The Realtime Assessment of Mental Workload by Means of Multiple Bio-Signals

Master thesis Report

Methodology and Statistics for the Behavioral, Biomedical and Social Sciences

Utrecht University

Bart-Jan Boverhof, 6000142

Thesis Supervisor

Prof.dr.ir. B.P. Veldkamp

Date

October 14, 2020

1 Introduction

The topic of mental workload has received widespread attention across a variety of different fields, amongst others the field of ergonomics (Young, Brookhuis, Wickens, & Hancock, 2015), human factors (Pretorius & Cilliers, 2007) and neurosciences (Shuggi, Oh, Shewokis, & Gentili, 2017). A commonly utilized definition of mental workload, hereafter referred to as simply workload, is the demand placed upon humans whilst carrying out a certain task. As pointed out by de Waard (1996), such a definition is too shallow, for it defines workload solely in external sense. It is of importance to acknowledge that workload is person-specific construct, for the amount of experienced workload ushered by a given task may differ across people (De Waard & te Groningen, 1996). Hence, when referring to workload throughout this research, workload in its person-specific sense is meant.

A commonly employed method for assessing workload is the NASA-Task Load Index (or TLX) questionnaire, operationalizing workload in clusters of six different dimensions (Hart, 2006). Such measurements are usually conducted post experiment, which might be unpractical in certain situations. For example, consider an experiment with the objective of the assessing workload for pilots during a flight. Only well after the flight a measurement in the form of a questionnaire can occur, potentially generating bias. An example of such a bias is the observer bias, prescribing that actors participating in an experiment tend to over-exaggerate the treatment effect when having to report it themselves post-experiment (Mahtani, Spencer, Brassey, & Heneghan, 2018).

An alternative method for assessing workload is the measurement of bio-signals during the experiment, and use these to classify the degree workload. Examples of such bio-signals, hereafter referred to as modalities, include techniques such as electroencephalogram, eye tracking, galvanic skin response, functional near-infrared spectroscopy, etc. The advantage of this approach is that complementary information streams, each stemming from a different modality, can all be interpreted simultaneously (Ramachandram & Taylor, 2017). This yields an objective and rich multifaceted classification of a mental construct, such as workload. Additionally, a separate model for each individual can be trained, catering towards the personal perception of workload for that specific individual. This approach comes however at the cost of an increase in complexity, residing in the need to construct a complex framework that inputs the data from each of the utilized

modalities, and ultimately outputs a single classification outcome.

The current research builds upon previous research conducted by Dolmans, Poel, van 't Klooster, and Veldkamp (in press), who proposed a framework for multi-modular classification of workload with deep-learning. The current research differs from this previous endeavor in that it utilizes different modalities and data, but most importantly because it investigates the feasibility of a real-time approach. Real-time in this sense reflects the real-time classification of workload, that is classification whilst the experiment takes place. Doing so enables the possibility to alter the state of the experiment whilst it takes place, by responding towards the perceived workload of a participant. Such an approach may enable a wide range of possibilities: consider for example a flight simulation with the objective of educating pilots. Altering an experiment dynamically after the individual could substantially enhance the learning-experience, namely by anticipating on the perceived workload during a certain segment.

In order to build both frameworks, three modalities are included. These modalities include the techniques of electroencephalogram hereafter referred to as EEG, galvanic skin response hereafter referred to as GSR and photoplethysmography hereafter referred to as PPG. It is important to stress that the objective of the current research is not to gain insight into the most optimal model design for each of the previously delineated modalities. It is rather to construct a framework with which real-time classification can be managed, and to which modalities of choice can easily be added. Consequently, one of two design principles on which the architecture of the framework reclines is the principle of modularity. Modularity refers the extent to which different modalities can freely be added and/or removed towards the framework, without the necessity to re-architect and rebuild it entirely. The second adhered design principle is the principle generalizability, prescribing that the framework should not solely be utilizable in the context of workload, but also for the measurement of other mental constructs.

A deep-learning approach towards the construction of the real-time multi-modular framework is endeavored. Considering the complexity and sheer size of such a deep neural network, an important point of attention is speed. In order for real-time classification to be a possible, a swift and thus optimized network is required. The problem in real- time classification with deep learning is often not the accuracy of classification, but rather the speed of classification. Deep neural networks easily involve thousands of calculations to be made simultaneously, which is even more true for a multi-modular approach as the current.

