2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Investigation on Robustness of EEG-based Brain-Computer Interfaces

Aarthy Nagarajan, Neethu Robinson, Cuntai Guan

Abstract— **Electroencephalogram (EEG)-based brain-computer interface (BCI) systems tend to suffer from performance degradation due to the presence of noise and artifacts in EEG data. This study is aimed at systematically investigating the robustness of state-of-the-art machine learning and deep learning based EEG-BCI models for motor imagery classification against simulated channel-specific noise in EEG data, at various low values of signal-to-noise ratio (SNR). Our results illustrate higher robustness of deep learning based MI classification models compared to the traditional machine learning based model, while identifying a set of channels with large sensitivity to simulated channel-specific noise. The EEGNet is relatively more robust towards channel-specific noise than Shallow ConvNet and FBCSP. We propose a preliminary solution, based on activation function, to improve the robustness of the deep learning models. By using saturating nonlinearities, the percentage drop in classification accuracy for SNR of -18 dB had reduced from 10.99% to 6.53% for EEGNet and 14.05% to 3.57% for Shallow ConvNet. Through this study, we emphasize the need for a more precise solution for enhancing the robustness, and thereby usability of EEG-BCI systems.**

I. INTRODUCTION

Brain-computer interface (BCI) is a special communication protocol being used in a variety of application domains ranging from entertainment [1] to health [2]. A typical BCI procedure involves collection of brain data, which is used to decode user's intent and then translate it to an action command to be executed by the connected external device [3]. There are various methods of collecting data from the brain, out of which non-invasive EEG is predominantly used due to its high temporal resolution, ease of use and cost-effectiveness. Nevertheless, EEG data is high-dimensional, nonstationary and is highly susceptible to artifacts and noise. In addition, hardware related issues such as channel disconnections and displacements can introduce external noise factors in the signal, leading to poor performance of EEG-BCI systems. Channel disconnections, in particular, are quite common yet hard to detect during the experimental study [4]. The resulting noise in the signal is, therefore, identified only after the data has been collected. As re-recording EEG is not always viable, it is essential for EEG-BCI classification systems to remain extremely robust to unexpected data perturbations that may occur during data collection. Amongst the different EEG-based BCI paradigms, motor imagery (MI) [5] is the most widely researched, due to its connection with important clinical applications used for communication [6] and rehabilitation purposes [7]. EEG-BCI classification methods for MI have predominantly been using machine learning based algorithms, such as the common spatial patterns (CSP) [8] and the filter-bank CSP (FBCSP) [9]. Due to the recent progress of deep learning [10], several neural network based EEG-BCI classification models have been introduced and are reported to be performing better than the machine learning counterparts [11]. In particular, the Deep ConvNet & Shallow ConvNet [12], and the EEGNet [13] are considered as the benchmark networks for MI classification. Nevertheless, the sensitivity of deep learning models to data perturbations is a well-known and widely researched issue in computer vision [14]. As our dependency on deep learning methods for building EEG-based BCIs increase, the need of the hour is to ensure robustness in these methods for demonstration of stable classification performance even in the presence of unwanted noise in EEG data.

A. Related Work

Few recent studies have discussed the robustness of machine learning and deep learning based EEG-BCI models to perturbations in data. In [15], Zhang et al. investigated the robustness of EEG-BCI models developed using convolutional neural networks (CNN), such as EEGNet [13], Deep ConvNet and Shallow ConvNet [12], to adversarial attacks. The idea exposed the susceptibility of CNN models to hard-to-detect adversarial noise, and their results illustrated that even small amounts of noise in data can be detrimental to the classification performance of these models. In [16], Nakagome et al. performed a comparative analysis of eight different machine learning and neural network based algorithms under different conditions of data pre-processing such as downsampling, tap size, and usage of various frequency bands. In addition, they also conducted a channel perturbation analysis wherein they randomly perturbed the channel data one-by-one, and used the performance drop to identify the channel importance and model robustness. These studies provide evidence on the vulnerability of machine learning and deep learning based EEG-BCI classification methods to noisy data. They also highlight the need for a more systematic investigation of network sensitivity to channel-specific noise in EEG, which is the purpose of our study. Such an investigation will pave the way for finding solutions to improve the robustness of EEG-BCIs so as to enhance their usability in real-world applications.

