Detecting Mental Fatigue in Vessel Pilots Using Deep Learning and Physiological Sensors

Thiago Gabriel Monteiro[1], Houxiang Zhang[1], *Senior Member, IEEE,* Charlotte Skourup[2], and Eduardo Aoun Tannuri[3]

***Abstract*— Nowadays, human related issues are the main causes of accidents in the maritime domain. Among these issues, mental fatigue is responsible for reducing cognitive capabilities, situational awareness, and decision-making skills. Early detection and assessment of mental fatigue can be used to reduce the number of causalities, to the benefit of crewmembers, ship owners, and the maritime environment. Although the use of physiological sensors is the most trusted approach for measuring mental fatigue, it is a complex task due to the different ways mental fatigue can manifest in different people. In this paper, we present the application of deep learning techniques and physiological sensors to assess mental fatigue in the maritime domain, using a vessel piloting task as case study. The results demonstrate that because of their ability to extract features otherwise hard to recognize from in data, deep learning techniques in special convolutional neural networks can achieve high levels of mental fatigue classification accuracy, although cross-subject classification performance is still not sufficient for real-life applications.**

I. INTRODUCTION

Human related issues and human errors greatly contribute to accidents in several domains, such as driving [1], aviation [2], air traffic control [3], nuclear power plants control [4], and maritime operations [5]. This role of bad cognitive performance in unsafe behaviors, near-misses, and accidents is particularly well-established in the maritime domain [6]–[8].

The concept of operator functional state (OFS) offers a good approach for analyzing which factors can lead to human errors. OFS characterizes how well human operators react to both internal and external operational demands, according to their psychological and cognitive capabilities [9]. The concept can be divided into three main areas, namely, situation awareness, mental workload, and mental fatigue (MF). MF accumulates as the operation progresses and is responsible for decreasing the operator's capacity to react to unexpected events and understand and solve problems.

This work was supported by the project "SFI Offshore Mechatronics" funded by Norway Research Council, Norway (Project No: 237896) and INTPART Subsea - Subproject USP/NTNU (NFR no. 261824) and partially supported by the project SFI Marine Operations funded by Norway Research Council, Norway (Project No: 237929). The last author also thanks the CNPq (process 304784/2017-6) for the research grant.

[1]Thiago Gabriel Monteiro and Houxiang Zhang are with the Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology, Aalesund, Norway {thiago.g.monteiro, hozh}@ntnu.no

[2]Charlotte Skourup is with ABB As, Products and Services R&D; Oil, Gas and Chemicals, Oslo, Norway charlotte.skourup@no.abb.com

[3]Eduardo Aoun Tannuri is with the Department of Mechatronics Engineering and Mechanical Systems, University of São Paulo, São Paulo, Brazil eduat@usp.br

We currently lack the tools we need to track and fully address the role of MF in maritime operations. International organizations such as the International Maritime Organization provide direction as to how to mitigate the effects of MF, but they do not provide objective ways to track and control MF [10]. Established MF assessment methods such as Crew Status Survey [11], the Karolinska Sleepiness Scale (KSS) [12], and the Chalder Fatigue Scale [13] are subjective. To the best of our knowledge, there is currently no method to detect and measure MF in real time in maritime operations without relying solely on subject evaluation. This paper represents a step toward using objective methods such as monitoring operators behavior, reaction time, and physiological signals, such as electrocardiogram (ECG), electromyogram (EMG), electrooculogram (EOG) and electroencephalogram (EEG).

Using physiological sensors to detect MF presents complications. The multivariate time series data obtained from these sensors is usually noisy, highly dimensional, and non-stationary, have an explicit dependency on the time variable, and features extracted from the data must be invariant to translations in time [14]. Also, physiological data can manifest differently in different individuals, making pattern recognition and feature extraction complicated. The need for real time implementation is also a serious constraint and limits which kind of algorithm and preprocessing can be performed. The use of deep learning techniques can provide reasonable accuracy levels together with high efficiency, reducing the need for data preprocessing.

In this paper we investigate the use of a deep learning algorithm to perform MF assessment in maritime operations using physiological sensors. The rest of the paper unfolds as follows. Section II briefly describes our experimental approach and the methodology used to MF assessment. Section III presents and discusses the obtained results for both single- and cross-subject classification problems. Section IV concludes the paper and presents directions for future work.

