advances in Technology

Adv. Technol. **2022**, *2(3), 277-290*

Full Paper

Deep Residual Learning-Based Convolutional Variational Autoencoder for Driver Fatigue Classification

Sameera Adhikari* and Senaka Amarakeerthi

Department of Information and Communications Technology, Faculty of Technology, University of Sri Jayewardenepura, Sri Lanka

Email Correspondence: sameerabadhikari@gmail.com (S. Adhikari)

Received: 15 March 2022; Revised:14 June 2022; Accepted: 18 June 2022; Published: 25 August 2022

Abstract

Driving under the influence of fatigue often results in uncontrollable vehicle dynamics, which causes severe and fatal accidents. Therefore, early warning on the fatigue onset is crucial to avoid occurrences of such kind of a disaster. In this paper, the authors have investigated a novel semi-supervised convolutional variational autoencoder-based classification approach to classify the state of the driver. A convolutional variational autoencoder is a generative network. The authors have proposed a discriminative model using convolutional variational autoencoders and residual learning. This approach calculates an intermediate loss base on deep features of the network in addition to the label information in training. The loss obtained by this method helps the training to be more effective on the model and leads to better accuracy in driver fatigue classification. The trained model has managed to classify driver fatigue with higher accuracy (97%) than the other successful models taken into comparison, proving that the proposed method is more practical for computing classification loss for driver fatigue to currently available methods.

Keywords: Autoencoder, brain-computer interface, driver fatigue classification, electroencephalography, residual learning

Introduction

Variational Autoencoder (VAE) is a lucid and coherent method used in data reconstruction. These methods implement a latent variable model which assumes prior distribution over the latent space to be Gaussian. The VAE neural network parameterizes the mean and variance of a distribution. Here, the latent space is a continuous, low dimensional parameterized representation of the input space optimized by the reconstruction loss, which represents one-half of the VAE loss function. Moreover, comparing to standard autoencoders, VAE's latent space gets the continuity specialty with the help of having an additional prior distribution layer over its latent space.

VAE-based methods have been utilized with astounding results for many tasks. These methods can be found in supervised [4], unsupervised [2, 3] and semi-supervised [1] learning categories. Semi-supervised learning category includes EEG-based motor imagery [5], text classification [1], intrusion detection [6], image classification [7], etc. However, a VAE-based semi-supervised method for EEG driver fatigue detection is yet to be tested.

EEG-based fatigue and drowsiness detection [8] and driver fatigue detection and mitigation [9, 10] have been significantly investigated using various methods over the years. However, the authors of this paper have investigated novel ways to improve the performance of driver fatigue classification. In this paper, the

https://doi.org/10.31357/ait.v2i3.5545

authors have proposed a method that consists of an architectural collaboration of VAE's encoder/decoder models and a deep learning classifier model.

Experimental Section/Materials and Methods

Due to the heterogeneous and high dimensional nature of data, an unsupervised method for recognizing patterns in time-series data like driver fatigue can be an extra demanding task. The existence of noise and artifacts makes it even harder to reach the classification benchmarks. Convolutional Variational Autoencoders (CVAE) is one solution for the classification problem as it addresses the issues comes under low accuracy, making it a proficient proxy for an ideal classifier [5]. However, this approach is untested with driver fatigue EEG data. Also, a reconstruction model like the CVAE itself is incapable of predicting classes (classification) without the help of an implementation for a classifier.

A Novel Approach for Driver Fatigue Classification

CVAE is a bipartite model consisting of two deep learning networks, an encoder (inference network) (Equation 1) and a decoder (generative network) (Equation 2) [11]. The encoder is responsible for the sampling of random variable z (latent variable) from the input variable x. The latent space is a compressed, low dimensional representation of a vector space for the input space, whereas, the decoder generative network, is responsible for reconstructing the original data with approximation.

$$z = f_{\phi}(x) \tag{1}$$

$$\hat{x} = f_{\theta}(z) \tag{2}$$

Moreover, the method adopted by the authors utilizes a separate structure for classifier alongside the encoder and decoder models for the classification of fatigue and alert stages of drivers. The implementation for the complete classification model is described in Method Outline section.

Variational Inference

The latent variable model can be denoted by the joint distribution of $p_{\theta}(x,z)$. Here, the marginal distribution of the latent variable model over the input variables $p_{\theta}(x)$ is given as:

$$p_{\theta}(x) = \int p_{\theta}(x,z)\, dz \tag{3}$$

The simplest form of the latent variable model is as follows:

$$p_{\theta}(x,z) = p_{\theta}(z)\, p_{\theta}(x \mid z) \tag{4}$$

Here, the distribution $p_{\theta}(z)$ is the prior distribution because it is unconstrained by any input observation (x). Also, $p_{\theta}(x \mid z)$ is called the stochastic decoder. The posterior distribution $p_{\theta}(z \mid x)$ can be described as follow.

$$p_\theta(z|x) = \frac{p_\theta(x,z)}{p_\theta(x)} \tag{5}$$

The $p_\theta(x|z)$ is computationally tractable, but the posterior distribution is intractable because of the marginal likelihood $p_\theta(x)$ being intractable. Therefore, techniques for the approximation of inference have to be used in order to enable the approximation of posterior likelihood $p_\theta(z|x)$ and marginal likelihood $p_\theta(z|x)$.

