Cognitive workload monitoring in virtual reality based rescue missions with drones⋆

Fabio Dell'Agnola[1][0000−0002−3556−2578], Niloofar Momeni[1][0000−0002−9435−9772], Adriana Arza[1][0000−0002−4190−7878], and David Atienza[1][0000−0001−9536−4947]

Embedded Systems Laboratory of Swiss Federal Institute of Technology Lausanne, Switzerland. https://www.epfl.ch/labs/esl
`{fabio.dellagnola, niloofar.momeni, adriana.arza, david.atienza} @epfl.ch`

Abstract. The use of drones in search and rescue (SAR) missions can be very cognitively demanding. Since high levels of cognitive workload can negatively affect human performance, there is a risk of compromising the mission and leading to failure with catastrophic outcomes. Therefore, cognitive workload monitoring is the key to prevent the rescuers from taking dangerous decisions. Due to the difficulties of gathering data during real SAR missions, we rely on virtual reality. In this work, we use a simulator to induce three levels of cognitive workload related to SAR missions with drones. To detect cognitive workload, we extract features from different physiological signals, such as electrocardiogram, respiration, skin temperature, and photoplethysmography. We propose a recursive feature elimination method that combines the use of both an eXtreme Gradient Boosting (XGBoost) algorithm and the SHapley Additive exPlanations (SHAP) score to select the more representative features. Moreover, we address both a binary and a three-class detection approaches. To this aim, we investigate the use of different machine-learning algorithms, such as XGBoost, random forest, decision tree, k-nearest neighbors, logistic regression, linear discriminant analysis, gaussian naïve bayes, and support vector machine. Our results show that on an unseen test set extracted from 24 volunteers, an XGBoost with 24 features for discrimination reaches an accuracy of 80.2% and 62.9% in order to detect two and three levels of cognitive workload, respectively. Finally, our results are open the doors to a fine grained cognitive workload detection in the field of SAR missions.

Keywords: Cognitive Workload Monitoring· Physiological Signals · Machine Learning · Search and Rescue Missions · Simulator · Drones.

1 Introduction

The use of drones in search and rescue (SAR) missions can be very cognitively demanding. Rescuers often operate in extreme conditions, under time pressure,

---

and dealing with the scarcity of human resources. Moreover, they have to find strategies to face all types of unexpected events that a SAR mission may hide. Since high levels of cognitive workload can negatively affect human performance [10,13], there is a risk of compromising the mission and leading to failure with catastrophic outcomes. Therefore, a Cognitive Workload Monitoring (CWM) system that detects elevate workload levels could be used to notify the rescues and prevent them from taking dangerous decisions.

Nowadays, there are three main methods to measure cognitive workload; that is, subjective surveys, performance metrics, and assessing the human physiological response [6,3]. Due to the needs of a continuous monitoring and the unpredictability of SAR missions, neither surveys nor performance metrics can be used for cognitive workload monitoring of rescues. In contrast, the cognitive workload monitoring from physiological signals, which can be acquired with wearable sensors, is suitable for this application.

Continuous workload monitoring is a challenging task because is a multi-dimensional problem. Many factors affect cognitive workload; that is mental, physical, and temporal demands, the overall performance, the frustration level, and the effort [5]. The perceived workload is both environmental and subjective dependent, it changes among individuals according to their learning skills and their ability to address and perform a particular task. Despite that, state-of-the-art studies that target cognitive workload detection from physiological signals showed promising results [3,12,4,9].

The objective of this work is to detect cognitive workload levels from physiological signals in the field of SAR missions with drones. However, due to the unpredictable conditions and the difficulties of having a cognitive workload reference value during real SAR missions, the data collection becomes really difficult. Therefore, we rely on Virtual Reality (VR) to emulate immersive SAR mission. Our main contributions are as follows:

- We induce three levels of cognitive workload using a VR simulator for SAR missions with drones.
- We detect high and low levels of cognitive workload with an accuracy, precision and recall of 80.2%, 79.6%, 71.7%, respectively on an unseen dataset, which prove the generalization of our model.
- We explore a three-class cognitive workload detection, thus identifying high, medium, and low levels of cognitive workload with an accuracy of 62.9%, which is a promising results considering the complex identification of cognitive workload levels generated in our virtual SAR mission.