B. Problem Statement

The objective of this study is to examine the response of EEG-BCI models to channel-specific noise, thereby, evaluate the relation between the performance of these models and negative signal-to-noise ratio (SNR) of noise. To accomplish this, we introduce three scenarios of channel disconnection:

Aarthy Nagarajan, Neethu Robinson, Cuntai Guan are with the Nanyang Technological University, 50 Nanyang Avenue, Singapore (aarthy001@e.ntu.edu.sg, nrobinson@ntu.edu.sg, ctguan@ntu.edu.sg) Corresponding author: Cuntai Guan

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

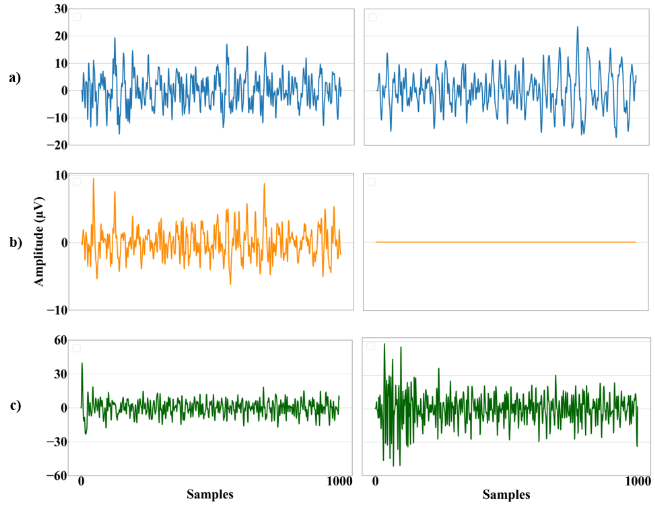

Fig. 1: Initial and concluding trials of a a) good channel b) disconnected channel and c) noisy channel from real EEG data.

1) there is no signal flow to the electrode, 2) the electrode is detecting background noise rather than EEG, 3) the SNR of background noise deteriorates gradually. To demonstrate these scenarios with examples, we selected some channels from a real EEG dataset, whose data illustrate either a sudden loss of signal or noise with increasing variance over time. The data was bandpass filtered in the frequency range of 4-40 Hz and the channel-wise percentage change in variance across trials was estimated. The channel whose change in variance over time was within $\pm50\%$ was chosen to represent the good channel. The channel exhibiting a sudden loss of signal in the final trials was chosen to be the disconnected channel. The noisy channel is the one showing an increasing trend in variance over time. Fig. 1 contains the plots of sample initial and concluding trials of the good channel, the disconnected channel, and the noisy channel from the EEG data.

Motivated by the observations of our analysis and to further explore how the aforementioned real scenarios impact the performance of the MI classification models, we design experiments in which each channel in the EEG data is replaced by a constant 0V signal, a Gaussian noise signal with SNR of 0 dB, and Gaussian noise signal with SNR further reduced to -6 dB, -12 dB, -16 dB, and -18 dB, to represent each of the three scenarios of channel disconnection, and the models are evaluated using the simulated noisy data. This experimental setup of replacing channel-wise EEG data with noise is unique in comparison with the studies cited in the related work section, which have performed analysis using additive noises or random perturbations in EEG data.

As an outcome of this robustness analysis, we quantify the accuracy deviation of the models for each scenario of channel disconnection and also highlight the channel-wise impact on performance. Using the analysis results, we propose a preliminary solution to improve the robustness of the deep learning models and evaluate the proposed approach. We conclude our analysis by indicating the need to explore further on the varied robustness behavior of the different models and to bring forth a more precise solution to boost their robustness. We believe that this is the first study to perform a methodical channel-wise robustness analysis of EEG-BCI systems.