Inference Model

The intractable posterior inference of the previous latent variable model is converted into a tractable problem by formulating a parametric inference model $q_\phi(z|x)$. This model is designated as an encoder or recognition model. The ϕ is the model parameter known as the variational parameter. The focus of the encoder model is to approximate the posterior by optimizing the variational parameter ϕ.

$$q_\phi(z|x) \approx p_\theta(z|x) \tag{6}$$

Commonly, the variational parameter (ϕ) is composed of weights and biases of the deep learning model in the encoder.

$$\begin{aligned}(\mu, \log\sigma) &= Encoder_\phi(x)\\ q_\phi(z_n|x_n) &= \mathcal{N}(z|\mu, diag(\sigma^2))\end{aligned} \tag{7}$$

The μ and σ, mean and standard deviation of Gaussian distribution, are derived from the encoder output. A more robust strategy of sharing variational parameters is *amortized variational inference*, which approximates ϕ_n for each observation x_n.

$$q_\phi(z_n|x_n) = \mathcal{N}(z_n|\mu_\phi(x_n), diag(\sigma^2_\phi(x_n))) \tag{8}$$

Objective Function

From Equation 5, the marginal likelihood $p_\theta(x)$ can express as:

$$p_\theta(x) = \frac{p_\theta(x,z)}{p_\theta(z|x)} \tag{9}$$

Therefore,

$$\begin{aligned}\log p_\theta(x) &= \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x)]\\ &= \mathbb{E}_{q_\phi(z|x)}\left[log\left[\frac{p_\theta(x,z)}{p_\theta(z|x)}\right]\right]\\ &= \mathbb{E}_{q_\phi(z|x)}\left[\log\left[\frac{p_\theta(x,z)}{q_\phi(z|x)}\frac{q_\phi(z|x)}{p_\theta(z|x)}\right]\right]\end{aligned}$$

$$= \mathbb{E}_{q_\phi(z|x)}\left[\log\left[\frac{p_\theta(x,z)}{q_\phi(z|x)}\right]\right] + \mathbb{E}_{q_\phi(z|x)}\left[\log\left[\frac{q_\phi(z|x)}{p_\theta(z|x)}\right]\right] \quad (10)$$

The first term in Equation 10 is the *evidence lower bound (ELBO)* $\mathcal{L}_{\theta\phi}(x)$.

$$\mathcal{L}_{\theta\phi}(x) = \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x,z) - \log q_\phi(z|x)\right] \quad (11)$$

The second term in Equation 10 is the non-negative *Kullback-Leibler (KL) divergence* ($D_{KL}(.)$). KL divergence quantifies the difference between distributions $q_\phi(z|x)$ and $p_\theta(z|x)$.

$$D_{KL}(q_\phi(z|x)\|p_\theta(z|x)) = \mathbb{E}_{q_\phi(z|x)}\left[\log\left[\frac{q_\phi(z|x)}{p_\theta(z|x)}\right]\right]$$
$$= \mathbb{E}_{q_\phi(z|x)}\left[\log q_\phi(z|x) - \log p_\theta(x|z) - \log p_\theta(z)\right] + \log p_\theta(x) \quad (12)$$

Finally, the ELBO can be formulated as:

$$\mathcal{L}_{\theta\phi}(x) = \log p_\theta(x) - D_{KL}(q_\phi(z|x)\|p_\theta(z|x))$$
$$= \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right] - D_{KL}(q_\phi(z|x)\|p_\theta(z)) \quad (13)$$

The objective here is to maximize the ELBO in Equation 13. The maximization effect can be found in two things. First, it will produce better approximation results out of the generative model by maximizing $p_\theta(x)$. Second, minimize the divergence between the approximate posterior $q_\phi(z|x)$ and $p_\theta(x|z)$, hence, produce better $q_\phi(z|x)$. In practice, ELBO can be simplified by getting the Monte Carlo estimator of expectation in Equation 11[12].

$$\mathcal{L}_{\theta\phi}(x) = \log p_\theta(x|z) + \log p_\theta(z) - \log q_\phi(z|x) \quad (14)$$

Reparameterization

A general way to create sample space $p_\phi(z)$ is by using the factorized Gaussian encoder definition in Equation 7.

$$z \sim q_\phi(z|x) \quad (15)$$

However, in the training process, the gradients of the backpropagation algorithm cannot propagate through z due to its randomness. Therefore, z is reparameterized from externalizing randomness to another independent random variable (ε) [13].

$$z = g(\varepsilon, \phi, x) \quad (16)$$

Here, the ε is independent of x and ϕ.

$$\varepsilon \sim \mathcal{N}(0,1) \quad (17)$$

After the reparameterization, the latent variable:

$$z = \mu + \sigma \odot \varepsilon \tag{18}$$

Where $\odot$ represents the element wise product operation [14]. The altered Equation 13 can be,

$$\mathcal{L}_{\theta\phi}(x) = \frac{1}{L} \sum_{l=1}^{L} \log p_{\theta}(x \mid z^{(l)}) + \frac{1}{2} \sum_{j=1}^{J} (1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2) \tag{19}$$

In this study, Equation 19 is the objective function. The first term of the objective function is the reconstruction error which calculates the mean square error of the reconstructed signal and original signal. The second term is the KL divergence between the inference model and the prior distribution over z [2].

