on multimodal applications by integrating the cost of each feature and sensor without decreasing the detection performance.

III. COST-AWARE MULTIMODAL ACUTE STRESS MONITORING FOR WEARABLE DEVICES

The framework design of our proposed low-power, wearable and multimodal stress monitoring methodology is shown in Fig. 1. The biosignals used are ECG, RSP, PPG, and EDA. Firstly, we present a method for the multimodal stress monitoring model design (shown in blue) that has three conceptual steps namely, signal processing and feature extraction, feature selection, and model training, which are detailed in Section IV. First, signals are filtered, delineated and different parameters are obtained to extract several physiological features that represent the physiological stress response of the human body. Next, in the feature selection step, we explore the contribution of a wide variety of physiological features to the stress detection model. Finally, a machine learning algorithm is trained and tuned to obtain the stress monitoring model.

In the feature selection step, features are selected by eliminating the irrelevant and redundant ones for training a good predictor, firstly without cost-constraints (highlighted in grey) and then, including energy-constraints in our design (highlighted in blue). The design with no constraints uses a common feature selection algorithm to obtain the full and complex stress classification mode. Whereas the cost-aware design uses our novel CAFS methodology that allows the design of monitoring modes with different complexity, depending on the weak/strong cost-constraints. CAFS trades-off between features' cost and prediction-power by solving an integer linear programming problem, see Section V-A.

Using CAFS, weak/no energy-constraints assumption allows free selection of most important features (that can be either energy-expensive or not) and yields a more confident and accurate acute stress detection model. Whereas, stronger constraints on the energy consumption restrict the selection

of features and, as a consequence, lead to more errors in the model, see Section V-B. In the CAFS methodology, we start with a no constraint features selection and then tighten the constraints using an iterative algorithm in every step until we obtain the least expensive set of features for stress detection. Thus, we obtain different energy-scalable monitoring modes from a more accurate mode but that demands more energy to a budget mode that is less accurate but saves more energy.

Finally, we make use of the energy-scalable stress monitoring modes and apply the self-awareness concept as in [32]–[34] to improve both the energy efficiency and stress detection confidence of the proposed system. The overall flow of our self-aware stress detection is shown in red in Fig. 1. In the real-time self-aware monitoring of acute stress, the system triggers between a budget mode and a complex mode for the next segmentation window ($i+1$) whenever a predefined confident threshold is not reached. The selected mode for the next window defines which signals and processing algorithms are going to be activated. In fact, a more complex mode is not used unless the previous prediction is not confident enough using a simple budget mode. As a result, the system is running a simpler but confident mode with less energy consumption. A detailed description of the self-aware stress detection mechanism and its design is provided in Section VI.

Our proposed framework is trained and validated on an experimental physiological data of subjects who went under a virtual reality acute stress experiment, see Section VII-A. The machine learning stress detection models for every monitoring mode, from 1 to n , and also in self-aware modes are evaluated using an unseen test set from the experimental dataset. Moreover, we interpret and describe the decisions taken by our stress detection models to make predictions.

IV. DESIGN OF A NON-CONSTRAINED MULTIMODAL ACUTE STRESS MONITORING MODEL

We first look at a general machine learning algorithm that can reliably detect acute stress with no energy constraints. Its designing steps are described hereafter.

A. Signal Processing and Feature Extraction

For each of the aforementioned physiological signals, we extract several physiological parameters based on the latest state-of-the-art results [8], [16], presented in Fig. 2. These parameters are segmented in 60-second-length sliding windows.

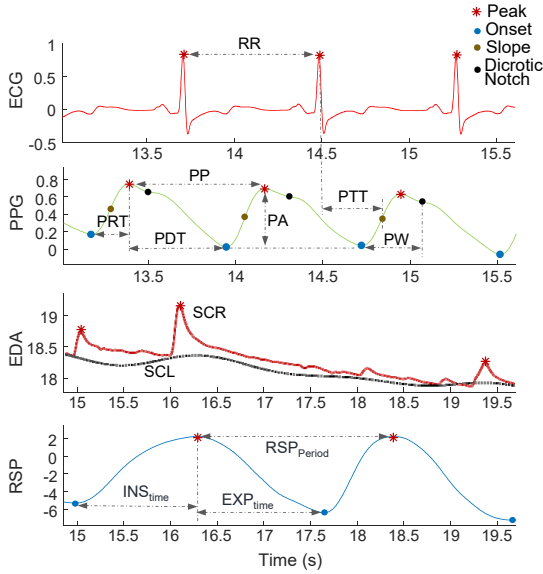


Figure 2: Bio-parameters extracted from ECG and PPG signals

Then, for each window, an initial set of features is obtained using time-domain and frequency-domain analysis as in [8] and [16]. The initial considered parameters and features from each biosignals are as follows:

1) *ECG*: *RR* intervals are extracted from the ECG signals. From the *RR* interval, several features on time and frequency domain are extracted based on the Heart Rate Variability (*HRV*) analysis [35]. Non-linear features are also extracted from Poincaré plot indicating the vagal and sympathetic functions from both *RR* interval series and its second-order difference (RR'). They are the following: the length of the transverse axis ($SD1$); the length of the longitudinal axis ($SD2$); the ratio $SD2/SD1$, called Cardiac Sympathetic Index (*CSI*); and the Cardiac Vagal Index (*CVI*) as $\log_{10}(SD2 \cdot SD1)$ [35]. Also, a modified *CSI* as RR'_{SD2}/RR_{SD1} .

2) *PPG*: Several biomarkers are computed, and they are shown in Fig. 2. They are pulse period (*PP*), pulse amplitude (*PA*), pulse wave rising time (*PRT*). Slope of the Pulse Wave (k): The slope transit time divided by the difference in amplitude between 1/4 and 3/4 of the pulse wave. From these biomarkers, same as the *RR*, several features are extracted.

3) *RSP*: We compute the respiration rate (F_R), respiration period (RSP_{Period}), inspiration and expiration time (INS_{time} , EXP_{time}), ratio of inhalation to exhalation duration (*IE* Ratio). Then, in the frequency domain, we compute the segmented signal power in the classical ECG high frequency (HF_{ecg}) band (0.15 – 0.40 Hz) and in five different bands with equal bandwidth within the HF_{ecg} band range ($HF_{ecg} - pF_{mn}$). These power values are normalized by the total power in the 0 – 1Hz band. Furthermore, we extract two features from the HF band (0.15 – 0.5 Hz). First, the mean frequency of a Gaussian distribution used to fit the power spectral density (*PSD*) estimated in the HF band (HF_{gauss}). This feature describes the *PSD* shifting in frequency mainly caused by respiratory activity [36]. The second one is the ponderate frequency mean of the HF band (HF_{pond}). Moreover, for each segmentation window, we compute as proposed in [37] the estimated respiratory frequency eRF being the largest peak power (P_k) of the Lomb-Scargle *PSD*

of respiration. The power around the peak is computed in a 0.04 Hz bandwidth divided by the total power ($power_{tot}$). This normalized respiratory peak power ($power_{norm}$) represents the RF variability within the interval.

4) *EDA*: The EDA signal is divided into two main components, namely, the skin conductance level (*SCL*) and the skin conductance response (*SCR*).

B. Feature Selection

The feature selection techniques aim to reduce the possible overfitting and the model's complexity. Hence, memory usage and power consumption are both minimized [33], [38]. In this regard, we use Recursive Feature Elimination with Cross-Validation (RFECV) proposed in [39], which recursively eliminates the least important features in a loop without losing classification performance. In RFECV the features are ranked by a permutation importance measure, we propose to use instead SHAP features importance [40]. The SHAP importance value is able to enforce consistency and accuracy more than the permutation approach [41]. The most informative features are selected when the RFECV-SHAP curve arrives at the best accuracy and enters in a relatively steady state.

C. Model Training, and Generalization

Features extracted from the physiological signals have variety of interactions, which requires a strong learner model that captures such interactions. Hence, for both classification and feature selection, we consider XGBoost algorithm since it outperforms alternative ones used for similar problems [10], [42]. XGBoost [43], is a tree-based model and is an optimized version of the gradient boosting machine.

