A Decentralized Sensor Fusion Approach to Human Fatigue Monitoring in Maritime Operations

Mateus Sant' Ana, Guoyuan Li* and Houxiang Zhang

***Abstract*— Human fatigue is one of the main causes of accidents in maritime domain. How to use physiological data to estimate degree of human fatigue without medical domain knowledge is significant to the safety of tasks in maritime operations. In this paper, a decentralized sensor fusion approach is proposed. Various sensor data used to monitor brain wave, heart rate, muscle tension, body temperature, visual focus and head movement, together with subjective measurement of Karolinska Sleepiness Scale (KSS) values are selected as the data source for this study. Convolutional neural networks are adopted in the approach to extract local features of each individual data channel. The local features are further fused into a 5-layer fuzzy neural network for classification of the KSS values. A case study of fatigue monitoring test of ship maneuvering in simulator has been carried out. Through a comparative study with a centralized fusion approach, the proposed method is verified to be able to provide high accuracy up to 96.08% for fatigue level classification, and in particular, robust enough to maintain the accuracy to 88.42% in case of sensor failure.**

***Index Terms*— Human fatigue, Sensor fusion, Maritime operation.**

I. Introduction

There is an increasing demand to address human factors in maritime applications [1]. Fatigue as one of the most surveyed human factors, defined by the International Maritime Organization (IMO) as the reduction in both physical and mental capacity, has been receiving a lot of attention recently [2]. The working status for the crews is a significant issue due the fact that the crews often have an intensive workload and the tasks to be performed have rigorous instructions. Fatigue often leads to or might create serious accidents. In [3], fatigue is considered a causal factor in 82% of the groundings in the United Kingdom waters. A similar study from [4] detects that 1/3 of the vessels accidents with personal injuries and 16% of fatal accidents were caused by fatigue. In order to prevent potential human factor accidents, it is necessary to establish efficient methods to monitor and evaluate fatigue during maritime operation.

In the literature, there have been attempts to measure human fatigue either qualitatively or quantitatively. Qualitative measures like the Karolinska Sleepiness Scale (KSS) [5], [6] or the Crew Status Survey (CSS) have been used in maritime domain and have obtained satisfactory results [7]. In this type of method, the surveyed subject is questioned, and the fatigue is measured in a subjective scale. Though this type of measurement is reliable to detect fatigue, they lack the ability to monitor fatigue, and thus to prevent it without compromising the work schedule.

The quantitative methods have not been much explored in the maritime domain, but they provide a more objective approach by measuring of physiological and behavioral mechanism that is related to fatigue. Several efforts have been made to understand this mechanism and based on them, many security systems to monitor the human fatigue have been designed [8]–[11]. Common solutions are sensor fusion methods based on the measuring of electrical brain waves (EEG) [8], [9], eyes movements (EOG) [11], electrocardiogram (ECG) [12] and accelerometers [13]. Other methods like psychomotor vigilance test and muscle tension (EMG) have also been studied, though they often interfere with the task. All the studies have emphasized that human fatigue comes from complicated physiological mechanism and many factors can influence it.

According to the types of measurement, most of the presented monitoring methods are based on limited sources of sensor data. In fact, acquiring distinct types of measurements and combining the data together and applying data fusion technology can achieve better performance compared to using a single sensor alone [14]. Our project aims to measure human fatigue quantitatively in a real-time manner, to support the on-board decision-making system for maritime operations. In this paper, we propose a decentralized sensor fusion approach to combine EEG, EOG, EMG, body temperature and head movement data together, and map the result to the KSS levels. The highlight of the paper is that the decentralized approach not only provides a quantitative measurement of human fatigue, but also shows how important of the sensor data is in terms of accuracy for monitoring human fatigue.

The rest of the paper is organized as follows. Section II describes the decentralized sensor fusion framework using both convolutional neural network (CNN) and fuzzy neural network. In Section III, the sensor setup, together with a case study of navigation operation, is presented in detail. Section IV introduces the performance comparison with a centralized sensor fusion method, as well as sensory deprivation analysis on individual sensor input. Conclusion and future work are given in Section V.