II. Experimental Setup

To investigate and evaluate the robustness performance of the EEG-based BCI models, we conducted a simulated experimental study by exposing three different EEG-BCI models that operate using the MI paradigm, to artificially introduced channel-specific noise. The following subsections describe in detail about the composition of our experiment and our inquiry into the results obtained thereby.

A. *Classification of MI-EEG*

For the robustness analysis, we focus mainly on subject-specific models, as they are simpler than the subject-independent models in terms of not considering the EEG variabilities between individuals. We have evaluated the robustness performance of FBCSP [9], which is the machine-learning based benchmark algorithm for MI classification, along with other latest deep learning models such as Shallow ConvNet [12] and EEGNet [13]. Out of the two models, Deep ConvNet and Shallow ConvNet introduced by Schirrmeister et al. [12], we chose to work with Shallow ConvNet for our analysis as it is a simpler model with fewer parameters.

B. *Data*

We performed the analysis using two-class (left and right hand) MI data from the Korea University EEG dataset [17] that contains MI-EEG data collected from 54 healthy people. For every subject, data was obtained from two sessions collected using 62-channel EEG at 1000 Hz sampling frequency. The data consists of 200 MI trials from each session, of which 100 trials belong to each class. We used 0-4 s post-cue data from 20 channels in the motor region (FC-5/3/1/2/4/6, C-5/3/1/z/2/4/6, and CP-5/3/1/z/2/4/6) and down-sampled it by four for the analysis.

C. *Experiment using Simulated Channel-Specific Noise*

1) Model Training: We first trained all the models - FBCSP, EEGNet and Shallow ConvNet, using hold-out analysis. The hold-out analysis was performed using session 1 of the dataset for training and session 2 for testing. The trained subject-specific model parameters were saved for the robustness analysis to be performed later.

2) Noise Generation: To perform a simulated experiment that will mimic a practical channel disconnection scenario, we perturbed the EEG data, channel by channel, by using two types of noise. As discussed in the introduction section, a channel disconnection can affect the data in two ways. It can either obstruct the flow of signal, in which case the EEG data has zero amplitude or can cause background noise. For no signal condition, we simply replaced the channel data with zeros (zero channel) and to mimic background noise we substituted the channel data with a Gaussian noise distribution (Gaussian channel) generated using subject-specific mean and standard deviation obtained from EEG data. It is to be noted that the simulated noise used in this study does not resemble the noise found in real EEG signals. Nevertheless, the simulated Gaussian noise, generated using the characteristics of real EEG, is used for contaminating channel-wise EEG data for the purpose of analysis.

Algorithm 1: Channel-Wise Robustness Analysis

```
Input: EEG data and pre-trained subject-specific
       model parameters
Output: Avg. subject-specific accuracy of the model
        for zero channel and Gaussian channel
foreach subject s do
    load data $x_s \in \mathbb{R}^{N \times C \times T^*}$;
    load model $m_s$;
    $acc_{zc} = []$;
    $acc_{gc} = []$;
    foreach channel $c \in C$ do
        $x'_s = x_s$;
        $x'_s(c) = 0$;
        $acc_{zc} \overset{+}{\leftarrow} m_s(x'_s)$;
        generate 100 samples of $X \sim \mathcal{N}(\mu, \sigma)$,
                 where $\mu = mean(x_s), \sigma = std(x_s)$;
        $acc_{gc\_100} = []$;
        foreach sample $X_n \in X$ do
            $x''_s = x_s$;
            $x''_s(c) = X_n$;
            $acc_{gc\_100} \overset{+}{\leftarrow} m_s(x''_s)$;
        end
        $acc_{gc} \overset{+}{\leftarrow} mean(acc_{gc\_100})$;
    end
    $acc_{zero\_channel} \overset{+}{\leftarrow} mean(acc_{zc})$;
    $acc_{gaussian\_channel} \overset{+}{\leftarrow} mean(acc_{gc})$;
end
Avg. subject-specific classification accuracy for,
1) $Zero\ channel = mean(acc_{zero\_channel})$
2) $Gaussian\ channel = mean(acc_{gaussian\_channel})$
```

*N = No. of trials, C = No. of channels, T = No. of time samples

3) Model Evaluation with Noisy Data: For every model, we performed a subject-specific analysis by simulating channel-specific noise and then evaluating the model performance on the resulting noisy data. The proposed experimental procedure is summarized in Algorithm 1.