The rest of the paper is organized as follows. Section 2 describes the methods applied to induce and detect different levels of cognitive workload. Section 3 describes the setup of the experiment and Section4 reports our results. Finally, in Section 5 we draw the main conclusions of this work.

2   Cognitive Workload Characterization and Detection Methods

The general idea of the method applied for estimating cognitive workload, including sensor placement, is show in Figure 1. A VR based simulator is used to induce different levels of cognitive workload by changing the difficulty of the tasks [4]. Then, the cognitive workload is detected from physiological signals by following a standard methodology, which includes signals acquisition, preprocessing, feature extraction, and classification. Finally, an elicitation technique based on a multidimensional assessment tool, the NASA Task Load Index (NASA-TLX) [6], and a questionnaire to evaluate the difficulty of the tasks are used to label the different levels of induced cognitive workload.
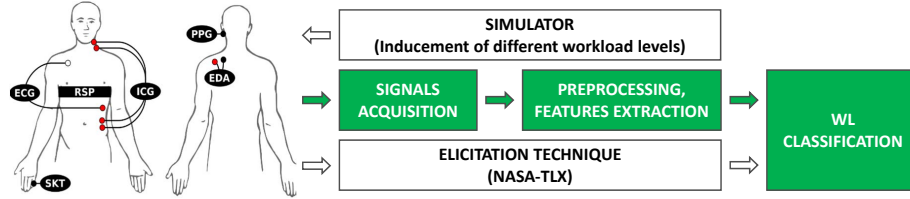

Fig. 1. Cognitive workload (WL) estimation from physiological signals, such as Electrocardiogram (ECG), Respiration (RSP), Impedance Cardiogram (ICG), Skin Temperature (SKT), Photoplethysmography (PPG), and Electrodermal Activity (EDA)

2.1   Simulator for Search and Rescue Mission with Drones

The proposed VR based simulator emulates a simplified SAR mission, where the pilot needs to fly a drone through a pathway and map a damage situation of a disaster area. The simulator was implemented with Unity3D [15] and used to track the workload influence on SAR missions with drones in [4]. In the simulator, the flight pathway is shown by 90 waypoints (black rings) distributed every 20 m along a randomly generated trajectory over a village. The speed of the drone is fixed at 6 m/s for all the tasks. The damage situation is represented by cubes that randomly appear over the flying pathway, as shown in Figure 2. The colors of the cubes are chosen according to the regulation of the Swiss Firefighters; that is, yellow to indicate rescue situations, red for fire, blue for water damages, and green for the accidents.

To induce different levels of workload, both flying and mapping activities are combined yielding in the following tasks:

- *Training (T):* A training sequence is proposed to let the user familiarize himself with the simulator. The training sequence has a combination of both flying and mapping activities. The user is asked to fly as close as possible
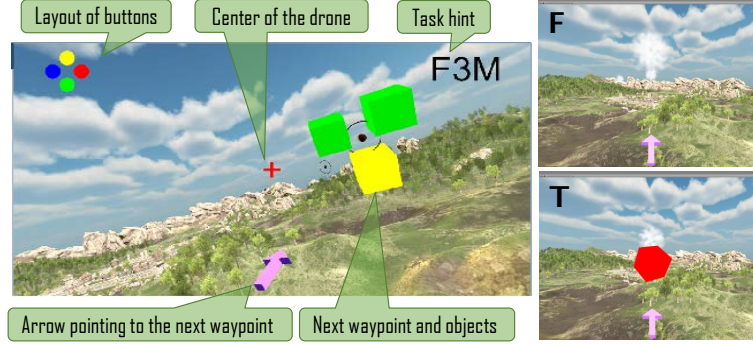
Fig. 2. Simulator for search and rescue mission with drones. Overview of different conditions: flying and mapping (F3M), flying (F), and training (T).

through the center of the waypoints. In addition, the user needs to press the button of the controller relative to the color of the objects that are randomly displayed. The number of objects to be mapped are 60 per session, i.e., 15 per color.