II. MATERIALS AND METHODS

In order to evaluate the efficiency of different deep learning methods for MF classification, we performed experiments in a vessel simulator. The experiment applies a mixed method approach, where qualitative and quantitative approaches are used simultaneously, in the form of questionnaires and

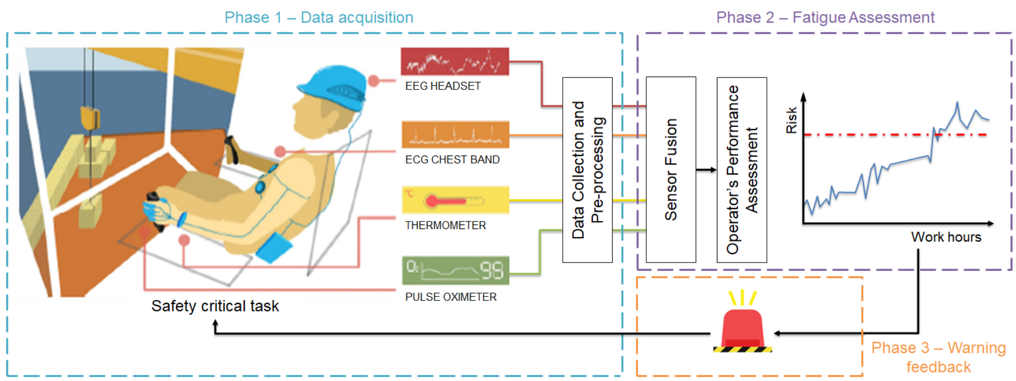

Fig. 1. Proposed sensor framework for fatigue assessment.

scenario-based experiments. This section describes our experimental setup and the proposed sensor framework for MF assessment.

A. Experimental setup

This experiment was performed in a partnership with University of São Paulo/TPN [15], using their simulator facilities to conduct the experiments. The case study consisted of a harbor navigation task with a large vessel (oil tanker, container ship). The operation required moderate levels of attention due to sea conditions and vessel traffic near the harbor. The task took between 60 and 90 minutes to complete and the operator was monitored with physiological sensors during the whole duration of the experiment.

The simulated operations were conducted by trained personnel from the Brazilian navy, in order to make the results more reliable. In total nine navigation runs were collected. Six different pilots performed the runs: all of them performed one simulation run during the morning and three performed an additional run during the afternoon on a different day. We ensured that no participant took any kind of drug or stimulant eight hours prior to the experiment and all participants had at least eight hours of sleep the previous night. The test subjects were all men between 19 and 48 years old.

For the qualitative portion of the experiment the Karolinska Sleepiness Scale (KSS) questionnaire was used as a self-assessment tool. Each participant indicated their MF state at the beginning and at the end of each of their experiments. In this research we are not considering intermediate levels of MF. We are only interested in how the final MF state compares to the initial one. This information will help us to label the data obtained during the experiments.

B. Sensor framework

Figure 1 shows the sensor framework used here to assess the fatigue state of human operators during demanding maritime operations. Its work-flow has three main phases: data acquisition, MF assessment, and warning feedback. For this work we are focusing on the first and second phases.

During the data acquisition phase we used a set of sensors to collect physiological data from the operator. The data is collected from disparate sensors and is centralized by a micro-controller. The possible physiological sensors include the ECG, EMG, body temperature sensor, EEG, EOG, and eye tracker. During this phase the data is also preprocessed to remove noise and unwanted artifacts that can disturb the data fusion and classification processes. In this experiment, 1 ECG channel and 6 EEG channels were recorded. We opted for portable sensors in order to interfere as little as possible in the simulated task. The ECG of choice was the Electrocardiogram Sensor PRO for MySignals (eHealth Medical Development Platform) [16] and we opted for the 14 channels EEG headset Emotiv Epoc+ [17].

During the MF assessment phase a sensor fusion algorithm was applied to the preprocessed data. This algorithm fused the disparate data channels and outputting a fatigue indicator. This fatigue indicator was registered and compared to a specified threshold in order to evaluate the risk level in the operation. In case the current risk level exceeds the specified threshold, a feedback system is activated to warn the operator on the high risk of accident. This alert system can be visual, sonorous, or tactile. Extreme levels of risk may require the interruption of the operation or replacement of the operator.