The research is supported by the projects "Digital Twins For Vessel Life Cycle Service" (Project No: 280703), and "SFI MOVE" (Project No: 237929), funded by Norway Research Council, Norway.

Mateus Sant' Ana, Guoyuan Li and Houxiang Zhang are with Department of Ocean Operations and Civil Engineering, Faculty of Engineering, Norwegian University of Science and Technology, Postboks 1517, N-6025, Aalesund, Norway. mateussa@stud.ntnu.no, guoyuan.li@ntnu.no, hozh@ntnu.no

*Corresponding author.

II. Decentralized Sensor Fusion Structure for Human Fatigue Monitoring

As mentioned in Section I, our goal of human fatigue monitoring is to map raw sensor data to KSS levels scaled from level 1 (very alert) to level 9 (great effort to keep awake) [6]. In machine learning, the mapping is commonly considered to be a classification problem. The inputs are mapped to different classes(“KSS levels” in our case) with the help of adaptive algorithm like artificial neural network (ANN), by optimizing its weights to minimize the output error between expected and actual classes [15]. The problem of such methods is that the increase of input dimension causes an exponentially increase of number of parameters to be tuned. The process of designing this pattern recognition method is the careful feature selection that can represent the meaningful data with minimal amount of inputs. Thus, good solutions often require a domain expertise over the variables to be recognized.

The following will introduce a CNN-based centralized sensor fusion structure, together with the proposed counterpart based on a decentralized structure.

A. Centralized Sensor Fusing using CNN-based Classification

Conventional classification methods require the features to be carefully crafted and extracted from the data [16]. However, for human fatigue detection, it is difficult to interpret the features of sensor data, even for people who have the domain knowledge [17].

Deep learning methods, such as the CNN classifier, can help to classify the desired output without domain knowledge expertise. The CNN classifier exploits these property by using convolutional layers. The layer slides a certain amount of small filters into the inputs represented in a matrix form. Each filter is made of trainable weights and mathematically corresponding to apply a discrete convolution to the data with the filter, as shown in Eq. (1), where x represents the input; and g denotes the filter. This convolution is able to extract features and the size of filters controls how the filter convolves around the feature map. Apart from convolutional layers, the pooling layers in CNN are used to merge the semantically similar features into one, as well as to make the features less sensitive to the precise locations in the structure of its original input.

$$x * g = \sum_{j=-N}^{N} \sum_{i=-M}^{M} x[i,j] g[m-i, n-j] \qquad (1)$$

Fig. 1a illustrates a CNN-based sensor fusion structure. It is a type of low-level fusion, which uses several sources of raw data such as the data from EEG, EMG, ECG and eye tracker, to estimate fatigue level in forms of KSS values. In principle, it will be more informative and synthetic than the original inputs. However, from robust viewpoint, the structure will result in inferior result in case of sensor failure.

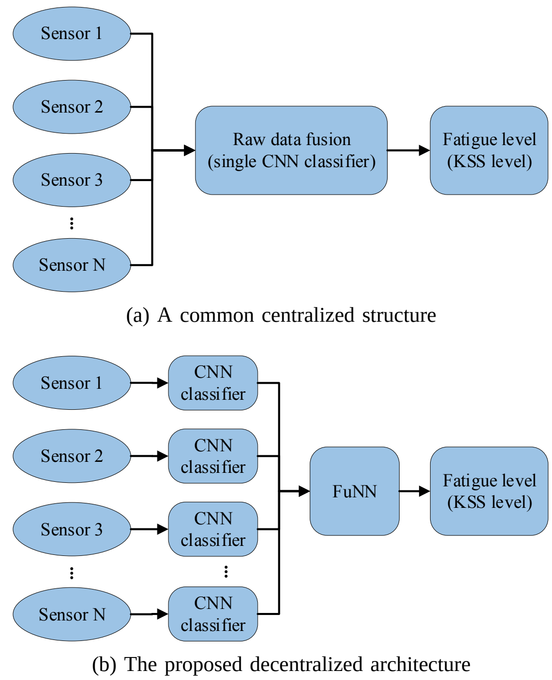


Fig. 1. Sensor fusion structures for human fatigue monitoring.

