Leveraging Transdisciplinary Engineering in a Changing and Connected World
P. Koomsap et al. (Eds.)

doi:10.3233/ATDE230679

EEG-Based Stress Recognition Through the Integration of Convolutional Neural Networks and Mixture of Experts Ensemble Modelling

Ziqing XIA [a], Justin JiXi WEN [a], Chun-Hsien CHEN [a,1], Meng-Hsueh HSIEH [a] and Jo-Yu KUO [b]
[a] *School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 639798, Singapore*
[b] *Department of Industrial Design, National Taipei University of Technology, Taipei City, 10608 Taiwan*

Abstract. Vessel Traffic Service Operators (VTSOs) are responsible for ensuring the safe and efficient operation of waterways. They use a Vessel Traffic Management System (VTMS) to provide real-time information and ensure the smooth flow of vessel traffic. However, the demanding nature of their work can lead to stress, which can impact their performance and lead to physical and mental health problems. To alleviate stress on VTSOs and reduce the risk of maritime accidents, investment in stress management technologies is important. In this study, we propose a novel machine learning model called 3D Mixture of Experts Convolutional Neural Network (3DMoEConvNet) to predict stress level for VTSOs based on their EEG signals. The 3DMoEConvNet model combines the strengths of 3D Convolutional Neural Network (3D CNN) and Mixture of Experts (MoE) architecture to effectively capture both the spatial and temporal features of EEG data, while also addressing the issue of individual differences in EEG data. The 3DMoEConvNet achieved accuracies of 99.80%, 99.80% and 99.84% for 2-Class, 3-Class and 4-Class predictions, respectively. The proposed model provides a basis for the advancement of EEG-based stress detection systems.

Keywords. Stress, electroencephalogram (EEG), neural network, individual difference, mixture of experts

Introduction

Vessel Traffic Service Operators (VTSOs) play a crucial role in ensuring the safe and efficient operation of waterways. They are responsible for providing essential services that ensure the smooth flow of vessel traffic. To perform their duties, VTSOs use a Vessel Traffic Management System (VTMS) that provides real-time information on vessel movements, weather conditions, and other important information[1]. However, the demanding nature of their work and the associated working environments can take a toll on VTSOs, leading to stress. Stress is a major concern for VTSOs because it can have a significant impact on their performance. When VTSOs are under stress, they are more

[1] Corresponding Author, Mail: mchchen@ntu.edu.sg

likely to make mistakes, and their decision-making ability may be compromised. Additionally, stress can lead to physical and mental health problems, such as fatigue, headaches, and depression, which can further impact their ability to perform their duties effectively.

To prevent operators from experiencing excessive stress, recognising stress in its early stages is key. Based on literature review, there is currently no stress detection research specifically targeting vessel traffic service (VTS) or air traffic control (ATC). Existing stress management research in VTS/ATC settings tends to approach from a human factors analysis perspective. For instance, Borghini et al. [2] used multimodal signals to investigate the impact on air traffic controllers during stressful events. Kutilek et al. [3] studied heart rate variability of air traffic controllers in long-term stressful contexts. However, neither of these studies utilized these physiological signals for stress detection. Interventions are often implemented at the organizational level, such as the development and updating of training programs. This indicates a gap in stress detection research for VTS scenarios, and the experimental contexts in existing stress detection research differ substantially from VTS operations.

Implementation of stress detection in high-stake industries such as VTS requires the consideration of several factors. Primarily, the ability to perform real-time stress detection with high accuracy is paramount. Given that emergency situations can arise unexpectedly in the VTS context, timely and accurate detection of stress levels among VTSOs is crucial to preventing potential catastrophes caused by stress. Electroencephalography (EEG), due to its high frequency and rich information content, is recognized to be frequently used in affective computing for mental state detection, including stress recognition. However, a lack of publicly available EEG datasets specifically targeting stress recognition has been identified. Consequently, the decision to design our own experiment has been made. To fill this gap, the Stroop Test reflecting the continuous operation context of VTSOs was chosen in our experiment design, data collection, and model validation.

