Jurnal Teknik Industri, Vol. 24, No. 2, December 2022
ISSN 1411-2485 print / ISSN 2087-7439 online

DOI: 10.9744/jti.24.2.83-94

Machine Learning Models for the Cognitive Load Detection Using Heart Rate Variability Signals

Nailul Izzah[1], Auditya Purwandini Sutarto[1*] and Mohammad Hariyadi[1]

Abstract: Cognitive domains play a critical role in daily functioning. The prediction of cognitive load state is essential to better monitor work performance. This study aims to explore machine learning models to detect cognitive load or state using heart rate variability (HRV) signals. HRV data were recorded from thirty subjects during rest, two cognitive tasks (d2 Attention and Featuring Switcher task), and recovery. Seven HRV indexes from both time and frequency domains, extracted from raw R-R intervals, were used to identify whether subjects performed cognitive tasks. Five classifier models: linear support vector machine (LSVM), kernel SVM radial basis function, k-nearest neighbor (KNN), and random forest (RF), were trained and evaluated using a leave-one-out cross-validation approach. The accuracies and F-1 score range from 0.54 to 0.62, with LSVM, showing the best. These acceptable performances indicate that the machine learning approach could be used to further distinguish between rest and cognitive state. With the ubiquity of non-invasive and low-cost wearable devices, this finding offers insight to be incorporated into personal work performance monitoring in the digital age.

Keywords: Cognitive, heart rate variability (HRV), work performance, machine learning.

Introduction

Modern digital work relies much more heavily on cognitive functioning than physical demands. Cognitive functioning refers to the operation and interaction of and between mental processes involved in information processing, such as attention, working memory, decision-making, and learning [1]. Furthermore, individual cognitive performance fluctuates throughout the day, affected by health status, affective state, and other stressful conditions such as workload, time pressure, fatigue, sleep physical deprivation environment [2]. Therefore, maintaining cognitive function is crucial for work performance and daily functioning. Of particular concern is how to predict its current state, which will benefit various work applications encompassing critical systems, office work, operational environments, and others [3,4,5,6].

Traditionally, the cognitive state is measured subjectively by self-report or objectively inferred from a reduced behavioral performance over time, such as reaction time and errors. These assessments pose some limitations in terms of subjectivity (user's bias), disruptiveness (users need to stop during the test administration), timeliness (a lead time between the assessment time and the results are known), and generalizability (the ability to be applied in various population or situations) [7].

On the other hand, cognitive activation elicits autonomic nervous system (ANS) reactions which are commonly reflected by electrodermal activity (EDA) or skin conductance response (SCR), skin temperature (ST), and heart rate variability (HRV) [7]. The last two decades have seen a growing trend toward using bio-signals HRV across various settings, perhaps because of its desirable characteristics: non-invasive, cost-efficient, and straightforward. In addition, the advance in technology and computer science have made HRV collection and analysis very accessible.

Heart rate variability refers to the variation in time between adjacent heartbeats or RR intervals (see Figure 1) and is measured in milliseconds (ms). HRV reflects the complex interaction between heart-brain interactions and dynamic non-linear autonomic nervous system (ANS) processes [8]. ANS consists of two branches, the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS), which work in a dynamic balance in a healthy organism. PNS activity predominates at rest, resulting in an average HR of 75 beats per minute (bpm). Conversely, when a person faces a stressor (e.g., excessive workloads, conflicting demands, tight deadlines, job insecurity, etc.), the SNS releases specific hormones to equip the person with the necessary resources to manage the stressor [8]. An optimal HRV level is associated with better general health status as it allows high self-regulatory capacity and adaptability or resilience to external and internal stimuli. Individuals with higher levels of resting vagally-mediated HRV are associated with higher performance of executive functions like attention and emotional processing by the prefrontal cortex [9].

[1] Faculty of Industrial Technology, Department of Industrial Engineering, Qomaruddin University, Jl. Raya Bungah 1, Bungah, Gresik 61152 Indonesia. Email: auditya@uqgresik.ac.id

* Corresponding author

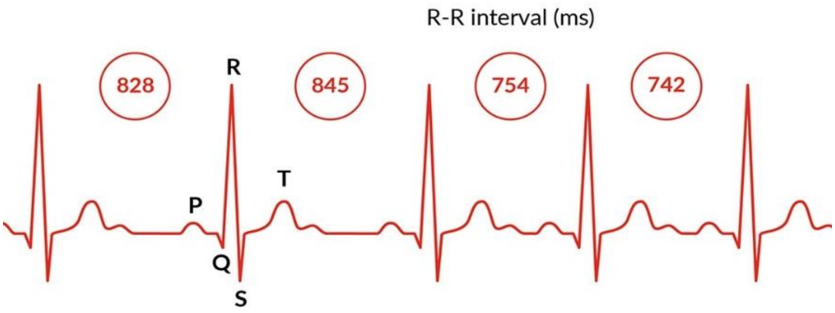

Figure 1. The calculation of HRV is obtained from the R–R intervals (in milliseconds or ms) of the QRS complex, extracted from the electrocardiogram (ECG) signal [10]

Table 1. Summary of the HRV features and their interpretation

Parameter	Description	Unit	Origin
SDNN	Standard deviation of all NN* intervals	ms	Reflects the cyclical components responsible for variability or respiratory-related to the parasympathetic nervous system
RMSSD	The square root of the mean of the sum of the squares of differences between successive NN intervals	ms	Reflects the vagally mediated changes reflected in HRV
pNN50	Percentage of differences between successive adjacent NN intervals that are > 50 ms	%	Reflects variability in parasympathetic or vagal tone
Total power	Sum of the energy in all frequency bands		Variance of all NN intervals
VLF	Total spectral power of all NN intervals (frequency power 0 – 0.04Hz)	ms^2	Reflects changes in the combination of sympathetic and vagal activity; baroreflex activity
LF	Total spectral power of all NN intervals (frequency power 0.04 – 0.15Hz)	ms^2	Reflects changes in mix sympathetic and vagal activity; baroreflex activity
HF	Total spectral power of all NN intervals (frequency power 0.15 – 0.4Hz)	ms^2	Reflects changes in parasympathetic or vagal tone
LF/HF	The ratio of low to high-frequency power	%	A mix of sympathetic and vagal activity

Notes. The NN (normal to normal) intervals represent RR intervals which are filtered or free from artifacts

Research in HRV is mainly based on time or frequency domain analysis [11]. Time-domain indices measure the variability in normal-to-normal beat intervals, while frequency-domain indices quantify the power in frequency bands via spectral analysis. Among these indices, the most frequently reported in response to acute stress in time domains are decreases in SDNN, pNN50, and RMSSD, while in the frequency domain are reduced in high frequency (HF-HRV) and increases in low frequency (LF-HRV) [12]. Table 1 summarizes the time and frequency-domain HRV parameters and their physiological origin.