B. Decentralized Sensor Fusing with Neuro-fuzzy System

In Fig. 1b, we propose a decision level fusion structure that combines several sub-decisions from different sources of sensor data, to yield a final decision. The decentralized architecture is a competitive fusion that achieves high reliability and accuracy [18]. This structure also separates the problem of fatigue classification from the fusion. Each classifier is directly trained for fatigue level classification. The separated classifiers with fatigue estimation grant the final decision to the fusion unit. The fusion of homogeneous type of data is natural for human and requires less domain expertise, which can be explored by the fusion model. Hybrid expert system can fuse the estimation of the classifier by “rules of thumbs” that are human understandable, and permit the user to learn and analyze the individual estimation, their accuracy, errors and relation among them.

The proposed design utilizes a fuzzy neural network (FuNN) system [19] together with CNNs for the final estimation. The FuNN is built in three main components: the basic “if-then” logic rules; a function of fuzzy set membership; and the reasoning which maps from the rules to the respective outputs [20]. The fuzzy system uses a multilayer perceptron architecture, allowing both the rules and the membership functions to be optimized by the backpropagation and training of the network. Therefore, fuzzy inference system will be used to further fuse the data together.

Though the individual estimator have limited precision, the output accuracy can be increased by the decision fusion of the FuNN. Compared to the centralized structure which handles all the parameters all together, the separated optimization of the CNNs and the Neuro-Fuzzy system might find a local minimum solution. Nevertheless, the architecture,

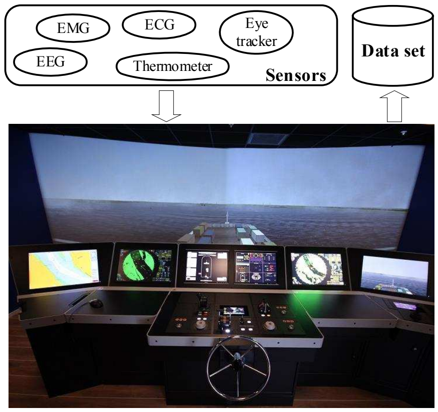


Fig. 2. Sensor setup for data acquisition in simulator.

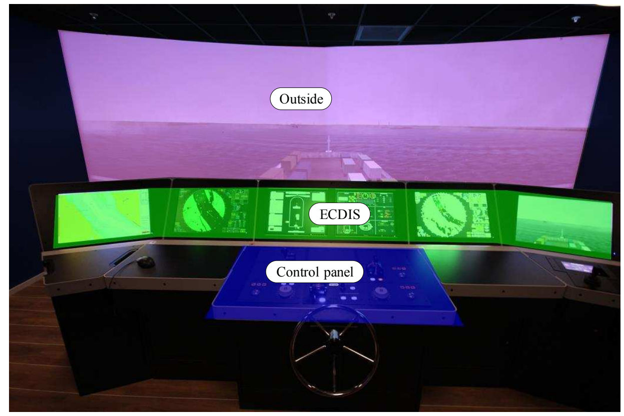


Fig. 3. AOI definition for the fatigue monitoring experiment during ship maneuvering.

once trained, will be able to give insight into the importance of each individual sensor input, and even provide fatigue estimation when not all sensors are available.

III. Experiment for Fatigue Monitoring of Ship Maneuvering

An experiment of fatigue monitoring for ship maneuvering in simulator was carried out. The following introduces how the sensors were setup, what data were collected and the fusion structure setup.