D. Further Analysis

1) Channel-Wise Sensitivity: In order to identify the most sensitive channels for all models, we evaluated the significance of accuracy deviation per channel for all subjects and for all models, using the permutation test. The permutation test was performed for each subject using the set of accuracies obtained by applying 100 different Gaussian noise samples per channel.

2) Evaluation with Decreasing SNR in the Gaussian Channel: As the SNR of background noise may gradually worsen over time as indicated in Fig. 1, we further repeated the experiment by decreasing the SNR of the generated Gaussian noise to -6 dB, -12 dB, -16 dB, and -18 dB. As mentioned earlier, the baseline Gaussian noise is generated using subject specific mean and standard deviation derived from the original signal. The SNR with reference to the baseline noise is -6 dB when the amplitude of noise is increased by a factor of 2. Similarly, the SNR is -12 dB, -16 dB, and -18 dB, when the noise amplitude is increased by factors of 4, 6, and 8, respectively. The model accuracies obtained for each value of SNR and for every channel were recorded for further analysis.

3) Changing the Activation Function to Improve Robustness: The ensuing challenge is to robustify the MI classification models such that they are better able to tolerate noisy signal with negative SNR. A core component of a deep learning model is its activation function, which decides the activation of neurons and controls the stability of the network. The EEGNet uses the Exponential Linear Unit (ELU) activation function [18], and the Shallow ConvNet uses two nonlinearities in the architecture, a squaring nonlinearity followed by a logarithmic nonlinearity that together mimic the log-variance computation.

The ELU activation, which is similar to the Rectified Linear Unit (RELU) activation [19] except for a more gradual saturation on the negative part, has achieved good results in deep learning architectures pertaining to computer vision [18]. In spite of its success, ELU and the family of related activation functions, such as RELU and Leaky RELU [20], follow a linear function on the positive part. Given this characteristic, these functions can produce large, unstable activations when exposed to high amplitude EEG data. Similarly, the squaring nonlinearity used by the Shallow ConvNet model can also lead to large activations as it doubles the input values.

Hence, we propose to replace the activation functions used by the EEGNet and Shallow ConvNet models with a function that has larger stability to high amplitude input while maintaining the baseline. The Sigmoid [21] and the Tanh [22] are saturating activation functions that may help to maintain the stability of the network when exposed to high amplitude EEG data. To verify the effectiveness of using saturating nonlinearities as a solution to the robustness issue, we repeated the analysis of the two deep learning networks after replacing their existing activation functions with Sigmoid and Tanh activation functions.

III. Results

The robustness analysis results of FBCSP, EEGNet, and Shallow ConvNet models indicate an overall performance drop with channel-wise absence of signal (zero channel) and presence of Gaussian noise (Gaussian channel). In addition, the performance continues to degrade as the SNR of the Gaussian channel is reduced, thereby illustrating the effect of decreasing SNR on accuracy. Statistical analysis of the obtained results points out the most sensitive channels for every model. Usage of saturating activation functions seems to boost the robustness of the deep learning models considerably, yet it does not completely eradicate their sensitivity to noise. The following sections describe in detail about our analysis results.

A. Baseline Results vs Analysis Results with Simulated Channel-Specific Noise

The baseline accuracy of all models obtained from hold-out analysis is presented in column 2 of Table I. The baseline accuracy is 61.20% for FBCSP, 63.54% for EEGNet, and 60.44% for Shallow ConvNet.