Model's hyper-parameters that have more effect on the performance and complexity of the XGBoost model are the number of trees n_{tree} , and the maximum depth of trees d_{tree} [43]. A larger n_{tree} means a more complex model and more time to run but results in a smooth and better prediction [42]. Deeper trees result in a fewer number of trees but may lead to overfitting [43]. Therefore, for further deployment of the model on a resource-constrained system, we perform a model's hyper-parameter tuning. Grid-Search-CV [39] is carried out on the n_{tree} and d_{tree} to search for their optimized values that correspond to the maximum CV accuracy.

The maximum number of nodes n_{node} is calculated by $n_{node} = n_{tree} * (2^{d_{tree}} + 1)$. Note that after tuning hyper-parameters, some features can get zero importance, specifically if the model has high variance and a tendency to overfit. Therefore, a second RFECV-SHAP is applied after tuning to remove the zero important ones. Finally, the generalization power of the stress detection model on our experimental conditions is evaluated on the unseen test set.

V. COST-AWARE FEATURE SELECTION METHODOLOGY FOR MULTIMODAL STRESS MONITORING: CAFS

In the previous section, we present the general design of a non-constrained multimodal stress monitoring model using a non-constrained feature selection algorithm, i.e., RFECV-SHAP. The model decides which features to take to accurately classify stress without considering the energy cost of each one. Therefore, hereafter, we propose the CAFS methodology to consider how likely each feature can accurately detect stress

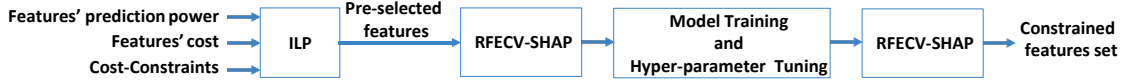


Figure 3: Cost-aware features selection methodology

while also limiting the energy consumption.

To strike the right balance between the cost of selected features and their overall prediction power, an exhaustive search is generally impractical and computationally intractable. Moreover, we need to consider the cost dependencies of the features. Therefore, our proposed CAFS methodology consists of a four-phase process, as shown in Fig. 3. First, an initial pre-set of the more important features that satisfies the cost constraints and maximize the prediction power are selected by solving an Integer Linear Programming (ILP) [44] problem. The ILP allows to formulate our problem considering the benefit, cost and cost dependencies of the features.

After solving the ILP problem, a pre-set of the more important features that satisfy the constrained cost is obtained. Then, from this pre-set we select a subset of the most important features that gives the highest possible CV-accuracy score using the RFECV-SHAP algorithm. Next, a new dataset with a selected subset of features is considered for the model training and hyper-parameter tuning. Finally, a second RFECV-SHAP is applied to avoid having features with zero importance, as described in Section IV-C.

In our CAFS methodology, we need to measure two feature's attributes i.e., cost and benefit. The feature's cost is the energy it consumes to be extracted from a physiological signal, and the benefit is its contribution to the model prediction power. The details of these measurements are as follows.

1) Energy-Cost of Features

The total energy consumption of the embedded stress detection model is composed of the energy consumption of signal acquisition by sensors, bio-parameters extraction algorithms, and features calculation. A large part of the energy is consumed by the sensors, and is determined by their type and sampling rate [6]. The energy-cost calculation of the bio-parameters and features extraction depends on their complexity, state-of-the-art implementation of the algorithms and the processing unit selected.

The energy consumed to acquire a specific signal from a sensor, e.g., ECG sensor, should be counted only once for all the features extracted from that signal, e.g., ECG features. Similarly, the energy consumed to get a bio-parameter should be accounted for all the features derived from that parameter. Therefore, the total energy consumption E_{total} consists of three energy values: sensors' energy, parameter extraction algorithms' energy, and features computation's energy. They are respectively defined as: $E_s = [E_s^{(1)} \dots E_s^{(k)}]$, $E_p = [E_p^{(1)} \dots E_p^{(l)}]$, and $E_f = [E_f^{(1)} \dots E_f^{(m)}]$, where k is the number of sensors used. l is the number of different parameters extraction steps performed (e.g., psd, poincaré plot, scl,...), and m is the features number.

2) Features' Prediction-Power

We choose SHAP importance values (I) to measure the feature's benefit to the target prediction in our CAFS algorithm. Since I is the average contributions of a feature across all the possible combinations of features, it is neither

dependent on the other features nor model specific. According to [40], the I of feature j is the collective of absolute Shapley values ϕ of that feature over the data, which is written by $I_j = \sum_{i=1}^n |\phi_j^{(i)}|$. Since ϕ values tell us how the prediction is distributed among the features for each observation, their collective value over all the observations, which is I , shows the prediction power (benefit) of each feature. Hence, adding up I of selected features represents model's prediction-power. Maximizing prediction power subject to the total costs be less than the budget energy describes our optimization problem.

A. Cost-Aware Feature Selection Algorithm

Our goal is to maximize the prediction-power without exceeding the energy budget limits of features' cost. The general form of our problem is written as follows:

$$\begin{aligned} & \text{maximize} && \text{prediction-power} \\ & \text{subject to} && \text{total energy consumption} \leq \text{energy budget}, \end{aligned} \quad (1)$$

In the first phase of our proposed CAFS approach, we assume we can either take a feature or not. Therefore, this phase is expressed as an ILP problem in which a variable is restricted to be an integer. Our ILP problem has three parts:

1) *The selection vector:* $x_f = [x_f^{(1)} \dots x_f^{(m)}]^T$, a m -dimensions (number of features) vector of 1's and 0's, i.e., when the j^{th} feature is selected or not, respectively.

2) *The constraints:* We only need to ensure that the sum of our energy does not exceed the limit E_{budget} . Thus, we compute the total energy consumption of our features with the dot product of the selection vector x_f and the energy vector $E_f = [E_f^{(1)} \dots E_f^{(m)}]^T$ with m dimensions.

3) *The objective function:* We aim to select a subset of features that maximizes our prediction-power. The total prediction-power of any given selection is the dot product of the selection vector x_f and the importance vector $I = [I_1 \dots I_m]^T$ with m dimensions, where I_j is the SHAP features importance of the j^{th} feature.

Thus, our problem can be formulated as follows:

$$\max I^T x_f \quad \text{s. t.} \quad E_f^T x_f \leq E_{budget}, \quad (2)$$

To have the total energy consumption of the model, in addition to the energy of features (E_f), we need to count the energy of sensors (E_s) and parameters (E_p) as well. To do so, we add the sensor' selection vector $x_s = [x_s^{(1)} \dots x_s^{(k)}]^T$ and the parameters' selection vector $x_p = [x_p^{(1)} \dots x_p^{(l)}]^T$. In addition, as explained before, in our application we need to consider the cost dependencies of the features. Thus, when the j^{th} feature ($f^{(j)}$) extracted from the k^{th} sensor and l^{th} parameter is selected ($x_f^{(j)} = 1$), then its corresponding sensor and parameter should be selected as well ($x_s^{(k)} = 1$, $x_p^{(l)} = 1$). Since the selection variable x is binary, sensors and parameters are selected only once. For instance, if two features $f^{(i)}$ and $f^{(j)}$ are selected and both are extracted from the RSP sensor and the PSD parameter, their total energy consumption is $E_f^{(i)} + E_f^{(j)} + E_p^{(psd)} + E_s^{(rsp)}$, where $x_f^{(i)} = x_f^{(j)} = x_s^{(rsp)} = x_p^{(psd)} = 1$. This is equivalent to

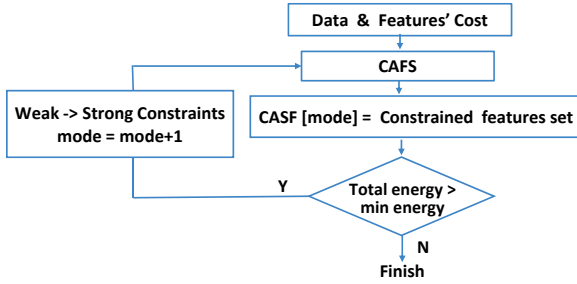


Figure 4: Applying CAFS for different energy-scalable modes by varying the energy-budget constraints.