A. Sensor Setup

Fig. 2 depicts all the sensors applied to the experiment. In order to monitor the changes of physiological effects from human body, especially potential signal related to fatigue like heart rate, muscle activity and temperature, a box containing the corresponding sensors for EMG, ECG and body temperature has been developed. Measuring the brain waves is another potential important indicator for getting knowledge about tiredness of human body. An EEG sensor was utilized to take measurements of brain waves in 14 channels [21]. In addition, visual focus is considered a valid clue for fatigue detection. Here we utilized the professional wearable eye tracking device — Tobii pro glasses 2 with gyro and accelerometer integrated for head movement analysis [22] and our developed eye-tracking software to collect the data [23].

B. Experiment Description and Data Collection

A navigation operation test in simulator was selected as the case study for verification. Four participants took part in the experiment. The operator was asked to wear on the sensors and navigate a vessel in a narrow canal area at a speed of 22 knots, as shown in Fig. 2. There were marine traffic nearby during the maneuvering. Despite this, the task is quite simple. The test was lasted more than one hour to increase the possibility of fatigue occurrence. KSS questionnaire was used before and after the test respectively to label the fatigue level for the supervised learning method. Three different fatigue levels were recorded according to the participants' answers.

TABLE I

Simplified Dataset for the Experiment

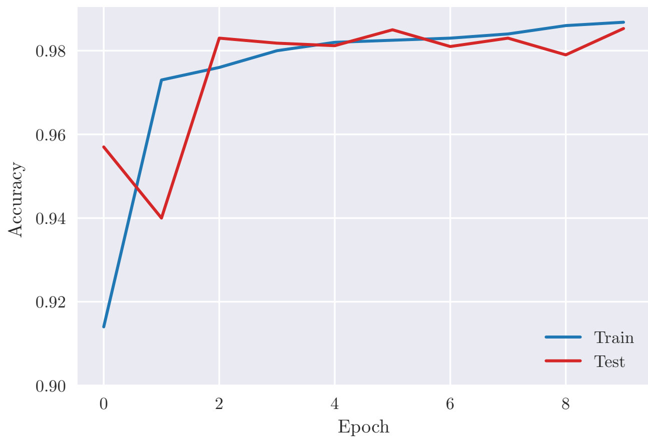
Data source	Number of channel	Measurement
AOI hit	3	Hit number for AOIs of outside, ECDIS, and control panel
Body sensors	3	ECG, EMG and thermometer
Accelerometer	3	Acceleration in axis X,Y and Z
Gyro	3	Angular velocity in axis X,Y, and Z
Gaze	4	Left Z, right Y, right Z, and duration
Pupil	6	Diameter left and right, position left and right in Y and Z directions
Eye movement	2	Mapped eye movement in axis X and Y

There were over 60 channels of data collected from the experiment. The EEG data contained too much noise and was not used for analysis. In addition, to decrease computational complexity, a countermeasure was adopted to remove the highly correlated channels with correlation over 90%. The resulted dataset was reduced to 24 channels, as listed in Table I. In total, there are 21 channels in the dataset from the eye tracker. Note that the hit in area of interests (AOIs), as defined in Fig. 3, was also considered an indicator for the evaluation.

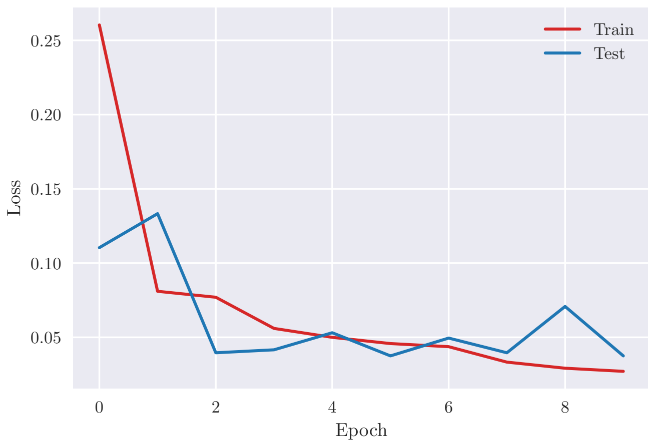
C. Sensor Fusion Setup

For the centralized architecture, a cooperative and complementary fusion was realized via a CNN classifier. It receives a segment of data every second, i.e., approximately a 300×24 array, and produces an estimation of KSS value. The classifier consists of two convolutional layers with 5×5 and 3×3 kernels respectively. Max-pooling was selected so to down-sample by a factor of 2.

