A Novel Convolutional Neural Network for Emotion Recognition Using Neurophysiological Signals

Marc Tunnell, Huijin Chung, and Yuchou Chang, *Member, IEEE*

***Abstract*—Non-invasive brain-computer interfaces (BCIs) provide us with the unique ability to classify the psychological state of a person using only neurophysiological signals, such as those captured with an electroencephalogram (EEG). With this ability, new avenues for innovation in the field of healthcare arise, especially as it is used for robotics. EEGNet is a novel deep learning technique for the classification of EEG data with a limited training set that generalizes well to a variety of BCI paradigms, and the performance thereof can further be improved. We propose the use of Thomson Multitaper Power Spectral Density estimation in the EEG-BCI classification pipeline as well as a novel convolutional neural network (CNN), which extends EEGNet with sparse feature maps produced by efficient regularized separable convolutions. Further, we test the efficacy of interspersed Gaussian noise as a data augmentation technique. To show the improvements found with this new pipeline, we test on a widely used public EEG dataset related to emotion classification, then perform an ablation study to determine the most contributing factors. The accuracy on this public dataset was 77.16%. These results show that our pipeline improved the classification accuracy by 10.86% when compared with the state-of-the-art.**

I. Introduction

Emotion is a profound part of the human experience, dictating our perception of the world and interactions with those around us. The use of non-invasive Brain-Computer Interfaces (BCIs), which provide a link between the human and computer through electroencephalogram (EEG) signals [1], allow for classification of emotion without regard to the physical characteristics associated with it. This may allow for healthcare professionals to better understand the needs of patients living with severely debilitating disease [2] such as locked-in syndrome, where traditional methods of emotion classification are inadequate and largely inaccurate [3].

For people living with motor-related disabilities, assistive technologies in the form of robotics are increasing yearly [4]. Despite this rapid growth, the implementation of human emotion recognition in robots has not progressed as quickly [5, 6]. This may be due to higher accuracy requirements, ensuring correct responses to stimulus [5], conversely causing undue stress to the user when incorrectly applied [7]. Even simple movements performed by the robot may cause error through the addition of artefacts or noise [5], further complicating accurate classification. Affective computing, as applied to the field of robotics, opens novel possibilities in improving quality of life for the people who need social support. Given that the ratio of those requiring assistive care services to care givers is expected to grow steadily over the first half of this century [6], emotionally aware robotics becomes an enticing prospect.

Despite the need for advancement, the nature in which EEG data are collected leads to exorbitant costs, effectively constraining the size of our training datasets. Considering this, extremely deep and complex CNNs with many parameters to fit may be less suitable for use on the smaller datasets associated with EEG research. Reference [8] attempts to solve this with EEGNet, a compact CNN intended for the classification and interpretation of a wide range of EEG-based BCI paradigms. Furthermore, the non-stationary nature of brainwave data [9] may lead to worse accuracy by CNNs [4]. Reference [10] attempts to improve the classification accuracy by adding a pre-processed feature extraction method to their pipeline and employing a modified version of EEGNet (S-EEGNet) with an additional offset to the convolutions of the original architecture.

EEGNet is a state-of-the-art CNN with strong feature extraction that is competitive for many EEG-BCI tasks [8]. However, EEGNet does not perform as well when applied to emotion classification on the selected dataset, with similar average performance to classical feature extraction and classification pairs. The Hilbert Spectrum, derived from the Hilbert-Huang Transformation (HHT) [11], as was done in [10] produces a robust feature set and improves the accuracy, but in a memory-constrained contexts may not be as appropriate when compared to more lightweight representations of the data. The resolution of the Hilbert Spectrum is bounded by the number of frequency bins used, whereas the optimal number of frequency bins increases both with the sampling rate and length of the original data [11].

In this study, we proposed a method for improving EEGNet based on sparse feature vectors produced by regularized separable convolutions. We further improve EEGNet with the use of Thomson's multitaper method of power spectral density (PSD) estimation, and Gaussian data augmentation. This PSD estimation method was chosen particularly for the reduction in signal artefacts when compared to other common spectral estimations as was noted in [12]. We directly compare these improvements together against the state-of-the-art method, EEGNet and its automatic feature extraction. Additionally, we test the efficacy of Layer-wise Adaptive Moments optimizer for Batch training (LAMB)

Marc Tunnell is with the School of Computing and Engineering, Grand Valley State University, Allendale, MI 49401 USA (e-mail: tunnellm@mail.gvsu.edu).