In order for real-time classification to work, these calculations cannot take to long, for if they do the purpose of conducting a real-time approach in general renders redundant. The focus of the current research is therefore to architect and optimize a multi-modular deep neural network, which is capable of real-time classification, whilst still ensuring ample accuracy in classification. A total of four different network variations, each inhibiting a different degree of complexity, are assessed in their ability to classify in real-time, as well as their performance. Ultimately, this line of research pursues the ability to conduct a dynamic experiment for multiple people simultaneously, the state of which can be altered in real-time.

2 Methods

2.1 Related Work

The current section will provide an overview of previous research on the network architecture for each modality separately. Additionally, the most feasible architecture for the multi-modular framework in is entirety will be explored. In particular, attention is placed upon the data fusion, the real-time component and several model optimization techniques.

2.1.1 First modality: Electroencephalogram (EEG)

The first utilized modality is EEG, which constitutes a technique that detects electrical activity in the brain using electrodes. EEG is a widely utilized method for classifying workload. An overview of the complete literature on EEG classification with deep learning has been made by Craik, He, and Contreras-Vidal (2019), who reported a total of 16 % of all available papers to constitute with workload, lending credence to the ability of EEG as classifier of workload. Additionally, the usage of deep learning techniques for a range of different EEG application was investigated upon. Summarizing, it was reported that studies mostly found deep belief networks and convolutional neural networks to perform best when classifying workload, and advice one of these approaches as a consequence (Craik et al., 2019).

Research by Schirrmeister et al. (2017) contrasted the performance of several convolutional neural networks, hereafter referred to as "ConvNets", against the widely acknowledged baseline method for EEG classification, filter bank common spatial pattern,

hereafter referred to as "FBCSP". The investigated networks included a deep, a shallow, a deep-shallow hybrid and a residual ConvNet. Both the deep and shallow ConvNets were found to reach at least similar, and in some regard better classification results, as compared with the FBCSP baseline. Altogether, a deep ConvNet with four convolutional-max-pooling blocks was found to perform best, displaying an accuracy of 92.4 % (Schirrmeister et al., 2017).

A different approach is proposed by Tabar and Halici (2016), combining a ConvNet with a Stacked auto-encoder network, hereafter referred to as "SAE". Within this network, the input layer feeds into a convolutional layer with the objective of learning the filters and network parameters. The output of this convolutional layer subsequently feeds into SAE part of the network, constituting an input layer, 6 hidden layers and an output layer. A classification accuracy of 90 % was acquired with this model (Tabar & Halici, 2016) .

2.1.2 Second modality: Galvanic Skin Response (GSR)

The second utilized modality is GSR, measuring sweat gland on the hands and hereby inferring arousal. GSR readings have been found to significantly increase as a consequence of an increase in task workload, hence constituting to be an objective predictor (Shi, Ruiz, Taib, Choi, & Chen, 2007).

Sun and colleagues explored the most optimal deep learning network for classifying six different emotional states by means of GSR data. Several models were investigated upon, amongst others a support vector machine, a ConvNet, a long-short-term-memory model and a hybrid model combining the ConvNet and long-short-term-memory, herefater referred to as "LSTM", approaches. The hybrid model was found to perform best, exhibiting an accuracy of 74% (Sun, Hong, Li, & Ren, 2019).

A variant on the CovNet LSTM model was employed by Dolmans et al. (in press), who aimed to classify workload by means of, amongst other modalities, also GSR. The performance of this model was contrasted with a network consisting solely of fully connected dense layers. Conform with findings by Sun et al. (2019), the hybrid model was found to perform best, displaying an accuracy of 82 % (Dolmans et al., in press). The model architecture as utilized by Dolmans et al. (in press) deployed two convolutional max-pooling blocks and two LSTM layers.

2.1.3 Third modality: Photoplethysmography (PPG)

The third modality constitutes PPG, which is a technique utilized to measure heart rate. PPG is, not undeservedly, a widely deployed technique within the field of workload classification. Zhang and colleagues investigated several approaches for measuring workload, comparing in total four modalities, out of which one was PPG. PPG was found to display both the highest sensitivity and reliability for measuring workload, lending credence to the feasibility of PPG as a method for classifying workload (Zhang et al., 2018).