$x_f^{(i)} \leq x_s^{(rsp)}$ and $x_f^{(j)} \leq x_s^{(rsp)}$, similarly, $x_f^{(i)} \leq x_p^{(psd)}$ and $x_f^{(j)} \leq x_p^{(psd)}$. Consequently, we defined the conditions for all $f^{(j)} \in \{\text{features extracted from the } k^{\text{th}} \text{ sensor}\}$ and for all $f^{(j)} \in \{\text{features extracted from the } l^{\text{th}} \text{ parameter}\}$, respectively:

Therefore, to include cost-dependencies between features in our problem formulation, we extend (2) as follows:

$$\begin{aligned}
 & \max \quad I^T x f \\
 & \text{s. t.} \quad E_s^T x_s + E_p^T x_p + E_f^T x f \leq E_{budget}, \\
 & \quad x_f^{(j)} - x_s^{(k)} \leq 0 \quad \forall f^{(j)} \in \{\text{features from } k^{\text{th}} \text{ sensor}\}, \\
 & \quad x_f^{(j)} - x_p^{(l)} \leq 0 \quad \forall f^{(j)} \in \{\text{features from } l^{\text{th}} \text{ parameter}\},
 \end{aligned} \quad (3)$$

The ILP problem in (3) is solved using the GLPK_MI solver [45] yielding a pre-selected features set. Then, the RFECV-SHAP is applied, the model is trained and hyperparameters are tuned. Finally, a second RFECV-SHAP is applied as described in Section IV-C resulting in the constrained features set.

B. Weak and Strong Constraints Variations

The pre-selected set of features obtained by solving the constrained ILP problem in (3) can change depending on the available resources, i.e., E_{budget} . On the one hand, weak/no energy-constraints assumption (i.e., an unlimited battery lifetime) allows to freely select the most important features regardless of how costly they are, hence yielding a more accurate stress detection model. On the other hand, stronger constraints on the resources restrict the selection of features and allow for errors in the model. Thus, we propose a dynamic CAFS methodology that provides the best possible set of features according to the target energy constraints.

As shown in Figure 4, CAFS starts with no budget constraints and results in a full mode, called $Mode_1$, a stress detection model with a complex set of features. In a second step, the energy constraint is updated to be less than the total features' cost of the previous mode. Then, the algorithm continues n times until it reaches a model with the lowest possible cost ($Mode_n$ or budget mode). As a result, the optimization finishes with an energy-scalable stress detection model builder that automatically trades-off feature cost and accuracy giving different output modes. It starts from a complex full mode that is more accurate but demands more energy, until a simple budget mode that is less accurate but saves more energy.

An application we adopt for different energy-scalable modes ($Mode_1$ to $Mode_n$) obtained using the proposed CAFS method is self-aware real-time monitoring of stress. In the following Section VI, the details of our proposed self-aware stress monitoring system are described.

VI. SELF-AWARE STRESS MONITORING SYSTEM

Real-time monitoring of stress can be done on a multimodal platform and in a self-aware fashion [33], [38], so that the battery lifetime is prolonged without compromising the detection quality. Although stress has a multimodal nature, it is not always required to have a complex and highly sensitive model running on a wearable device to detect stress events. People usually are in normal conditions and occasionally experience stress. Therefore, to improve the energy-efficiency of the wearable system, often a simple model with a small subset of physiological features is sufficient. However, the system should be aware of a stress event presence and be able to switch to a more complex mode to detect it.

In this work, we propose a two-mode (S_2) and a three-mode (S_3) self-aware classifiers. The S_2 toggles between a simple mode ($Mode_S$) and a medium mode ($Mode_M$), and S_3 toggles between $Mode_S$, $Mode_M$, and a full mode ($Mode_F$). Each of these self-aware classifiers always runs on a simple mode, e.g. $Mode_S$, while other features, parameters and sensors from a more complex mode ($Mode_M$ or $Mode_F$) are inactive until the output prediction is not confident enough according to the desired confidence threshold. Then, the system selects a more complex mode for running in the next segmentation window ($i+1$). The selected mode for the next time window defines which sensors and feature extraction algorithms are needed to become active.

To measure the desired confidence threshold, we analyze the classification performance in the CV set by varying the decision threshold, th_S . It is selected when the total performance is relatively high respect to the energy the model consumes and before the probability of using the simple mode becomes very small. In the S_3 , th_M is chosen similarly by comparing the CV performance between a S_2 and $Mode_F$. Then, we estimate the energy consumption of the whole self-aware system by having the probability of triggering each mode [33]. The simpler mode $Mode_S$ is always running for every decision; however, the more complex mode is activated with the probability of P_M and P_F . The S_2 total energy consumption is as follows:

$$E_{total} = E_S + P_M \cdot E_{M \setminus S}, \quad (4)$$

where E_S and $E_{M \setminus S}$ are the energy consumption of $Mode_S$ and $Mode_M$ without features from $Mode_S$, respectively. In S_3 , the total energy is the following:

$$E_{total} = E_S + P_M \cdot E_{M \setminus S} + P_F \cdot E_{F \setminus (M \cup S)}, \quad (5)$$

where $E_{F \setminus (M \cup S)}$ is the energy consumption of $Mode_F$ without features from $Mode_M$ and $Mode_S$.

VII. EXPERIMENTAL SETUP

To evaluate the feasibility of our proposed techniques, we build and test the generalization of the CAFS techniques and the self-aware stress monitoring with experimental data from [46]. Additionally, a multi-sensor platform with an ultra-low-power microcontroller unit is used to analyze and estimate the energy consumption [47]. The CAFS algorithm and the self-aware stress monitoring source codes are available on our public [repository](#).

A. Stress Database: Experiment Protocol and Setup

Physiological signals of 60 male participants ($Age_{mean} = 20.43 \pm 2.17$) were recorded using the Biopac System [48] (sampling rate of 1kHz) while performing the experimental tasks in a virtual reality (VR) environment [46], [49]. ECG was recorded following the lead II positioning. EDA was recorded on the thenar/hypothenar surface of the non-dominant hand. The PPG was recorded on the middle finger from the non-dominant hand. The stress experiment was approved by Vaud Ethics Committee of Switzerland (2017-00449). Participants were divided into two groups performing either a control or a stress task lasting in total 10 minutes each [46], [49].

The stress task exposed participants to an uncontrollable social-evaluative and timed problem-solving task with negative feedback in the VR challenge. Participants were immersed in an empty room with tiled flooring, on which they could move around. Mental arithmetic questions appeared briefly in the heads-up display. Incorrect responses caused a tile on the floor to break and disappear, leaving an open hole where participants could fall into. The control task has an equivalent condition but without the stressful elements of the stress task. More details about the experimental protocol are in [46], [49].

B. Energy Measurement Platforms

To estimate the energy consumption of the features, we adopt a single-chip system-on-chip platform similar to the one used in [6], i.e., having an ARM Cortex M3 as a processing unit. The energy consumption for signal acquisition sensors is assumed to be the same as reported in the measurements from [6], [50]. For measuring the energy of the different algorithms for filtering, delineations, parameter extraction, feature computation and inference, we use the Simplicity Studio software energy profiler on EFM32LG-STK3600 board [47]. This board contains a 48 MHz ARM Cortex-M3 CPU, 32 kB of RAM, and 256 kB of flash memory. Every algorithm is run using the -O3 compiler optimization level, which is the best optimization tolerated by the EFM32. The energy profiler measures the execution time and energy consumed by the algorithms within a specified execution window, in our case, a 60s-window length.