In recent years, using machine learning techniques such as Artificial Neural Network (ANN) for predicting stress has become increasingly popular. There has been a wide use of Electroencephalogram (EEG) for stress prediction as they are hard to conceal, which is advantageous for eliciting genuine emotions. This is helpful to measure the true state of mind of humans since EEG data directly represents the brain's activity [4]. Common methods such as Convolutional Neural Networks (CNN), Graph Neural Networks (GNN), Long-short term memory networks (LSTM) have been used to perform EEG stress prediction. The aim of this study is to assess the effectiveness of a CNN-based stress prediction model using EEG data. This study will go beyond prior research by addressing the limitations of current stress prediction models.

1. Related Work

Recent advancements in EEG technology have prompted interest in deep learning techniques for stress or emotion prediction from EEG data [5]. This is driven by the improved quality and volume of EEG data and the success of deep learning in areas like computer vision and speech recognition. A primary obstacle in EEG analysis is the variability in EEG signals between individuals due to factors like personality traits, genetic variations, and brain chemical levels. This leads to substantial inter-subject variability in EEG signals [6].

Fu et al. [7] devised a Symmetric Deep Convolutional Adversarial Network (SDCAN) blending CNN and adversarial theory for stress level classification in EEG signals. The adversarial aspect extracts invariant features from raw EEG signals, aiming to improve classification accuracy and address individual differences. SDCAN achieved 87.62% and 81.45% accuracies in four-class and five-class stress predictions, respectively.

Song et al. [8] proposed two models. The Dynamical Graph Convolutional Network (DCGNN), capable of dynamically learning EEG channel relationships, achieved an accuracy of 79.95% for three-class emotion classification using the SEED dataset. The Variational Instance-Adaptive Graph (V-IAG) technique recognises interconnections between different EEG electrodes and measures the underlying uncertain information. It addressed individual differences and uncertain relationships between EEG channels, achieving an 88.38% accuracy for a three-class emotion classification [4].

Zhang et al. [9] introduced an architecture combining deep recurrent and 3D convolutional neural networks (R3DCNNs), achieving an average accuracy rate of 88.9%. This method responds to the limitation of most mental workload assessment research focusing on a single task.

EEG data representation is a crucial aspect of deep learning analysis. The common method of converting EEG data into a 2D array neglects important spatial correlations between electrodes. Zhao et al. [10] proposed a 3D EEG data representation preserving both temporal and spatial information, potentially providing a more accurate analysis. Another study proposed Multiscale CNNs and a biologically inspired decision fusion model for multimodal affective states recognition [10]. While they didn't predict stress, their 3D EEG representation and High Scale CNN architecture provide valuable insights. Their 9x9 2D matrix representing the 10-20 international system achieved an accuracy of 98.52% for the DEAP dataset and 99.89% for the AMIGOS dataset.

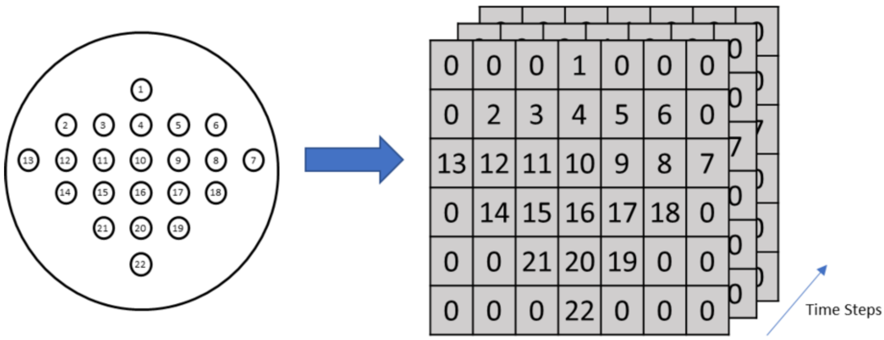

Figure 1. Left: 10-20 System, Right: 3D Representation.

Previous studies have independently addressed the challenges of individual differences and loss of spatial information and correlations between electrodes. However, there is a need for a stress prediction model that effectively tackles both issues simultaneously.