- *Baseline (B):* To set a physiological baseline avoiding as much as possible the effect of uncontrollable variables, a passive task with the same framework as for the entire experiment is presented. This task is a flying sequence controlled by an auto-pilot. During this task, the user just needs to watch the sequence presented to him without any other additional activity.

- *Flying (F):* Flying task is considered as a medium level of cognitive workload. During this task, the subject solely needs to fly the drone as close as possible to the center of the waypoints.

- *Mapping 3 object (3M):* This task is also considered as a medium level of cognitive workload. During this task, the subject has to identify three objects of different color that appear at the same time on the screen and has to press press the button of the controller corresponding to the color of the object. The number of objects to be mapped are 240 per session, i.e., 60 per color.

- *Flying and Mapping 3 objects (F3M):* This task is considered as a high level of the cognitive workload. It is the same as training, including the same combination of both the flying and mapping activities but with a more demanding mapping activity. The objects displayed simultaneously on the screen are three as in 3M task, and not only one as in the training phase.

2.2 Cognitive Workload Classification

The proposed methodology to build a model for CWM has three main steps. The first step is the acquisition of different signals, such as Electrocardiogram (ECG), Respiration (RSP), Impedance Cardiogram (ICG), Skin Temperature

(SKT), pulse wave through Photoplethysmography (PPG), and Electrodermal Activity (EDA), as selected in [4,9].

The second step is the preprocessing and feature extraction, where we filter the physiological signals and we extract several features to capture the functionality of the human physiology as we did in [9,1]. For the feature extraction we use a 1-minute sliding window without overlap. Then, we apply a Recursive Feature Elimination with Cross Validation (RFECV) to select the most important physiological features that better characterize different levels of cognitive workload. Our RFECV is based on the eXtreme Gradient Boosting (XGBoost) algorithm and uses the SHapley Additive exPlanations (SHAP) framework for a more detailed prediction interpretation [8].

Finally, the third and last step is the cognitive workload classification. In this step, we first investigate several machine learning algorithms to select the model that best fits our classification problem. To this aim, we compare XGBoost, Random Forest (RF), Decision Tree (DT), k-Nearest Neighbors (kNN), Logistic Regression, Linear Discriminant Analysis (LDA), Gaussian Naïve Bayes (GNB), and Support Vector Machine (SVM).

Both RFECV and the machine learning algorithms are trained and validated on the training set using a shuffled Leave-P-Groups-Out Cross-Validation (LP-GOCV) with five iterations considering each time 20% of the groups as validation set. Each group is composed of all observations recorded in one day that belong to one subject. If not otherwise specified, the machine learning algorithms are trained and tested with [11] using default parameters.

Thus, we have applied the same procedure to address both binary and three-class classification problems. Finally, to prove the generalizability of our model, a final test is done on an unseen test set, which includes data from 30% of the groups, that was initially putted apart for this purpose only.

3 Experimental Study

To build a database and investigate the cognitive workload characterization and detection from physiological signals, we perform an experimental study applied a designed protocol as in Figure 3 using the VR simulator for SAR mission with drones. To control the simulator it was used a Gamepad from Logitech [7]. We recorded the physiological signals with the Biopac acquisition set [2], which is CE certified for medical monitoring.

3.1 Study Protocol

Our study protocol starts with a setup phase, where we provide information about the experiment and we place the sensors. Additionally, we propose a training sequence to let the participant get familiar with the simulator. Then, we proceed with two phases of data collection, as shown in Figure 3.