C. Machine Learning Training and Generalization

We consider both low and high-stress levels as a two-class problem corresponding to Control and Stress tasks. The data are segmented into 60-second-length windows without overlap and each segment is labelled according to its associated task. The dataset composes of 415 observations after removing the outliers and 66 features. Since we get more than one segment of data from the signals of each participant, the training, cross-validation, and test sets should not have any data overlap to avoid bias in evaluation. Therefore, we divide the data based on the subject-data where all data from the same subject is always grouped in either the training, validation or test set. Hence, 67% of the subjects, i.e. 40 (23 in Control and 17 in Stress groups), are considered for the training and cross-validation, yielding 282 observations. 33% of the subjects, i.e. 20 (14 in Control and 6 in Stress groups), are considered for the test set yielding 133 observations.

A Group-Shuffle-Split-CV with 10 iterations is used to get smooth mean validation scores. On each iteration, 20% of

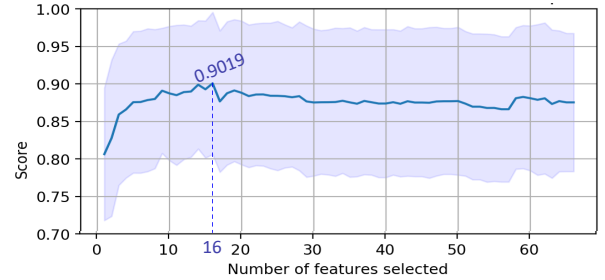


Figure 5: RFECV-SHAP: CV mean and STD accuracy score.

subjects, i.e. 8, are randomly selected as a validation set. Finally, the generalization power of the stress detection model is evaluated on the unseen test set composed of 20 subjects.

VIII. EXPERIMENTAL RESULTS

A. Exploration of Machine Learning Models for MultiModal Acute Stress Detection on Wearables

We train the XGBoost classification algorithm obtaining a CV accuracy and F1-score of 87.55% and 83.61%, respectively. Next, we implement the RFECV-SHAP algorithm to obtain the optimized number of features. Fig. 5 illustrates that the CV accuracy increases with the number of selected features until 16 features when it arrives at the maximum accuracy of 90.19% (F1-score of 86.81% and Gmean of 89.99%). After the 16 features, it enters in a relatively steady-state, which implies that there is no information lost by removing the last 50 features. Consequently, the complexity of the XGBoost model is reduced. The sudden changes in the curve trend are due to the subject-based CV since the physiological response to the stress stimulus is subject-specific. Moreover, features are not independent and their combination contributes to the output prediction, see Section VIII-C.

These 16 features are ranked according to their SHAP features importance values, as shown in Fig. 6. It can be seen that $ECG_{RR_{mean}}$ is the most important feature followed by SCL_{mean} in the stress detection model. Even with much less importance value, the other features are needed when the values of $ECG_{RR_{mean}}$ and SCL_{mean} are in certain ranges that do not provide a confident classification. In Section VIII-D, we illustrate these cases.

Moreover, our XGBoost model's hyper-parameters are tuned by performing a Grid-Search-CV. The maximum of the mean CV accuracy within that region corresponds to 41 trees with a maximum depth of 6. Finally, we obtain a mean CV accuracy of 90.93%, F1-score of 87.80%, and Gmean of 90.66%.

B. Energy-Cost Characterization of Physiological Features

The total energy cost of each feature has several common steps including signal acquisition, filtering, delineation, signal processing to extract bio-parameters, and a final step that is specific for each feature, as explained in Section V-1. Since each physiological signal and its extracted features vary in processing, we group signal acquisition and filtering as our first step in our approach (including delineation for ECG and PPG). Next, we perform the parameter extraction in the second step by including the common processing algorithm, such as the power spectral estimation (PSD). Feature extraction steps and their energy consumption per 60s-segmentation window are presented in Table I. Moreover, in Fig. 6, the energy

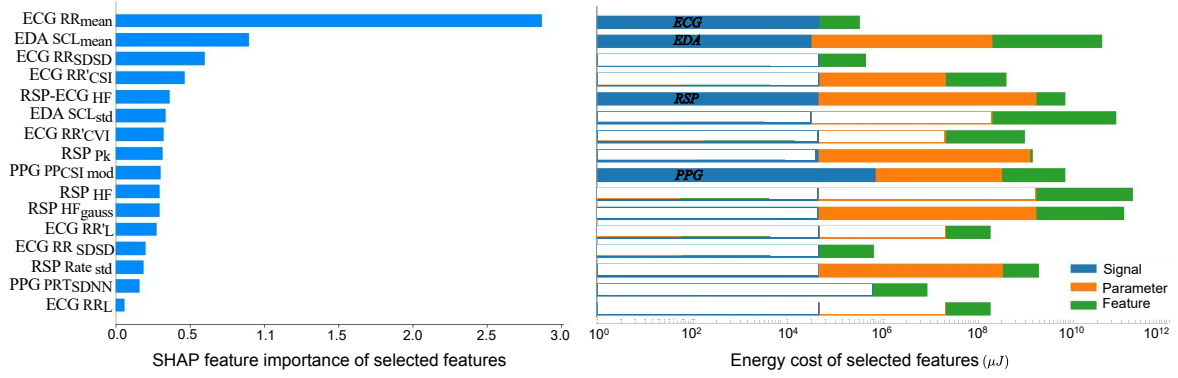


Figure 6: The 16 most important features selected. Left: features ranking based on their SHAP importance. Right: energy consumption of signal acquisition, parameter extraction, and feature extraction for each of the features

Table I: Feature extraction algorithms steps and their energy consumption per segmentation window

Sensor + Preprocessing (Energy mJ*)	Parameter Processing (Energy mJ*)	Feature Extraction Processing (Energy μ J*)	Description
ECG (2.4) + filtering + delineation (2.38)	Poincaré Plot of RR (0.04)	RR_{SD2} (0.88)	Longitudinal axis length of RR poincaré plot.
		RR_{CVI} (2.69)	Cardiac Vagal Index (CVI).
		RR_{CSI} (1.85)	Cardiac Sympathetic Index (CSI).
		$RR_{CSI_{modified}}$ (1.85)	Modified CSI of RR and RR' (second-order RR difference).
	Poincaré Plot of RR' (0.04)	RR'_{SD2} (0.88)	Longitudinal axis length of RR' poincaré plot.
		RR'_{CSI} (1.85)	CSI of the RR' poincaré plot.
		RR'_{CVI} (2.69)	CVI of the RR' poincaré plot.
	PSD (0.09)	$RR_{VLF}, RR_{LF}, RR_{HF}$ (90.65), $RR_{LF/HF}$ (181.30)	Heart rate variability (HRV) indexes in frequency domain. Low Frequency (LF), High Frequency (HF), and Very Low Frequency (VLF)
		RR_{mean} (0.69), RR_{SDNN} (1.39), RR_{SDSD} (1.46), RR_{PNN50} (0.06)	HRV indexes in time domain. Standard Deviation of NN intervals (SDNN), Standard Deviation of Successive Differences (SDSD), Percentage of successive NNs that differ by more than 50 ms (PNN50).
PPG (60.36) + filtering + delineation (2.43)	Poincaré Plot of PP (0.04)	PP_{SD2} (0.88)	Longitudinal axis length of the PP poincaré plot.
		PP_{CVI} (2.69)	CVI of PP
		PP_{CSI} (1.85)	CSI of Pulse Period (PP)
		$PP_{CSI_{modified}}$ (1.85)	Modified CSI of PP and PP' (second-order PP difference).
	Poincaré Plot of PP' (0.04)	PP'_{SD2} (0.88)	Longitudinal axis length of PP' poincaré plot.
		PP'_{CSI} (1.85)	CSI of PP' poincaré plot.
		PP'_{CVI} (2.69)	CVI of PP' poincaré plot.
	PSD of PP (0.09)	$PP_{VLF, LF, HF}$ (90.65), $PP_{LF/HF}$ (181.30)	HRV indexes in frequency domain.
		$PRT_{VLF, LF, HF}$ (90.65), $PRT_{LF/HF}$ (181.30)	Pulse rising time (PRT) indexes in frequency domain.
	PSD of PRT (0.09)	PP_{mean} (0.69), PP_{SDNN} (1.39), PP_{SDSD} (1.46), PP_{PNN50} (0.06), PP_{PNN50} (0.06)	HRV indexes in time domain.
PRT_{mean} (0.69), PRT_{std} (1.39), PRT_{SDSD} (1.46)		PRT indexes in time domain.	
PA_{mean} (0.69), PA_{std} (1.39), PA_{SDSD} (1.46)		Pulse amplitude (PA) indexes in time domain.	
k_{mean} (0.69), k_{std} (1.39), k_{SDSD} (0.96)		Pulse rising speed (k) indexes in time domain.	
RSP (4.44) + filtering (1.26)	Delineate (0.75)	$Rate_{mean}$ (0.17), $Rate_{std}$ (0.58)	Respiration rate indexes in time domain.
		Peakness estimation (2.84)	$Pk, power_{tot}, power_{norm}$ (0.09)
	PSD (3.74)	$RSP_{HF_{ecg}}$ (10.91)	Respiration power of the HF_{ecg} band.
		$RSP_{HF_{ecg-pF_{mn}}}$ (11.31)	Normalized power of five equally distributed bands in the HF_{ecg} band
		$RSP_{LFp1/Hf}$ (56.35)	$LF_{norm} + (1/HF_{norm})$
		$RSP_{HF_{gauss}}$ (340.8)	Mean frequency of a Gaussian distribution fitted in the HF band.
		$RSP_{HF_{pond}}$ (11.31)	Ponderate frequency mean of the HB band (0.15 – 0.50Hz).
EDA (2.34) + filter (0.86)	SCL (0.64)	mean, gradient (20.00), std (40.00)	SCL indexes in time domain.
		power (70.00)	power of SCL
		power phasic (70.00)	Phasic power of SCR.