2. Method

2.1. Data Collection Experiment

To collect stress dataset, we designed an experiment based on the Stroop Test [12]. In the experiment, ten participants (9 males and 1 female) aged between 25 and 33 years (M = 25.42, SD = 3.66), who were students or research staff from Nanyang Technological University, were involved. The study was approved by the Institutional Review Board of Nanyang Technological University (NTU-IRB), Singapore. The reference number is IRB-2022-015. All participants had normal or corrected visual acuity and were not suffering from any mental or physical conditions. This study chose the Stroop Test as it is a classic stress-inducing paradigm, proven through numerous experiments to reliably evoke stress. Other paradigms like the Trier Social Stress Test (TSST) [13] and the Montreal Imaging Stress Task (MIST) [14] use psychosocial pressure and images, respectively, to induce stress. TSST and MIST do not resemble the working scenarios of VTSO. The parallels between Stroop Test and VTSO scenarios are the operation task, continuous monitoring of a screen, and maintaining of situation awareness.

The experiment contains 4 blocks, which aims to induce increasing stress levels. In block 1, subjects were required to watch a trial demonstration video on the designed Stroop test, in which the queries were automatically answered. In block 2, the participants were presented with colored-words displayed in a congruent color ink (e.g., the word "green" was printed in green color). In block 3, incongruent colored-words were presented, (e.g., the word "green" was displayed in blue color). In block 4, incongruent colored-words were presented and the participants were required to answer the queries in very limited time. The participants were required to reflect on their stress levels and answer a questionnaire after the practice session and after each block. The EEG signals were recorded using a 14-channel headset (Emotiv Epoc+ Pro) and filtered using a highpass filter of 1 Hz before being used for stress prediction tasks. The data was collected in a moving window of 10 seconds with a stride of 1 second.

2-class. For the binary classification task, we chose data from Block 2 over Block 1 as low-stress data to eliminate task-discriminative features. We selected data from Block 4 as high-stress data as it has a higher likelihood of containing discriminative features related to mental stress. Hence, we utilised the data from Block 2 and Block 4 for the binary stress prediction task. This resulted in 6150 samples with a data size of 14 channels and 1024 time steps per sample ([6150, 14, 1024]).

3-class. For 3-class classification dataset, we selected data from Block 2, 3 and 4. This resulted in 9321 samples with a data size of 14 channels and 1024 time steps per sample ([9321, 14, 1024]).

4-class. For 4-class dataset, we selected data from all 4 Blocks. This resulted in 12331 samples with a data size of 14 channels and 1024 time steps per sample ([12331, 14, 1024]).

2.2. Data pre-processing

For data pre-processing, we first perform Z-score normalisation on all the 3 datasets, this is to provide numerical stability, preventing the gradients of the neural network optimiser from exploding. To prevent the loss of spatial information of the EEG electrodes, we adopt the data pre-processing methodology used in [11]. This will help to retain the 2D electrode's topological structure. Firstly, for each sample, the dataset is reshaped into [samples*1024, 14]. After which, for each time step, the 14 EEG channels' data are mapped into a 9x9 matrix, giving the shape of [samples*1024, 9, 9]. Then, the 9x9 matrices are stacked into cubes of [128, 9, 9], this gives rise to a data shape of [samples*8, 128, 9, 9]. In order for the dataset to be processable by 3D CNN, we have to reshape it into [samples*8, 1, 128, 9, 9]. This is the final shape of the EEG dataset, ready for prediction. Figure 2 gives a visual illustration of the mapping process.

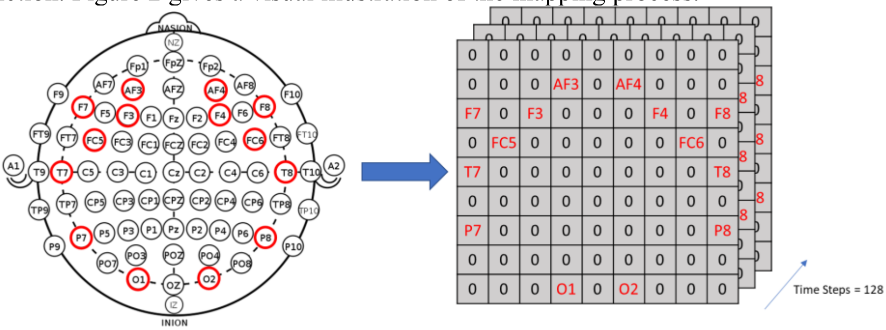

Figure 2. Matrix Mapping [15].