Phase 1 is composed of four different tasks (i.e., B, F, 3M, and F3M), where the order of F, 3M, and F3M is randomized to avoid any bias related to the order
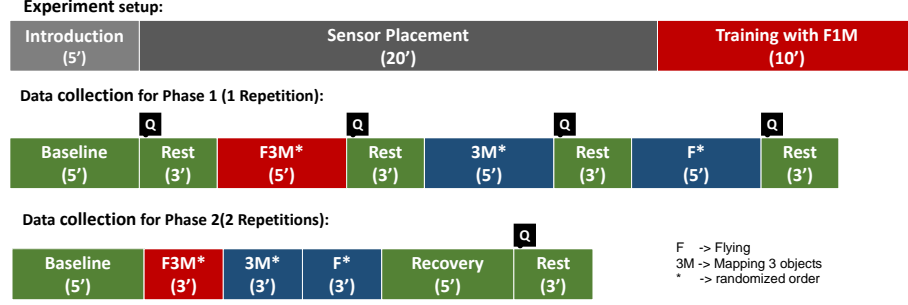
Fig. 3. Protocol of the experiment, which includes different conditions, such as baseline (B), mapping activity (3M), flying activity (F), flying and mapping performed simultaneously (F3M), and filling a questionnaire (Q) to evaluate the NASA-TLX.

of these activities. Before each resting period, the subjects were asked to rate the perceived level of cognitive workload by evaluating the NASA-TLX. The time duration for each task in Phase 1 is five minutes.

Phase 2 consists of a concatenated sequence of the tasks presented in Phase 1 (i.e., B, F, 3M, and F3M). As in Phase 1, the order of F3M, 3M, F is randomized. However, in Phase 2 there is no resting period between the tasks. Phase 2 is executed twice in a row, and the NASA-TLX is evaluated only at the end of each sequence. The tasks duration in Phase 2 is limited to three minutes. Except the data collected from the second repetition of Phase 2 that is kept apart for testing our final machine-learning algorithm, the rest of the data is selected only for training the machine-learning algorithms.

Participants were asked not to talk, and to avoid as much as possible any kind of unnecessary movements during the main tasks, but they were free to rest and move otherwise.

3.2 Participants

In our study participated 24 volunteers (6 females and 18 males), aged between 21 and 39 years old (27.7 ± 4.8). All but one participated twice in two different days. The participants were healthy, free of any cardiac abnormalities and were receiving no medical treatment. The ethical approval for this study was obtained from the Cantonal Ethics Commissions for Human Research Vaud and Geneva; namely, ethical approval application number PB2017-00295.

3.3 Database

From the 47 experimental sessions performed, we build our datasets based on the workload reported by the participants in the NASA-TLX questionnaire. Values higher than 55% and lower than 5% are labeled as low and high workload, respectively. Following this criteria, 259 observations are labeled as low workload and

249 are labeled as high, which makes a dataset of 508 observations. By including the medium level of workload, we extend our dataset to 748 observations.

Both datasets are split into two subsets of 70% and 30% of groups of observations for both training and testing, respectively. Thus, the test set includes all the observation of 14 groups randomly selected from the second day, while the training set includes all the remaining groups of observations acquired the first day.

4   Results

Our results include the cognitive workload perceived by the volunteers and used as reference. Moreover, we provide results of a binary CWD and results of a three-class CWD.

4.1   Cognitive Workload Reference

Both the tasks' difficulty perceived by the participants and the cognitive workload level reported with the NASA-TLX procedure after each task are shown in Figure 4. Each box includes 144 data points, which were averaged across all the trials.
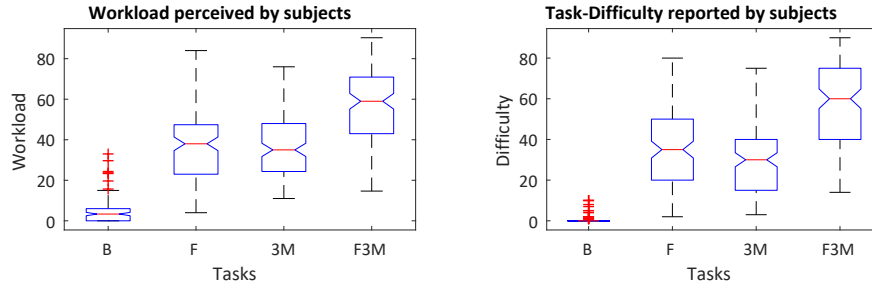

Fig. 4. Perceived workload per task. Different conditions: baseline (B); mapping (3M); flying (F); and flying and mapping (F3M).