Work by Biswas et al. (2019) investigated upon a deep learning approach towards PPG classification, with the objective to perform both bio-metric identification and obtain heart rate information. Exceptional results were attained with a neural network, attaining an average accuracy of 96 % (Biswas et al., 2019). This performance was managed with a hybrid model, incorporating two convolutional max-pooling blocks, followed with two LSTM layers.

The previously delineated model as proposed by Biswas et al. (2019) was adopted by Dolmans et al. (in press), and subsequently applied towards the MWL case. Batch normalisation and max pooling was performed in both contrasted networks (Dolmans et al., in press).

2.1.4 Fusion strategy

When architecting a multi-modular network, information streams stemming from different modalities are required to be combined, i.e. "fused", at one point in the network in order to ultimately result in a single classification. Fusion can be done at different locations, and in different ways. Several fusion strategies as proposed by Ramachandram and Taylor (2017) will be considered.

Early, or data-level, fusion focuses on how to optimally combine data sources, before being fed into the network. Techniques that realize this include for example principle component analysis or factor analysis. Early fusing is usually challenging, especially for a multi-modular situation such as the current. This resides in the fact that data stemming from different modalities often differ with regards to dimensionality and sampling rate. Another disadvantage of early fusing, is that usually the oversimplified assumption of conditional independence is made. This assumption is unrealistic in practice, for data stemming from different modalities are expected to be correlated (Ramachandram &

Taylor, 2017).

Late, or decision level, fusion on the other hand refers to the process of aggregating the decisions from multiple models, each separately applied on each modalitiy separately. In case the data sources stemming from the various modalities are either correlated or inhibit a different dimensionality, late fusion is a much more feasible approach (Ramachandram & Taylor, 2017).

Lastly, intermediate fusion is the most widely employed fusion strategy for multi-modal deep learning problems. Modalities are fused by concatenation and adding a higher order layer, to which the individual networks, separately defined for each modality, feed into. This need not be a single layer, but could be multiple layers, as long as each modality ultimately feeds into the highest order output layer. The depth of the fusion (i.e. the amount of fusion layers) can be chosen conform the specific situation, posing intermediate fusion to be the most flexible, and therefore the most widely utilized fusion strategy (Ramachandram & Taylor, 2017).

Indeed, when consulting the literature, intermediate fusion strategies are the most prevailing for multi-modular deep neural networks. When taking such an approach, one is required to consider the design of this higher order network. Rastgoo, Nakisa, Maire, Rakotonirainy, and Chandran (2019) utilize a multi-modular ConvNet approach, and fuse the modalities by concatenation, followed with two LTSM layers, two dense layers and a softmax layer. A simpler approach is utilized by Han, Kwak, Oh, and Lee (2020), who utilized an intermediate fusion approach solely consisting of several fully connected dense layers, and ending with a soft-max layer. Lastly, Dolmans et al. (in press) used a relatively deep intermediate fusion approach, consisting of two dense layers, two convolutional layers followed by another two dense layers.

2.1.5 Model optimization

The technique of batch normalization was proposed by Ioffe and Szegedy (2015), and is often applied in deep learning with the objective of enhancing the stability of a network. This is endeavored by including a batch normalization layer after each convolutional layer, re-centering and re-scaling the input feeding into this layer. If incorporating a batch normalization layer, it is recommended to do so before feeding into the activation function (Ioffe & Szegedy, 2015). An increase in accuracy for EEG classification was attained by Dolmans et al. (in press) and Schirrmeister et al. (2017) by specifying a batch

normalization layer after each convolutional layer. Equally so, the best performing model for PPG data as proposed by Biswas et al. (2019) included a batch normalization layer after each convolutional layer.

Pooling layers are often used in ConvNets, following a convolutional layer with the attempt to decrease the dimensionality. The objective of such layers are to merge similar features into one (for a more extensive elaboration see: LeCun, Bengio, and Hinton (2015)). Considering the earlier delineated EEG ConvNets, both Schirrmeister et al. (2017) and Tabar and Halici (2016) specified a max-pooling layer after each convolutional layer. The network as proposed for GSR by Sun et al. (2019) incorporated a max-pooling layer after each, but one of the convolutional layers. Lastly, the network as proposed for PPG by (Biswas et al., 2019) specified a max pooling layer after each convolutional layer.