2.3. Proposed Mixture of Experts (MoE) model

In this study, we propose a novel machine learning model for EEG stress level prediction named 3D Mixture of Experts Convolutional Neural Network (3DMoEConvNet). The use of a 3D CNN is appropriate for this task as it is capable of retaining the spatial and temporal information present in the EEG data. The 3D CNN architecture allows the model to capture the patterns and relationships in the EEG signals over time and across different regions of the brain. In addition to the 3D CNN, the proposed model also employs a Mixture of Experts (MoE) architecture. The MoE architecture is designed to tackle the issue of individual differences in EEG data, which can lead to varying levels of stress among individuals. The MoE architecture allows the model to learn different experts for different individuals, which can lead to more accurate stress predictions for each individual. The proposed 3DMoEConvNet model combines the strengths of both the 3D CNN and MoE architecture to effectively capture both the spatial and temporal features of EEG data, while also addressing the issue of individual differences.

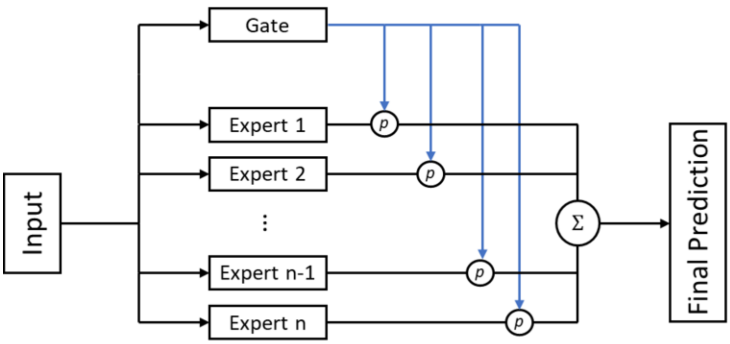

Figure 3. Mixture of Experts (MoE) model.

2.4. 3D CNN Architecture

In this study, the base model, also known as expert, is a 3D CNN. We adopt the same 3D CNN architecture as used by Zhao et al. [11] for their High Scale CNN due to its demonstrated effectiveness. By utilising the same architecture, we aim to leverage the strengths of the previous study and build upon its success. The details of the 3D CNN are shown in Table 1.

2.5. Mixture of Experts (MoE)

The MoE model is a machine learning technique that is introduced to address the issue of individual differences in EEG data stress prediction. Recently, the MoE model has seen remarkable success in the field of deep learning [15]. Typically, the MoE approach decomposes predictive modelling tasks into sub-tasks, trains a separate expert model on each one, creates a gating model that determines which expert to rely on based on the input to be predicted, and combines the resulting predictions. There are various methods for combining the predictions, including selecting the expert with the highest confidence from the gating model or taking a weighted sum prediction. In this study, we opt for the weighted sum prediction approach. This method involves combining the predictions of each expert and the confidence estimated by the gating network into a weighted sum. The model then selects the class with the highest weighted sum as its final prediction output.

Table 1. 3D CNN Architecture.

Layer Type	Output Size	Kernel	Stride	Padding
Input	[120, 1, 128, 9, 9]	-	-	-
Convolutional	[120, 32, 128, 9, 9]	(4, 3, 3)	(1, 1, 1)	same
ReLu	[120, 32, 128, 9, 9]	-	-	-
Pooling	[120, 32, 64, 9, 9]	(2, 1, 1)	(2, 1, 1)	-
Convolutional	[120, 64, 64, 9, 9]	(4, 3, 3)	(1, 1, 1)	same
ReLu	[120, 64, 64, 9, 9]	-	-	-
Pooling	[120, 64, 32, 9, 9]	(2, 1, 1)	(2, 1, 1)	-
Flatten	[120, 165888]	-	-	-
FC	[120, 1024]	-	-	-
Dropout	[120, 1024]	-	-	-
FC	[120, num_class[2]]	-	-	-

3. Result

3.1. Model Training and Hyperparameters

The training of the 3DMoEConvNet uses the Adam optimiser with an initial learning rate of 0.001 and follows the ReduceLROnPlateau schedule. This scheduler will reduce the learning rate when a metric has stopped improving. The model is trained using a batch size of 120. To ensure numerical stability, the epsilon of the optimiser is set to 1e-2. The loss function we used is CrossEntropyLoss. The number of experts selected in this study was 4, in accordance with the methodology described by Chen et al. [15]. The

[2] Number of prediction classes

training is conducted using 10-fold cross validation mechanism. The training process is divided into two parts, with the first part consisting of 30 epochs and the second part consisting of 10 epochs. To validate the effectiveness of the proposed model, it is compared with the base model without the MoE module.