From columns 3 and 4 of Table I, we can observe the performance drop in all models with channel-specific no signal (zero channel) as well as Gaussian noise (Gaussian channel). EEGNet is the most robust to channel-specific noise amongst the three models considered. EEGNet shows a 2.52% relative drop (63.54% vs 61.94%) in accuracy with zero channel and 2.79% relative drop in accuracy (63.54% vs 61.77%) with Gaussian channel, both of which are not significant. The performance drop of EEGNet with Gaussian channel is the lowest of all models considered. While EEGNet is the most robust model in the presence of Gaussian noise, Shallow ConvNet is the most robust to the absence of signal, for which the model shows a performance drop of 1.47% (60.44% vs 59.55%) which is the lowest amongst all models and is not significant. Shallow ConvNet also suffers an accuracy drop of 4.22% (60.44% vs 57.89%) with Gaussian channel, which is not significant as well. The accuracy of FBCSP declines by 5.54% (61.20% vs 57.81%) which is not significant and 8.37% (61.20% vs 56.08%, $p < 0.05$) which is significant, with zero channel and Gaussian channel, respectively. The results of analysis with Gaussian channel illustrate that the deep learning models EEGNet and Shallow ConvNet are relatively more robust to noise than FBCSP. Fig. 2 compares the subject-specific classification performance of the different models under simulated channel-specific noise.

B. Analysis Results with Decreasing SNR in the Gaussian Channel

From columns 5-8 of Table I, we see the overall trend of decreasing performance with decreasing SNR of Gaussian channel for all models. EEGNet, which is the most robust model to channel-specific noise, also seems to be relatively robust to the worsening SNR of the incoming noisy signal. The initial accuracy drop of EEGNet with Gaussian channel (2.79%) increases up to 10.99% as the SNR reduces to -18 dB. FBCSP and Shallow ConvNet models begin with 8.37% and 4.22% drops in performance in the presence of Gaussian channel, respectively, reaching up to 17.53% and 14.05% drops, respectively, as the SNR drops to -18 dB. Results of all models obtained with SNR of -12 dB, -16 dB, and -18 dB, are significantly different from their respective baseline accuracies ($p < 0.05$). Specifically, all models show the most significant performance deviation when the SNR is -18 dB ($p < 0.001$).

Fig. 3 illustrates the model-specific performance deviation with decreasing SNR of simulated noise. We have sorted the three plots based on the subject-specific baseline accuracies of the respective model, for better clarity. We observe an overall trend of declining performance with decreasing SNR in the Gaussian channel, where the performance is pushed down to chance-level. This is especially visible for FBCSP. EEGNet is clearly the most robust to negative SNR, and the model's overall deviation in performance for all subjects is not as large as that seen for the other two models. Shallow ConvNet is the next most robust model to negative SNR, illustrating stable performance for some of the subjects, including those with higher baseline accuracy. Nevertheless, the baseline accuracy of Shallow ConvNet for several subjects is close to chance-level to begin with, hence making it difficult to assess the robustness of the model for these subjects.

C. Channel-Wise Sensitivity of the Models

As all twenty channels considered for this study belong to the motor region, which is highly relevant for MI classification, the subject-specific performance deviation is significant for most of the channels to start with and for all channels when the SNR is reduced to -18 dB. Hence, we considered Gaussian noise with SNR of 0 dB to identify the channels that are most sensitive to noise across subjects. For each model, those channels with significant ($p < 0.0001$) deviations, as indicated by the permutation test described in section II. D 1, were identified for every subject. The number of subjects for whom each of these channels caused significant deviation in the presence of noise was then calculated.

FBCSP is most sensitive to noise in C6, CPz, CP1 and CP6, showing significant performance deviation in these channels for more than 45 out of 54 subjects. EEGNet shows significant performance deviation with noise in C6, CP1, CP3, FC2, and FC3 for more than 45 subjects. Shallow ConvNet shows high sensitivity to all channels for more than 43 out of 54 subjects. In particular, Shallow ConvNet illustrates significant performance deviation with noise in channels FC3 and FC5 for more than 50 out of 54 subjects. Channels C6, CP1 and FC3 appear in the list of most sensitive channels for two out of the three models considered.

