Deep Learning with Convolutional Neural Network for detecting microsleep states from EEG: A comparison between the oversampling technique and cost-based learning

Venkat Krishnamoorthy, *Student Member, IEEE*, Reza Shoorangiz, *Member, IEEE*, Stephen J. Weddell, *Senior Member, IEEE*, Lutz Beckert, Richard D. Jones, *Fellow, IEEE*

***Abstract*— Any occupation which involves critical decision making in real-time requires attention and concentration. When repetitive and expanded working periods are encountered, it can result in microsleeps. Microsleeps are complete lapses in which a subject involuntarily stops responding to the task that they are currently performing due to temporary interruptions in visual-motor and cognitive coordination. Microsleeps can last up to 15 s while performing a particular task. In this study, the ability of a convolutional neural network (CNN) to detect microsleep states from 16-channel EEG data from 8 subjects, performing a 1D visuomotor was explored. The data were highly imbalanced. When averaged across 8 subjects there were 17 responsive states for every microsleep state. Two approaches were used to handle the CNN training with data imbalance – oversampling the minority class and cost-based learning. The EEG was analysed using a 4–s epoch with a step size of 0.25 s. Leave-one-subject-out cross-validation was used to evaluate the performance. The performance measures used for assessing the detection capability of the CNN were: sensitivity, precision, phi, geometric mean (GM), AUC_{ROC}, and AUC_{PR}. The performance measures obtained using the oversampling and cost-based learning methods were: AUC_{ROC} = 0.90/0.90, AUC_{PR} = 0.41/0.41 and a phi = 0.42/0.40, respectively. Although the performances were similar, the cost-based learning method had a considerably shorter training time than the oversampling method.**

I. Introduction

According to a survey conducted by Ministry of Transport in New Zealand in the year 2016, fatigue was identified as a contributing factor in 43 fatal crashes, 199 serious injury crashes and 450 minor injury crashes and the total cost of crashes involving fatigued drivers was about $363 million. During these accidents, the drivers themselves did not recognize the brief moment when they lost their attention or consciousness or the adrenaline rush after the accident disguised their drowsiness. This brief period during which the driver has absence of sensory-motor and cognitive performance are termed as microsleeps. They can range anywhere from 0.5 s to 15 s [1],[2]. Microsleeps occurs generally without warning. It is a light sleep state in which the person having microsleep won't be aware of the state they are in. These kinds of lapses can even occur in a perfectly healthy non-sleep-deprived subjects performing a repetitive task without any prior indications such as drowsiness [3]. Thus it is very important to detect these states and warn the subject before any serious mishap occurs.

Davidson et al. [1] used log-power spectral measures as features, principal component analysis (PCA) for dimensionality reduction and a long short-term memory (LSTM) neural network for classification. An AUC_{ROC} of 0.84 and a phi of 0.38 was achieved. Peiris et al. [4] used 6 LDA classifiers with the same features and dimensionality reductions and achieved an AUC_{ROC} of 0.86 and a phi measure of 0.39. Ayyagari et al. [5] approach was similar. They used a stacked echo state networks (ESN) with leaky neurons to achieve a phi of 0.44, and AUC_{ROC} of 0.88. In all these studies, an epoch length of 2 s with 50% overlap was used.

Shoorangiz et al. [6] used power spectral features to predict microsleep states, for a prediction time of 1 s ahead and achieved a best performance of AUC_{ROC} of 0.94, and a phi of 0.44, using a single LDA classifier. Buriro et al. [7] investigated the effectiveness of seven pairwise interchannel feature sets: covariance, Pearson's correlation coefficient, wavelet cross-spectral power, wavelet coherence, joint entropy, mutual information and phase synchronization index in the prediction of microsleep states. A 5-s window of EEG was used with a step size of 0.25 s for extracting features. The joint entropy features with LDA classifier resulted in the best performance of phi 0.47, AUC_{PR} 0.50, and AUC_{ROC} 0.95, for a prediction time of 0.25 s ahead.

Conventional machine-learning techniques need to manually go through the process of feature extraction, feature selection and feature reduction. Recent exploration in the field of deep learning have led to the application of several popular methods, especially convolutional neural network (CNN) in the field of image processing, video processing,

*This work was supported by University of Otago and the New Zealand Brain Research Institute.