3.2. Prediction Results

The verification results can be seen in the following tables.

Table 2. *2-Class* Performance Comparison.

Model	Epochs	Accuracy		Recall		Precision		F_1 Score	
		Value	S.D.	Value	S.D.	Value	S.D.	Value	S.D.
3D CNN (Base)	30	0.9981	0.0004	0.9981	0.0004	0.9981	0.0004	0.9981	0.0004
3DMoEConvNet	30	0.9977	0.0005	0.9977	0.0005	0.9977	0.0005	0.9977	0.0005
3D CNN (Base)	10	0.9962	0.0024	0.9961	0.0024	0.9962	0.0024	0.9962	0.0024
3DMoEConvNet	10	0.9952	0.0038	0.9951	0.0038	0.9952	0.0037	0.9952	0.0038

Table 3. *3-Class* Performance Comparison.

Model	Epochs	Accuracy		Recall		Precision		F_1 Score	
		Value	S.D.	Value	S.D.	Value	S.D.	Value	S.D.
3D CNN (Base)	30	0.9981	0.0004	0.9981	0.0004	0.9981	0.0004	0.9981	0.0004
3DMoEConvNet	30	0.9977	0.0005	0.9977	0.0005	0.9977	0.0005	0.9977	0.0005
3D CNN (Base)	10	0.9962	0.0024	0.9961	0.0024	0.9962	0.0024	0.9962	0.0024
3DMoEConvNet	10	0.9952	0.0038	0.9951	0.0038	0.9952	0.0037	0.9952	0.0038

Table 4. *4-Class* Performance Comparison.

Model	Epochs	Accuracy		Recall		Precision		F_1 Score	
		Value	S.D.	Value	S.D.	Value	S.D.	Value	S.D.
3D CNN (Base)	30	0.9983	0.0016	0.9983	0.0012	0.9983	0.0012	0.9983	0.0012
3DMoEConvNet	30	0.9983	**0.0003**	0.9983	**0.0003**	0.9983	**0.0003**	0.9983	**0.0003**
3D CNN (Base)	10	0.9964	0.0029	0.9964	0.0029	0.9964	0.0029	0.9964	0.0029
3DMoEConvNet	10	**0.9980**	**0.0005**	**0.9980**	**0.0005**	**0.9980**	**0.0005**	**0.9980**	**0.0005**

Further analysis can be performed based on the results presented in Tables 2, 3, and 4.

4. Discussion

For 2-Class and 3-Class prediction, the results obtained in Tables 2 and 3 shows that the base model and 3DMoEConvNet performed equally well in terms of all four metrics. The performance metrics did not provide any valuable trends or insights. However, for 4-Class prediction, the 3DMoEConvNet demonstrated faster convergence to its optimal accuracy compared to the base model. The standard deviation across all the metrics for both 10 and 30 epochs are also lower for the 3DMoEConvNet as compared to its base model. The results shows that the MoE architecture can be a useful tool for more complex and comprehensive EEG analysis, has a higher capacity to learn more complex features than the base model and more reliable as the standard deviation is lower. This is evidenced by the fact that the 3DMoEConvNet was able to converge to its optimal accuracy using fewer epochs compared to the base model. This suggests that the MoE architecture allows the model to learn faster and reach optimal accuracy more quickly. The faster convergence of the 3DMoEConvNet can be credited to its model complexity and capability to handle complex data structures (4-Class is more complexed and harder to predict as compared to 2-Class and 3-Class) and extract meaningful features from the EEG data. The use of multiple experts in the 3DMoEConvNet is a key feature that sets it apart from the base model. The multiple experts in the 3DMoEConvNet are designed to learn from different clusters of the EEG data, each specialising in a specific cluster of the EEG signal. This allows the 3DMoEConvNet to handle complex data structures and extract meaningful features from the EEG data. The use of multiple experts in the 3DMoEConvNet has several advantages for EEG analysis. Firstly, it allows for better generalisation of the model. By having multiple experts, the 3DMoEConvNet can learn from different aspects of the EEG data, making it more robust to variations in the EEG data. Secondly, the utilisation of multiple experts in the 3DMoEConvNet makes it easier to find the optimal parameters for each individual model. This is because each expert can be optimised separately, allowing for a more fine-tuned model that can better handle the EEG data.