Both cognitive workload and tasks' difficulty show an increasing trend in the order of B, 3M, F and F3M. Then, a two-tail t-tests yielded a significant difference between most of the tasks (p-value < 0.001, n $= 47$), except for the cognitive workload between 3M and F where the difference is not significance . The significance difference of the difficulty between tasks 3M and F is limited to a p-value < 0.05. A significant Pearson's correlations between cognitive workload and tasks' difficulty is observed (p-value < 0.001, n $= 47$). The high correlation confirms the validity of the assumption that cognitive workload increases with the difficulty of the tasks.

4.2   Binary Cognitive Workload Detection

Following our methodology to develop a model for cognitive workload monitoring (see Sec. 2) we first process the physiological signals. In particular we consider in this study the RSP, ECG, PPG and SKT from where we extracted 156 features as in [9]. Then, we performed the feature selection step.

The result of the RFECV based on SHAP values and using a XGBoost model is shown in Figure 5. The score increases by adding features until it reaches its maximum with 24 features. Then, it decreases and reaches a plateau if more than 45 features are used. Finally, we selected 24 features to train an XGBoost classifier, which reaches a LPGOCV accuracy of 79.5%. The 24 most important physiological features characterizing high levels of cognitive workload level of the participants are shown in Table 1.
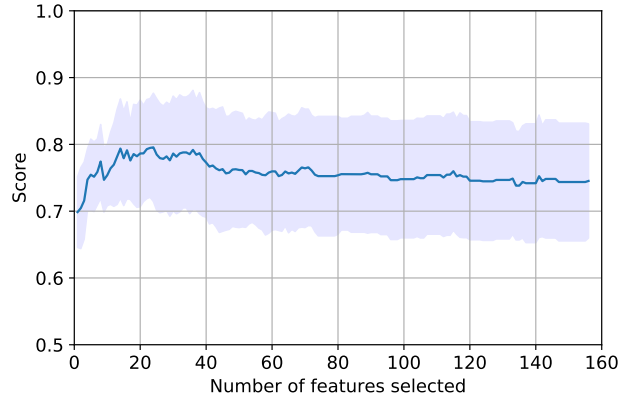

Fig. 5. SHapley Additive exPlanations (SHAP) score used for feature selection.

The comparison of different classification algorithms for the binary classification between low and high workload levels are summarized in Table 2. In particular, we report both training and LPGOCV accuracy before and after feature selection. As expected, considering the large amount of features compared with the limited size of our dataset, the models trained with all the features show an important gap between train and LPGOCV accuracy, which is a clear sign of overfit. This overfitting problem was addressed by using simplest models (e.g., linear instead of non-linear models) and models with less parameters (i.e., lower capacity). Indeed, the use of LDA drastically reduces the gap between training and LPGOCV accuracy, especially when a limited amount of features are used. Although the training data is not sufficient to find an optimum, the highest LP-GOCV accuracy is given by XGBoost (79.5%). Thus, the results provided next are based on XGBoost only.

Table 1. Description of the selected features.