*Measures with Gecko Board per 60s-window length and ECG, PPG, RSP and EDA's sampling frequency of 256Hz, 128Hz, 128Hz and 4Hz, respectively.

consumption distribution of the 16 most important selected features is shown. The colored steps represent the values used to calculate the energy consumption considering the cost-dependencies among the selected features.

C. Cost-Aware Feature Selection (CAFS) for Energy-aware Stress Detection Model

Considering unlimited available resources, 16 features are selected, which yields a complex stress detection model named $Mode_1$. $Mode_1$ uses data of four physiological signals ECG, RSP, PPG, and EDA to detect stress with a total energy-cost of 85.0 mJ and CV accuracy of 90.93%. Then, to reduce the total cost while still has good performance, we apply

the CAFS methodology. In this case, we constrain the energy consumption to be less than $Mode_1$, which results in $Mode_2$.

Table II shows that twelve features are selected in $Mode_2$ from three physiological signals ECG, RSP, and SCL with a total cost of 22.12 mJ, and CV accuracy of 90.60%. In $Mode_2$ features from the PPG signals are not selected. PPG features have a high computation for the energy cost metrics, yet they have a little overall value, according to Fig. 6. Therefore, their energy cost is not justified. On the other hand, ECG features are also costly, but they appear to be highly valued, as shown in Fig. 6. Then, the CAFS methodology is repeated this time subject to energy consumption being less than the total energy of $Mode_2$, which results in $Mode_3$ with 7 features from ECG

Table II: XGBoost Acute Stress Detection Modes

Mode	XGBoost				Energy (mJ)	CV scores (%)							Selected features
	feat.	depth	trees	nodes		Acc.	F1	Prec.	Rec.	Gmean	Spec.	Conf.	
1	16	6	41	2665	85.00	90.93	87.80	89.74	87.91	90.66	94.38	94.35	ECG: RR_{mean} , RR_{SDNN} , RR_{SDSD} , RR_L , RR'_L , RR'_{CSI} , RR'_{CVI} RSP: $Rate_{std}$, Pk , HF , $HF_{gaussF1}$, HF_{pF2n} PPG: $PP_{CSI_{mod}}$, PRT_{std} EDA: SCL_{mean} , SCL_{std}
2	12	3	69	621	22.12	90.60	87.47	88.81	88.30	90.51	93.68	95.46	ECG: RR_{mean} , RR_{SDNN} , RR_{SDSD} , RR'_{CSI} , RR'_{CVI} RSP: $Rate_{std}$, Pk , HF , $HF_{gaussF1}$, HF_{pF2n} EDA: SCL_{mean} , SCL_{std}
3	7	5	1	33	8.75	89.49	86.39	89.46	85.01	89.12	94.18	62.26	ECG: RR_{mean} , RR_{SDSD} , RR_{PNN50} , RR'_{CSI} , RR'_{CVI} EDA: SCL_{mean} , SCL_{std}
4	1	1	1	3	4.78	88.80	84.32	91.51	81.44	87.56	96.02	60.46	ECG: RR_{mean}
5	1	6	1	65	3.86	64.68	59.81	57.34	70.72	64.33	64.20	56.95	EDA: SCL_{mean}

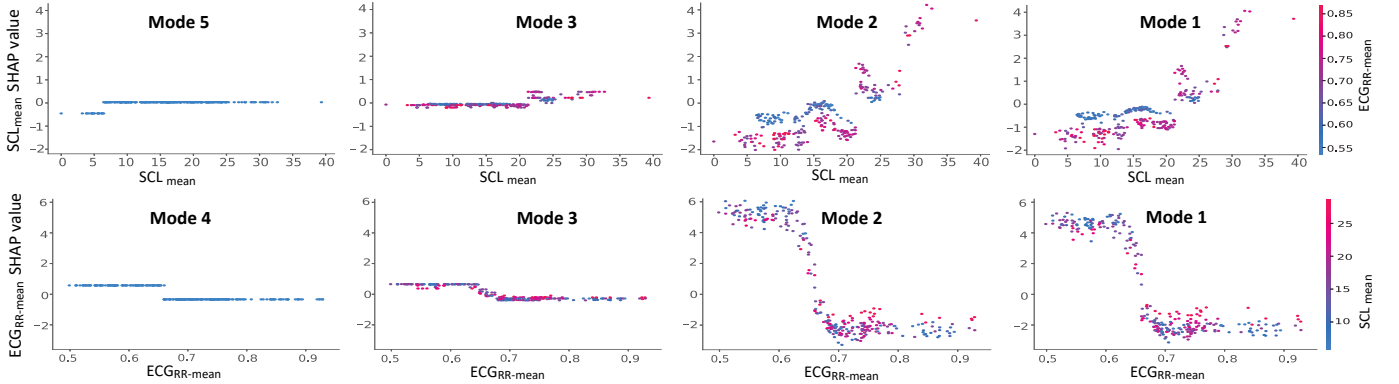


Figure 7: Effect of SCL_{mean} and RR_{mean} variations on their SHAP values contributions to the acute stress prediction on each observation of the training set. Positive and negative SHAP values express a prediction toward the stress and control class, respectively, and around zero means that the model cannot predict the class.

and EDA sensors. Thus, it consumes 8.75 mJ energy in total and has an 89.49% CV accuracy. The results of $Mode_3$ prove that RSP features are not as valuable as ECG and EDA ones when considering all factors (i.e., cost and prediction power).

By further tightening the available energy budget, we end up with $Mode_4$ and $Mode_5$ using only one-modality of ECG and EDA, respectively. The CV accuracy of 88.80% in $Mode_4$ shows that the mean value of RR interval in the ECG signal is the most important physiological feature for one-modality stress detection. Conversely, the SCL_{mean} provides poor accuracy and insignificant energy improvements. Therefore, it is not suitable for one-modality stress detection.

Table II shows the CV results of the XGBoost stress detection models with tuned hyper-parameters. $Mode_1$ uses the largest number of nodes (2665) among other modes due to its large number of deep trees used for voting. In $Mode_5$, the XGBoost model has one tree with a maximum depth of six, which is too deep to predict when the model has only one tree and one feature. Hence, the model in $Mode_5$ is overfitting. However, by decreasing the maximum depth of the model to one, we reduce the overfitting problem in $Mode_5$.