Convolutional Variational Autoencoders as a Classifier

As mentioned in section 2, utilizing CVAE as a solitary classification method is fruitless because of the autoencoder is only tasked with regenerating input [18]. In order to implement discriminative modeling with the help of a deep generative model CVAE, the authors of this article have used a semi-supervised learning approach based on the study done in the article [19].

Method Outline

The driver fatigue classification model (DFCM) has three main components; encoder, decoder, and classifier (Figure 1). Here, the authors have trained the encoder and decoder independently. It is a unique method compared to the standard way of training autoencoders. The encoder training phase starts with applying prepared EEG driver fatigue data to the encoder model and obtaining the latent space (Figure 2). The results gives the KL loss ($\mathcal{L}_{kl}$) mentioned in section 2 from the latent output. After that, the latent space formed is inputted into the classifier model for the prediction of the class associated with the input. The categorical cross entropy is applied to the predicted and true classes to obtain the categorical loss ($\mathcal{L}_{cat}$). Finally, the encoder model is optimized based on encoder loss ($\mathcal{L}_{en}$) which is the summation of ($\mathcal{L}_{kl}$) and ($\mathcal{L}_{cat}$). The weights of the encoder and classifier are saved for use in the decoder training phase.

The decoder training is done using the same training dataset used in the first phase. As illustrated in Figure 3, the decoder model and the encoder is used to form a variational auto-encoder. The output of the VAE is fed into a second encoder model to obtain the reconstruction loss ($\mathcal{L}_{rec}$). The reconstruction loss gives the mean square error (MSE) between two sets of deep features from two encoders (Figure 3). A single set of features contains the outputs from the first three intermediate layers of an encoder model (Figure 1).