Hyper-parameter optimization, hereafter referred to as "HPO", is a technique that can be used to optimize hyper-parameter, such as for example learning rate, dropout probability and momentum. Substantial advancements within the deep learning community have been attained by utilizing HPO, especially considering the performance of ConvNets (Bergstra & Bengio, 2012). The Optuna toolbox provides a method for creating a parameter search space, from which values for the hyper-parameters can be sampled, and thus HPO can be performed (Akiba, Sano, Yanase, Ohta, & Koyama, 2019).

2.2 Data

The current section will provide an overview of the utilized data. Special attention is placed on the experimental setup, the description of the participants and the data collection / synchronization process.

2.2.1 Experimental setup

The experimental setting for data collection is the open-source spaceship video-game Empty Epsilon, in which the respondent is required to carry out tasks on a virtual spaceship (Daid & Nallath, 2016). This experiment is instituted by the Brain Computer Interfaces (BCI) testbed lab, hosted by the University of Twente (UT) and carried out in cooperation with Thales group Hengelo. The experiment constituted three different segments, during each of which the respondent had to carry out different tasks, all of which constructed to entail varying degrees of perceived workload. Each segment consists of six

small sessions of roughly 5-10 minutes. These sessions varied in difficulty, including two easy, two intermediate and two hard sessions per segment. A schematic overview of the experimental structure is delineated as table 1. After each of the 18 sessions, respondents filled in the TLX questionnaire consisting of 6 questions each, resulting in 18 filled in questionnaires. Each questionnaire inquired upon the degree to which the respondent experienced workload during the previous session. These ratings have been use as labels in later training of the network. Within each segment, the order in which the sessions (varying in difficulty) were presented have been randomized. The order in which the segments were administrated were not random. Between every three sessions, respondents were requested to take a short 2 minute break.

Table 1

Experimental setup: Number of sessions per segment and difficulty setting

	Easy	Intermediate	Hard
Segment 1	2	2	2
Segment 2	2	2	2
Segment 3	2	2	2

The first segment emulated a scenario in which hostile spaceships approach the respondent's spaceship. The respondent is required to quickly react by defusing hostile spaceships in order to survive. The increment in difficulty caused the process of defusing hostile spaceships to be more challenging, and hereby to take longer, aiming to increase workload as a consequence. The second segment emulated a scenario in which the respondent had to navigate their spaceship trough space, gathering as many way-points as possible. Obstacles around which the respondent had to carefully navigate were introduced in the intermediate difficult scenario, and hostile spaceships the respondent had to decimate were introduced in the hard scenario. Both increase difficulty, and hereby aim to increase workload as a consequence. The third and final segment emulated a machine room, in which respondents had to control the power based on randomly generated requests. Variables that could overheat the spaceship were introduced as a consequence

of an increase in difficulty, demanding the respondent to multi-task, hereby aiming to increase workload.

2.2.2 Participants

In total, twenty-five respondents are participating in the study. Currently, the data is still in the process of being collected, for which no additional descriptive statistics can be presented in the this section as of yet. The respondents are students, recruited from the University of Twente. Recruitment has been conducted with Sona, which is a cloud-based participant management system. Requirements were that respondents didn't have any constraints that might interfere with the utilized sensors, such as for example a pacemaker. This was assessed by means of a short demographic questionnaire prior to the experiment. Additionally, the respondents were made aware of ethical consent prior to the experiment, with the objective to ensure completely voluntary participation. Respondents were able to draw back from the experiment at any time.

2.2.3 Devices and Synchronization

The Shimmer3 GSR+ sensor was used for both PPG and GSR measurements. The device is worn on the wrist, and is able to communicate the signal wirelessly. An ear-clip was utilized for measuring PPG, and converting this to estimate heart rate. Skin conductivity, or GSR, was monitored by two electrodes attached to the fingers (Shimmer-Research, n.d.). Both PPG and GSR were measured on a sample rate of 256 Hz. EEG measurement is conducted with the shimmer 2, equally so on a sampling rate of 256 Hz.

Data streams stemming from the different modalities were required to be properly synchronized such that they are parallel. This was accomplished by means of an application called Lab-Streaming Layer, hereafter refered to as "LSL", to which the different data are streamed during the experiment. LSL properly synchronized these data streams, such that they refer to the same points in time, and subsequently record all data into a single file per participant (Kothe, Medine, & Grivich, 2018).