D. Exploration of Self-Aware Stress Detection on Wearables

From the analysis of the different modes in Table II, we observe that a simple model like $Mode_4$ using only one feature is often sufficient to detect stress. However, in other situations, a small subset of physiological features ($Mode_3$) or a complex mode ($Mode_1$) is needed. Indeed, Fig. 7 shows the effect of changing SCL_{mean} and $ECGRR_{mean}$ on the stress classification in the training dataset for the different modalities' modes. First, it can be seen that $Mode_5$ predicts

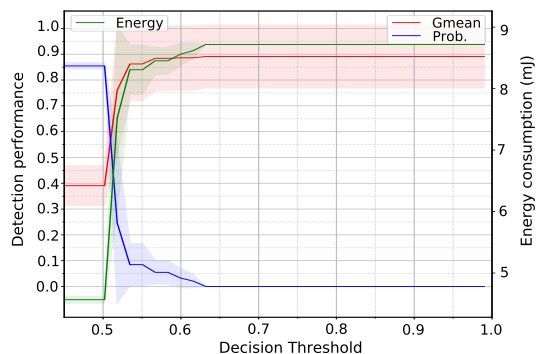
low values of SCL_{mean} , i.e., less than $6.42 \mu S$, but it fails to predict for larger values. Conversely, in $Mode_4$, also one-modality mode using $ECGRR_{mean}$, the SHAP values show that short RR interval, i.e. smaller than $0.65 sec$, (having fast heart rate) is predicted as stress and larger than this value (having slow heart rate) is predicted as no stress.

Next, Fig. 7 shows that by fusing SCL_{mean} and $ECGRR_{mean}$ in $Mode_3$, we can classify for larger SCL_{mean} values. In this case, SHAP values above zero are for the majority of observations having SCL_{mean} larger than $21.26 \mu S$, hence classified as the stress group. However, this model has difficulties discriminating between classes when the RR interval is between 0.65 and $0.68 sec$ and SCL_{mean} is between 6.42 and $21.26 \mu S$. Finally, using more modalities (PPG and RSP), as in $Mode_1$ and $Mode_2$, we better discriminate between stress and control prediction. For SCL_{mean} values between 6.42 and $21.26 \mu S$, a slower heart rate increases the prediction tendency toward the control class.

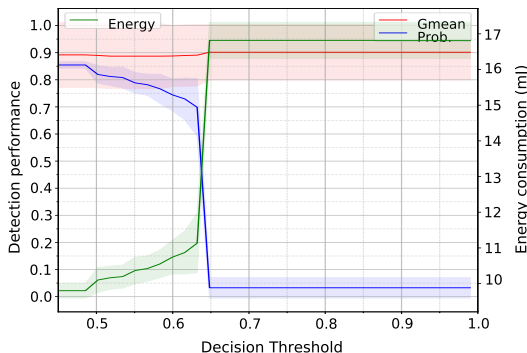
Based on the previous observation, we propose a self-aware system that toggles to a more confident mode whenever the prediction confidence (i.e., decision threshold) is not good enough. In this study, we choose to combine three monitoring modes on our self-aware stress detection. In particular, we combine $Mode_5$ (the simplest mode), $Mode_3$ (medium mode having features from the previous mode), and $Mode_2$ (the full mode), which were obtained from our CAFS methodology. Although $Mode_1$ is more complex and might perform better in a real environment, we chose $Mode_2$ since it performs high and consumes less energy, for our lab-based dataset. First, we cross-validate the decision threshold for the S_2 monitoring mode. Then, a second decision threshold is chosen between

Table III: Self-aware Stress Detection Models

Self-aware	Probability (%)			Energy (mJ)	CV scores (%)						
	P_2	P_3	P_5		Acc.	F1	Prec.	Rec.	Gmean	Spec.	Conf.
S2	–	96.7	3.3	8.56	88.97	85.55	88.38	84.63	88.61	93.69	62.07
S3	30.1	66.6	3.3	11.13	89.37	85.88	88.29	85.54	89.08	93.71	71.56



(a) Two-mode self-aware classifier



(b) Three-mode self-aware classifier

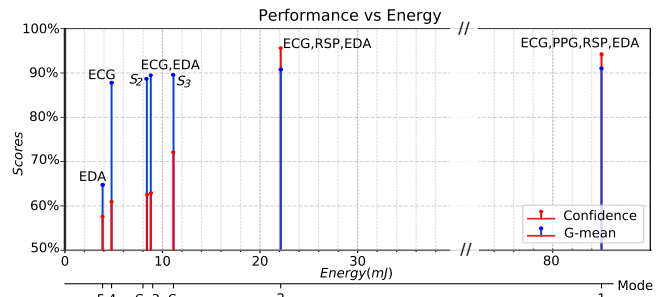
Figure 8: Self-aware classifier: Gmean, probability of using a simple mode and E_t versus decision thresholds.

S_2 and $Mode_2$ for the S_3 monitoring mode.

Fig. 8a shows the CV Gmean and energy consumption (E_{total} defined in (4)) of S_2 when changing the probability of using the simplest mode ($Mode_5$) by varying the decision threshold. The higher the decision threshold is, the lower the probability of using the simple mode is. Consequently, it is more probable to use a more confident mode, i.e., $Mode_3$ in S_2 , which increases the performance, and also the energy consumption. The decision threshold to stop using $Mode_5$ on S_2 is chosen to be 0.6 (60%), namely, when the performance is relatively high 88.61% and the E_{total} is reduced.

Fig. 8b shows the CV scores of the S_3 , which includes $Mode_5$, $Mode_3$, and $Mode_2$. The desired confidence threshold between S_2 and $Mode_2$ is chosen to be 0.63 (63%), where the Gmean is high (89.64%). Also, the energy consumption is reduced from 22.12 mJ to 11.13 mJ just before the system almost stops the probability of using the lower energy modes. Our CV results of the S_2 and S_3 are reported in Table III, showing that the energy consumption is reduced in the self-aware modes without sacrificing the performance.

The CV performance vs all the modes' energy consumption is presented in Fig. 9. Although the detection Gmean for all modes, except $Mode_5$, is similar ranging from 87.5–90.6%, the confidence levels vary from 95% for the complex modes ($Mode_2$ and $Mode_1$) to 61% for simpler models ($Mode_3$,

Figure 9: CV confidence and Gmean versus the total energy-cost of all five modes and S_2 and S_3 self-aware models.

$Mode_2$ and S_2). Our proposed self-aware mode S_3 improves the confidence levels of the simpler modes by 12% with little increase of the energy consumption. Furthermore, S_3 reduces the energy consumption significantly compared to $Mode_2$ and $Mode_1$, but it still has a good detection performance.

E. Generalization

Finally, we assess the generalization of our proposed acute stress monitoring modes for our experimental conditions with 20 subjects' data were not seen before. The results are presented in Table IV. It shows that the test scores in every mode, except $Mode_5$, are high and the gaps between test scores in Table IV and the CV scores in Table II and Table III are small. Therefore, we can conclude that the generalization power of our proposed stress monitoring modes is high. Moreover, we can conclude that $Mode_1$, $Mode_2$, as well as the S_3 self-aware mode, perform accurate and confident stress monitoring.