Huijin Chung is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: hchung95@gatech.edu).

Yuchou Chang is with the Computer and Information Science Department, University of Massachusetts Dartmouth, North Dartmouth, MA 02747 USA (e-mail: ychang1@umassd.edu).

for use on smaller datasets. We also perform an ablation study to determine the contribution of each of these components.

This paper is organized as follows. First, the introduction is presented. Then, the related technical background is given in the second part. The proposed method is elaborated in the third section, and the experimental results are presented in the fourth part. Finally, a conclusion is given in the last section.

II. Related Work

EEGNet is a CNN with three convolutional layers for feature extraction and one fully connected, dense layer for classification [8]. The network was designed with the intent of automatic robust feature extraction that works on a wide-range of EEG-BCI tasks [8]. The number of filters and kernel lengths of the original architecture were formulated specifically for use on 128 Hz data [8]. This architecture is illustrated in Fig. 4a.

S-EEGNet modifies the EEGNet architecture with the addition of an offset convolution, which was achieved by using bilinear interpolation on the input to the convolution layers [10]. They further improved the classification accuracy of EEGNet using HHT in their pre-processing pipeline [10]. Reference [13] was able to improve the accuracy by using an averaging of an ensemble of EEGNet with different kernel size inputs. Moreover, it was found that the optimal kernel size differed based on which subject the classification was performed on [13].

The depthwise separable convolution (separable convolution) is a two-part convolution in which a depthwise convolution is first performed, and then a pointwise convolution is performed afterward. In other studies, it has been empirically shown that the separable convolution can produce more efficient feature maps, even if the network contains a similar number of parameters [14, 15]. Reference [15] applied separable convolutions to the domain of machine translation, showing the versatility of separable convolutions. Further, the offset convolution found in [10] is also based on the separable convolution.

Adam is a memory-efficient optimizer [16, 17] that is still in wide use today. LAMB is an optimizer designed to work with large batch sizes without a degradation in performance [18]. LAMB uses layerwise normalization, which has similarity to that which is found in the LARS optimizer [16, 19]. This is further combined with dimensional normalization with respect to the square root of the second moment, as is found in the Adam optimizer [16-18].

III. Proposed Method

A. Thomson Multitaper Method of Power Spectral Density Estimation

The PSD of a signal is an estimation of its distribution of power in the frequency domain [20]. In this study, we use Thomson Multitaper Method of PSD estimation (PMTM) for its consistency and resolution of the extracted features [21]. This consistency and resolution can be attributed to the averaging that the method utilizes [12, 21-23]. PMTM uses Discrete Prolate Spheroidal (Slepian) sequences, first described in [21], to decompose the signal and extract its power concentration which is then averaged over multiple tapers [21-23]. This method has been shown to reduce the bias and variance of the EEG signal when applied to short segments of data [12, 23].

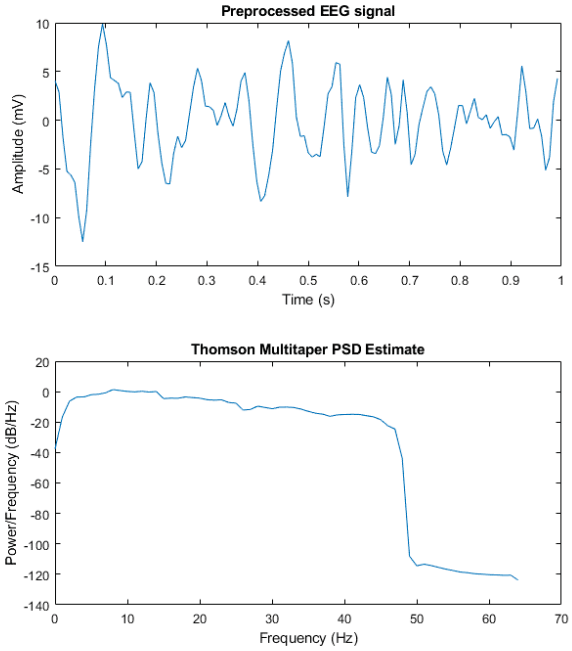


Fig. 1 Thomson's Multitaper Method of Power Spectral Density Estimation.