Beyond the clinical settings, it indicates the potential use of HRV analysis in several domains of organizational and management theory and practice [13]. With advances in technology and computer science, HRV analysis has been shifting toward a more predictive approach rather than an explanation-focused strategy using traditional statistical modeling [14]. The development of stress detection systems using HRV indices and machine learning techniques has shown moderate to good accuracies both in laboratory [15, 16, 17, 18, 19] and field settings [20, 21, 22]. This approach has been successful in differentiating non-stress and stress conditions that were commonly induced by various acute stressors such as arithmetic tests, Stroop tests, trier-social threat tests, Montreal imaging, and horror viewing (e.g., [17, 18, 23, 24]. Concepts of stress, emotions, and cognition load overlap and have complex interrelationships [7]. A recent review suggests that HRV should be more strongly correlated with cognitive fatigue than stress or mental workload [7]. While a growing body of literature also predicts cognitive state through HRV parameters and machine learning approaches, only a few studies have used a single wearable HRV sensor and modalities. Tsunoda et al. [25] used traditional and established ECG system devices (BIOPAC 3 channels) to predict when the cognitive performance started to decrease. Some scholars (e.g., [26, 27, 28] also employed three ECG channels to detect mental stress and cognitive task. One of the few studies that collected HRV data using one wearable sensor is Huang *et al.,* [29]. The authors recorded through the 'LaPatch' wearable ECG, but the validity of this measure is still unclear. Besides, their goal was to predict mental fatigue using a set of mathematic quizzes which did not mimic cognitive tasks at work.

Table 2. Related works with the main attributes

Authors	Signals	Device, # Sensors	Subject	Cognitive stressor	Features	Target	Classifiers (Accuracy)
McDuff *et al.,* [33]	HRV (PPG)	infrared digital cameras, contactless	10	Mental arithmetic	HRV: HR, LF (nu), HF (nu), LF/HF BR	Cognitive stress and rest	NB (70%) LSVM (70%)
Tsunoda *et al.,* [25]	HRV (ECG)	BIOPAC, 3 channels	45	ATMT	AVNN, CVNN SDNN, pNN50, LF, HF, LF/HF, L, T, HF peal, LF peak	Changes in cognitive performance	SVM (60%) RF (57.8%)
Quintero *et al.,*[28]	HRV (ECG)	HP 78354A ECG monitor (Hewlett-Packard), 3 channels	16	PVT, n-back, visual search	HRV: LF, LF (nu), HF, HF (nu) EDA: SCL, SCR	Cognitive task	KNN (66%) LSVM (62%) GSVM (56%) LDA (62%)
Castaldo *et al.,* [27]	HRV (ECG)	Easy ECG Pocket Atas Medica, 3 channels	42	University examination	HR, Mean RR, SDNN, pNN50, LF, HF, LF/HF, TP, SD1, SD2, Entropy	Stress and rest	SVM (88%) LDA (94%)
Wang *et al.,* [26]	HRV (ECG)	HP 78354A ECG monitor (Hewlett-Packard), 3 channels	160 (adolescents)	Arithmetic test	AVNN, SDNN, RMSSD pNN50, SDANN, TP, VLF, LF, HF, LF/HF, L, T, HF, SD1/SD2	Mental stress and rest	SVM (80.2%) K-NN (72.8%) RF (84.6%) DT (84.6%) XGBoost (93.4%)
Huang, *et al.,* [29]	HRV (1 channel)	LaPatch Wearable (one channel)	35	Mathematic Quiz (spatial imagination, computation , reasoning, and memory)	AVNN, pNN50, RMSSD, TP, VLF, LF	Mental fatigue and rest	SVM (57.08%) K-NN (65.37%) NB (48.84%) Logistic (59.71%)

Notes. PPG= Photoplestymograph, BVP = Blood Volume Pressure, BR= Breathing rate. EDA= Electrodermal activity. ATMT = Advanced Trial Making Test. SCL = Skin conductance level, SCR = Skin conductance response. AVNN = Average NN internal, TP = Total power, SD1, SD2= non-linear HRV features based Poincaré plot.

Meanwhile, other researchers [4, 28] have used other physiological modalities such as facial features, EDA, and skin temperature to improve the accuracy of cognitive performance prediction in various tasks. The use of multiple sensors and physiological signals in cognitive state detection has also been emphasized in a recent review [3, 30], which are more challenging if applied in natural conditions. A summary of related work in HRV-based for detecting cognitive load or cognitive performance is displayed in Table 2. Although these studies produced models with relatively high accuracies, it is desired to implement a minimal and unobtrusive sensor setup for comfort in daily life applications.

Furthermore, due to the nature of the machine learning algorithms, the spurious relationship among features can be misinterpreted because the theoretical underpinning has been ignored [14]. For example, a short-term five minutes recording is required to get reliable frequency-domain features [11], whereas using the machine learning approach, an ultra-short duration of around 10-60 seconds is commonly utilized [3, 31]. Of particular concern, hence, is to detect cognitive load using a single HRV sensor and HRV features only according to the HRV measurement in psychophysiological research, which can be easily further deployed into portable wearable devices. Moreover, it allows for real-time physiological data collection, an essential feature of personal productivity management in the digital age [32].

Therefore, this study aims to explore machine learning classification models of cognitive load based on HRV. Cognitive load refers to the number of working memory resources used. We trained different models to classify the state of cognitive function and evaluated the performance of the best model. Our main contribution is to show that HRV measures derived from one wearable sensor ECG have the potential to detect cognitive load or stress state. We also employed two cognitive tasks, measured by attention and speed processing, strongly associated with cognitive domains during work but less studied [24]. The present study explores the development of a system to measure and observe cognitive states that influence individual productivity and help workers to manage challenges at work better.