For the decentralized architecture, the separated classifications were designed based on the group of measures listed in Table I. The classifiers have the same CNN structure as the one in the centralized architecture, but with different size of kernel. The classifiers for data source containing 3 channels



(a) Model accuracy



(b) Model loss

Fig. 4. Training and validation test of CNN for raw data fusion.

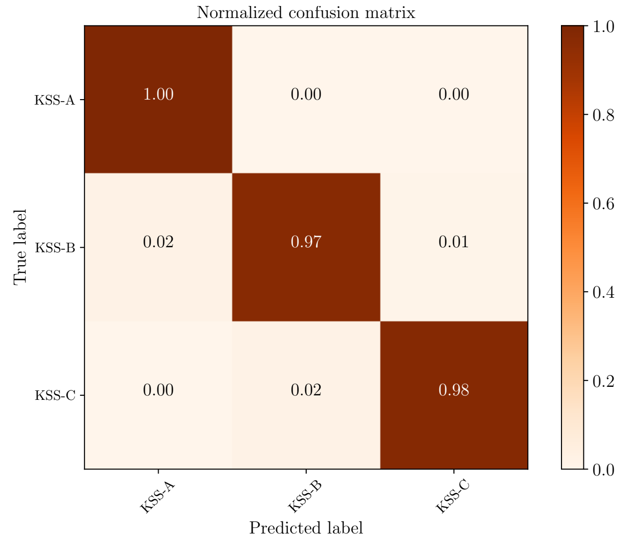


Fig. 5. Confusion matrix of raw data fusion.

TABLE II
10-FOLD CROSS-VALIDATION TEST

Data source (classifier)	Accuracy mean $\pm$ standard deviation (%)
AOI hit	55.32 ± 0.56
Body sensors	77.40 ± 1.79
Accelerometer	67.85 ± 1.25
Gyro	63.89 ± 0.83
Gaze	63.21 ± 1.69
Pupil	80.98 ± 1.18
Eye movement	78.91 ± 4.41

were designed to have a 2×2 kernel for convolution and a 2×1 kernel for the pooling layer. The classifiers for the data sources from "gaze" and "eye movement" have a 2×1 pool in both max-pooling layers, whereas the one from "pupil" has a 2×2 pool size. After the individual classifiers were tuned, a 5-layer FuNN with two Gaussian functions as the membership function was constructed for decision fusion.