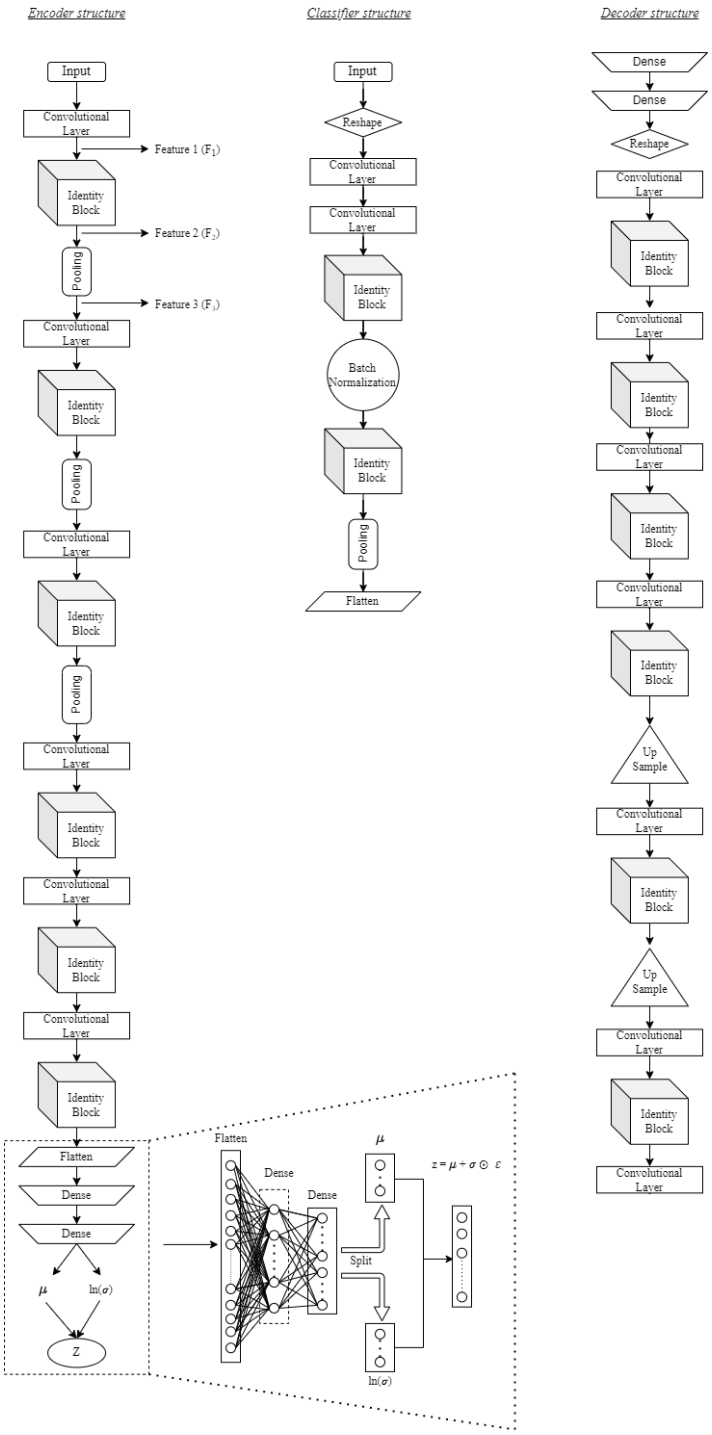


Figure 1. Components of the proposed driver fatigue classifier

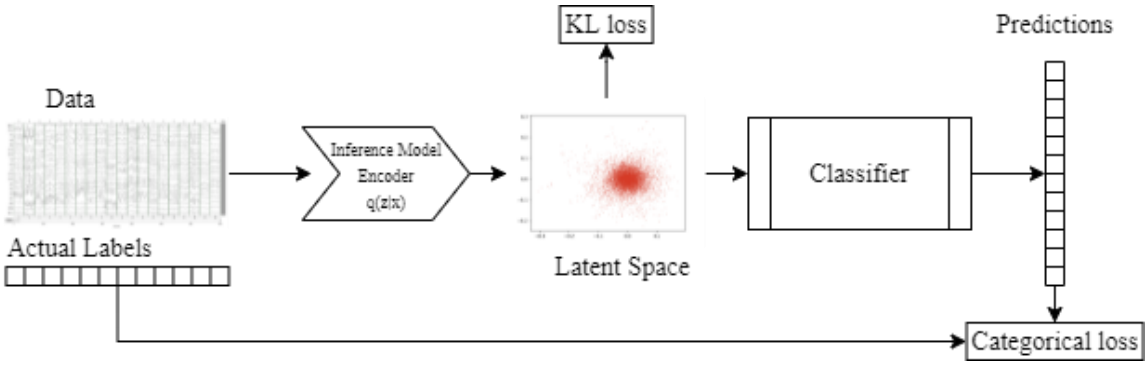


Figure 2. Encoder Training

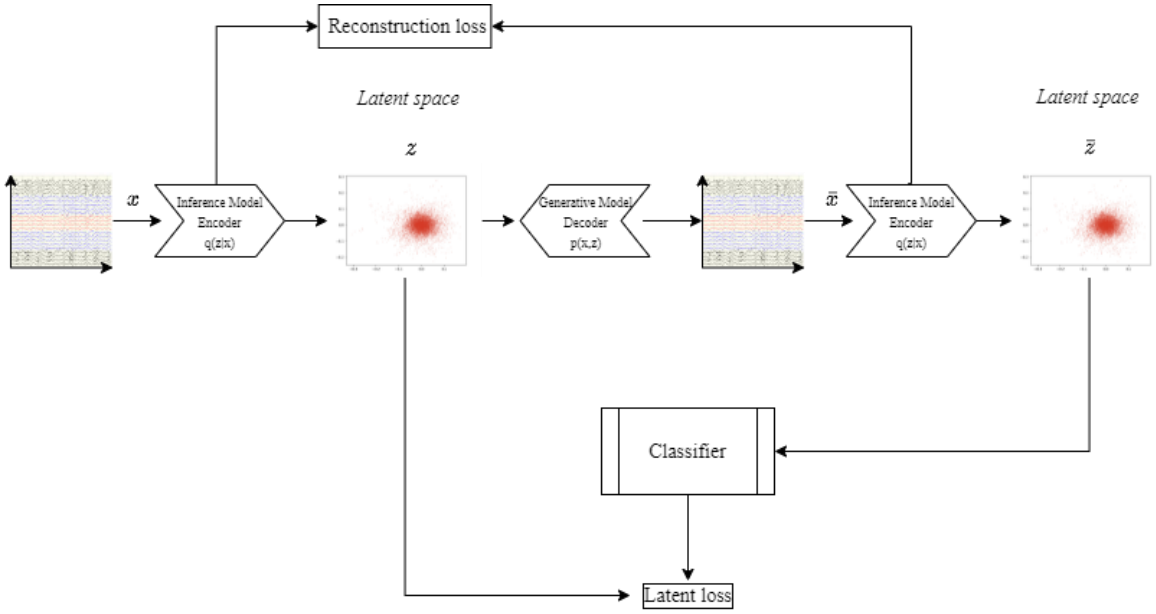


Figure 3. Decoder Training

$$\mathcal{L}_{rec} = \sum_{i=1}^{3} MSE(\mathcal{F}_i - \tilde{\mathcal{F}}_i) \tag{20}$$

$\mathcal{F}_i$ and $\tilde{\mathcal{F}}_i$ are respectively the features from the first and second encoders illustrated in Figure 3. The classifier gets the input latent space from the encoder. The loss between the predicted class of the classifier and the true class is the latent loss. ($\mathcal{L}_{latent}$).

$$\mathcal{L}_{latent} = MSE(y_{true} - y_{pred}) \tag{21}$$

Finally, the total loss for the decoder ($\mathcal{L}_{dec}$),

$$\mathcal{L}_{dec} = \alpha\, \mathcal{L}_{rec} + \beta\, \mathcal{L}_{latent} \tag{22}$$

Here, α and β are fitting parameters.