Features	Description
$RR_{\mathrm{CSI}}{}^{2}$	Cardiac Sympathetic Index, L/T ratio of the RR Lorenz plot [14]. L and T are the longitudinal and transverse axis' lengths, respectively.
$RR_{\mathrm{Lnz2L}}{}^{2,3}$	Longitudinal axis length of the second-order RR difference Lorenz plot.
$RR_{\mathrm{Lnz2LoT}}{}^{2,3}$	L/T ratio of the second-order RR difference Lorenz plot.
$RR_{\mathrm{Lnz2log_{10}LT}}{}^{3}$	$\log_{10} LT$ of the second-order RR difference Lorenz plot.
$RR_{\mathrm{HFpond}}{}^{2,3}$	Mean of the HF band frequencies weighted by their corresponding power.
$PP_{\mathrm{LnzT}}{}^{3}$	Length of the transverse axis T of the PP Lorenz plot.
$PA_{\mathrm{Std}}{}^{2}$	Standard deviation of the pulse amplitude.
$PA_{\mathrm{Median}}{}^{3}$	Median value of the pulse amplitude.
$PRT_{\mathrm{Mean}}{}^{3}$	Mean value of the pulse wave rising time.
$PRT_{\mathrm{Median}}{}^{3}$	Median value of the pulse wave rising time.
$PRT_{\mathrm{Std}}{}^{2,3}$	Standard deviation of the pulse wave rising time.
$k_{\mathrm{Median}}{}^{3}$	Median value of the pulse rise speed.
$k_{\mathrm{Std}}{}^{2}$	Standard deviation of the pulse rise speed.
$Inh_{\mathrm{Time\ Median}}{}^{2,3}$	Median value of the inspiration time from RSP.
$Exh_{\mathrm{Time\ Median}}{}^{2,3}$	Median value of the expiration time from RSP.
$Inh_{\mathrm{Time\ RMS}}{}^{2}$	RMS value of the inspiration time from RSP.
$Exh_{\mathrm{Time\ Mean}}{}^{2}$	Mean value of the expiration time from RSP
$Inh_{\mathrm{Amp\ Median}}{}^{2,3}$	Median value of the inspiration amplitude from RSP.
$Exh_{\mathrm{Amp\ Median}}{}^{2,3}$	Median value of the expiration amplitude from RSP.
$RSP_{\mathrm{Prd\ Mean}}{}^{2,3}$	Mean value of respiration period.
$RSP_{\mathrm{Prd\ Median}}{}^{3}$	Median value of respiration period.
$RSP_{\mathrm{Rate\ Median}}{}^{2}$	Median value of respiration rate.
$RSP_{\mathrm{Rate\ Mean}}{}^{3}$	Mean value of respiration rate.
$RSP_{\mathrm{Rate\ Std}}{}^{2}$	Standard deviation of respiration rate.
$RSP_{\mathrm{Rate\ RMS}}{}^{2,3}$	RMS value of respiration rate.
$RSP_{\mathrm{PSD2n}}{}^{2,3}$	Normalized Power Spectral Density (PSD) of RSP in the second frequency band (i.e., 0.25-0.5).
$RSP_{\mathrm{PSD3}}{}^{2,3}$	PSD of RSP in the third frequency band (i.e., 0.5-0.75 Hz).
$RSP_{\mathrm{PSD3n}}{}^{3}$	Normalized PSD of RSP in the third frequency band.
$NN50^{3}$	Number of interval differences of successive NN-intervals greater than 50 ms.
$RSA_{\mathrm{Shift\ Rad}}{}^{2,3}$	Phase shift between RSP and RR.
$RSA_{\mathrm{R0\ Mean}}{}^{2}$	Mean value of the Respiration Sinus Arrhythmia (RSA) curve.
$RSA_{\mathrm{Rn\ Std}}{}^{2}$	Standard deviation of RSA curve.
$SKT_{\mathrm{Power}}{}^{2}$	Total Power of the SKT.
$SKT_{\mathrm{Gradient}}{}^{2,3}$	Gradient of the SKT.

[2] and [3] refer to the binary and three-class problems, respectively.

Table 2. Comparison of different classification models in term of accuracy.

	All features		24 features	
Model	Training	LPGOCV	Training	LPGOCV
XGBoost	100%	74.5% (0.09)	100%	79.5% (0.08)
Random Forest	100%	73.3% (0.08)	100%	76.8% (0.08)
Decision Tree	100%	68.1% (0.09)	100%	73.2% (0.07)
k-Nearest Neighbors	92.35%	71.4% (0.06)	90.7%	76.5% (0.08)
Logistic Regression	86.89%	74.7% (0.07)	83.3%	76.1% (0.05)
Linear Discriminant Analysis	95.9%	71.6% (0.06)	86.6%	78.7% (0.05)
Gaussian Naïve Bayes	81.6%	70.4% (0.09)	82.7%	76.3% (0.06)
Support Vector Machine	89.6%	72.8% (0.07)	84.4%	75.4% (0.04)