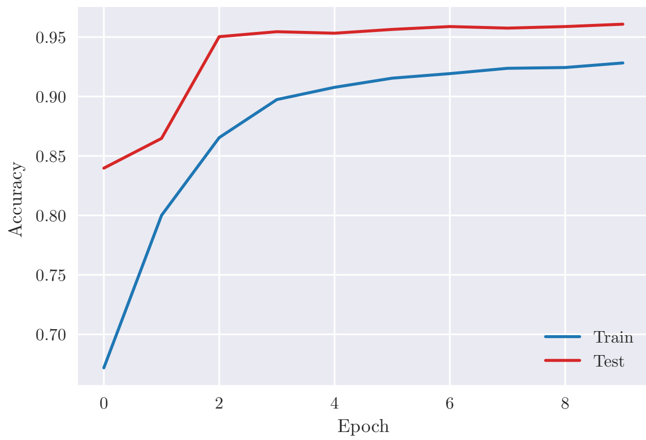
IV. Results

A. Performance Comparison

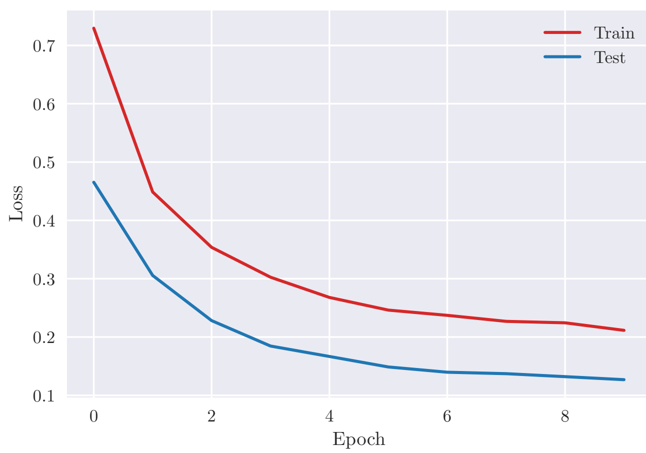
The training of CNN for raw data fusion was performed with 10 epochs and the progression of the accuracy and loss of the model is shown in Fig. 4. The maximum accuracy can reach 98.68% with a corresponding loss of 0.04 for the validation data. The classifier has a better grasp of the overall information, thus making better adjustment of parameters of the network to optimize the loss function. Fig. 5 evaluates the network using confusion matrix. Note the three recorded KSS values are labeled as KSS-A, KSS-B and KSS-C, respectively. The result shows the architecture is able to classify the labels in high precision, which is consistent with expected result of the centralized structure due to the completeness of sensor information.

The training of CNNs for decision fusion was similar to the training of CNN for raw data fusion. Moreover, the classifiers were further evaluated by k-fold cross-validation test, i.e., dividing data into k equal parts, and utilizing $k-1$ parts for training and the last part for validation. Table II lists the result for $k=10$. It is noted that the classifier for "AOI hit" only has 55% mean accuracy, which means "AOI hit" cannot provides valid information for fatigue monitoring. In contrast, the classifiers for "pupil", "body sensor" and "eye movement" have much higher mean precision, playing key roles in fatigue level estimation. Nevertheless, the classifier for "eye movement" has a relative high standard deviation, which indicates the classifier is sensitive to the quality of the data provided for training.

Fig. 6 illustrates the training result after fusing the classifiers into the FuNN. As expected, the FuNN can obtain a higher accuracy compared to individual classifiers up to 96.08% with a loss of 0.13. The confusion matrix for the decision fusion is shown in Fig. 7. There are false probability of 0.03, 0.04 and 0.06 for "KSS-A", "KSS-B" and "KSS-C", respectively. The result shows decision fusion has an little inferior performance compared to that of raw data fusion shown in Fig. 5. However, the decentralized architecture will provide more robust estimation. We will demonstrate it in



(a) Model accuracy



(b) Model loss

Fig. 6. Training and validation test for decision fusion.

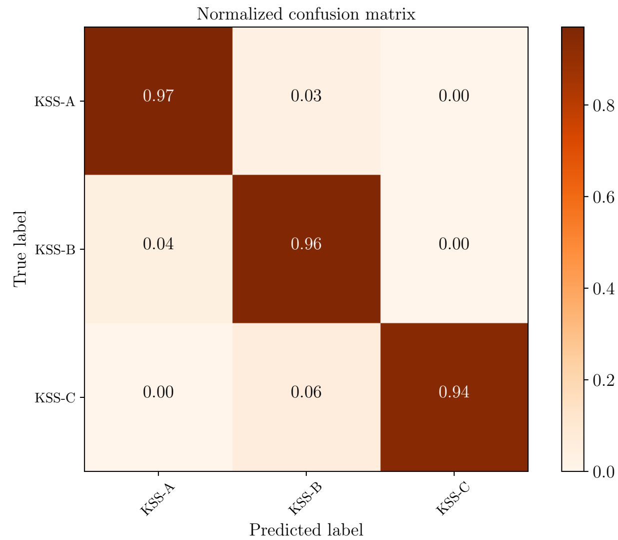


Fig. 7. Confusion matrix of decision fusion.