C. Data acquisition and preprocessing

The first phase of our framework consists of recording sensor data and preprocessing it to input in the MF state classification algorithm. The preprocessing phase includes filtering, smoothing, resampling, and discrete wavelet transform (DWT). During the filtering phase, artifact removal algorithms can be applied to remove unwanted perturbation from the desired signal. Any noise contaminating the data can be removed during the smoothing phase. If necessary, the channels from different sensors can be resampled to a desired frequency and aligned to ensure temporal correlation between different signals.

The ECG channel and the six EEG channels were sampled at 128 Hz. The EEG channels were preprocessed using DWT to decompose them into their main frequency bands of clinical interest for MF detection: delta (δ), theta (θ), alpha (α), and beta (β) [18]. Working with wavelets is advantageous in this case since they merge time and frequency domain characteristics, decomposing the original time-series into several time-series on the desired frequency bands.

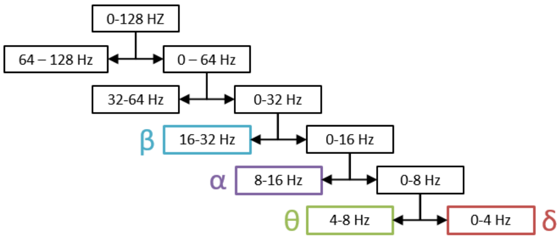

Fig. 2. DWT filter bank.

This approach allows a frequency domain analysis while conserving the temporal characteristics of the data.

The DWT of a signal is calculated by passing it through a series of filters known as a filter bank. First the input samples are passed through a pair of quadrature mirror filters, one high pass and one low pass. The convolution with the high pass filter provides the detail coefficients that represent the upper half frequencies of the original input. The convolution with the low pass filter provides the approximation coefficients that represent the lower half frequencies of the original input. The process is repeated with the output of the low pass filter and the signal is further decomposed. This process is repeated successively until the desired frequency decomposition is reached. The filter bank applied on this work is shown in Figure 2, where the location of the desired EEG sub band coefficients are shown.

D. Input data preparation

We decided to use only four input channels in order to make the input dimension manageable for the neural network (NN). We chose the alpha sub-band of the electrodes AF4, F4 and O2, as they have the greatest relevance for fatigue detection [18]. We didn't apply any artifact removal in the EEG data, since deep learning approaches can extract very complex set of features from the input data. Although not as relevant for MF detection as the EEG, the ECG was chosen as the last input channel, since it brings a completely new set of features that can help classify the MF states in a situation where only EEG channels are not enough.

The input data for the deep learning methods was composed using low-level (or raw data) fusion in the selected data channels from ECG and EEG, after preprocessing. The preprocessed data is segmented in four-second segments with no overlap. The segments from different channels were concatenated as a one-dimensional input vectors for the NN algorithms.

E. Mental fatigue state classification

In this section we briefly introduce some traditional NN methods we applied in the MF state classification task.

1) Feed forward neural network (FFN): The FFN is a type of artificial NN where the connection between nodes (and layers) is not recursive. The information transmission is unidirectional, moving only forward, going from the input layer through the hidden layers (if any) and arriving at the output layer. An FFN can be a single-layer perceptron or a multi-layer perceptron with at least one input, one hidden and one output layers.

2) Convolutional neural network (CNN): The CNN is a type of deep NN commonly used for imagery classification. It is composed by one input and one output layer as well as multiple hidden layers that provide the network its depth. The hidden layers include convolutional layers, pooling layers, and fully connected layers. The combination of convolutional and pooling layers form the feature extraction portion of the network and a CNN can have as many of those pairs as desired. A CNN can also have several fully connected layers and the last one is responsible for the classification task.

3) Deep belief neural network (DBN): A DBN is a type of NN that can learn relevant features from datasets in an unsupervised way (with no labeled data), by mean of an encoding and reconstruction process performed by Restricted Boltzmann Machines (RBMs). These RBMs are stacked together to provide depth to the network, allowing it to learn very complex features. In an ordinary DBN, a fully connected layer tops the stacked RBM layers, which are responsible for performing data classification. Each RBM layer is trained individually using a contrastive divergence algorithm. Then the whole network (including the fully connected layer) is fine-tuned using a supervised back propagation algorithm.