Table IV: Generalization of Acute Stress Detection Models

Mode	Energy (mJ)	Test scores tuned models						
		Acc.	F1	Prec.	Rec.	Gmean	Spec.	Conf.
1	85.00	89.47	81.08	83.33	78.95	86.00	93.68	94.92
2	22.12	88.72	79.45	82.86	76.32	84.55	93.68	95.74
3	8.75	90.98	84.21	84.21	84.21	88.82	93.68	62.02
4	4.78	82.71	69.33	70.27	68.42	77.78	88.42	60.48
5	3.86	63.91	46.67	40.38	55.26	61.02	67.37	57.40
S2	8.13	88.72	80.52	79.49	81.58	86.43	91.58	61.42
S3	9.14	88.72	79.45	82.86	76.32	84.55	93.68	72.58

IX. DISCUSSION

Real-life multimodal stress monitoring applications need highly accurate detection algorithms that allow a longer battery lifetime on wearable devices. Despite several efforts in multimodal stress monitoring [4]–[7], the energy consumption of wearable monitoring devices has received little attention. Here, our proposed CAFS methodology is designed to develop highly accurate multimodal stress models with low energy usage, tested in particular for acute stress detection within our experimental framework. Moreover, our proposed CAFS approach is model-agnostic and applies to any supervised classification problem that uses cost-dependent features, considered for the first time, to the best of our knowledge, on a

cost-aware feature selection solution.

Using an experimental dataset, CAFS provides five interpretable acute-stress detection modes from complex to simple (more confident to less confident) depending on the energy budget constraints. Our results show that this new CAFS methodology considerably reduces the energy requirements of the stress detection model, while maintaining high classification performance. All the five modes defined using CAFS and the two self-aware modes are compared and the interpretability of the models are further discussed in the next subsections.

Finally, concerning the limitations of this work, we first consider that the acute stress models are only generalizable for our particular experimental conditions. Further research should be done in a wider setting and real-life conditions for a more general understanding of stress and cognitive workload. On the other hand, due to the limited amount of data available for each participant in our dataset, here in this article, we do not consider the individual profile and differences of stress reactivity, instead we have used a homogeneous group of participants (i.e., healthy, $Age_{mean} = 20.43 \pm 2.17$ and male participants). For a heterogeneous population sample, a personalized stress detection models to obtain individualized prototypical signatures of stress for each participant could be developed. This will allow extending our presented findings by proposing personalized models.

A. Stress Detection Models: Scores, Confidence, and Cost

In Fig. 9 and Table IV, our results show that applying CAFS with no or weak energy-constraints yields highly-accurate and confident models (i.e., $Mode_{1,2}$). These models use more physiological signals and features and consequently, a larger number of trees to make a prediction. Hence, they are more complex having a higher number of nodes. However, having more trees for voting, the collective absolute SHAP value gets larger. Thus, the prediction-power gets higher and, consequently, a more confident prediction is achieved.

Indeed, the confidence level in $Mode_1$ and $Mode_2$ is about 95%, whereas it is only 62% in $Mode_3$ even when the detection accuracy for these modes is similar, ranging from 88.72% to 90.98%. Regarding the cost, using $Mode_2$ and $Mode_3$ the energy consumption is reduced by 73.98% and 89.71% concerning $Mode_1$, respectively. It is important to note that we employ a particular dataset, where the stress and control tasks are only 10 minutes long and where subjects perform the same activity. In a real environment with a wider range of stress responses, we expect to have a scalable detection accuracy and confidence with an increase of the model complexity. Having a very similar induced acute stress response explains our similar performance results for $Mode_1$ and $Mode_2$, although they have different complexity.

On the other hand, strong budget constraints modes ($Mode_{4,5}$) are simpler using a few features and few shallow trees. It, in turn, reduces the complexity and energy expenditure of the machine learning models but results in a less confident prediction. Using $Mode_4$ and $Mode_5$, the energy is reduced by 94.38% and 95.46%, with only a 6.76% and 25.56% drop in accuracy for $Mode_4$ and $Mode_5$. Due to poor performance in $Mode_5$, this mode is not suitable for stress monitoring but rather is useful for a self-aware stress monitoring by taking advantage of its low energy consumption and

the ability of confident classification for specific SCL_{mean} values. Using $Mode_4$, one-modality model with ECG signal, we are still able to predict stress, but this prediction is not as confident as using a multimodal model, such as, in $Mode_{1,2}$.

The prediction scores (accuracy, precision) and confidence levels comparisons among these five modes show that we cannot rely only on high prediction scores for a confident and reliable stress prediction. We corroborate that by using more modalities at the expense of energy, we have a more comprehensive physiological response that makes the predictions between stress and no stress more discriminable and hence more confident even if the prediction scores are same. That is the stress has a multimodal nature [4], [7], [12], [13].

Therefore, considering the multimodal nature of stress, we aim to reduce energy consumption while having an accurate and confident acute stress prediction by employing our new self-aware stress classifier. According to our results, S_3 reduces the total energy consumed for a stress prediction by 89.24% with a 0.75% drop in accuracy and 22.34% drop in the prediction confidence respect to $Mode_1$ on the unseen test data. In contrast, compared with $Mode_3$ that has a similar energy consumption and accuracy, the confidence increases by 10.56%. On the other hand, although using the S_2 mode the energy consumption is reduced by 90.44% and keeping high accuracy, its confidence level is not improved enough since no complex and confident mode is used in its design.

Note that for S_2 and S_3 the energy consumption varies according to the needs of using the different modes, which depends on the physiological stress response induced. Self-aware classification brings more energy savings while keeping high stress detection accuracy and confidence in a real-life application. On the other hand, higher detection accuracy variations can be expected when using different modes.

B. Interpretability of Acute Stress Detection Models

In agreement with previous studies [4], [5], [7], the RR_{mean} in the ECG signal is the most important feature followed by the SCL_{mean} from EDA signal for acute stress prediction. Our results show that a low SCL value ($SCL_{mean} < 6.42\mu S$) and also a long RR interval ($0.68sec < RR_{mean}$) represent no stress status. While a high SCL ($21.26\mu S < SCL_{mean}$) and a fast HR ($RR_{mean} < 0.65s$) suggest a stress status. However, for other ranges of RR_{mean} and SCL_{mean} (e.g., $0.65 < RR_{mean} < 0.68$ and $6.42 < SCL_{mean} < 21.26$) the model requires more features for a confident stress classification. Although RR_{mean} provides a pretty accurate stress prediction in one-modality $Mode_4$, using SCL_{mean} in one-modality $Mode_5$ does not provide an accurate and confident stress prediction. Nevertheless, the contribution of the SCL_{mean} with RR_{mean} in the multi-modality models, as in $Mode_{1,2,3}$, provides a more confident stress prediction. Furthermore, although the RSP and PPG features are not as valuable as adding ECG and EDA features to the model, which improves up to 35% the stress discrimination confidence

X. CONCLUSION

In this work, we have proposed CAFS, a comprehensive solution for the design of energy-aware multimodal acute stress monitoring on wearable devices. We have addressed the problem of selecting an optimized feature subset in the

presence of individual feature costs with cost-dependency and a total cost constraint. Our CAFS methodology automatically explores the trade-off between energy-cost and the benefits of adding more features for multimodal stress prediction.

Using CAFS on our acute stress experimental database, four generalizable and energy-scalable stress detection models from full-complex multimodal to simple low-budget unimodal were obtained, keeping the accuracy as high as the non-constrained model and reducing the energy cost by 10 times. The best accuracy obtained on the unseen data is 90.98%, which outperforms the state-of-the-art studies. Finally, our self-aware stress detection model achieves an accuracy of 88.72% on the unseen dataset, having the same energy level as the simple model but with a 11.16% more confident prediction.