Data-set and Network Architecture

The authors have used an original EEG driver fatigue dataset [20] for the training, validation, and testing the models. This dataset consists of data from 12 subjects. The data has been recorded using a 40-channel Neuroscan amplifier (36 EEG/ 4 EOG). EOG monitors horizontal and vertical eye movement of both eyes in order to detect ocular artifacts.

EEG signal is band-locked with a band-pass filter (40Hz upper and 4Hz lower cutoff frequencies). The signals are preprocessed by applying Independent Component Analysis (ICA). The prepared data has 29,658 samples. A sample (epoch/window) has data from 18 channels and 240-time slots, hence, the dimension of a sample is $18 \times 240 \times 1$. Sixty percent (60%) of data has been allocated for training. The remaining forty percent (40%) have been divided into two equal sets for validation and testing.

As illustrated by Figure 1, the encoder, decoder, and classifier are deep residual convolutional neural networks. Here, all residual blocks are identity blocks. The identity mapping created by the residual block contains two convolution layers. The additional layers in the deep network helps the model to learn more to improve the performance in contrast to shallow networks. Also, the residual block will preserve the advantage gained in shallow layers in the deep network by preventing them from vanishing or exploding [21, 22].

Results and Discussion

The authors have studied two-dimensional (2D) (Figure 4) and 100-dimensional latent space for this study and found that latent space with 100 dimensions has better performance over 2D latent space.

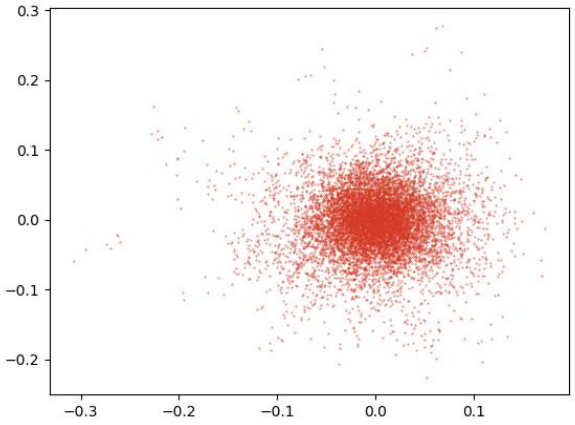
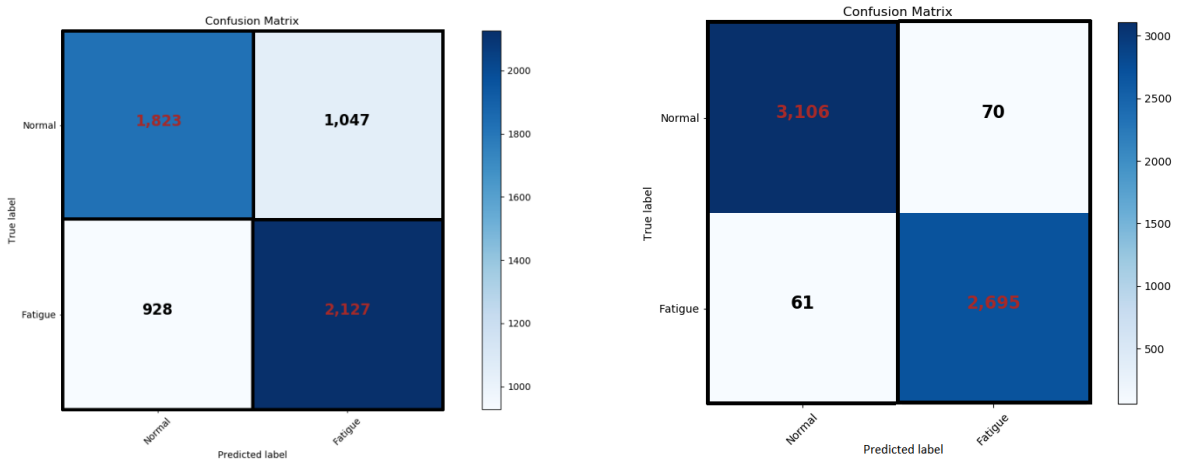