Venkat Krishnamoorthy is with the Department of Medicine, University of Otago, Christchurch, the Christchurch Neurotechnology Research Programme, and the New Zealand Brain Research Institute, Christchurch, New Zealand. (venkat.krishnamoorthy@nzbri.org)

Reza Shoorangiz is with Christchurch Neurotechnology Research Programme, the New Zealand Brain Research Institute, Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand (reza.shoorangiz@nzbri.org)

Stephen J. Weddell is with the Department of Electrical and Computer Engineering at University of Canterbury and the Christchurch Neurotechnology Research Programme, Christchurch, New Zealand. (steve.weddell@canterbury.ac.nz)

Lutz Beckert is with the Department of Medicine, University of Otago, Christchurch, New Zealand (lutz.beckert@otago.ac.nz)

Richard Jones is with the Department of Medicine, University of Otago, Christchurch, New Zealand, the Christchurch Neurotechnology Research Programme, the New Zealand Brain Research Institute, Christchurch, Department of Electrical and Computer Engineering, University of Canterbury, Christchurch, New Zealand. (richard.jones@nzbri.org)

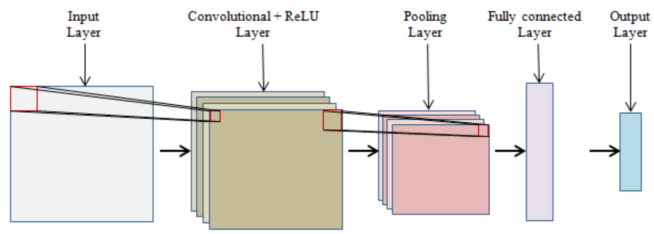

Fig. 1: Basic Structural and functional representation of a CNN.

and speech processing. CNNs are now one of the most sought deep-learning approaches, especially for image processing. The CNN is a biologically-inspired architecture emulating the visual processing in the brain. It consists of several layers including at least one convolution layer, rectified linear unit (RELU) layer, pooling layer, and a fully-connected layer at the end (Fig. 1). Unlike the neural network architecture which is fully connected at all layers, the neurons in the convolutional layers are only connected to a set of local adjacent neurons in the next layer. This is analogous to the local receptive field processing of the visual cortex. The convolutional mask or the filter connecting the sub-regions of the adjacent layers is common to the entire input data, thus helping us to extract shift invariant features. A feature map is obtained by repeated application of a function across sub-regions of the entire image, in other words, by convolution of the input data with a linear filter, adding a bias term and then applying a non-linear function.

Several works have been carried out which have shed more insight into CNN and how efficient the model is concerning classification or prediction on biosignals, especially EEG. Schirrmeister et al. [8] designed several CNN architectures ranging from a 2-layer shallow architecture to a 31-layer deep architecture and analyzed the impact of CNN design choices and training strategies on decoding accuracies. Their work gives insights into CNN's perception and how they extract discriminative feature maps to understand the EEG signals. Bashivan et al. [9] used deep recurrent-CNN to obtain effective learning representations that are invariant to inter- and intra-subject differences and also to the inherent noises associated with the EEG. Supratak et al. [10] designed a CNN model, DeepSleepNet, to score sleep stages automatically.

Inspired by the CNN's ability in the above-mentioned biomedical applications, this paper explores the effectiveness of CNN in detection of microsleeps states from EEG, taking into account the resource requirement, the substantial data imbalance (**??**), and time involved in training the CNN.

II. Methodology

A. Data

Fifteen normal healthy subjects aged 18-36 years, who had no neurological or sleep disorders and had an average previous night sleep of 7.8±1.2 h [11], were observed during two 1 hr sessions performing a 1D continuous tracking task. The tracking task required subjects to control the horizontal position of a cursor using a steering wheel over a 175° range. EEG, tracking performance and a video of the subject were recorded for each session. EEG was obtained from 16 electrodes – Fp1, Fp2, F3, F4, F7, F8, C3, C4, O1, O2, P3, P4, T3, T4, T5, and T6, placed according to the international 10-20 system, at a sampling frequency of 256 Hz. Of the 15 subjects, 8 subjects had one definite microsleep in at least one of the sessions. The Institution's Ethical Review Board approved all experimental procedures involving human subjects.