PMTM, using Slepian sequences, is based around the results of a series of eigenvalue expansion equations [20-22]. Per the references [21, 22], the Slepian sequence is given by equation (1):

$$\sum_{m=0}^{N-1} \frac{sin(2\pi W(n-m))}{\pi(n-m)} g_k(m) = \lambda_k g_k(n), \quad (1)$$

$$\underline{k, n = 0,1,2, \ldots N-1,}$$

where the Slepian sequence g_k, is the k[th] eigenvector corresponding to the k[th] eigenvalue λ_k, of non-increasing order, of a given frequency band spanning [-W, W] [21, 22].

As an example of PMTM, the finished extraction from one electrode, 128 points of data equating to 1 second, is shown in Fig. 1. The original preprocessed data are depicted above the feature extracted version. Note that the size of the PMTM feature extracted data is decreased from 128 points of data to 65. This algorithm is provided by [22].

B. Regularization

Regularization attempts to reduce the probability of over-fitting the training data by imposing penalties on the model during the training process. In this study, we focus on L_1 regularization, which penalizes the model by adding an extra cost that is proportional to the absolute value of each parameter [24]. In theory, the optimal solution to minimize a cost function with added penalties is to gravitate towards the smallest vector that solves the problem [25]. In this case, practice follows theory, and the addition of L_1 regularization decreases the probability of the model accumulating large parameters and in fact trends towards many parameters being equal to zero [24], producing

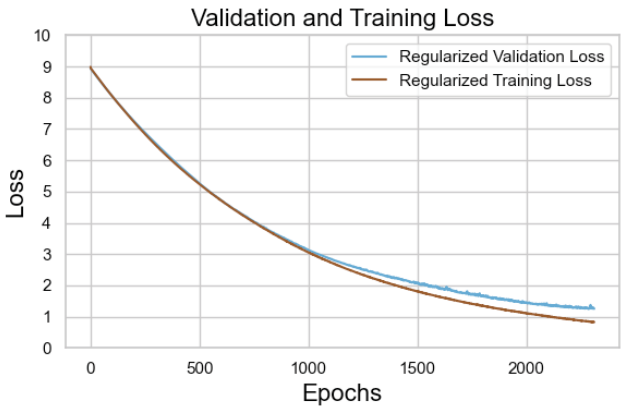


Fig. 2 The validation curve of the regularized model decreased smoothly.

sparsely populated feature maps. The equation for L_1 regularization is given in (2), as follows:

$$penalty = \lambda \sum_{i=0}^{N-1} |w_i|, \tag{2}$$

where w is the set of model parameters that the regularization is to be applied to, $w_i \in w$, $\lambda \in [0, 1)$, and N is the cardinality of w. This penalty is then added to the cost function in the loss calculation stage of training. An example of the loss curve when fitting a regularized model is shown in Fig. 2.

C. Gaussian Noise

Gaussian noise is random statistical noise with a probability density function that is equal to the Gaussian distribution [26]. For our purposes, this noise is centered with a mean of zero. The probability distribution, p of random variable $z \in \mathbb{R}$, is given by equation (3):

$$p(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-z^2}{2\sigma^2}}, \tag{3}$$

where σ is the standard deviation of the noise, and σ^2 is its variance [26]. To show that the distribution of random numbers is normal, we set $\sigma = 1$ and generated 10^7 random numbers. These numbers were plotted in a histogram, illustrated in Fig. 3, which shows that the distribution of the random variable z is highly close to a truly normal distribution.

D. Improved EEGNet Architecture

The improved EEGNet architecture, illustrated in Fig. 4b, is designed to capture only the most relevant features from the PMTM feature extracted data, leading to improvements in the generalization capabilities of the network. This is accomplished in two main ways. Through the use of highly efficient separable convolutions, and the addition of L_1 regularization that is active during training, bringing the weights of the least relevant parameters to zero.