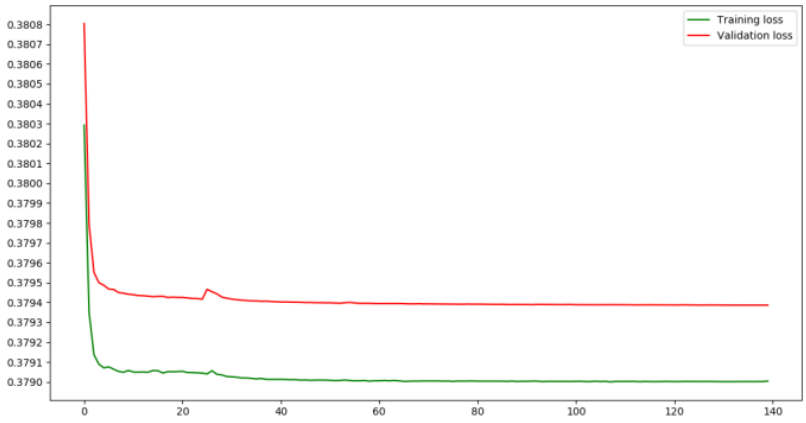
Figure 4. Two-dimensional Latent Space

Moreover, the authors have investigated the effects of α and β on the testing performance of the network. According to [19], the best result has been obtained by the combination of values of $\alpha = 1$ and $\beta = 0.1$. Authors have tested this combination against the identical combination (i.e. $\alpha = 1$ and $\beta = 1$), and the results show that the model trained with the values of $\alpha = 1$ and $\beta = 0.1$ has scored 67% testing accuracy (Figure 5(a)), whereas the model with the identical combination has shown 97% testing accuracy (Figure 5(b)) for the inter-subject data set (i.e., all the samples of the subjects). Also, the rate of learning is higher on models with an identical one to one combination of fitting parameters (Figure 6). The model used for the prediction is shown in Figure 7.

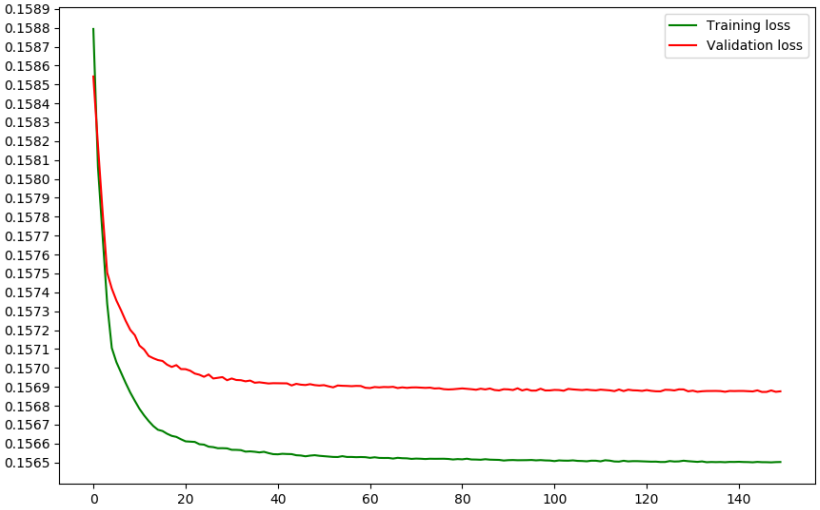


(a). α=1, β=0.1 **(b).** α=1, β=1

Figure 5. The effects of α and β on accuracy



(c). α=1, β=0.1



(c). α=1, β=1

Figure 6. Loss function

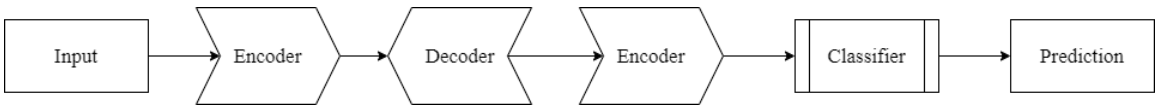


Figure7. Prediction model

Apart from that, the authors have tested the EEG driver fatigue dataset using another similar type of model, a Bi-directional Long-short Term Memory Network-based CNN model (BLSTM-CNN). The BLSTM-CNN structure can be found in Figure 8.

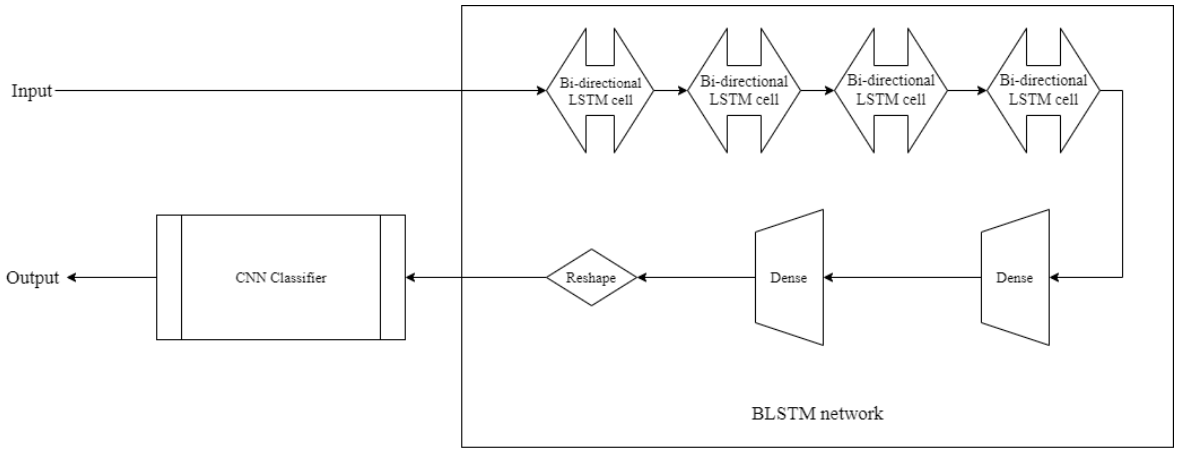


Figure 8. BLSTM-CNN model

A comparison of the results from this study and three previous studies is summarized in Table 1. From the summarized results, it is clear that DFCM with α = 1 and β = 0.1 has shown superior accuracy among the other models.