Methods

Respondens

Thirty right-handed students (male: 17, female: 13, age range: 19–24 years old; age Mean = 21.0 with SD = 1.38 years) participated in this study without a history of psychiatric diseases or other medical problems. All participants were asked to provide informed written consent before the experiment was initiated. The study was approved by the Local Research Ethics Committee and followed the Helsinki Declaration.

Cognitive Load Tasks

In this study, an original d2 Attention Test was executed as a paper-and-pencil version [34]. This task assesses selective and sustained attention, a critical cognitive function in everyday life [35]. Participants have to scan for target stimuli among a variety of distractors. The target is defined as the letter *d* with two apostrophe marks which may be located above or below the letter. There are a total of 14 rows consisting of 47 letters. The participants have 20 seconds to cancel out as many target symbols as possible on each row. She or he should immediately move on to the next row after the time has elapsed. The test duration lasts approximately 5 minutes.

The second task is the Feature Switching Task which assesses cognitive flexibility involved in repeatedly switching between rule dimensions [36]. The task was executed using Psychology Experiment Building Language (PEBL), an open-source psychological test battery [37]. Participants viewed a screen with ten colored shapes, five different colors, and five different shapes. Each object only had a single dimension in common with another object (color, shape, or letter). A participant was asked to choose a matching object based on a shape, color, or letter displayed at the top of the screen after one object was circled. After successfully matching the object, the participant was required to "switch" to a different feature, try to match the object based on that feature, and then return to the previous feature. The task consisted of three sessions, each of which had nine alternative configurations and ten responses from the participants (following a practice round). More detail about the task execution can be found in [36].

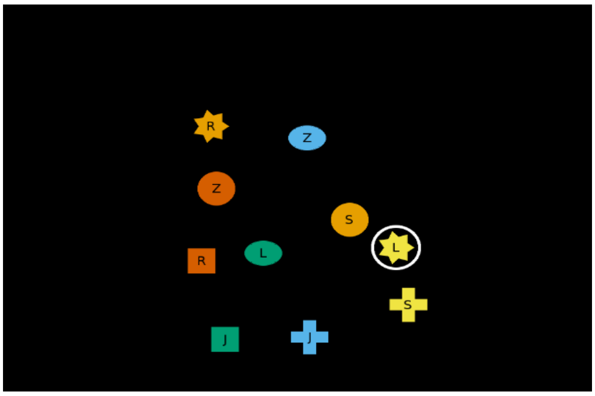

Figure 2. Screenshot of Switcher Task (PEBL Battery [37]

HRV Recording

The inter-beat intervals (IBI) or RR of HRV were collected from a chest-strap device Polar H10 (Polar Electro, Kempele, Finland), which had an excellent validity when being compared to a three-lead ECG Holter monitor [38]. The elite HRV app [39] was employed to store the recording IBI on a smartphone. This app exported raw data as a text file and imported it into Kubios HRV software [40]. Pre-processing was conducted by filtering out the artifact in the IBI time series at the medium correction level. Subsequently, the software provided the time and frequency-domain analysis of the IBI. In this study, the time-domain HRV features consisted of the mean of the heart rate (Mean HR), the standard deviation of the R-to-R intervals (SDNN), the root mean square of successive differences (RMSSD), and the percentage of successsive normal sinus RR intervals more than 50 ms (pNN50). The frequency-domain indexes included the low-frequency (LF), high frequency, and ratio of low and high frequency (see Table 1). The LF and HF were quantified in normalized units (nu), representing the proportion of the power for a specific frequency band and the summed power of the LF and HF bands.

Procedure

Upon obtaining consent, the Polar H10 device was securely placed on the participant's chest just below the chest muscles. Participants were asked to abstain from caffeine, smoking, and heavy meal consumption two hours before the data acquisition, following the methodological consideration for n HRV research [11]. This study emphasized controlling these transient variables influencing HRV. The measurement was taken while participants were sitting in front of a 15" laptop screen. The baseline and recovery measurements were collected by having participants remain stationary for five minutes. The cognitive tasks d2 attention and feature switching task were performed sequentially. To distinguish recording between two consecutive sessions, we did a manual timestamp. It allowed us to align later with HRV data analyzed via Kubios HRV software [40] and minimize the possibility of interruption between sessions. Figure 3 shows the experimental protocol.