III. RESULTS AND DISCUSSION

EEG characteristics are not the same for every person. Each individual presents different patterns in the EEG data as the MF state develops. Even for the same person, these patterns can change depending of factors such as level of excitement, period of the day, consumption of stimulants, etc. This makes the classification of fatigue across different session and across different subjects very hard. In this section we will present the experimental results obtained by the deep learning methods when considering single session and single subject cases and single session and multiple subject cases.

A. Single subject classification

For the single subject case, we trained each NN on the data for each subject individually. During the training process, a 20-fold cross-validation approach was used. We divided each dataset into five equal size parts. These portions were combined in 20 different sets of training, validation, and test sets in the proportion 60%-20%-20%. The NN were trained using the back propagation algorithm for a minimum of 10 epochs. After every epoch the validation set was used to adjust the dynamic learning rate policy. If the validation classification didn't improve after four epochs, the learning rate was reduced by 20%. If the validation classification didn't improve after 10 epochs (indicating trapping in local minimum or overfitting), the training process was terminated and the NN internal parameters that yielded to the best validation accuracy were reloaded in the NN model. Then the test set was used as new data for the network in order to assess its classification and generalization capabilities.

TABLE I
HYPERPARAMETERS SELECTION (SINGLE SUBJECT CASE)

Network	Parameters range	Best configuration
MLP	# layers = [1, 2, 3, 4] # nodes = [256, 512, 1024, 2400]	{2400, 1024, 512}
CNN	# layers = [1, 2, 3, 4] # nodes = [32, 64, 128, 256] filters size = [3, 5, 8, 10]	{64 (3), 128 (5)}
DBN	# layers = [1, 2, 3, 4] # nodes = [256, 512, 1024, 2400]	{2400, 1024, 512}

The good performance of NN models is intrinsically related to the correct selection of their hyperparameters. Hyperparameters are responsible for controlling the learning process in the NN, and, unlike, for example, nodes weights, are not learned during the training process. Important hyperparameters include the number of layers, number of nodes per layer, learning rate, etc. Using a grind search approach, we investigated different sets of hyperparameters for each NN model applied in our case study since they are network dependent. The tested values for each hyperparameter for each NN model and the best configuration selected for the single subject case study is presented in Table I.

The results obtained with each kind of NN and for each subject are shown in Table II, where the average accuracy and its standard deviation obtained in the 20-fold cross-validation process for each case is presented. The best result for each subject is highlighted in bold. Analyzing Table II it is possible to see that none of the networks achieved outstanding results. The CNN and DBN presented an overall better classification performance than the FFN. For some subjects such as Subject 1, all the NNs were capable of achieving good classification results. For others, such as Subject 6, all the NNs performed badly. Usually the causes of this problem are two-fold: bad input data quality or bad data labeling. These factors affect the classification performance of the NN by making it difficult to differentiate correctly between two MF states. These points are discussed in the next sections.

B. Improving input data

One of the main sources of problems on sensor data are sources of noise during the data acquisition process. These sources include hardware-related noises such as power line interference and electrodes contact and process related noises such as eye and muscle artifacts in EEG data. We are handling noise by using DWT to decompose the original data in frequency bands of interest, as discussed in Section II. Keeping only slower frequency bands in the data reconstruction process filters out mid- to high-frequency noises. Here we ignore the presence of artifacts, since rejecting than would incur extra computational cost and would require the acquisition of extra data during the experiment, for example, EOG data for eye artifact rejection. This could make real-time applications inviable.

The input data can also be made better by improving the representation of the desired phenomena to the NN. Since we are handling time-series data, the temporal dependency of observations is relevant. We previously defined the representation of the MF state at a specific time interval as a vector composed by sections of the desired sensor channels concatenated as a single input. This conserves the time dependent aspect of each channel inside that specific interval, but it may cause loss of information in the inferior and superior fringes of the interval if the representation of an important phenomena is divided between two consecutive intervals.

We used two approaches to ensure better representation of the MF state in the input vectors. First, we increased the size of the time interval evaluated from four seconds to six seconds. Second, we implemented an overlap of two seconds between consecutive intervals. The increase in size of the input vector imposed a significant computational constraint to FFN and DBN, since the input vector size increase significantly increases the number or trainable parameters in these NN. Despite the extra computational requirement, no significant increase in classification performance was found when using these two NNs.

On the other hand, the CNN performance wasn't much affected, since the size and number of trainable kernels don't increase with the increase of the input vector size. The CNN classification results with the new input configuration is presented in Table III. With this new configuration the CNN presented high level of accuracy and consistency across all test subjects.