Although our architecture is heavily based on the EEGNet architecture [8], there is considerable difference between each. As illustrated in Fig. 4, our improved architecture shares the first and third convolutional layers with EEGNet, a spatial and separable convolution. A pointwise convolution is added to the depthwise convolution that is found in EEGNet. We increase the depth of the

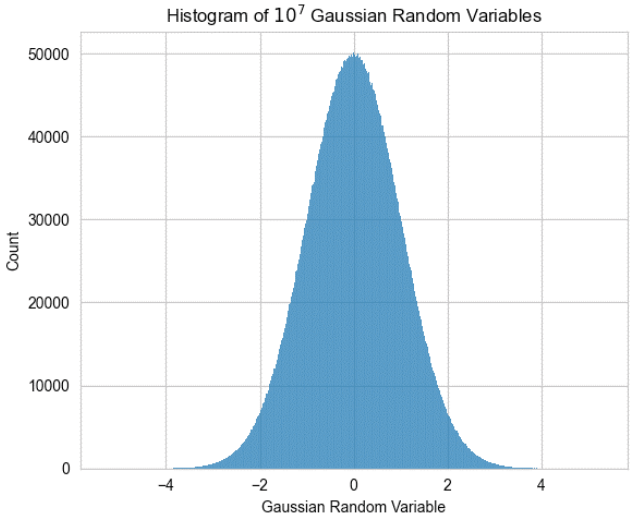
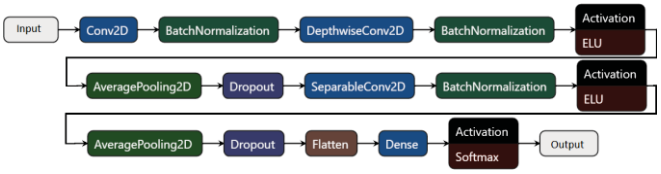


Fig. 3 The bell-shaped curve of Gaussian random variable z shows that its distribution is approximately normal.

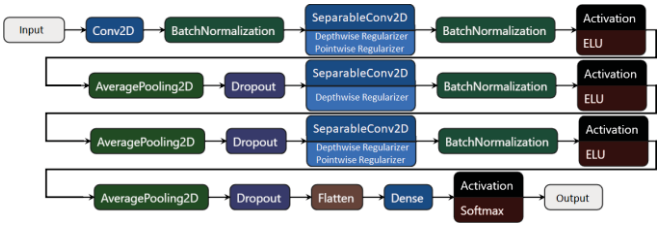
network with an additional separable convolution, batch normalization, activation, and average pooling layers to the architecture following the location of the separable convolution found in EEGNet. The batch normalization accelerates training by standardizing the inputs, whereas the average pooling layer downsamples the feature map, helping to reduce the number of parameters to fit. Further, we place L_1 regularization parameters on each of the separable convolutions. To the second separable convolution, only the depthwise convolution is regularized. To the first and third separable convolutions, both the depthwise and pointwise convolutions are regularized.

IV. Experimental Results

The model was allowed to train for 3500 epochs with early stopping based on validation loss enabled at 25 epochs unless otherwise stated explicitly. The exact methods of data handling are explicitly described in the subsequent subsection. For all testing, we set the batch size equal to the size of



(a)



(b)

Fig. 4 The unmodified EEGNet architecture depicted in (a). The improved EEGNet architecture depicted in (b).

the training set when testing with the LAMB optimizer and otherwise equal to 64 and 128 when testing with the Adam optimizer, reporting the better performing result. We used the default hyperparameters from the original paper for all testing on the 128 Hz data with EEGNet.

When testing with the original EEGNet with the non-feature extracted data, we found that the Adam optimizer was unable to converge on certain subjects. To solve this, we tested a variety of different optimizers, finding the LAMB optimizer to be most suitable. Using a subject that has poor performance with EEGNet, we plot an example of this observed abnormality in Fig. 5. The validation and training loss of the Adam optimizer is plotted over 3500 epochs without early-stopping and shows that, from the start of training, the optimizer is unable to converge towards a solution and the gradient explodes. Note that this was performed using subsampled non-feature extracted data with the original EEGNet.