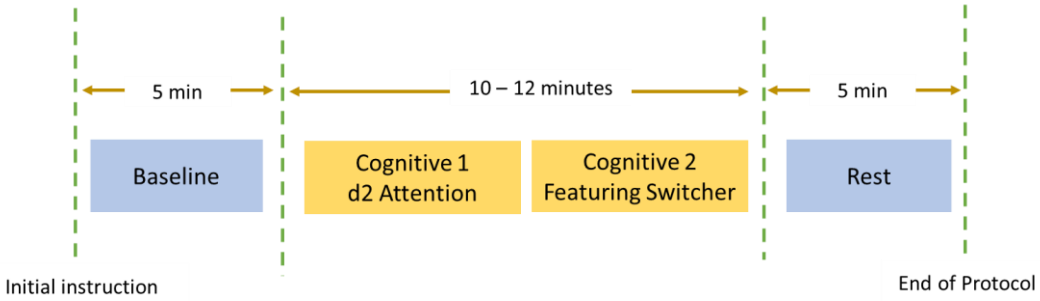

Figure 3. Experimental protocol

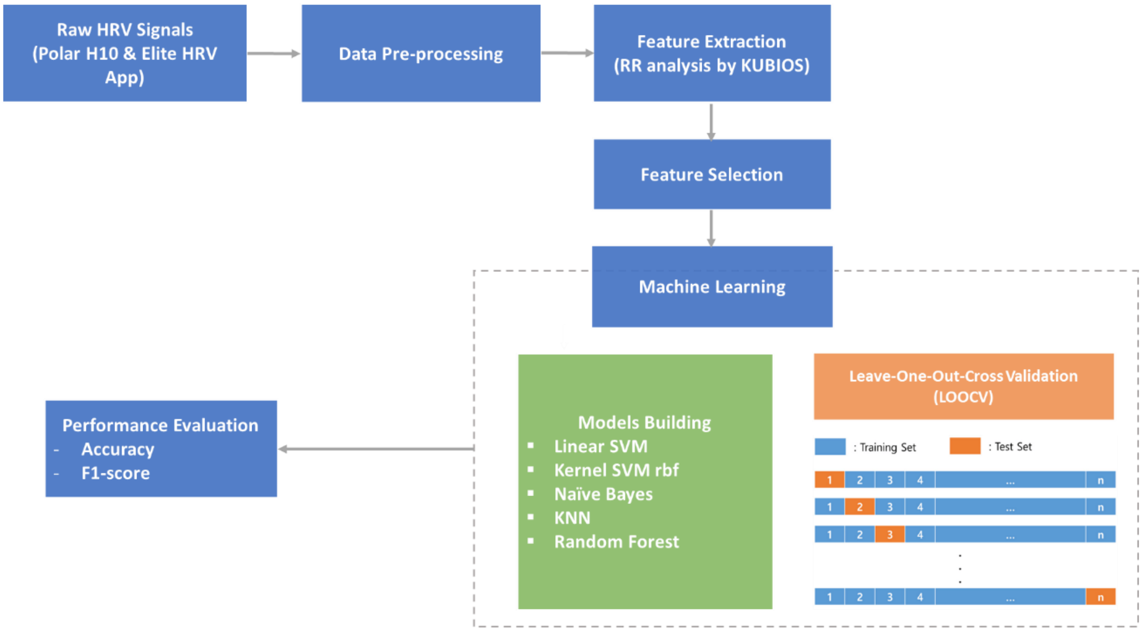

Figure 4. An architecture of HRV cognitive recognition system

Data Analysis, Feature Selection, and Machine Learning Models Development

Before classifier models were built, the normality of all HRV variables was analyzed with the Kolmogorov–Smirnov statistic. When data violates the normality assumption, the median and interquartile range (IQR: percentile 25th and 75th) are reported. The architecture of the proposed HRV-based cognitive recognition system is depicted in Figure 4. The input is raw HRV signals, and the output is the automatic decision on the two cognitive states under investigation (baseline and rest as the non-cognitive class, d2 attention, and feature switching task as the cognitive load class).

We developed a machine learning model using binary or two-classes labeling because this exploratory study focused on distinguishing whether subjects experienced cognitive load rather than classifying between cognitive tasks. We labeled baseline and rest sessions as the non-cognitive class (class 1) and the d2 attention and switcher task as the cognitive class (class 2).

As there is no clear consensus on HRV metrics in the field of HRV-based prediction through the ML approach, we selected the features based on the knowledge domain in cognitive and HRV literature [8, 11, 41] and related ML studies (see Table 2). We did not utilize purely data-driven approaches to give inputs to a machine learning algorithm. We selected HR, SDNN, RMSSD, and pNN50 from the time domain and HF, LF, and LF/HF from the frequency domains. The role of these features in the ANS and cognitive has pertained in the prior section, while a more detailed discussion can be found in [12], [24]. Furthermore, we conducted a Spearman rank correlation analysis to check the correlation between all features. Table 3 shows the moderate to high coefficient correlation for all pairs of HRV features used in this (all $p<0.001$).

We trained binary classifiers on five predictive algorithms: Linear Support Vector Machine (SVM), kernel radial basis function SVM, Naïve Bayes (NB), *k*-nearest neighbor (kNN), and Random Forest (RF).

Table 3. Spearman correlation coefficient between HRV features

	Mean HR	SDNN	RMSSD	pNN50	LF (nu)	HF (nu)	LF/HF
Mean HR	1.						
SDNN	-0.62	1					
RMSSD	-0.74	0.84	1				
pNN50	-0.73	0.84	0.89	1			
LF (nu)	0.37	-0.33	-0.60	-0.62	1		
HF (nu)	-0.38	0.34	0.60	0.61	-0.87**	1	
LF/HF	0.36	-0.33	-0.60	-0.62	0.88	-0.89	1

These classifiers were chosen because they are well established and because there is a good trade-off between their complexity and computational cost [42], [43]. The SVM, or the maximum margin classifier, is one of the best-performing predictive methods widely implemented in HRV-based classification models. This study applied linear and radial basis kernels using the default kernel functions provided by the sci-kit-learn library [44]. The kNN approach classifies new data points using their similarity (e.g., distance function) to the available data. We utilized $k = 50$ (number of nearest neighbors) and a Euclidean distance function with no distance weighting. A naïve Bayes classifier is built based on Bayes' Theorem that assumes independencies among features. Again, we used the default settings for the parameters. The random forest is an ensemble learning technique that combines the output of multiple decision trees (in this study 100) to obtain a more stable and accurate prediction than individual decision trees. It is usually trained with the bagging method [43].

Models are validated using leave-one-out cross-validation (LOOCV). This method is recommended in the affective computing domain [45] because it provides a less biased measure of the test mean square error compared to a single train-test set. It is fitted n (sample size) times by leaving out one subject for the testing and $n - 1$ for training the model.

Performance evaluation is conducted by accuracy and F-1 score. Accuracy was calculated as the number of correct predictions (sum of true positive and true negative predictions) into the binary groups (no cognitive and cognitive), divided by the total number of predictions. The F-1 score reflects the overall ability to make a correct classification. It is a harmonic mean of precision and recall or sensitivity. Precision refers to the proportion of identifications divided by the total number of classified positive samples, either correctly or incorrectly. At the same time, recall is calculated as the proportion of positive identifications divided by the total number of positive samples). The analysis was completed using sci-kit learn; a library specialized in machine learning from Python [44].

Results and Discussions

Descriptive Statistics

Table 4 displays descriptive statistics of heart rate and six HRV features within two and four classes. Since not all HRV variables met the normality assumption (e.g., pNN50), we also reported the respective medians and interquartile ranges (IQRs). Although non-normality data did not affect the ML models' performance, it is not mandatory for the classification models. We have also employed feature scaling using the standardized scale as suggested by [46], such that data have the properties of a standard normal distribution with a mean of zero and a standard deviation of one. Our data suggested that the differences in all HRV between any cognitive task and non-cognitive task were in the expected direction and magnitude. When individuals face stressors, their bodies release adrenaline and cortisol hormones which elevate their heartbeat and raise blood pressure as a way of coping with the situation [9, 12]. An increase in LF when performing cognitive tasks indicated the activation of the sympathetic system as a natural response to cope with external stressors [24]. Furthermore, reduced RMSSD, pNN50, and HF reflect the withdrawal of parasympathetic or vagal tone activity [47]. At the same time, lower SDNN is also associated with weaker cognitive performance in both global and specific cognitive domains [24]. These feature trends are in line with studies reviewing the relationship between HRV and external stimuli, such as stress [12, 48], executive functioning [47], and emotions [49]. The LF/HF, however, did not consistently increase during the cognitive load task. Although many HRV-based machine learning models have used this metric as a predictor of certain behavioral outcomes [17, 50, 51], there is a consensus among HRV scholars that the role of LF/HF as an index of balance between sympathetic and parasympathetic system is ambiguous, thus lowering its predictive value [11].