C. Mental fatigue state analysis using Functional Principal Component Analysis

The approach we used for labeling the experiment data consisted of using the KSS self-assessment fatigue questionnaires. The obtained levels of MF at the beginning and at the end of the experiment were used to define a relation between the initial and final MF for every test subject. The absolute relation between the two states was not important, only which states were more likely to represent non-fatigue and fatigue states. These labels were assigned to data on the first and last thirds of each dataset. The flaw in this approach is that it assumes no variation in the subject's MF state during these periods. The classification results of Subject 6 suggest that this can be a too strong assumption in some cases.

A more reliable way to label the MF state observations would be to increase the classification performance in datasets with complex MF state distribution. For this end, one can apply principal component analysis (PCA) to reduce a large set of variables to a smaller one, while still retaining most of the information present in the original dataset. Besides working as a dimensionality reduction tool, PCA can help with interpreting complex data correlation and underlying phenomena hidden in the data.

Despite its usefulness, PCA doesn't perform well with time-series data, since it doesn't consider temporal correlation among observations. Functional PCA (FPCA) can be

TABLE II

SINGLE SUBJECT CLASSIFICATION RESULTS. (ACCURACY ± STD)

	FFN			CNN			DBN		
	train	validation	test	train	validation	test	train	validation	test
Subject 1	0.94 ± 0.02	0.83 ± 0.03	0.82 ± 0.03	0.93 ± 0.09	0.88 ± 0.07	**0.86 ± 0.08**	0.95 ± 0.02	0.86 ± 0.02	0.84 ± 0.03
Subject 4	0.94 ± 0.03	0.87 ± 0.02	0.85 ± 0.02	0.93 ± 0.04	0.88 ± 0.03	0.86 ± 0.03	0.95 ± 0.02	0.89 ± 0.02	**0.89 ± 0.02**
Subject 5	0.92 ± 0.04	0.75 ± 0.04	0.71 ± 0.03	0.73 ± 0.16	0.71 ± 0.11	0.68 ± 0.13	0.95 ± 0.03	0.75 ± 0.04	**0.75 ± 0.05**
Subject 6	0.60 ± 0.09	0.56 ± 0.03	0.52 ± 0.03	0.81 ± 0.15	0.72 ± 0.09	**0.71 ± 0.10**	0.57 ± 0.07	0.54 ± 0.04	0.50 ± 0.04
Subject 7	0.90 ± 0.04	0.74 ± 0.04	0.73 ± 0.04	0.80 ± 0.14	0.73 ± 0.09	0.70 ± 0.09	0.93 ± 0.02	0.78 ± 0.02	**0.76 ± 0.03**
Subject 8	0.78 ± 0.06	0.71 ± 0.04	0.68 ± 0.04	0.91 ± 0.11	0.81 ± 0.06	**0.78 ± 0.04**	0.62 ± 0.06	0.60 ± 0.07	0.58 ± 0.06
Subject 9	0.88 ± 0.04	0.73 ± 0.03	**0.71 ± 0.04**	0.71 ± 0.20	0.62 ± 0.09	0.61 ± 0.11	0.92 ± 0.03	0.73 ± 0.04	0.70 ± 0.04

TABLE III

SINGLE SUBJECT CLASSIFICATION RESULTS FOR CNN USING NEW INPUT CONFIGURATION. (ACCURACY ± STD)

	train	validation	test
Subject 1	0.94 ± 0.03	0.91 ± 0.02	0.89 ± 0.02
Subject 4	0.99 ± 0.01	0.97 ± 0.02	0.96 ± 0.02
Subject 5	0.97 ± 0.11	0.89 ± 0.10	0.88 ± 0.09
Subject 6	0.87 ± 0.03	0.83 ± 0.03	0.82 ± 0.02
Subject 7	0.97 ± 0.06	0.88 ± 0.05	0.85 ± 0.07
Subject 8	0.99 ± 0.01	0.94 ± 0.02	0.92 ± 0.02
Subject 9	0.99 ± 0.01	0.91 ± 0.02	0.89 ± 0.03

used to handle functional data such as time-series [19]. FPCA works like the traditional PCA, but instead of handling discrete observations it takes approximation of variables observations as functions. For performing the FPCA, we opted to use the MATLAB package PACE [20].