We will present our results in a manner similar to what was done in [10]. To evaluate the efficacy of our model, we use accuracy, defined by the equation (4):

$$accuracy = \frac{TP+TN}{TP+TN+FP+TN} \times 100\% \,, \qquad (4)$$

where *TP* is true positive, *TN* is true negative, *FN* is false negative, and *FP* is false positive.

A. A dataset for emotion analysis using eeg signals

The dataset is A Database for Emotion Analysis Using Physiological Signals (DEAP) [27]. This dataset contains 32 volunteers of 40 trials each, with each trial consisting of EEG and other physiological signals recorded over the duration of a 60 second music video extract [27]. For 22 subjects, face recordings were also available [27]. After viewing each extract, the participants rated their levels of arousal, valence, and dominance using a 9-point scale [27]. These data were originally recorded with a sampling rate of 512 Hz, with 32 EEG channels and 12 peripheral channels [27]. Participants 1-22 were recorded in Twente, and the remaining in Geneva–the authors note minor differences in the hardware used for each [27].

The dataset provides a pre-processed version, which was used. This pre-processed version rearranges the Twente EEG channels to coincide with the ordering used in Geneva, and the units used to measure were converted as well. The data were then downsampled to 128 Hz, bandpass filtered from 4.00 - 45.00 Hz, and segmented into 60 second trials with a 3 second pre-trial baseline. Our team discarded all of the channels not corresponding to EEG signals, then removed the first 3 seconds, corresponding to 384 points of data.

Prior to testing, we produced three sets of data. The first set was created by subsampling the data at 1 second intervals. The remaining two sets were created by taking the PSD estimation of the subsampled set, the first using Welch's method of PSD estimation (PWELCH), and the second used PMTM. Each set was then randomized with respect to the length of the original 40 trials.

From the subject rating data, we created a set of binary labels for the BCI classification task of emotion recognition. We threshold the valence data at greater than or equal to 5

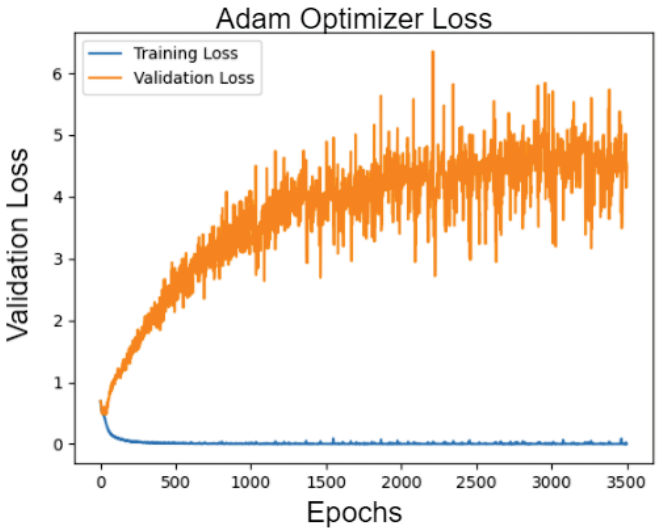


Fig. 5 The validation loss of EEGNet, using the Adam optimizer and the original pre-processed as applied to a non-performant subject.

as high valence and less than or equal to 5 as low valence. High valence represents a state of pleasant emotion whereas low valence represents a state of unpleasant emotion [28]. Each subsample was then assigned a binary label corresponding to the value of its original time series.

We then further threshold the subjects into two categories, skewed and balanced. This is done by first computing the ratio of low valence to high valence for each subject. If this ratio was greater than or equal to 60%, or less than or equal to 40%, we placed the subject in the skewed category, otherwise we placed the subject in the balanced category. This is done so that we may effectively use accuracy as the metric of efficacy such that our results on the balanced data are not positively affected by the skewed nature of the data. After applying this threshold, 65.63% of subjects were non-skewed and the remaining skewed.

We find that at approximately this threshold, we may effectively separate only the subjects that positively affected the results in a manner not relating to the efficacy of the model. In the Classification Results section, we report the results of both the skewed and balanced sets, empirically showing the erroneous advantage produced by the skewed set of subjects.