TABLE III
ACCURACY LOSS BY FAILURE OF DATA SOURCE

Data source failure	Accuracy loss (%)	
	Centralized architecture	Decentralized architecture
AOI hit	1.40	1.87
Body sensors	0.44	5.72
Accelerometer	4.45	0.53
Gyro	8.60	0.66
Gaze	30.37	0.39
Pupil	2.73	0.52
Eye movement	34.65	7.66

Section IV-B.

B. *Sensory Deprivation Analysis*

Beside the overall accuracy, it is also important to study the robustness of the fusion results to sensory deprivation. Here we focus on investigate how the loss of certain of measures, e.g., loss of acceleration data, affects system accuracy.

Table III shows the accuracy loss caused by failure of each individual classifier used in the two fusion architectures. The centralized architecture depends on most of the measures from the eye tracker device and its embedded sensors including accelerometer and gyro. In particular, the lack of data source from either “gaze” or “eye movement” drops the accuracy to less than 70%. The accuracy reduction shows that the centralized architecture is sensitive to lack of data source. This is because the network itself needs all inputs to modify its weight during training, to achieve high accuracy at an expense of adaptability.

In contrast, the decision fusion has much better performance when sensor failure happens. Even though the individual classifiers for “body sensors”, “pupil” and “eye movement” have high accuracy for fatigue level estimation (as shown in Table II), the loss of one of them only results in accuracy loss to 5.72% , 0.52% and 7.66%, respectively. In other words, the network, even in the worst case of data source failure, can still reach a precision of 88.42%. This is because the use of fuzzy logic in the network allows for better addressing uncertainty. Here the uncertainty of lack of data can be represented by the two membership function for each label in the FuNN, corresponding to the probability of likely and unlikely of data source failure for each of the classifiers.

C. *Discussion*

The case study has shown that both architectures are valid options to classify the fatigue level, with high accuracy over 96%. The raw data fusion as a complementary type of fusion allows the network to have better performance (about 2%) in the classification problem than the performance of decision fusion architecture. However, the decision fusion architecture using a competitive type of fusion offers superior robustness when sensor failure happens.

For most tasks in maritime operations like navigation, deck operation and crane lifting, the sensors we used for fatigue monitoring may be intrusive and interfere with the task. Furthermore, due to the movement of the stuff crew during the task, the acquisition of sensor data may be paused or even get interrupted completely. From the case study results, the decision fusion architecture is preferred in maritime operations, taking advantage of the adaptability of the structure.

V. Conclusion

This paper presents a decentralized approach to realize human fatigue monitoring system in maritime operations. The aim is to establish a mapping relationship between the objective data from sensors including ECG, EMG, EEG, thermometer and eye tracker, and the subjective measurement of KSS values. Highly correlated data channels are removed and the rest are grouped via CNNs for feature extraction and preliminary fatigue classification. A 5-layer FuNN is used to fuse individual CNN output to further improve the classification accuracy. To investigate the performance of the approach, it is compared with a centralized fusion approach in an experiment of fatigue monitoring during ship maneuvering in simulator. The result shows both method can achieve a classification accuracy over 96% and the proposed method can even maintain an accuracy of 88.42% in case of sensor failure. This implies the decentralized approach is more applicable to fatigue monitoring in maritime operations.

For future work, we will focus on (1) a thorough analysis of individual measurement on how important it is in the classification problem, and (2) a new fatigue monitoring experiment of ship maneuvering with more KSS values recorded for classification.

Acknowledgment

The authors would like to thank Mr. Arnt Håkon Barmen who provided expertise that greatly assisted the research, and all participants who involved in the experiment.