The results obtained in this study supports the findings by Chen et al. [15]. In their experiment, they found that the Mixture of Experts (MoE) model performed similarly to their base model when using the original CIFAR-10 dataset but outperformed their base model when using the more complex CIFAR-10 Rotate dataset. The MoE model was able to achieve a higher accuracy than the base model when trained for the same number of epochs, indicating its ability to handle complex data structures. Similarly, in our experiment, when we trained both the base and MoE models for 10 epochs, the MoE model achieved a higher accuracy than the base model. This suggests that the MoE model is capable of handling complex EEG data structures and extracting meaningful features from the EEG data. It is important to note that we cannot determine whether the base models in Chen et al.'s experiment would perform equally well as their MoE counterparts if the number of training epochs was increased for the CIFAR-10 Rotate dataset. However, our results indicate that when trained for the same number of epochs, the MoE model outperforms the base model in terms of accuracy. These results align with the findings of Chen et al. [15].

5. Conclusion

In conclusion, our study aimed to explore the potential of the MoE architecture in EEG stress level prediction and at the same time preserve the spatial information and correlations between electrodes. The results showed that the proposed 3DMoEConvNet outperformed the base model in terms of accuracy and convergence speed. The 3DMoEConvNet's ability to handle complex data structures and extract meaningful features from EEG signals was a key factor in its superior performance. The results of this study are in line with previous findings that indicate the MoE architecture can improve performance in deep learning tasks. The proposed 3DMoEConvNet demonstrates the potential of the MoE architecture for EEG stress analysis and highlights its effectiveness in handling individual differences in EEG signals. Further research is needed to fully understand the potential of the MoE architecture for EEG analysis and to investigate its applications in other EEG-based tasks.

One area of future work is the selection of the number of experts in the 3DMoEConvNet. In this study, we chose to use 4 experts, following the methodology described by Chen et al. [15]. However, there are other methods that can be used to determine the optimal number of experts, such as clustering algorithms or domain knowledge. Further studies could explore these methods to determine the optimal number of experts for different types of EEG datasets.

An additional area of exploration could be the type of pooling or aggregation mechanism the MoE architecture employs to make the final prediction. The weighted sum prediction, also known as soft gating method, used in this study is solely dependent on the training optimiser's algorithm, and may not be the best choice for all scenarios. If the algorithm is unable to learn the features in the EEG dataset, it is very difficult for the developer to debug as it trains and updates the MoE model's parameters as one single complex model. Other gating and training methods, such as expert routing[3] and training each expert on a subset of the data, may be more reliable and suitable for different EEG data structures. These methods require the use of more machine learning algorithms such as KNN clustering algorithm [16], to determine which subset the particular data belongs to. Further studies could explore these alternative gating and training methods to determine their effectiveness for EEG analysis.

One significant constraint is the limited number of subjects included in the collected dataset. Further, the design of the experiment does not completely emulate the complex real-world VTS context, marking another limitation. However, this study aimed to verify the feasibility of our algorithm before moving to actual VTS settings. The intention was not to represent the entire population of VTSOs, but to provide a proof of concept for our proposed model. The number of subjects will be significantly increased in the next phase of the study when we collect data in actual VTS scenarios.

This study has proven that the 3D CNN base model is able to retain sufficient spatial and correlation information among the EEG electrodes and there is significant potential for further exploration and improvement in the use of the MoE architecture for EEG stress analysis using deep learning. Future works should focus on optimising the number of experts, optimising the training time and to explore the effectiveness of other pooling or aggregation mechanism to fully realise the potential of the MoE architecture. As for future work, while the focus has been on improving the algorithm, we will also consider more deeply how we can address stress management for VTSOs, which was our original question. We plan to collect data in real VTS settings and test the proposed model to further improve our research's relevance and impact on the maritime industry.