C. Data Augmentation

Two matrices of Gaussian noise were generated. The first matrix was of equal dimension to the training set and the second was of equal dimension to the validation set. To these generated matrices, the training and validation sets were added, respectively. Additional sets were allowed to be made as an iterative process. No noise was added to the test set. The augmented data were then concatenated to their respective sets prior to training. The respective sets were then shuffled independently to evenly intersperse the augmented data by stacking every 60th subsample in a respective new matrix, repeating until all subsamples and augmented data were contained in the respective new matrices.

Table I. Accuracy on skewed data, comparing different methodologies.

Classifier	Skewed Accuracy
EEGNet	81.18%
Improved EEGNet	79.06%
PMTM + SVM	86.20%

D. *Classification Results*

For the purpose of the binary BCI task of emotion classification on the DEAP dataset, we compare our method against the state-of-the-art EEGNet with its automatic feature extraction, using LAMB as the optimizer for both. This study makes use of classical feature extraction and classifier pairs that are still in wide use today, namely: PSD and SVM, and PSD and LDA. For the PSD estimation, we tested these pairs with PMTM and PWELCH, reporting the better performing pair.

We split the data into four even folds using four-fold cross validation. Each fold was created with respect to the length of the original trial, such that no two folds contain a subsample from the same trial. For each iteration of the four-fold testing on CNNs, 80% of the fold in question was used as the testing set. Of the remaining data, 81.25% were used as training set, and 18.75% as validation set. All sets were created with respect to the length of the original trial. For the testing of the classical pairs, each iteration of the four-fold testing uses the same 80% of the given fold as the testing set, and the remaining data as the training set. We first show Table I, in which the results of testing on the skewed data are displayed. Note that the highest skewed classification accuracies of 81.18%, 79.06%, and 86.20% were performed on data skewed to 75.00%.

For the sake of brevity, the remainder of testing shown is on the balanced data.The results for the binary classification task are shown in Fig. 6 and show that the highest accuracy on the balanced data for our improved EEGNet is 77.16%, up from 66.30% and 67.32% with EEGNet using the original data and PMTM respectively. These results represent an uplift in accuracy over EEGNet of 10.48% by the improved EEGNet pipeline.

Using the LAMB optimizer, we then perform an ablation study to determine empirically which components of our method contribute the most to our improved performance. The checkmark means that the item in the header is present in the testing and the "X" mark means that it is not. For example, row 4 of Table II means that when using Improved EEGNet with PMTM, No Regularization, No Gaussian Noise, the accuracy is 63.05%. The results of this are shown in Table II and show that the new architecture benefits greatly from the use of both PMTM and regularization and has an additional minor uplift from the interspersed Gaussian augmentation.

Table II. Improved EEGNet results from the ablation study, showing which components provide the most uplift in accuracy, performed on the DEAP dataset.

PMTM	Regularization	Gaussian	Accuracy
✗	✗	✗	48.29%
✗	✓	✗	45.96%
✗	✗	✓	46.23%
✓	✗	✗	63.05%
✓	✗	✓	67.54%
✓	✓	✗	75.65%
✓	✓	✓	**77.16%**

V. Conclusion

This study proposed the use of PMTM in the BCI-EEG emotion classification pipeline, as well as a novel CNN architecture based on EEGNet. The results of our testing show that implementing PMTM as a feature extraction method may have large improvements, as was seen in the high accuracy of the SVM + PMTM combination as well as with the improved EEGNet. As shown in the marginal gain with the original EEGNet, some changes to an architecture that is designed to work on data in the time domain may be required in order to receive the most effect from using PMTM.

As shown in the ablation study in the previous section, the improved EEGNet architecture would appear to have difficulty working with data that is not pre-processed and feature extracted using PMTM. This is likely due to differences in the number of filters learned between EEGNet and the improved EEGNet. This shows that the new model has some limitations that should be investigated in future work.