Table 1. Inter-subject accuracy comparison with diver fatigue classification model

Model	**DFCM [α = 1,β= 1]**	**DFCM [α = 1,β= 0.1]**	**BLSTM-CNN**	**LSTM [15]**	**k-NN [16]**	**EEG-Conv-R [17]**
Accuracy	97.12%	66.67%	86.15%	73.41%	94.25%	92.68%

Furthermore, the authors have investigated the intra-subject performance of the classifier. Here, the DFCM model with parameter values of $\alpha = 1$ and $\beta = 1$ has been trained on a per subject basis and obtained the accuracy for each subject. A subject's data is split into training and testing sets where 80% of the data is used for training and 20% for testing. The result for the intra-subject classification is shown in Figure 9. This average accuracy from this experiment reaches 95.037%. The statistical analysis results can be found in Table 2.

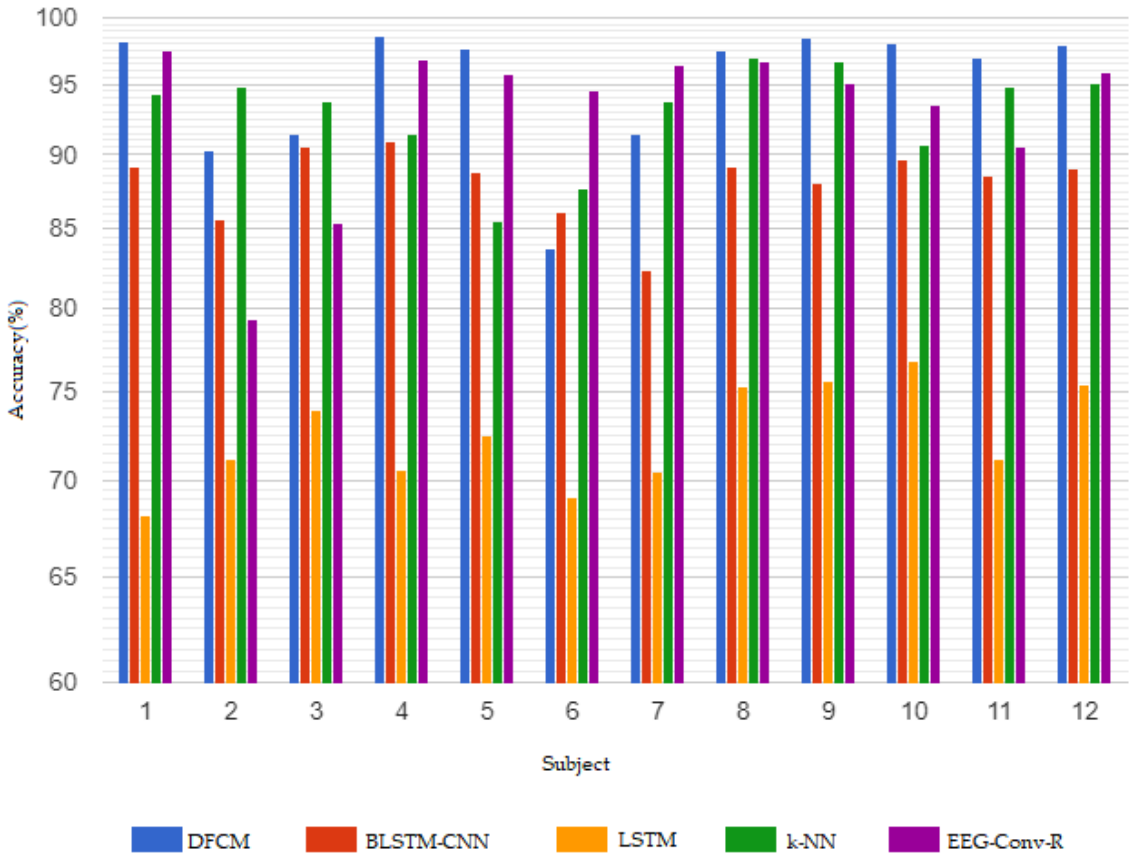


Figure 9. Intra-subject accuracy

Table 2. Statistical analysis of intra-subject accuracy for DFCM

	Our approaches		**Excising approaches**		
Model	**DFCM [$\alpha = 1, \beta = 1$]**	**BLSTM-CNN**	**LSTM**	**k-NN**	**EEG-Conv-R**
Mean (%)	95.0367	88.1533	72.5408	92.9608	93.13
Standard Deviation	4.5200	2.27086	2.69213	3.36116	5.2801
Variance	20.4308	5.1568	7.2475	11.2974	27.8798

From the results obtained in the experiment done on intra-subject data, it is evident that the DFCM model is highly capable of learning features of each distinct subject despite of the considerable variances among

the EEG data. Even though the statistical analysis shows some significant variances for the accuracy values, our proposed DFCM model yields excellent classification results compared to the other tested model.