Table 4. Descriptive statistics of Heart Rate and HRV indices

			Mean	**Stdev**	**Median**	**IQR**	
HR	4 classes	Baseline	82.73	9.41	82.00	78.50	85.25
		d2 Attention	89.20	13.25	88.00	80.25	96.25
		Switcher Task	86.07	10.69	84.50	80.00	91.00
		Rest	83.50	9.25	84.50	78.75	87.50
	2 classes	No Cognitive	83.12	9.26	83.12	79.00	87.00
		Cognitive	87.63	12.04	86.00	80.25	95.50
SDNN	4 classes	Baseline	40.69	17.59	34.25	29.08	45.93
		d2 Attention	31.87	14.29	28.50	20.28	43.73
		Switcher Task	34.19	12.80	30.90	25.05	41.90
		Rest	40.26	16.51	35.45	29.98	42.18
	2 classes	No Cognitive	40.47	16.92	35.35	29.68	42.78
		Cognitive	33.03	13.50	30.05	23.83	42.30
RMSSD	4 classes	Baseline	40.60	24.82	34.15	27.23	43.50
		d2 Attention	33.42	20.62	29.55	20.18	39.80
		Switcher Task	33.07	16.43	30.85	23.20	39.23
		Rest	35.54	20.47	28.75	24.35	40.33
	2 classes	No Cognitive	38.07	22.70	30.10	25.15	41.98
		Cognitive	33.24	18.48	29.80	21.95	39.03
pNN50	4 classes	Baseline	18.66	18.22	14.67	5.82	22.90
		d2 Attention	14.63	17.79	9.17	1.58	23.44
		Switcher Task	13.31	14.65	9.79	3.37	19.32
		Rest	13.93	16.52	7.53	4.31	21.57
	2 classes	No Cognitive	16.29	17.41	9.48	4.60	21.71
		Cognitive	13.97	16.17	9.39	2.14	20.01
LF (nu)	4 classes	Baseline	53.98	20.20	57.25	35.82	71.94
		d2 Attention	53.83	17.74	57.53	38.62	69.59
		Switcher Task	58.85	14.75	60.26	47.40	70.48
		Rest	63.73	16.83	61.44	47.52	77.48
	2 classes	No Cognitive	58.86	19.08	59.25	43.65	73.15
		Cognitive	56.34	16.37	59.32	43.00	69.71
HF (nu)	4 classes	Baseline	45.89	20.20	42.68	27.56	63.83
		d2 Attention	45.93	17.67	41.97	30.37	60.92
		Switcher Task	51.07	55.22	39.74	29.33	54.08
		Rest	36.21	16.82	38.56	22.50	52.10
	2 classes	No Cognitive	41.05	19.06	40.74	26.78	56.34
		Cognitive	48.50	40.73	40.70	30.22	57.50
LF/HF	4 classes	Baseline	1.75	1.50	1.35	0.56	2.62
		d2 Attention	1.57	1.20	1.37	0.63	2.29
		Switcher Task	1.75	1.01	1.52	0.90	2.40
		Rest	2.98	3.11	1.60	0.91	3.46
	2 classes	No Cognitive	2.37	2.50	1.45	0.78	2.73
		Cognitive	1.66	1.10	1.46	0.79	2.31

Machine Learning Classifiers

The performance of each algorithm on binary classes with is presented in Table 5 with the leave-one-out cross-validation. In addition to accuracy and F-1 score, we also reported the precision and recall metrics.

The LSVM model exhibited the highest performance of the other classifiers, with around 61% of accuracy, 62% precision, 62% recall, and a 0.62 F-1 score. The lowest accuracy was exhibited by the k-NN (accuracy: 0.54, precision: 0.55, recall: 0.54), accompanied by a 0.51 F-score. The ability of LSVM to act as the best classifier in HRV analysis has also been recognized in prior studies [15, 41]. This finding supports the nature of linear correlations among time and frequency domains of HRV indices. Although k-NN was among the most used algorithm in building HRV-based ML models, its lower performance compared to other algorithms was also reported in other studies [26], [53]. This indicated the variance of performance across some supervised classification algorithms. Meanwhile, some algorithms had accuracies relatively similar to F-1 scores (e.g., Naïve Bayes). This might be due to our balance data for each class ($n = 60$ for either cognitive or non-cognitive classes). For imbalanced data, the F-1 score provides more robust results because it evaluates both recall and precision.

While the classification accuracy in this study is slightly lower when compared to prior studies [3, 25, 54], our models can be considered acceptable. Using the same dataset to infer cognitive load from a wrist-worn physiological sensor's data, Gjoreski *et al.,* [54] reported the accuracies of 13 machine-learning methods ranging from 0.50 to 0.69. Using a wearable device, Tervonen *et al.,* [3] compared six ultra-short window length measurements (5, 10, 25, and 30 s) to detect cognitive load. The authors utilized 82 – 93 selected physiological features and found that the highest accuracy was 67.6% at 25 s window length. Nevertheless, those previous studies employed multi-

Table 5. Leave-one-subject-out cross-validation accuracy for each classification model

Algorithm	Accuracy	Precision	Recall	F-1 Score
Linear SVM (LSVM)	0.607	0.620	0.615	0.617
Kernel SVM RBF	0.575	0.580	0.575	0.577
Naïve Bayes	0.542	0.545	0.540	0.542
k-Nearest Neighbor (kNN)	0.500	0.500	0.500	0.509
Random Forest (RF)	0.558	0.560	0.555	0.575

physiological modalities and sensors, including EDA, skin temperature, and HRV, while this study focused on a single sensor that extracted only HRV indices. A more specific device can be easily deployable outside laboratory settings. Another explanation may be due to the within-subject differences and individual-specific factors, which might lead to samples being misclassified. Furthermore, as mentioned earlier, we used the knowledge domain to select HRV features instead of feeding all HRV features provided by the software. However, our selected HRV features were also among the most commonly used in building ML models (see Table 2). Nevertheless, we must consider non-linear features to improve the model's performance [17, 41]. Lastly, because this is our preliminary study, we neither utilized data-driven approaches to give inputs to a machine learning algorithm nor applied feature engineering. Further studies should address this issue, for example, by following the feature selection method proposed by Gjoreski *et al.*,. [55] and employing complex feature engineering (e.g., [4, 26]).