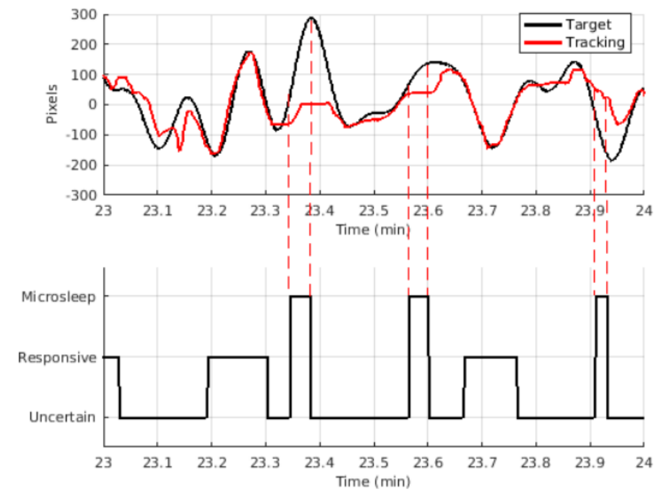

Fig. 2: Tracking performance and corresponding gold standard.

B. EEG pre-processing

The acquired raw EEG signals were pre-processed using a FIR bandpass filter with cut-off frequencies of 1 Hz to 45 Hz. To improve the SNR, the filtered signals was re-referenced to a common average. Artifact-subspace reconstruction (ASR) [12] with a 2-min window length and 50% overlap was used to remove any artefacts with a z-score>5.

C. Gold-standard

Using the tracking performance and video recording during the sessions, a gold-standard was generated. An expert examined the video, independent of the tracking performance and identified the behavioural clues [4],[13]. Tracking performance was used to identify 'responsive', 'deviated' and 'flat spot' regions [12]. Both the video and tracking analysis were used to generate the 'gold standard'. The gold standard has 3 states – 'responsive', 'microsleep', and 'uncertain'. 'Responsive' refers to data segments where the subject was closely tracking the target, 'microsleep' refers to data segment where the subject is not tracking the target and the video rating was either 'deep drowsy' or 'microsleep', and 'uncertain' refers to data segment which cannot be accurately defined as either a responsive or a microsleep state. (Fig. 2).

TABLE I: CNN Architecture.

Input – 16×1024 (16-channel EEG epochs sampled at 256 Hz)
Convolution Layer 1×40, 25 ReLu
Convolution Layer 16×1, 25 batch normalization ReLu maxpool 1×3, Stride 1×3
Convolution Layer 1×40, 50 ReLU maxpool 1×3, Stride 1×3
Convolution Layer 1×40, 100 ReLU maxpool 1×3, Stride 1×3
Fullyconnected Layer Softmax Classification Layer

D. CNN architecture

We adopted an architecture mimicking 'deepConvNet' used for EEG decoding and visualization [8]. The base architecture was modified to best suit the problem in hand. As described in Table I, our CNN has 16 layers. The convolutional layer parameters are denoted by 'filter height'×'filter width' and 'number of filters'. The pooling layer parameters are denoted by 'height×width'.

The input to CNN was EEG epochs of duration 4 s. The first convolutional layer convolves with the input of size 16×1024 (EEG sampled at 256 Hz). The resultant feature map is processed by the second convolution layer. The batch-normalization layer after the second convolution layer reduces the internal covariate shift among the multiple feature maps, thus making the weight initialization easier. The obtained feature map is down-sized by the maxpool layer before being passed on to the next set of layer. The dimensionality reduction is automatically taken care of as the layer deepens. Finally, the fully connected layer processes the 100 feature maps of size '1×19' after the final maxpool layer. A weighted cross-entropy was used to calculate the prediction error at the output stage. The NVIDIA GeForce GTX1070 GPU was used to train the CNN model.

The dataset on an average (all 8 subjects combined) had an imbalance ratio of 1:17 (microsleep:responsive). Fig. 3 gives the subject-wise class imbalance along with the number of uncertain states in each subject. The performance of the classifier will be affected by the class imbalance. While training the learning algorithms insensitive to class imbalance could classify all samples to the majority class in order to minimize the error rate [14].

The class imbalance problem was handled in two ways:

• An over-sampling technique was used, in which the minority class (microsleep) was replicated to equal the majority class (responsive). Although there are several over-sampling techniques available (e.g., SMOTE, ADASYN), they can only be used on features extracted from the EEG data. In our case since the actual EEG signal was used as the input, it was not possible to generate synthetic EEG epochs. Hence, the minority class was replicated to equal the majority class in numbers (**case 1**).