D. Impact of Activation Function

The EEGNet model using Sigmoid activation function produced a slightly lower baseline accuracy (62.35%, $p > 0.05$) compared to the original model (63.54%), nevertheless, demonstrated improved robustness to channel-specific noise. The performance drop of the EEGNet model with Sigmoid activation function for SNR of -18 dB is 6.53% ($p > 0.05$) as against the 10.99% drop of the original model with the ELU activation function. The results obtained for EEGNet using the Sigmoid activation function are not significantly different from those obtained using the original EEGNet model, as presented in Table I. The baseline performance of Shallow ConvNet with Sigmoid activation function is 58.27% (p >

TABLE I
AVERAGE SUBJECT-SPECIFIC CLASSIFICATION ACCURACY WITH CHANNEL-SPECIFIC NOISE

Model	Baseline_{orig}	Zero Channel	Gaussian Channel 0 dB	 -6 dB	 -12 dB	 -16 dB	 -18 dB
FBCSP [9]	61.20±15.14	57.81±12.11	$56.08\pm10.41^{*}$	$54.17\pm7.83^{*}$	$51.86\pm4.27^{**}$	$50.85\pm2.56^{**}$	$50.47\pm1.62^{**}$
EEGNet [13]	63.54±11.53	61.94±10.30	61.77±10.25	61.18±9.66	$59.43\pm8.26^{*}$	$57.80\pm7.34^{*}$	$56.56\pm6.67^{**}$
Shallow ConvNet [12]	60.44±14.97	59.55±13.43	57.89±11.42	55.94±9.14	$53.82\pm6.83^{*}$	$52.62\pm5.56^{**}$	$51.95\pm4.57^{**}$

The *, and ** represent that the accuracy is significantly different from the baseline accuracy, with *: p < 0.05 and, **:p < 0.001.

TABLE II
ROBUSTNESS OF DEEP LEARNING MODELS USING SIGMOID AND TANH ACTIVATION FUNCTIONS

Activation Function	Model	Baseline_{orig}	Baseline_{mod}	Zero Channel	Gaussian Channel 0 dB	 -6 dB	 -12 dB	 -16 dB	 -18 dB
Sigmoid	EEGNet [13]	63.54±11.53	62.35±11.34	60.91±10	60.90±10.00	60.72±9.81	59.95±9.16	59.04±8.57	58.28±8.06
	Shallow ConvNet [12]	60.44±14.97	58.27±8.69	57.68±7.99	57.72±8.14	57.70±7.94	$57.39\pm7.24^{*}$	$56.79\pm6.54^{**}$	$56.19\pm5.84^{**}$
Tanh	EEGNet [13]	63.54±11.53	62.90±10.73	61.69±9.74	61.61±9.71	61.32±9.57	60.51±9.06	59.55±8.55	58.69±8.12
	Shallow ConvNet [12]	60.44±14.97	58.56±9.32	58.04±8.50	58.13±8.55	58.13±8.48	$57.72\pm8.00^{*}$	$57.09\pm7.35^{**}$	$56.42\pm6.73^{**}$

The *, and ** represent that the accuracy is significantly different from the accuracy obtained using the original model for the respective noise condition (in Table I), with *: p < 0.05 and, **:p < 0.001. Baseline_{orig} contains the baseline accuracies obtained using the original models. Baseline_{mod} contains the baseline accuracies obtained using the models with replaced activation functions.