Conclusion

This study explores several machine learning models to classify cognitive load states based on HRV signals. We found that the prediction accuracies were acceptable using one single sensor modality. Furthermore, it was observed that the Linear Support Vector Machine performed the best for binary classification. These findings can help to understand and identify, from a physiological point of view, the current cognitive state of a person. However, this finding should be interpreted with caution since we have not compared it to prior similar works, particularly that used only one single sensor (Polar H10) and employed in specific cognitive functioning tasks (i.e., d2 attention and switcher feature). This certainly limits the generalizability of the models.

Moreover, one study limitation was that the HRV data were exported by default into an R-R file in milliseconds. Therefore, raw signals ECG (e.g., QRS) were needed to extract more features which allow for improved model accuracies either by more complex feature engineering or other machine learning approaches such as deep learning and neural network. Another strategy to improve model accuracy is using several epochs or window segments. Since the validity and reliability of frequency domains HRV required at least a five-minute recording [11], one epoch can be produced by overlapping every 30-60 seconds between the windows. However, it needs a longer recording duration; for instance, we need 10 minutes with overlapping windows of 30 seconds to obtain 11 epochs, which unfortunately could not be performed in this study due to resource constraints. Moreover, the self-report questionnaires filled by the subjects can be used to predict the affective state of that specific person and be utilized to label a class. Further research also needs to investigate whether a photoplethysmography (PPG)-the based wearable device could produce good models. For example, the Empatica e4 wristband could simultaneously gather bio-signals data using only one sensor. This device can record EDA, body temperature, and blood volume pulse from which heart rate, IBI, and HRV are derived [56]. Notwithstanding the limitations, our study offers insight into further development of personal performance monitoring using cognitive performance state.

Acknowledgement

This study was funded by the Directorate of Indonesian Higher Education / DIKTI (PDP Grant 159/E5/P6/02.00.PT/2022). The authors would like to thank Tsabit al-Banan, Alimas Nurhati, Hadiatul Aula, and Rahma Indah Kusumawardani, for their assistance during the data collection and analysis.

References

1. Weintraub, S., Dikmen, S.S., Heaton, R.K., Tulsky, D.S., Zelazo, P.D., and Bauer, P.J., Cognition Assessment Using the NIH Toolbox, *Neurology*, 80(11), 2013, pp. S54–S64. doi: 10.1212/WNL.0b013e3182872ded.
2. Shields, G.S., Sazma, M.A., and Yonelinas A.P., The Effects of Acute Stress on Core Executive Functions: A Meta-Analysis and Comparison with Cortisol, *Neuroscience Biobehavioral Review,* 68, 2016, pp.651-668. doi: 10.1016/j.neubiorev.2016.06.038.
3. Tervonen, J., Pettersson, K., and Mäntyjärvi, J., Ultra-short Window Length and Feature Importance Analysis for Cognitive Load Detection from Wearable Sensors, *Electronics (Switzerland).*, 10(5), 2021, pp.1-19. doi: 10.3390/electronics10050613.
4. Sharma, K., Niforatos, E., Giannakos, M., and Kostakos V., Assessing Cognitive Performance Using Physiological and Facial Features, *Proceeding ACM Interactive, Mobile, Wearable*

Ubiquitous Technol., 4(3), 2020, doi: 10.1145/3411811.