Finally, in order to minimize the errors, we fine tuned the decision threshold of our model on cross-validation. A plot of precision, recall, and accuracy as a function of different decision thresholds is shown in Figure 6.
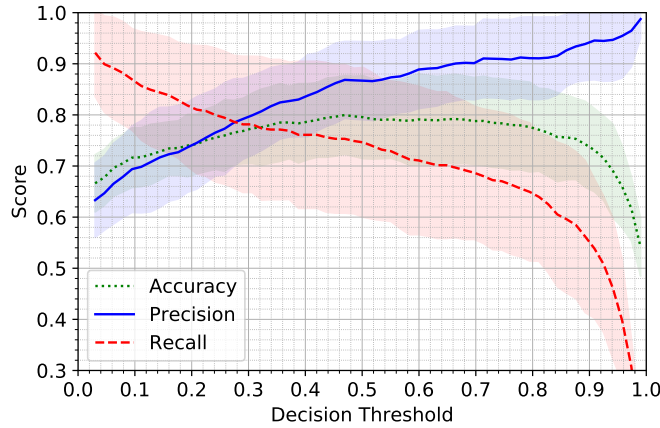
Fig. 6. Decision threshold as function of precision, recall, and accuracy.

Although the higher CV accuracy is reached with a threshold around 0.5, which is the default decision threshold in binary classification, the recall is poor. However, by choosing a threshold equal to 0.27 (where precision almost equals recall), we have an equal error rate and the accuracy is still acceptable. A detailed comparison between the use of a threshold at 0.5 and one at 0.27 is reported in Table 3.

Table 3. LPGOCV Performance of XGBoost with different decision thresholds.

Model	Features	Threshold	Accuracy	Precision	Recall
XGBoost	24	0.5	79.5%	86.5%	74.6%
		0.27	76.5%	78.4%	78.6%

The results on an unseen test set of the workload detection based on XGBoost with 24 selected features and a threshold at 0.27 that shows the generalization power of the final workload detection model are reported in Table 4.

Table 4. Generalization power of the workload detection model (XGBoost) on the unseen test set.

Model	Features	Threshold	Accuracy	Precision	Recall
	All	0.5	79.6%	94.3%	55.0%
XGBoost	24	0.5	76.8%	84.6%	55.0%
	24	0.27	80.2%	79.6%	71.7%

Although results obtained in CV suggest the use of an XGBoost with the 24 selected features and a threshold of 0.27, as a comparison we also report in Table 4 the results of an XGBoost before and after feature selection and a decision threshold at 0.5. As on the test set, an increase of accuracy exists and, as a result, a reduction of the difference between precision and recall after applying both feature selection and threshold tuning. Overall, these results show that our model generalizes well for cognitive workload detection.

4.3   3-Class Cognitive Workload Detection

As explained in Section 2, we also explore the three-class detection problem using the same procedure that we applied for the binary problem. The applied RFECV technique with XGBoost algorithm based on SHAP values also selected 24 features with a maximum accuracy of 54.7%, which is still clearly better than random (33%). The selected physiological features characterizing low, medium, and high levels of cognitive workload are reported in Table 1.

From the comparison of the different classification methods reported in Table 5, we can see that XGBoost is again slightly better than the others. The important difference between training and CV accuracy clearly indicates an overfitting problem. Acquiring more data will probably solve this problem and increase the performance of some models, such XGBoost, random forest, and decision tree. However, our results suggest to select again the XGBoost with 24 features to address our three-class problem.

Finally, on an unseen test set, the cognitive workload detection based on the 3-class XGBoost with 24 selected features shows an accuracy and both averaged precision and recall of 62.9%.

Table 5. Comparison of different three-class models in term of accuracy.

	All features		24 features	
Model	Training	LPGOCV	Training	LPGOCV
XGBoost	100%	53.5% (0.04)	98.7%	54.7% (0.04)
Random Forest	100%	52.9% (0.06)	100%	54.6% (0.07)
Decision Tree	100%	41.9% (0.07)	100%	45.9% (0.06)
k-Nearest Neighbors	69.8%	35.2% (0.08)	77.6%	53.4% (0.04)
Logistic Regression	74.2%	53.2% (0.05)	65.5%	54.4% (0.06)
Linear Discriminant Analysis	86.5%	53.5% (0.07)	69.3%	54.0% (0.05)
Gaussian Naïve Bayes	61.5%	49.3% (0.04)	62.5%	52.9% (0.05)
Support Vector Machine	75.7%	52.9% (0.05)	66.4%	54.3% (0.05)