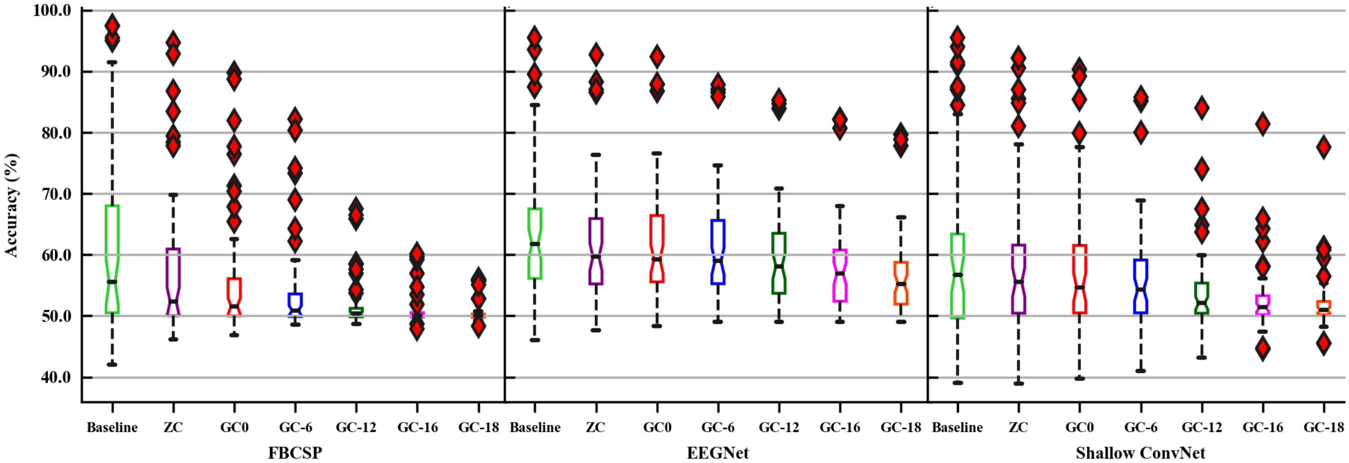

Fig. 2: Subject-specific baseline, zero channel (ZC) and Gaussian channel (GC) accuracies across models. Each box represents the first and the third quartile, and the horizontal line in the middle denotes the median accuracy. The SNR (in dB) of the Gaussian channels are indicated in the respective labels.

0.05) compared to 60.44% achieved by the original model. This deterioration in performance, which is not significant, can be attributed to the removal of the squaring and logarithmic nonlinearities, which was a fundamental characteristic of the model architecture that computes the log-variance of the input EEG. The accuracy of Shallow ConvNet with the Sigmoid activation function dropped by 3.57% (p < 0.001) when the SNR is -18 dB, thus showing an improvement in robustness compared to the original model for which the performance had dropped by 14.05% for the same value of SNR in the Gaussian channel. The accuracy drop in the absence of signal is 2.31% (p > 0.05) for EEGNet, and 1.01% (p > 0.05) for Shallow ConvNet, which is lower than their respective drops in accuracy using the original model. The results obtained for Shallow ConvNet using the Sigmoid activation function for SNR values of -12 dB, -16 dB, and -18 dB, are significantly different from the results obtained using the original Shallow ConvNet model (Table I).

The Tanh nonlinearity is usually preferred over Sigmoid nonlinearity due to its zero-centeredness and has been known to improve the training performance of deep learning models [23]. The robustness of EEGNet and Shallow ConvNet models to channel-specific noise had improved by the application of Tanh activation function, while showing a slight increase in the baseline accuracy when compared to the baseline accuracy obtained using the Sigmoid activation function. The baseline accuracies of EEGNet and Shallow ConvNet models with Tanh activation function are 62.90% and 58.56%, respectively, which are not significantly different (p > 0.05) from their original baseline accuracies. The performance of these models declined by 1.92% (p > 0.05) and 0.89% (p > 0.05), respectively, for zero channel, and 6.69% (p > 0.05) and 3.65% (p < 0.001), respectively, for Gaussian channel with SNR of -18 dB. The robustness analysis results obtained using Tanh function are not significantly different from the respective baseline results in Table I, except those observed for Shallow ConvNet for SNR values of -12 dB (p < 0.05), -16 dB (p < 0.05), and -18 dB (p < 0.001). The complete set of results with Sigmoid and Tanh activation functions are presented in Table II and the performance comparison of the two models is illustrated using box plots in Fig. 4.