5. Thielmann, B, Pohl, R., and Böckelmann, I., Heart Rate Variability as a Strain Indicator for Psychological Stress for Emergency Physicians During Work and Alert Intervention: A Systematic Review, *Journal of Occupational Medicine Toxicology,* 16(1), 2021, pp. 1-9. doi: 10.1186/s12995-021-00313-3.
6. Castaldo, R., Melillo, R., Bracale, U., Caserta, M. Triassi, M., and Pecchia, L., Acute Mental Stress Assessment via Short Term HRV Analysis in Healthy Adults: A Systematic Review with Meta-Analysis, *Biomedical and Signal Processing Control*, 18(April), 2015, pp. 370–377. doi: 10.1016/j.bspc.2015.02.012.
7. Lee, K. F. A., Gan, W.S, and Christopoulos, G., Biomarker-Informed Machine Learning Model of Cognitive Fatigue from a Heart Rate Response Perspective, *Sensors*, 21(11), 2021, pp. 1–16. doi: 10.3390/s21113843.
8. Shaffer, F., and Ginsberg, J.P., An Overview of Heart Rate Variability Metrics and Norms, *Frontiers in Public Health.*, 5(258), 2017, doi: 10.3389/fpubh.2017.00258.
9. Thayer, J. F., Hansen, A. L., Saus-Rose, E., and Johnsen, B. H. ., Heart Rate Variability, Prefrontal Neural Function, and Cognitive Performance: The Neurovisceral Integration Perspective on Self-Regulation, Adaptation, and Health, *Annals of Behavioral Medicine,* 37(2), 2009, pp. 141–153. doi: 10.1007/s12160-009-9101-z.
10. Hoffman, T., *What is Heart Rate Variability (HRV) & Why Does It Matter?*, *Firstbeat*, 2022, retrieved from https://www.firstbeat.com/en/blog/what-is-heart-rate-variability-hrv/ on 17 September 2022
11. Laborde, S., Mosley, E., and Thayer, J. F., Heart Rate Variability and Cardiac Vagal Tone in Psychophysiological Research – Recommendations for Experiment Planning, Data Analysis, and Data Reporting, *Frontiers in Psychology*, 8(213), 2017, doi: 10.3389/fpsyg.2017.00213.
12. Kim, H. G. , Cheon, E. J., Bai, D. S., Lee, Y. H., and Koo, B. H., Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature, *Psychiatry Investigation*, 15(3), 2018, pp. 235–245. doi: 10.30773/pi.2017.08.17.
13. Massaro, S., and Pecchia, L., Heart Rate Variability (HRV) Analysis: A Methodology for Organizational Neuroscience, *Organizational Research Methods*, 22(1), 2019, pp. 354-393
14. Yarkoni, T., and Westfall J., Choosing Prediction Over Explanation in Psychology: Lessons from Machine Learning, *Perspective in Psychology Science.*, 12(6), 2017, pp. 1100–1122. doi: 10.1177/1745691617693393.
15. Giannakakis, J., Marias, K., and Tsiknakis, M., A Stress Recognition System using HRV Parameters and Machine Learning Techniques, *Proceedings of 2019 8th International Conference on Affective Computing and Intelligent Interaction*, 2019, pp. 269–272. doi: 10.1109/ACIIW.2019.8925142.
16. Chen, C., Li, C., Tsai, C.-W., and Deng, X., Evaluation of Mental Stress and Heart Rate Variability Derived from Wrist-Based Photoplethysmography, *Proceedings of 2019 IEEE Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS),* 2019, pp. 65–68. doi: 10.1109/ECBIOS.2019.8807835.
17. Lee, S., Hwang, H.B., Park, S., Kim, S., Ha, J.J., Jang, Y., Hwang, S., Park, H., Lee, J., and Kim, I.Y., Mental Stress Assessment Using Ultra Short Term HRV Analysis Based on Non-Linear Method, *Biosensors (Basel)*, 12(7), 2022, doi: 10.3390/bios12070465.
18. Stewart, C. L., Folarin, A., and Dobson, R., *Personalized Acute Stress Classification from Physiological Signals with Neural Processes*, 2020, retrieved from http://arxiv.org/abs/2002.04176 on 23 June 2022
19. Ahmad, Z., and Khan, N. M., Multi-level Stress Assessment Using Multi-Domain Fusion of ECG Signal, *Proceedings of Annual International Conference IEEE Engineering, in Medicine and Biological Society*, 2020, pp. 4518–4521. doi: 10.1109/EMBC44109.2020.9176590.
20. Betti, S., Lova, R.M, Rovini, E., Acerbi, G., Santarelli, L., Cabiati, M., Ry, S.D., and Cavallo, F., Evaluation of an Integrated System of Wearable Physiological Sensors for Stress Monitoring in Working Environments by Using Biological markers, *IEEE Transaction on Biomedical Engineering,* 65(8), 2018, pp. 1748–1758. doi: 10.1109/TBME.2017.2764507.
21. Sanchez, W., Martinez, A., Hernandez, Y., Estrada, H., and Gonzalez-Mendoza, M., A Predictive Model for Stress Recognition in Desk Jobs, *Journal of Ambient and Intelligence of Humanized Computing*, 2018, doi: 10.1007/s12652-018-1149-9.
22. Can, Y. S., Arnrich, B., and Ersoy, C., Stress Detection in Daily Life Scenarios Using Smart Phones and Wearable Sensors: A Survey, *Journal of Biomedical Informatics,* 92(August), 2019, 103139. doi: 10.1016/j.jbi.2019.103139.
23. Hörmann, T., Hesse, M., Christ, P., Adams, M., Menßen, C., and Rückert, U., Fine-Grained Prediction of Cognitive Workload in a Modern Working Environment by Utilizing Short-Term Physiological Parameters, *Proceedings of BIOSIGNALS 2016 - 9th International Conference on Bio-Inspired System and Signal Processing*, 2016, pp. 42–51, 2016, doi: 10.5220/0005665000420051.
24. Forte, G., Favieri, F., and Casagrande, M., Heart

Rate Variability and Cognitive Function: A Systematic Review, *Frontiers in Neuroscience,* 13(July), 2019, doi: 10.3389/fnins.2019.00710.