Figure 3 presents the correlation surface between different MF states during the duration of the experiment. The yellow color represents a positive correlation while blue represents a negative correlation. In this case we can interpret these correlations as a degree of agreements between two MF states. Analyzing the correlation for the initial MF state we can see it is positively correlated with every state from 1 to 1,000 and from 2,500 to 3,500 and negatively correlated to each state from 1,000 to 2,500 and from 3,500 to 3,800.

Complementing the information obtained from the correlation surface with the information obtained from the KSS questionnaire during the experiment (MF State Initial > MF State Final) we can see that even though the final state represented less MF, it does not span the whole final section of the experiment. Using this new map of the MF states, we retrained the NN on the dataset for Subject 6. Table IV shows a significant improvement in the classification performance of all the NN after FPCA labeling.

D. Cross-subject classification

The way MF develops and presents itself is likely to vary from person to person depending on different factors such as age, specific physiological conditions and predispositions, period of the day, circadian rhythm, etc. Developing an algorithm capable of assessing MF state precisely across

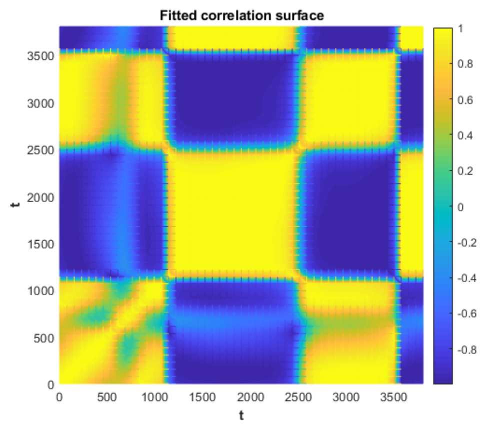

Fig. 3. Correlation surface for different MF states.

TABLE IV

CLASSIFICATION RESULTS FOR FPCA-LABELED DATA FOR SUBJECT 6. (ACCURACY ± STD)

	train	validation	test
FFN	0.84 ± 0.04	0.83 ± 0.04	0.80 ± 0.04
CNN	0.84 ± 0.02	0.84 ± 0.04	**0.84 ± 0.04**
DBN	0.86 ± 0.05	0.84 ± 0.03	0.83 ± 0.03

multiple subjects is a challenging task. Neural networks are a viable option to handle this kind of problem since they offer the possibility of automatic feature engineering. These features can capture very complex characteristics of the data, covering angles that could be impossible to approach by using hand-engineered features.

For the cross-subject classification problem, we applied a nested cross validation approach. We used one test subjects data as a test set and the other five datasets to train the NN using a six-fold cross validation. In each fold, data from one subject was used as validation and the remaining five datasets were used as the training set. The training set was shuffled to ensure that the data for the training process was subject-independent. This process was repeated until each subject was evaluated as a test set. We performed analysis using the three NNs presented previously, with a six second timespan and two second overlap for the input vectors. The

TABLE V

HYPERPARAMETERS SELECTION (CROSS SUBJECT CASE)

	Best configuration
FFN	{1024, 512, 256, 128}
CNN	{32 (3), 64 (5), 128 (8)}
DBN	{2400, 1024, 512, 256}

TABLE VI

CROSS-SUBJECT ANALYSIS. (ACCURACY ± STD)

	train	validation	test
FFN	0.93 ± 0.06	0.70 ± 0.13	0.66 ± 0.05
CNN	0.98 ± 0.17	0.82 ± 0.05	**0.75 ± 0.01**
DBN	0.89 ± 0.04	0.70 ± 0.02	0.69 ± 0.02

hyperparameters selection was made using grid search using the same variable and intervals for the single subject case. For the cross subject case the best configuration for each NN are deeper models which are more suitable to handle a more complex set of features from multiple subjects (Table V).

From Table VI, we can see that none of the NN had a trustworthy performance on the cross-subject case. Even when using longer input vector, overlap between consecutive inputs and the help of FPCA to label the data, the complexity added by multiple physiological response patterns was capable of confusing the NN algorithms. This shows how relevant the cross-subject problem is when considering the implementation of such a MF detection system in real life.

IV. CONCLUSION AND FUTURE WORK

In this work we presented a study about the use of a traditional deep learning algorithm to detect the MF in vessel pilots. We observed the importance of maintaining the temporal correlation on the sensor data when inputing it to the NN algorithm. This can be done by extending the observed interval when making a classification and by using time overlap between consecutive MF state observation.