Conclusions

In this article, the authors have proposed a novel method for driver fatigue classification. This method implements a dedicated classifier for classification together with an additional autoencoder network. The classifier is trained at first in conjunction with the encoder. The classifier and encoder have been optimized using the deep features from the encoder and classifier output. The decoder is trained and optimized at the next step based on the reconstruction loss. The overall model considers the classification error and intermediate feature errors in training. As a result, it delivers superior performance compared to a similar methods currently used for driver fatigue detection.

Conflicts of Interest

The Authors declare that there is no conflict of interest to disclose.

Funding

The research was funded by the University of Sri Jayewardenepura, Sri Lanka under the grant number ASP/01/RE/FOT/2019/60.

Supporting Information

The details for the available EEG data used in this article can be found in [20].

References

[1] W. Xu, H. Sun, C. Deng, Y. Tan, Variational Autoencoder for Semi-Supervised Text Classification In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA, 04 February 2017.

[2] M. Memarzadeh, B. Matthews, I. Avrekh, Aerospace. **2020,** 7(8), 115.

[3] T. Chen, X. Liu, B. Xia, W. Wang, Y. Lai , IEEE Access. **2020**, 8, 47072-47081.

[4] F. Berkhahn, R. Keys, W. Ouertani, N. Shetty, D. Geibler, arXiv: Learning(1908.03015). **2020.**

[5] M. Dai, D. Zheng, R. Na, S. Wang, S. Zhang, Sensors. **2019**, 19, 551.

[6] Y. Yang, K. Zheng, C. Wu, Y. Yang, Sensors. **2019**, 19, 2528.

[7] Y. Fan, G. Wen, D. Li, S. Qiu, M. D. Levine, F. Xiao, Computer Vision and Image Understanding. **2020**, 195, 102920.

[8] R. N. Khushaba, S. Kodagoda, S. Lal, G. Dissanayake, IEEE Transactions on Biomedical Engineering. **2011**, 58(1), 121-131.

[9] S. K. Lal, A. Craig, Psychophysiology. **2002**, 39(3), 313-321.

[10] Q. Wang, J. Yang, M. Ren, Y. Zheng, Driver fatigue detection: A survey In Proceedings of the 6th World congress on Intelligent Control and Automation, Dalian, China, 21 June 2006.

[11] J. Potratz, C. S. Arauco, J. Castro, A. Emerick, M. Pacheco, Large dimension parameterization with convolutional variational autoencoder: An application in the history matching of channelized geological facies models In Proceedings of the 20th International Conference on Computational Science and Its Applications, Cagliari, Italy, 01 July 2020.

[12] D. P. Kingma, M. Welling, Found. Trends Mach. Learn. **2019,** 12(4), 307-392.

[13] D. Kingma, M. Welling, Auto-encoding variational bayes In Proceedings of the 2nd International Conference on Learning Representations (ICLR 2014), Banff, Canada, 14-16 April 2014.

[14] T. Cui, G. Gou, G. Xiong, Gated convolutional variational autoencoder for IPv6 target generation In Proceedings

of the Advances in Knowledge Discovery and Data Mining, 24th Pacific-Asia Conference (PAKDD 2020), Singapore, 11–14 May 2020.

[15] Y. Ed-doughmi, N. Idrissi, Driver fatigue detection using recurrent neural networks In Proceedings of the 2nd International Conference on Networking, Information Systems & Security, Rabat, Morocco, 27 March 2019.

[16] Y. Jiao, Y. Deng, Y. Luo, B. L. Lu, Neurocomputing. **2020,** 408, 100-111.

[17] H. Zeng, C. Yang, G. Dai, F. Qin, J. Zhang, W. Kong, Cogn. **2018,** 12, 597–606.

[18] S. W. A. Canchumuni, A. A. Emerick, M. A. C. Pachecoa, Comput. Geosci. **2019,** 128, 87-102.

[19] R. Zemouri, Mach. Learn. Knowl. Extr. **2020**, 2, 361-378.

[20] J. Min, P. Wang, J. Hu, The original EEG data for driver fatigue detection, 2017, https://figshare.com/articles/The_original_EEG_data_for_driver_fatigue_detection/5202739/1, DOI: 10.6084/m9.figshare.5202739.v1.

[21] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , Las Vegas, United States, 26 Jun – 1 Jul 2016.

[22] W. Sun, Y. Su, X. Wu, X. Wu, Neurocomputing. **2020**, 404, 108-121.