25. Tsunoda, K., Chiba, A., Yoshida, K., Tomoki, W., and Mizuno, O., Predicting Changes in Cognitive Performance Using Heart Rate Variability, *IEICE Transacation on Information System*. E100.D(10), 2017, pp. 2411–2419. doi:10.1587/transinf.2016OFP0002.
26. Wang, C., and Guo, J., A Data-Driven Framework for Learners' Cognitive Load Detection Using ECG-PPG Physiological Feature Fusion and XGBoost Classification, *Procedia Computer Science*, 147, 2019, pp. 338–348. doi: 10.1016/j.procs.2019.01.234.
27. Castaldo, R., Montesinos, L., Melillo, P., James, C., and Pecchia, L., Ultra-Short Term HRV Features as Surrogates of Short Term HRV: A Case Study on Mental Stress Detection in Real Life, *BMC Medical Informatics and Decision Making*, 19(1), 2019, pp. 1–13. doi: 10.1186/s12911-019-0742-y.
28. Posada-Quintero, H. F., and Bolkhovsky, J. B., Machine Learning Models for the Identification of Cognitive Tasks Using Autonomic Reactions from Heart Rate Variability and Electrodermal Activity, *Behavioral Science (Basel).*, 9(4), 2019, doi: 10.3390/bs9040045.
29. Huang, S., Li, J., Zhang, P., and Zhang, W., Detection of Mental Fatigue State with Wearable ECG Devices, *International Journal of Medical Informatics*, 19(August), 2018, pp. 39–46. doi: 10.1016/j.ijmedinf.2018.08.010.
30. Alberdi, A., Aztiria, A., and Basarab, A., Towards an Automatic Early Stress Recognition System for Office Environments based on Multimodal Measurements: A Review, *Journal of Biomedical Informatics*, 59(February), 2016, pp. 49–75. doi: 10.1016/j.jbi.2015.11.007.
31. Castaldo, R., Montesinos, L., Wan, T. S., Serban, A., Massaro, S., and Pecchia, L., Heart Rate Variability Analysis and Performance during a Repeated Mental Workload Task, *Proceedings of IFMBE,* 65, 2018, pp. 69–72.
32. Lambusch, F., Weigelt, O., Fellmann, M., Hein, S., and Poppe, M., Personal Productivity Management in the Digital Age: Measures from Research and use of Conventional Tools, *Proceedings of 15th International Conference on Business and Information System WIRTSCHAFTSINFORMATIK*, 2020, doi: 10.30844/wi_2020_f5.
33. McDuff, D., Gontarek, S., and Picard, R., Remote Measurement of Cognitive Stress Via Heart Rate Variability, *Proceedings of 2014 36th Annual International Conference of the Engineering in Medicine and Biology Society,* 2014, pp. 2957–2960. doi: 10.1109/EMBC.2014.6944243.
34. Brickenkamp, R., *Test D2, Attentional Performance Test*, Hogrefe, Göttingen, Germany, 1994.
35. Steinborn, M. B., Langner, R., Flehmig, H. C., and Huestegge, L., Methodology of Performance Scoring in the D2 Sustained-Attention Test: Cumulative-Reliability Functions and Practical guidelines, *Psychological Assessment,* 30(3), 2018, pp. 339–357. doi: 10.1037/pas0000482.
36. Anderson, K., Deane, K., Lindley, D., Loucks, B**.,** and Veach, E., *The Effects of Time of Day and Practice on Cognitive Abilities: The PEBL Tower of London, Trail-Making, and Switcher Tasks*, 2012, retrieved from http://sites.google.com/site/pebltechnicalreports/home/2012/pebl-technical-report-2012-04 on 12 June 2022
37. Mueller, S. T., and Piper, B. J., The Psychology Experiment Building Language (PEBL) and PEBL Test Battery, *Journal of Neuroscience Methods*, 222(January), 2014, pp. 250–259. doi: 10.1016/j.jneumeth.2013.10.024.
38. Hinde, K., White, G., and Armstrong, N., Wearable Devices Suitable for Monitoring Twenty-Four Hour Heart Rate Variability in Military Populations, *Sensors (Switzerland)*, 21(4), 2021, doi: 10.3390/s21041061.
39. Perrotta, A.S., Jeklin, A. T., Hives, B.A., Meanwell, L.E. , and Warburton, D.E.R., Validity of the Elite HRV Smartphone Application for Examining Heart Rate Variability in a Field-Based Setting, *Journal of Strength and Conditions Research,* 31(8), 2017, pp. 2296–2302. doi: 10.1519/JSC.0000000000001841.
40. Tarvainen MP., JA, L., J-P, N., and Ranta-Aho, P., *Kubios HRV Software User's Guide*, 2021, retrieved from https://www.kubios.com/downloads/Kubios_HRV_Users_Guide.pdf. on 5 May 2022
41. Ishaque, S., Khan, N., and Krishnan, S., Trends in Heart-Rate Variability Signal Analysis, *Frontiers in Digital Health,* 3(February), 2021, pp. 1–18. doi: 10.3389/fdgth.2021.639444.
42. Borisov, V., Kasneci, E., and Kasneci, G., Robust Cognitive Load Detection from Wrist-Band Sensors, *Computer and Human Behavioral Reports*, 4(June), 2021, doi: 10.1016/j.chbr.2021.100116.
43. Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009.
44. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., and Thirion, B., Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, 12, 2011, pp. 2825–2830.
45. Schmidt, P., Reiss, A., Dürichen R., and Van Laerhoven, K., Wearable-Based Affect Recognition—a Review, *Sensors (Switzerland)*, 19(19),

2019, pp. 1–42. doi: 10.3390/s19194079.

46. Raschka, S., and Mirjalili, V., *Python Machine Learning - Second Edition: Machine Learning and Deep Learning with Python, Scikit-Learn, and TensorFlow*. Packt Publishing, 2020.
47. Pham, T., Lau, Z.J., Chen, S.H.A., and Makowski, D., Heart Rate Variability in Psychology: A Review of HRV Indices and an Analysis Tutorial, *Sensors*, 21(12) 2021, pp. 1–20. doi: 10.3390/s21123998.
48. Rodrigues, S., Paiva, J.S., Dias, D., Aleixo, M., Filipe, R. M., and Cunha, J. P. S., Cognitive Impact and Psychophysiological Effects of Stress Using a Biomonitoring Platform, *International Journal of Research Publich Health,* 15(6), 2018, doi: 10.3390/ijerph15061080.
49. Smith, T. W., Deits-Lebehn, C., Williams, P. G. , Baucom, B. R. W., and Uchino, B. N., Toward a Social Psychophysiology of Vagally Mediated Heart Rate Variability: Concepts and Methods in Self-Regulation, Emotion, and Interpersonal Processes, *Social and Personality Compass,* 14(3), 2020, pp. 1–24. doi: 10.1111/spc3.12516.
50. Al-Libawy, H., Al-Ataby, A., Al-Nuaimy, W., and Al-Taee, M. A., HRV-Based Operator Fatigue Analysis and Classification Using Wearable Sensors, *Proceedings of 13th International Multi-Conference on Systems, Signals, and Devices*, 2016, pp. 268–273. doi: 10.1109/SSD.2016.7473750.
51. He, J., Malinovic, A., and Jiang N., Fast Detection of Acute Cognitive Stress Measurement Via Heart Rate Variability, *Proceedings of the International of IEEE/EMBS Conference on Neural Engineering,* 2019, pp. 445–448. doi: 10.1109/NER.2019.8716939.
52. Géron, A., *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. Sebastopol, California: O'Reilly Media, Inc, 2019.
53. Gjoreski, M., Kolenik, T., Knez, T., Luštrek, M., Gams, M., Gjoreski, H., and Pejovic, V., Datasets for Cognitive Load Inference Using Wearable Sensors and Psychological Traits, *Applied Science*, 10(11), 2020, doi: 10.3390/app10113843.
54. Gjoreski, M., Mahesh, B., Kolenik, T., Uwe-Garbas, J., Seuss, D., and Gjoreski, H., Cognitive Load Monitoring with Wearables-Lessons Learned from a Machine Learning Challenge, *IEEE Access*, 9, 2021, pp. 103325–103336. doi: 10.1109/ACCESS.2021.3093216.
55. Gjoreski, M., Luštrek, M., Gams, M., and Gjoreski, H., Monitoring Stress with a Wrist Device Using Context, *Journal of Biomedical Informatics*, 73, 2017, pp. 159–170. doi: 10.1016/j.jbi.2017.08.006.
56. Milstein, N., and Gordon, I., Validating Measures of Electrodermal Activity and Heart Rate Variability Derived From the Empatica E4 Utilized in Research Settings that Involve Interactive Dyadic States, *Frontiers in Behavioral Neuroscience*, 14(August), 2020, doi: 10.3389/fnbeh.2020.00148.