• A weighted cost function (cross-entropy). This was inspired by prior work on using cost sensitive loss functions with imbalanced datasets in neural networks and deep learning (**case 2**) [15], [16].

The hyperparameters best suited for detecting microsleeps were determined as follows:

1. Of the 8 subjects, one subject was reserved for testing.
2. The hyperparameters of the model were tuned by training the model with the remaining 7 subjects, of which 6 were used for training and 1 for cross-validation.
3. The model with the best AUC_{ROC} was chosen.
4. The chosen model was trained using the 7 subjects and tested on the reserved test subject.
5. This process was repeated for all 8 subjects and the performance measures were averaged.

The above procedure was done for both cases 1 and 2. Finally, the model's performance to detect microsleep states and the average time utilised for training for cases 1 and 2 were compared.

III. Result

Performance measures for detection of microsleep states are presented in Table II. The performance measures between the two cases are similar. A model was also trained on the imbalanced training data It was found that training CNN with imbalanced dataset resulted in a bias towards majority class, giving high specificity but a low sensitivity. The mini-batch size was kept constant in both cases, thus the number of mini-batches would vary in both cases. The average time utilized for training the CNN in both cases 1 and 2 is illustrated in Fig. 4.

IV. Discussion

Performance of the CNN was essentially the same for the oversampling technique and cost-based learning. This

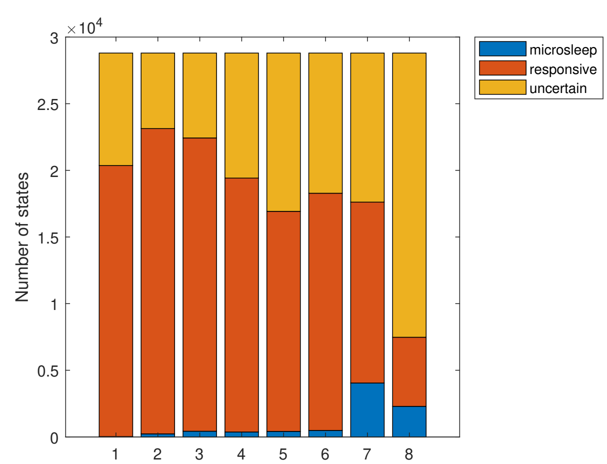

Fig. 3: Microsleep, responsive, and uncertain states in the 8 subjects ordered with respect to decrease in imbalance ratio.

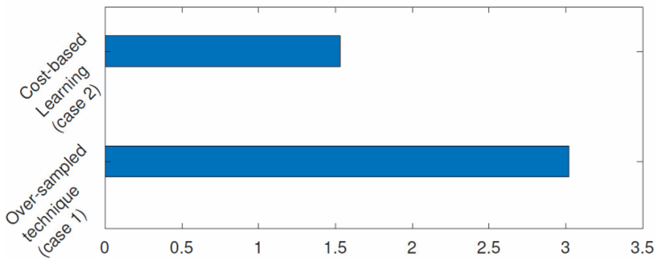

Fig. 4: Average training time for over-sampling technique Vs cost-based learning.

indicates that whether over-sampling the minority class to equal the majority class or by adding weights to the error function proportional to the ratio of minority class to majority class will yield almost the same performance. Thus in situations where computing power and is critical, the cost-based learning is advantageous.

Due to a memory constraint, when the minority class was over-sampled, the training data exceeded the GPU memory requirement. The training data had to be divided into 2 batches before the training process. Each training session was performed for 50 epochs on an NVIDIA GeForce GTX1070 GPU. The training process had to be repeated 2 times to ensure the model had seen all possible combinations at the end of training. This is time consuming as opposed to training the model with an imbalanced dataset. In cost-based learning, where the training dataset is imbalanced, it is accounted for by adding appropriate weights during the prediction error computation. This reduces the amount of data to be handled during training and, thus, reduces the time needed for training (Fig. 4).

The current measures obtained are on par with some of the prior work [1],[4], but not on par with the benchmark set by [7].