2.3 Framework Architecture

As was elaborated on in the introduction, several networks will be compared in their ability to classify workload in real-time, and the performance with which this is managed.

The upcoming section firstly describes the universal architecture that all networks variants have in common. Subsequently, the distinctions in each of the four varying networks are delineated.

2.3.1 Universal Framework Architecture

The architecture for the proposed framework is constructed by combining insights from the literature, whilst keeping in mind the previously delineated design principles (i.e. the principles of modularity and generalizability). No variations based on the general framework architecture will be made, for deviating from the architectures as verified in the literature could be detrimental with regards to classification accuracy. A visual representation of the complete network is depicted as figure 1.

The utilized network for the EEG modality is a ConvNet, as proposed by Schirrmeister et al. (2017). The network is designed to include four convolutional blocks, each constituting a convolutional layer, followed by a batch normalisation layer. The Exponential Linear Unit, hereafter referred to as "ELU", function is utilized as activation function, and is defined as equation 1. Lastly, each block is closed with a max pooling layer of stride three.

$$f(x) = \begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases} \tag{1}$$

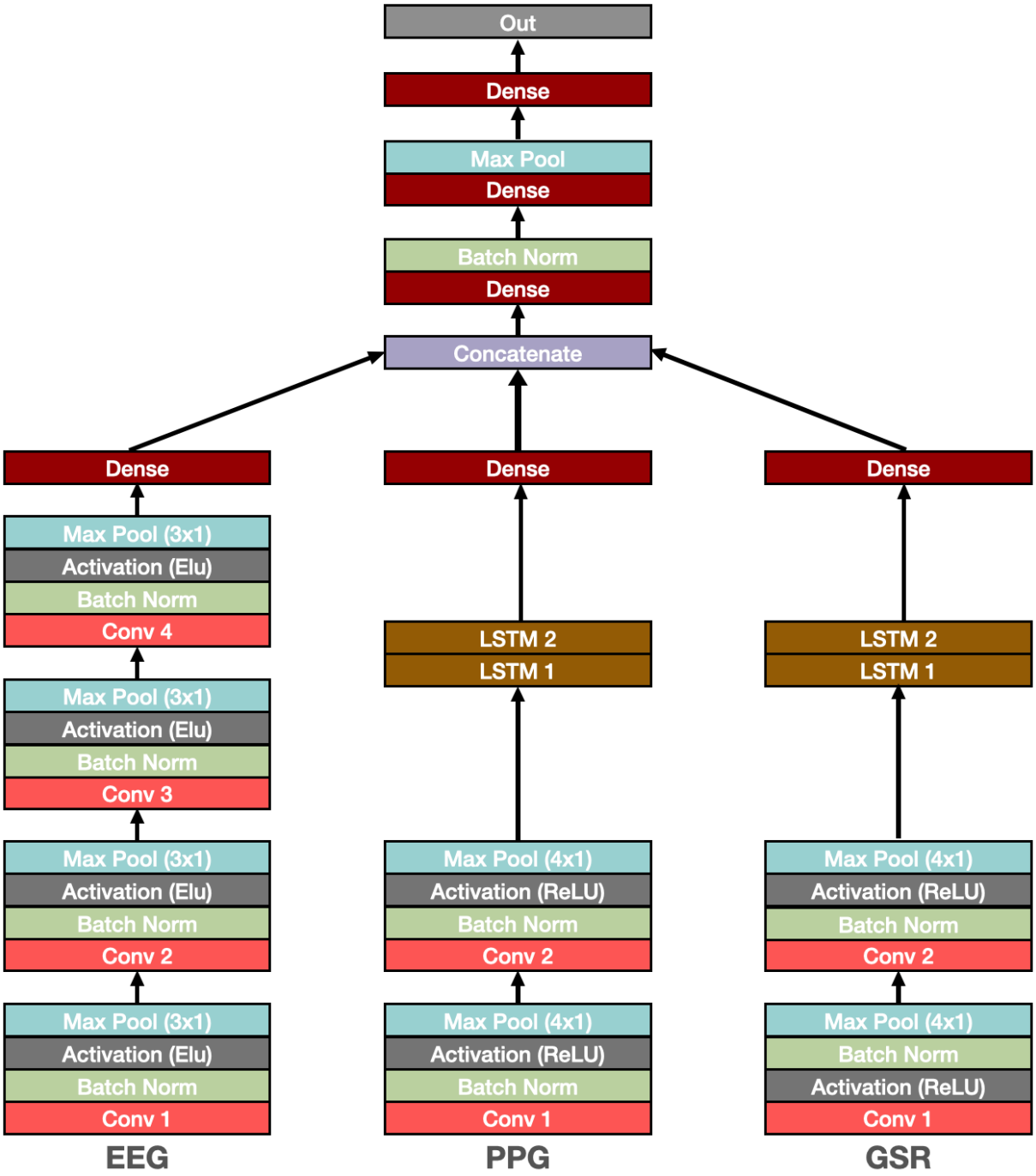
The utilized network for the GSR modality is a LSTM ConvNet network, inspired by the work of Sun et al. (2019) and Dolmans et al. (in press). The network is designed to include two convolutional blocks, each constituting a convolutional layer, followed by a batch normalization layer, an activation layer and closed with a max-pooling layer of stride four. The Rectified Linear Unit, hereafter referred to as "ReLU", function is utilized as activation function, defined as equation 2. Following these two convolutional blocks are two LTSM layers.