Acknowledgement

This work was supported by Singapore Maritime Institute Research Project (SMI-2021-MA-03).

References

[1] A. Napolitano, D. Di Crescenzo, and S. Gelli, A SelexSI solution to enable distributed decision making in vessels traffic management, in *2012 7th International Conference on System of Systems Engineering (SoSE)*, IEEE, 2012, pp. 203–208.

---

[3] The gating model will decide explicitly which expert will handle this particular input data.

[2] G. Borghini *et al.*, A multimodal and signals fusion approach for assessing the impact of stressful events on Air Traffic Controllers, *Scientific reports*, vol. 10, no. 1, pp. 1–18, 2020.
[3] P. Kutilek *et al.*, Heart rate variability during long-term air traffic control in cadets, in *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, IEEE, 2018, pp. 1–5.
[4] T. Song *et al.*, Variational instance-adaptive graph for EEG emotion recognition, *IEEE Transactions on Affective Computing*, 2021, Vol. 14, no. 1, pp. 343-356.
[5] A. Craik, Y. He, and J.L. Contreras-Vidal, Deep learning for electroencephalogram (EEG) classification tasks: a review, *Journal of neural engineering*, 2019, Vol. 16, no. 3, p. 031001.
[6] A. Kamrud, B. Borghetti, and C. Schubert Kabban, The effects of individual differences, non-stationarity, and the importance of data partitioning decisions for training and testing of EEG cross-participant models, *Sensors*, 2021, Vol. 21, no. 9, 3225.
[7] R. Fu *et al.*, Symmetric Convolutional and Adversarial Neural Network Enables Improved Mental Stress Classification From EEG, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022, Vol. 30, pp. 1384–1400.
[8] T. Song, W. Zheng, P. Song, and Z. Cui, EEG emotion recognition using dynamical graph convolutional neural networks, *IEEE Transactions on Affective Computing*, 2018, Vol. 11, no. 3, pp. 532–541.
[9] P. Zhang, X. Wang, W. Zhang, and J. Chen, Learning spatial–spectral–temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment, *IEEE Transactions on neural systems and rehabilitation engineering*, 2018, Vol. 27, no. 1, pp. 31–42.
[10] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, A multi-branch 3D convolutional neural network for EEG-based motor imagery classification, *IEEE transactions on neural systems and rehabilitation engineering*, 2019, Vol. 27, no. 10, pp. 2164–2177.
[11] Y. Zhao, X. Cao, J. Lin, D. Yu, and X. Cao, Multimodal affective states recognition based on multiscale cnns and biologically inspired decision fusion model, *IEEE Transactions on Affective Computing*, 2021, Vol. 14, no. 2, pp. 1391-1403.
[12] F. Scarpina and S. Tagini, The stroop color and word test, *Frontiers in psychology*, 2017, Vol. 8, 557.
[13] A. P. Allen, P. J. Kennedy, S. Dockray, J. F. Cryan, T. G. Dinan, and G. Clarke, The trier social stress test: principles and practice, *Neurobiology of stress*, vol. 6, pp. 113–126, 2017.
[14] K. Dedovic, R. Renwick, N.K. Mahani, V. Engert, S.J. Lupien, and J.C. Pruessner, The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain, *Journal of Psychiatry and Neuroscience*, 2005, Vol. 30, no. 5, pp. 319–325.
[15] Z. Chen, Y. Deng, Y. Wu, Q. Gu, and Y. Li, Towards Understanding the Mixture-of-Experts Layer in Deep Learning, *Advances in neural information processing systems*, 2022, https://proceedings.neurips.cc/paper_files/paper/2022/file/91edff07232fb1b55a505a9e9f6c0ff3-Paper-Conference.pdf, accessed July 6 2023.
[16] D. T. Larose and C. D. Larose, k-Nearest Neighbor Algorithm, in *Discovering Knowledge in Data: An Introduction to Data Mining*, Wiley, Hoboken, 2014, pp. 149–164. doi: 10.1002/9781118874059.ch7.