CNN yielded the best classification accuracy while presenting the best computational performance. FFN, a simpler algorithm, performed well when compared to CNN. DBN didn't performed as well as expected, probably due to loss of information during the RBM compression phase.

We also experimented with FPCA for a more precise labeling of MF states based on knowledge about the initial and final states of the operator. FPCA can provide insights in the dataset that are unavailable otherwise. This resource proved to be an important tool to improve classification performance when the original labeling is imprecise.

Future work might improve the performance of the classification algorithm by optimizing its hyperparameters. The classification of cross-subject cases needs special attention, and a reliable way to correlate MF states to different individuals reference frame needs to be developed. The next natural step is implementation of such an MF detection system in real-time, so it can see applications in real-life scenarios.

REFERENCES

[1] A. Williamson, D. A. Lombardi, S. Folkard, J. Stutts, T. K. Courtney, and J. L. Connor, "The link between fatigue and safety," *Accident Analysis & Prevention*, vol. 43, no. 2, pp. 498–515, 2011.

[2] P. Suraweera, G. I. Webb, I. Evans, and M. Wallace, "Learning crew scheduling constraints from historical schedules," *Transportation research part C: emerging technologies*, vol. 26, pp. 214–232, 2013.

[3] L. Giraudet, J.-P. Imbert, M. Bérenger, S. Tremblay, and M. Causse, "The neuroergonomic evaluation of human machine interface design in air traffic control using behavioral and eeg/erp measures," *Behavioural brain research*, vol. 294, pp. 246–253, 2015.

[4] L. Reinerman-Jones, G. Matthews, and J. E. Mercado, "Detection tasks in nuclear power plant operation: Vigilance decrement and physiological workload monitoring," *Safety science*, vol. 88, pp. 97–107, 2016.

[5] C. Chauvin, S. Lardjane, G. Morel, J.-P. Clostermann, and B. Langard, "Human and organisational factors in maritime accidents: Analysis of collisions at sea using the hfacs," *Accident Analysis & Prevention*, vol. 59, pp. 26–37, 2013.

[6] C. Chauvin, "Human factors and maritime safety," *The Journal of Navigation*, vol. 64, no. 4, pp. 625–632, 2011.

[7] C. Hetherington, R. Flin, and K. Mearns, "Safety in shipping: The human element," *Journal of safety research*, vol. 37, no. 4, pp. 401–411, 2006.

[8] J. U. Schröder-Hinrichs, "Human and organizational factors in the maritime worldare we keeping up to speed?" 2010.

[9] G. R. J. Hockey, *Operator functional state: the assessment and prediction of human performance degradation in complex tasks*. IOS Press, 2003, vol. 355.

[10] I. M. Association *et al.*, "Guidance on fatigue mitigation and management," 2001.

[11] M. R. Grech, A. Neal, G. Yeo, M. Humphreys, and S. Smith, "An examination of the relationship between workload and fatigue within and across consecutive days of work: Is the relationship static or dynamic?" *Journal of Occupational Health Psychology*, vol. 14, no. 3, p. 231, 2009.

[12] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *International Journal of Neuroscience*, vol. 52, no. 1-2, pp. 29–37, 1990.

[13] T. Chalder, G. Berelowitz, T. Pawlikowska, L. Watts, S. Wessely, D. Wright, and E. Wallace, "Development of a fatigue scale," *Journal of psychosomatic research*, vol. 37, no. 2, pp. 147–153, 1993.

[14] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters*, vol. 42, pp. 11–24, 2014.

[15] "Numerical offshore tank," http://tpn.usp.br/, accessed: 2018-12-15.

[16] "Mysignals - ehealth and medical iot development platform," http://www.my-signals.com/, accessed: 2018-12-15.

[17] "Emotiv epoc+," https://www.emotiv.com/epoc/, accessed: 2018-12-15.

[18] D. P. Subha, P. K. Joseph, R. Acharya, and C. M. Lim, "Eeg signal analysis: A survey," *Journal of medical systems*, vol. 34, no. 2, pp. 195–212, 2010.

[19] J. O. Ramsay and B. W. Silverman, *Applied functional data analysis: methods and case studies*. Springer, 2007.

[20] "Principalanalysis by conditional expectation package," http://www.stat.ucdavis.edu/PACE/, accessed: 2018-12-15.