5   Conclusion

The goal of this work was to detect cognitive workload levels from physiological signals in the field of SAR missions with drone. Although the volunteers were not exposed to the same stressful conditions as they would be in the case of a real SAR mission, with augmented reality we were able to inducing different level of cognitive workload that are related to SAR mission with drones. From the collected dataset we compared different machine-learning algorithms and we found that an XGBoost with only 24 features is enough to address both a binary and a three-class problem. The two class XGBoost was able to detect high and low levels of cognitive workload with an accuracy of 80.2%. The three-class XBGoost was able to detect high, medium, and low levels of cognitive workload with an accuracy of 62.9%. Considering the little difference between the levels of induced workload our results are promising and open the doors to a fine grained cognitive workload detection in the field of SAR missions.

References

1. Arza, A., Garzón-Rey, J.M., Lázaro, J., Gil, E., Lopez-Anton, R., de la Camara, C., Laguna, P., Bailon, R., Aguiló, J.: Measuring acute stress response through physiological signals: towards a quantitative assessment of stress. Medical and Biological Engineering and Computing pp. 1–17 (8 2018). https://doi.org/10.1007/s11517-018-1879-z, http://link.springer.com/10.1007/s11517-018-1879-z
2. Biopac: MP160 Data Acquisition Systems, https://www.biopac.com/product/mp150-data-acquisition-systems/
3. Cain, B.: A Review of the Mental Workload Literature. Defence Research and Development Toronto, Canada (2007), http://www.dtic.mil/dtic/tr/fulltext/u2/a474193.pdf
4. Dell'Agnola, F., Cammoun, L., Atienza, D.: Physiological characterization of need for assistance in rescue missions with drones. In: IEEE International Conference on Consumer Electronics (ICCE). pp. 1–6 (1 2018), https://infoscience.epfl.ch/record/261415

5. Evans, D.C., Fendley, M.: A multi-measure approach for connecting cognitive workload and automation. International Journal of Human-Computer Studies **97**, 182–189 (1 2017), https://www.sciencedirect.com/science/article/pii/S1071581916300623
6. Hart, S.G., Staveland, L.E.: Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, Advances in Psychology, vol. 52, pp. 139 – 183. North-Holland (1988), http://www.sciencedirect.com/science/article/pii/S0166411508623869
7. Logitech: Gamepad F130, http://support.logitech.com/en/product/gamepad-f310
8. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30, pp. 4765–4774. Curran Associates, Inc. (2017), http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf
9. Momeni, N., Dell'Agnola, F., Arza, A., Atienza, D.: Real-Time Cognitive Workload Monitoring Based on Machine Learning Using Physiological Signals in Rescue Missions. Tech. rep., EPFL (4 2019), https://infoscience.epfl.ch/record/265352
10. Moray, N.: Mental workload : its theory and measurement. Published in coordination with NATO Scientific Affairs, Plenum Press (1979), https://books.google.ch/books?id=SP3lBwAAQBAJ&lr=&source=gbs_navlinks_s
11. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
12. Ranchet, M., Morgan, J.C., Akinwuntan, A.E., Devos, H.: Cognitive workload across the spectrum of cognitive impairments: A systematic review of physiological measures. Neuroscience & Biobehavioral Reviews **80**, 516–537 (Sep 2017), https://linkinghub.elsevier.com/retrieve/pii/S0149763416305413
13. Teigen, K.H.: Yerkes-Dodson: A Law for all Seasons. Theory & Psychology **4**(4), 525–547 (11 1994), http://journals.sagepub.com/doi/10.1177/0959354394044004
14. Toichi, M., Sugiura, T., Murai, T., Sengoku, A.: A new method of assessing cardiac autonomic function and its comparison with spectral analysis and coefficient of variation of r–r interval. Journal of the Autonomic Nervous System **62**(1), 79 – 84 (1997), http://www.sciencedirect.com/science/article/pii/S0165183896001129
15. Unity Technologies, I.: Unity3D, https://unity3d.com/