Although all of the classifiers shown were able to receive a high accuracy on the skewed dataset, it should be noted that in some instances, the skew was extreme and the results received on this data should not be taken as a satisfactory representation of the performance of either model. In order to

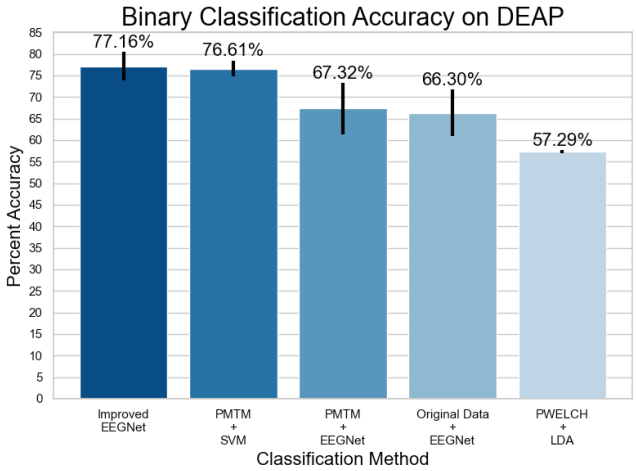


Fig. 6 The highest accuracy of 77.16% and 76.61% is achieved with classifiers augmented with PMTM as the feature extraction method. Error bars represent a 95% confidence interval.

emphasize on the skewed nature of this data, observe that on one such subject, the distribution of labels was skewed to 75.00% and it was not uncommon for the test set to be skewed upwards of 87.50% when producing the randomized folds. With early-stopping enabled at 25 epochs, EEGNet often stopped training on this subject at below 50 epochs, and with a training accuracy of less than 60.00% in the majority of cases. This would indicate that the model is not actually performing well, but rather obtaining a high accuracy due to highly imbalanced data.

The direction of our future work is largely dictated by what is previously discussed in this section. Additionally, we look to continue our research by implementing our pipeline on the arousal data in the DEAP dataset as well as a larger range of BCI-task paradigms, such as on an Error-Related-Negativity and Steady state visually evoked potential datasets. We will also continue our research on improving the efficacy and robustness of the model for interacting with emotional robot [29, 30] with other advanced optimization algorithms [31, 32].

ACKNOWLEDGMENT

This work was supported by the National Science Foundation under Grant No. 2050972.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[2] H. Huang, Q. Xie, J. Pan, Y. He, Z. Wen, R. Yu, and Y. Li, "An eeg-based brain computer interface for emotion recognition and its application in patients with disorder of consciousness," *IEEE Transactions on AffectiveComputing*, pp. 1–1, 2019.

[3] N. L. Childs, W. N. Mercer, and H. W. Childs, "Accuracy of diagnosis of persistent vegetative state," Neurology, vol. 43, no. 8, pp. 1465–1465, 1993.

[4] M. Aljalal, S. Ibrahim, R. Djemal, and W. Ko, "Comprehensive review on brain-controlled mobile robots and robotic arms based on electroencephalography signals," Intelligent Service Robotics, vol. 13, no. 4, pp. 539–563, Oct 2020.

[5] M. Spezialetti, G. Placidi, and S. Rossi, "Emotion recognition for human-robot interaction: Recent advances and future perspectives," Frontiers in robotics and AI, vol. 7, pp. 532 279–532 279, Dec 2020, 33501307[pmid].

[6] R. Bemelmans, G. J. Gelderblom, P. Jonker, and L. de Witte, "Socially assistive robots in elderly care: A systematic review into effects and effectiveness," Journal of the American Medical Directors Association, vol. 13, no. 2, pp. 114–120.e1, 2012.

[7] M. Alimardani and K. Hiraki, "Passive brain-computer interfaces for enhanced human-robot interaction," Frontiers in robotics and AI, vol. 7, pp. 125–125, Oct 2020, 33501291[pmid].

[8] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," Journal of Neural Engineering, vol. 15, no. 5, p. 056013, 2018.

[9] A. Y. Kaplan, A. A. Fingelkurts, A. A. Fingelkurts, S. V. Borisov, and B. S. Darkhovsky, "Nonstationary nature of the brain activity as revealed by eeg/meg: Methodological, practical and conceptual challenges," Signal Processing, vol. 85, no. 11, pp. 2190–2212, 2005, neuronal Coordination in the Brain: A Signal Processing Perspective.

[10] W. Huang, Y. Xue, L. Hu, and H. Liuli, "S-eegnet: Electroencephalo- gram signal classification based on a separable convolution neural network with bilinear interpolation," *IEEE Access*, vol. 8, pp. 131 636– 131 646, 2020.