$$f(x) = max(0, x) \tag{2}$$

Lastly, the PPG modality is analyzed by means of the network as proposed by Biswas et al. (2019). The nework opens with two convolutional blocks, each consisting of a

Figure 1

Full Network Architecture



convolutional layer, batch normalization layer, activation layer and closed with a max pooling layer of stride four. The utilized activation function is the ReLU, depicted as equation 2. Following these convolutional blocks are two LTSM layers, equal to the GSR model.

The networks for each of the modalities are finally closed with one fully connected dense layer, before feeding into the head network. This is done in order to flatten all inputs into a lower dimensional space, such that concatenation is possible. The head network consists of Four dense layers, alternated with some batch normalization and max-pooling layers, with the objective of stabilization.

2.3.2 Model Variations

Multi-modular classification in real-time by means of deep-learning requires an optimized network, as was elaborated upon in the introduction. Speed is a potential bottleneck, especially given the proposed network architecture is multi-modular, and substantially complex. Different network variations will be investigated upon as a consequence of this. The aim is to propose a network that is fast enough for proper real-time classification, whilst ensuring the maximum amount of accuracy. Variations based on two different parameters of the network as depicted in figure 1 will therefore be considered, resulting in a total of four variations.

The first parameter constitutes the sampling rate for each of the included modalities. As was elaborated on earlier, data collection for each modality was conducted on a sampling rate of 256Hz. Naturally, an increase in classification speed can be attained by feeding less data into the network, i.e. decreasing the sampling rate. Two out of the four networks will consequently be fed data measured on a sampling rate of 256Hz, whilst the other two networks will input based on a sampling rate of 128Hz.

The second parameter reflects the amount of utilized filters for convolutional layers, and the amount of neurons for all utilized layers. A decrease in the amount of filters / neurons constitutes a decrease in network size, and consequently an increase in speed. Two out of the four models will be the full sized network, whilst the other two models will be halved in size. An overview of all four network variations is provided in table 2.

Table 2

Model variation sizes

	EEG	GSR	PPG	Head
Model 1: 128Hz &	Conv1: 25	Conv1: 128	Conv1: 128	Dense: 712
Model 2: 256Hz	Conv2: 50	Conv2: 128	Conv2: 128	Dense: 356
	Conv3: 100	LSTM1: 256	LSTM1: 256	Dense: 178
	Conv4: 200	LSTM1: 256	LSTM2: 256	
	Dense: 200	Dense: 256	Dense: 256	
Model 3: 128 Hz &	Conv1: 13	Conv1: 64	Conv1: 64	Dense: 356
Model 4: 256Hz	Conv2: 25	Conv2: 64	Conv2: 64	Dense: 178
	Conv3: 50	LSTM1: 128	LSTM1: 128	Dense: 89
	Conv4: 100	LSTM1: 128	LSTM2: 128	
	Dense: 100	Dense: 128	Dense: 128	

Note: For all convolutional layers the depicted number reflects the amount of utilized filters, whereas for LTSM layers it reflects the amount of nodes.