V. Conclusion

The ability of CNN to detect microsleep states was investigated in this work, along with the model's performance when the minority class was over-sampled and when the imbalance was maintained during training but compensated using a cost factor while computing the prediction error. CNN being a new approach has numerous variations that needs to be explored in terms of hyperparameters, layers, and CNN structure (sequential/parallel). Future work will focus on experimenting with parallel CNN architectures, adding customized layers to reduce inter-subject variability, and using ensemble techniques to improve the performance for detecting and predicting microsleeps. In spite of a longer latency, multiple consecutive microsleep states will also be taken into consideration and experimented with to see whether it improves the performance of the model.

TABLE II: State detection performance.

	Over-sampled	**Cost-based learning**	**Imbalanced data**
Sensitivity	0.60	0.64	0.36
Specificity	0.95	0.92	0.97
Precision	0.39	0.33	0.48
phi	0.42	0.40	0.35
GM	0.73	0.76	0.54
AUC_{ROC}	0.90	0.90	0.86
AUC_{PR}	0.41	0.40	0.39

References

[1] P. R. Davidson, R. D. Jones, and M. T. R. Peiris, "EEG-based lapse detection with high temporal resolution," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 832–839, May 2007.
[2] C. R. H. Innes, G. R. Poudel, T. L. Signal, and R. D. Jones, "Behavioural microsleeps in normally-rested people," in *Ann. Int. Conf. of the IEEE Engg. in Med. and Bio.*, Aug 2010, pp. 4448–4451.
[3] R. D. Jones, G. R. Poudel, C. R. H. Innes, P. R. Davidson, M. T. R. Peiris, A. M. Malla, T. L. Signal, G. J. Carroll, R. Watts, and P. J. Bones, "Lapses of responsiveness: Characteristics, detection, and underlying mechanisms," in *32nd Ann. Int. Conf. of the IEEE Engg. in Med. and Bio.*, Aug 2010, pp. 1788–1791.
[4] M. T. R. Peiris, P. R. Davidson, P. J. Bones, and R. D. Jones, "Detection of lapses in responsiveness from the EEG," *J. Neural Eng.*, vol. 8, no. 1, p. 016003, jan 2011.
[5] S. Ayyagari, "Reservoir computing approaches to EEG-based detection of microsleeps." Ph.D. dissertation, University of Canterbury, Christchurch, New Zealand, 2017.
[6] M. Shoorangiz, "Prediction of microsleeps from EEG using bayesian approaches." Ph.D. dissertation, University of Canterbury, Christchurch, New Zealand, 2018.
[7] A. B. Buriro, R. Shoorangiz, S. J. Weddell, and R. D. Jones, "Predicting microsleep states using EEG inter-channel relationships," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 12, pp. 2260–2269, Dec 2018.
[8] R. Schirrmeister, J. Tobias Springenberg, L. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization: Convolutional neural networks in EEG analysis," *Hum. Brain Mapp.*, vol. 38, pp. 5391–5420, 11 2017.
[9] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks." *CoRR*, vol. abs/1511.06448, 2015.
[10] A. Supratak, H. Dong, C. Wu, and Y. Guo, "Deepsleepnet: a model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. PP, 03 2017.
[11] M. T. R. Peiris, R. D. Jones, P. R. Davidson, and P. J. Bones, "Detecting behavioral microsleeps from EEG power spectra," in *Int. Conf. of the IEEE Engg. in Med. and Bio. Soc.*, Aug 2006, pp. 5723–5726.
[12] R. Shoorangiz, S. J. Weddell, and R. D. Jones, "Prediction of microsleeps from EEG: Preliminary results," in *38th Ann. Int. Conf. of the IEEE Engg. in Med. and Bio. Soc. (EMBC)*, Aug 2016, pp. 4650–4653.
[13] M. T R Peiris, R. D. Jones, P. R. Davidson, G. J. Carroll, and P. J. Bones, "Frequent lapses of responsiveness during an extended visuomotor tracking task in non-sleep-deprived subjects," *J Sleep Res*, vol. 15, pp. 291–300, 10 2006.
[14] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Editorial: Special issue on learning from imbalanced data sets," *SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 1–6, Jun. 2004. [Online]. Available: http://doi.acm.org/10.1145/1007730.1007733
[15] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 1, pp. 63–77, Jan 2006.
[16] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, "Cost-sensitive learning of deep feature representations from imbalanced data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3573–3587, Aug 2018.