[11] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empiricalmode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Societyof London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[12] A. Delorme, T. Sejnowski, en S. Makeig, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis", *Neuroimage*, vol 34, no 4, bll 1443–1449, Feb 2007.

[13] Y. Zhu, Y. Li, J. Lu, and P. Li, "Eegnet with ensemble learning to improve the cross-session classification of ssvep based bci from ear-eeg," IEEE Access, vol. 9, pp. 15 295–15 303, 2021.

[14] F. Chollet, "Xception: Deep learning with depthwise separable convolu- tions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[15] L. Kaiser, A. N. Gomez, and F. Chollet, "Depthwise separable convo- lutions for neural machine translation," 2017.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[17] S. Ruder, "An overview of gradient descent optimization algorithms," CoRR, vol. abs/1609.04747, 2016.

[18] Y. You, J. Li, S. Reddi, J. Hseu, S. Kumar, S. Bhojanapalli, X. Song, J. Demmel, K. Keutzer, and C.-J. Hsieh, "Large batch optimization for deep learning: Training bert in 76 minutes," 2020.

[19] Y. You, I. Gitman, and B. Ginsburg, "Scaling SGD batch size to 32k for imagenet training," CoRR, vol. abs/1708.03888, 2017.

[20] D. Thomson, "Spectrum estimation and harmonic analysis," Proceedings of the IEEE, vol. 70, no. 9, pp. 1055–1096, 1982.

[21] D. Slepian, "Prolate spheroidal wave functions, fourier analysis, and uncertainty — v: the discrete case," The Bell System Technical Journal, vol. 57, no. 5, pp. 1371–1430, 1978.

[22] MATLAB, Signal Processing Toolbox version 9.10.0 (R2021a). Natick, Massachusetts: The MathWorks Inc., 2021.

[23] M. J. Prerau, R. E. Brown, M. T. Bianchi, J. M. Ellenbogen, en P. L. Purdon, "Sleep Neurophysiological Dynamics Through the Lens of Multitaper Spectral Analysis", Physiology, vol 32, no 1, bll 60–92, 2017.

[24] A. Y. Ng, "Feature selection, L1 vs. L2 regularization, and rotational invariance," in Proceedings of the Twenty-First International Conference on Machine Learning, ser. ICML '04. New York, NY, USA: Association for Computing Machinery, 2004, p. 78.

[25] A. Krogh and J. Hertz, "A simple weight decay can improve generalization," in Advances in Neural Information Processing Systems, J. Moody, S. Hanson, and R. P. Lippmann, Eds., vol. 4. Morgan-Kaufmann, 1992.

[26] A. Lyon, "Why are normal distributions normal?" The British Journal for the Philosophy of Science, vol. 65, no. 3, pp. 621–649, 2014. [Online]. Available: https://doi.org/10.1093/bjps/axs046

[27] S. Koelstra, C. Muehl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," Affective Computing, IEEE Transactions on, vol. 3, no. 1, pp. 18–31, 2012.

[28] V. Shuman, D. Sander, en K. Scherer, "Levels of Valence", Frontiers in Psychology, vol 4, 2013.

[29] Y. Chang, "Architecture design for performing grasp-and-lift tasks in brain–machine-interface-based human-in-the-loop robotic system," IET Cyber-Physical Systems: Theory & Applications, vol. 4, no. 3, pp. 198-203, 2019.

[30] Y. Chang and L. Sun, "EEG-based emotion recognition for modulating social-aware robot navigation," the 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021.

[31] H. Wang, J. Cheng, Sen Jia, Zhilang Qiu, Caiyun Shi, Lixian Zou, S. Su, Y. Chang, Y. Zhu, L. Ying, and D. Liang, "Accelerating MR imaging via deep Chambolle-Pock network," the 41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2019.

[32] C. Shi, J. Cheng, G. Xie, S. Su, Y. Chang, H. Chen, X. Liu, H. Wang, and D. Liang, "Positive-contrast susceptibility imaging based on first-order primal-dual optimization," Magnetic Resonance in Medicine, vol. 82, no. 3, pp. 1120-1128, 2019.