2.4 Model Evaluation

The performance of the four network variations will be contrasted by means of several performance metrics. The utilized metrics constitute six well known and widely applied metrics, all constructed from the confusion matrix, depicted as table 3.

Table 3

Confusion matrix

	True Positive	True Negative
Predicted Positive	a	b
Predicted Negative	c	d

The measures accuracy, sensitivity, specificity, PPV, NPV and F1 will be utilized in order to asses network performance. The network that performs best across these measures is considered to be the superior performing network. Table 4 depicts the constitution of these performance metrics, by partly referring to confusion matrix depicted as table 3.

Table 4

Performance Metrics

Accuracy:	$\frac{a+d}{a+b+c+d}$
Sensitivity:	$\frac{a}{a+c}$
Specificity:	$\frac{d}{b+d}$
Positive Predicted Value (PPV):	$\frac{a}{a+b}$
Negative Predicted Value:	$\frac{d}{c+d}$
F1-measure:	$\frac{2*Sensitivity*PPV}{Sensitivity+PPV}$

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining* (pp. 2623–2631).

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, *13*(1), 281–305.

Biswas, D., Everson, L., Liu, M., Panwar, M., Verhoef, B.-E., Patki, S., ... others (2019). Cornet: Deep learning framework for ppg-based heart rate estimation and biometric identification in ambulant environment. *IEEE transactions on biomedical circuits and systems*, *13*(2), 282–291.

Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (eeg) classification tasks: a review. *Journal of neural engineering*, *16*(3), 031001.

Daid, & Nallath. (2016). *Empty epsilon multiplayer spaceship bridge simulation.* URL: https://github.com/daid/EmptyEpsilon. GitHub.

De Waard, D., & te Groningen, R. (1996). *The measurement of drivers' mental workload.* Groningen University, Traffic Research Center Netherlands.

Dolmans, T., Poel, M., van 't Klooster, J.-W., & Veldkamp, B. (in press). Percieved mental workload detection using intermediate fusion multi-modal networks.

Han, S.-Y., Kwak, N.-S., Oh, T., & Lee, S.-W. (2020). Classification of pilots' mental states using a multimodal deep learning network. *Biocybernetics and Biomedical Engineering*, *40*(1), 324–336.

Hart, S. G. (2006). Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, pp. 904–908).

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Kothe, C., Medine, D., & Grivich, M. (2018). Lab streaming layer (2014). *URL: https://github. com/sccn/labstreaminglayer (visited on 02/01/2019)*.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Mahtani, K., Spencer, E. A., Brassey, J., & Heneghan, C. (2018). Catalogue of bias: observer bias. *BMJ Evidence-Based Medicine*, *23*(1), 23.

Pretorius, A., & Cilliers, P. (2007). Development of a mental workload index: A systems

approach. *Ergonomics*, *50*(9), 1503–1515.

Ramachandram, D., & Taylor, G. W. (2017). Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Processing Magazine*, *34*(6), 96–108.

Rastgoo, M. N., Nakisa, B., Maire, F., Rakotonirainy, A., & Chandran, V. (2019). Automatic driver stress level classification using multimodal deep learning. *Expert Systems with Applications*, *138*, 112793.

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., ... Ball, T. (2017). Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, *38*(11), 5391–5420.

Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007). Galvanic skin response (gsr) as an index of cognitive load. In *Chi'07 extended abstracts on human factors in computing systems* (pp. 2651–2656).

Shimmer-Research. (n.d.). *Shimmer3 gsr unit.* Retrieved from `https://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor`

Shuggi, I. M., Oh, H., Shewokis, P. A., & Gentili, R. J. (2017). Mental workload and motor performance dynamics during practice of reaching movements under various levels of task difficulty. *Neuroscience*, *360*, 166–179.

Sun, X., Hong, T., Li, C., & Ren, F. (2019). Hybrid spatiotemporal models for sentiment classification via galvanic skin response. *Neurocomputing*, *358*, 385–400.

Tabar, Y. R., & Halici, U. (2016). A novel deep learning approach for classification of eeg motor imagery signals. *Journal of neural engineering*, *14*(1), 016003.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17.

Zhang, X., Lyu, Y., Hu, X., Hu, Z., Shi, Y., & Yin, H. (2018). Evaluating photoplethysmogram as a real-time cognitive load assessment during game playing. *International Journal of Human–Computer Interaction*, *34*(8), 695–706.