REFERENCES

- [1] H. Selye, "Stress and the General adaptation syndrome," *British Medical Journal*, pp. 1383–1392, 1950.
- [2] N. Schneiderman *et al.*, "Stress and health: psychological, behavioral, and biological determinants," *Annu. Rev. Clin. Psychol.*, vol. 1, pp. 607–628, 2005.
- [3] E. S. Epel *et al.*, "More than a feeling: A unified view of stress measurement for population science," *Frontiers in Neuroendocrinology*, vol. 49, pp. 146–169, 2018.
- [4] A. Alberdi *et al.*, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *Journal of Biomedical Informatics*, vol. 59, no. C, pp. 49–75, 2 2016.
- [5] G. Giannakakis *et al.*, "Review on psychological stress detection using biosignals," *IEEE Transactions on Affective Computing*, pp. 1–1, 2019.
- [6] S. Song *et al.*, "A 769 μ W Battery-Powered Single-Chip SoC With BLE for Multi-Modal Vital Sign Monitoring Health Patches," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 6, pp. 1506–1517, 12 2019.
- [7] Y. S. Can *et al.*, "Stress detection in daily life scenarios using smart phones and wearable sensors: A survey," *Journal of Biomedical Informatics*, vol. 92, 4 2019.
- [8] A. Arza *et al.*, "Measuring acute stress response through physiological signals: towards a quantitative assessment of stress," *Medical & Biological Engineering & Computing*, pp. 1–17, 8 2018.
- [9] V. Montesinos Canovas *et al.*, "Multi-modal acute stress recognition using off-the-shelf wearable devices," in *International Engineering in Medicine and Biology Conference*, 7 2019, pp. 2196–2201.
- [10] N. Momeni *et al.*, "Real-time cognitive workload monitoring based on machine learning using physiological signals in rescue missions," in *International IEEE EMBC*, 4 2019.
- [11] H. Zhang *et al.*, "Detection of variations in cognitive workload using multi-modality physiological sensors and a large margin unbiased regression machine," in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 2985–2988.
- [12] A. Arza Valdés, "Measurement of acute psychological stress," Ph.D. dissertation, Universitat Autònoma de Barcelona, 9 2017.
- [13] S. Betti *et al.*, "Evaluation of an integrated system of wearable physiological sensors for stress monitoring in working environments by using biological markers," in *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 8, 8 2018, pp. 1748–1758.
- [14] N. Sharma and T. Gedeon, "Modeling a stress signal," *Applied Soft Computing*, vol. 14, pp. 53–61, 2014.
- [15] I. Goodfellow *et al.*, *Deep Learning*. MIT Press, 2016.
- [16] F. Dell'Agnola *et al.*, "Cognitive workload monitoring in virtual reality based rescue missions with drones," in *12th International Conference on Virtual, Augmented and Mixed Reality*, 7 2020.
- [17] A. Golgouneh and B. Tarvirdzadeh, "Fabrication of a portable device for stress monitoring using wearable sensors and soft computing algorithms," *Neural Computing and Applications*, pp. 1–23, 6 2019.
- [18] D. Sopic *et al.*, "Real-Time Event-Driven Classification Technique for Early Detection and Prevention of Myocardial Infarction on Wearable Systems," *IEEE TBioCAS*, vol. 12, no. 5, pp. 982–992, 10 2018.
- [19] G. Surrel *et al.*, "Online Obstructive Sleep Apnea Detection on Medical Wearable Sensors," *IEEE Transactions on Biomedical Circuits and Systems*, no. 99, pp. 1–12, 2018.
- [20] M. Tan, "Cost-sensitive learning of classification knowledge and its applications in robotics," *Machine Learning*, vol. 13, no. 1, pp. 7–33, 10 1993.
- [21] Q. Zhou *et al.*, "Cost-sensitive feature selection using random forest: Selecting low-cost subsets of informative features," *Knowledge-Based Systems*, vol. 95, pp. 1–11, 3 2016.
- [22] R. Sheikhpour *et al.*, "A Survey on semi-supervised feature selection methods," *Pattern Recognition*, vol. 64, pp. 141–158, 4 2017.
- [23] R. Jagdhuber *et al.*, "Cost-Constrained feature selection in binary classification: adaptations for greedy forward selection and genetic algorithms," *BMC Bioinformatics*, vol. 21, no. 1, p. 26, 12 2020.
- [24] A. Elsts *et al.*, "On-board feature extraction from acceleration data for activity recognition," *International Conference on Embedded Wireless Systems and Networks*, no. February, pp. 163–168, 2017.
- [25] H. Ghasemzadeh *et al.*, "Power-aware computing in wearable sensor networks: An optimal feature selection," *IEEE Transactions on Mobile Computing*, vol. 14, no. 4, pp. 800–812, 2015.
- [26] S. P. le Roux *et al.*, "Energy-aware feature and model selection for onboard behavior classification in low-power animal borne sensor applications," *IEEE Sensors Journal*, vol. 19, no. 7, pp. 2722–2734, 2019.
- [27] S. Lauwereins *et al.*, "Optimal resource usage in ultra-low-power sensor interfaces through context- and resource-cost-aware machine learning," *Neurocomputing*, vol. 169, pp. 236–245, 2015.
- [28] S. Karayev *et al.*, "Dynamic Feature Selection for Classification on a Budget," *ICML Workshop on Prediction with Sequential Models*, 2013.
- [29] J. Janisch *et al.*, "Classification with costly features as a sequential decision-making problem," *Machine Learning*, no. 0123456789, 2020.
- [30] S. Maldonado *et al.*, "Cost-based feature selection for Support Vector Machines: An application in credit scoring," *EJOR*, vol. 261, no. 2, pp. 656–665, 9 2017.
- [31] I. G. Lee *et al.*, "A mixed integer linear programming support vector machine for cost-effective feature selection," *Knowledge-Based Systems*, vol. 203, p. 106145, 9 2020.
- [32] F. Forooghifar *et al.*, "Resource-aware distributed epilepsy monitoring using self-awareness from edge to cloud," *IEEE transactions on biomedical circuits and systems*, vol. 13, no. 6, pp. 1338–1350, 2019.
- [33] G. Masinelli *et al.*, "Self-aware machine learning for multimodal workload monitoring during manual labor on edge wearable sensors," *IEEE Design and Test*, vol. 2356, pp. 1–7, 2020.
- [34] F. Forooghifar *et al.*, "Self-Aware Anomaly-Detection for Epilepsy Monitoring on Low-Power Wearable Electrocardiographic Devices," in *International Conference on Artificial Intelligence Circuits and Systems (AICAS)*. IEEE, 2021, pp. 1–4.
- [35] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 3 1996.
- [36] F. Shaffer and J. P. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Frontiers in Public Health*, vol. 5, p. 258, 9 2017.
- [37] A. Hernando *et al.*, "Inclusion of Respiratory Frequency Information in Heart Rate Variability Analysis for Stress Assessment," *IEEE J-BHI*, vol. 20, no. 4, pp. 1016–1025, 7 2016.
- [38] F. Forooghifar *et al.*, "A Self-Aware Epilepsy Monitoring System for Real-Time Epileptic Seizure Detection," *Mobile Networks and Applications*, pp. 1–14, 8 2019.
- [39] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [40] C. Molnar, *Interpretable Machine Learning*, 2019.
- [41] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in NIPS*, 2017, pp. 4765–4774.
- [42] S. M. Lundberg *et al.*, "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery," *Nature biomedical engineering*, vol. 2, no. 10, pp. 749–760, 2018.
- [43] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *22Nd ACM SIGKDD*, 2016, pp. 785–794.
- [44] A. Schrijver, *Theory of Linear and Integer Programming*. John Wiley & Sons, Chichester, 1986.
- [45] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2909–2913, 2016.
- [46] J. Rodrigues *et al.*, "Locomotion in virtual environments predicts cardiovascular responsiveness to subsequent stressful challenges," *Nature Communications*, vol. 11, no. 1, p. 5904, 2020.
- [47] "EFM32™ Leopard Gecko 32-bit Microcontroller," [url{ www.silabs.com/products/mcu/32-bit/efm32-leopard-gecko}](http://www.silabs.com/products/mcu/32-bit/efm32-leopard-gecko).
- [48] Biopac, "MP160 Data Acquisition Systems."
- [49] J. Rodrigues *et al.*, "IMVEST, an immersive multimodal virtual environment stress test for humans that adjusts challenge to individual's performance," *Neurobiol Stress*, vol. 15, p. 100382, 2021.
- [50] M. Konijnenburg *et al.*, "22.1 A 769 μ W Battery-Powered Single-Chip SoC with BLE for Multi-Modal Vital Sign Health Patches," in *Digest of Technical Papers - IEEE International Solid-State Circuits Conference*, vol. 2019-Febru. IEEE, 2 2019, pp. 360–362.