References

[1] J.-U. Schröder-Hinrichs, E. Hollnagel, M. Baldauf, S. Hofmann, and A. Kataria, "Maritime human factors and imo policy," *Maritime Policy & Management*, vol. 40, no. 3, pp. 243–260, 2013.

[2] IMO, "Guidance on fatigue mitigation and management," International Maritime Organization, Tech. Rep. MSC/Circ.1014, 2001. [Online]. Available: `https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/288812/imo_fategue_part_1.pdf`

[3] C. House and C. Place, "Bridge watchkeeping safety study," *Department for Transportation, Marine Accident Investigation Branch, Southampton*, 2004.

[4] A. P. Smith, "Adequate crewing and seafarers' fatigue: The international perspective," 2007.

[5] 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.

[6] A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro, "Karolinska sleepiness scale (kss)," in *STOP, THAT and One Hundred Other Sleep Scales*. Springer, 2011, pp. 209–210.

[7] M. Lützhöft, M. R. Grech, and T. Porathe, "Information environment, fatigue, and culture in the maritime domain," *Reviews of human factors and ergonomics*, vol. 7, no. 1, pp. 280–322, 2011.

[8] C.-T. Lin, C.-J. Chang, B.-S. Lin, S.-H. Hung, C.-F. Chao, and I.-J. Wang, "A real-time wireless brain–computer interface system for drowsiness detection," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 4, no. 4, pp. 214–222, 2010.

[9] M. Akin, M. B. Kurt, N. Sezgin, and M. Bayram, "Estimating vigilance level by using eeg and emg signals," *Neural Computing and Applications*, vol. 17, no. 3, pp. 227–236, 2008.

[10] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Information Fusion*, vol. 35, pp. 68–80, 2017.

[11] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE transactions on vehicular technology*, vol. 53, no. 4, pp. 1052–1068, 2004.

[12] M. Patel, S. K. Lal, D. Kavanagh, and P. Rossiter, "Applying neural network analysis on heart rate variability data to assess driver fatigue," *Expert systems with Applications*, vol. 38, no. 6, pp. 7235–7242, 2011.

[13] M. B. Kurt, N. Sezgin, M. Akin, G. Kirbas, and M. Bayram, "The ann-based computing of drowsy level," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2534–2542, 2009.

[14] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, 1997.

[15] S. Marsland, *Machine learning: an algorithmic perspective*. CRC press, 2015.

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[17] H. J. Michielsen, J. De Vries, G. L. Van Heck, F. J. Van de Vijver, and K. Sijtsma, "Examination of the dimensionality of fatigue: The construction of the fatigue assessment scale (fas)." *European Journal of Psychological Assessment*, vol. 20, no. 1, p. 39, 2004.

[18] W. Elmenreich, "An introduction to sensor fusion," *Vienna University of Technology, Austria*, 2002.

[19] N. K. Kasabov, J. Kim, M. J. Watts, and A. R. Gray, "Funn/2 – a fuzzy neural network architecture for adaptive learning and knowledge acquisition," *Information sciences*, vol. 101, no. 3-4, pp. 155–175, 1997.

[20] J.-S. R. Jang *et al.*, "Fuzzy modeling using generalized neural networks and kalman filter algorithm." in *AAAI*, vol. 91, 1991, pp. 762–767.

[21] B. T. Nugraha, R. Sarno, D. A. Asfani, T. Igasaki, and M. N. Munawar, "Classification of driver fatigue state based on eeg using emotiv epoc+." *Journal of Theoretical & Applied Information Technology*, vol. 86, no. 3, 2016.

[22] "Tobii pro glasses 2," (Date last accessed 01-03-2019). [Online]. Available: `https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/`

[23] G. Li, E. Homlong, H. Zhang, and B. A. Dimmen, "Integration of visual focus into marine operation simulator for behavior observation and analysis," in *IEEE OCEANS 2017-Aberdeen*, 2017, pp. 1–7.