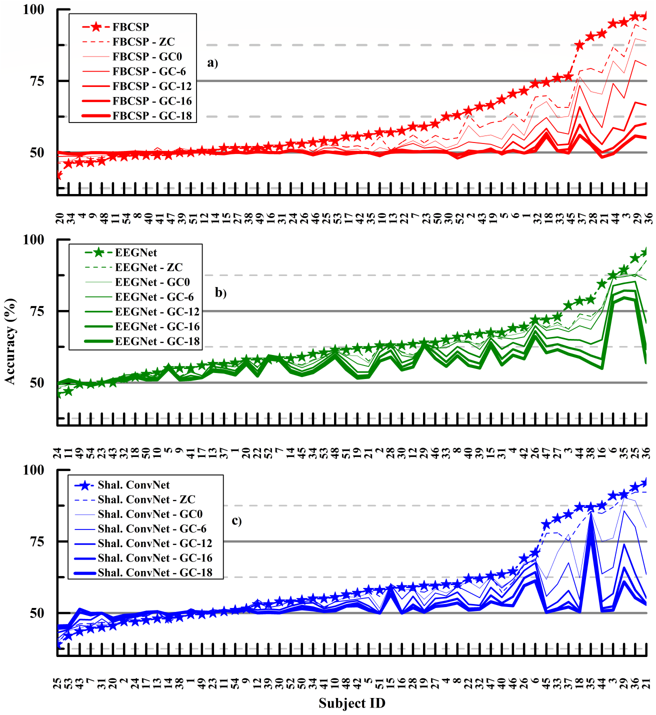

Fig. 3: Performance deviation with decreasing values of SNR in the Gaussian channel of a) FBCSP b) EEGNet, and c) Shallow ConvNet

IV. DISCUSSION

In this study, we evaluated the robustness of machine learning and deep learning based MI classification models under scenarios of channel disconnection such as absence of signal and presence of background noise. We further examined the performance deviation of these models by decreasing the SNR of simulated noise. By performing a statistical analysis of the results, we identified the most sensitive channels across subjects for each model. To the best of our knowledge, this is the first study that systematically investigates the channel-wise robustness of state-of-the-art EEG-based BCI systems. In addition to evaluating the robustness, we have also suggested a preliminary solution to enhance the robustness of the deep learning models by using appropriate activation functions. Nevertheless, a more optimal solution is necessary to stabilize the performance of these models when subjected to noisy data.

From the analysis results presented in Table I, Fig. 2 and Fig. 3, it is evident that all models considered in this study show poor response to zero channel and Gaussian channel conditions. In addition, the performance of these models continue to deteriorate, moving towards chance-level accuracy, when the SNR of Gaussian channel is further reduced. Our results indicate that EEGNet is the most robust to channel-specific noise. The performance drop of EEGNet is 10.99%, Shallow ConvNet is 14.05% and FBCSP is 17.53%, for an SNR of -18 dB. This study is a step forward towards understanding the robustness of deep learning based EEG-BCI models to noisy signal with negative SNR occurring due to unexpected experimental conditions.

Using the results of the robustness analysis, we identified the most sensitive channels for each model, by calculating the number of subjects showing significant accuracy deviation with Gaussian channel. Although different models are sensitive to noise in different sets of channels, we identified certain channels, such as C6, CP1 and FC3, that appeared to be the most sensitive for more than one model.

By replacing the activation functions used by the two deep learning models, EEGNet and Shallow ConvNet, we were able to boost their robustness such that their respective accuracy drops for SNR of -18 dB had reduced. Nevertheless, the impact of decreasing SNR on these models is still visible. This study indicates that the inherent network properties, including the choice of activation functions, can affect the robustness of deep learning based MI classification models. The analysis results using the original models (Table I) and the experiment with the two activation functions (Table II) have together led us to believe that there is a certain trade-off between robustness performance and training performance while designing the model architecture. While ELU activation function helps EEGNet model to achieve the best baseline performance, it produces large unstable activations when the model is exposed to noise, thus impacting its robustness. Similarly, the combination of squaring and logarithmic nonlinearities for log-variance computation improves the baseline performance of Shallow ConvNet, however, hurts the robustness of the model to channel-specific noise. On the other hand, saturating nonlinearities such as Sigmoid and Tanh, enhance the robustness of these models towards noise at the cost of deterioration in their training performance.

In spite of providing us with some important insights, this study is not complete in itself. We have considered negative SNR values, where the noise completely takes over the signal, to test the robustness of the BCI models, however, we have not performed an evaluation of the models by gradually varying the signal-to-noise ratio of the input EEG data, which may help us understand the behavior of these models in the presence of small amounts of noise in the signal. Similarly, unlike the examples of disconnected channels shown in Fig. 1 where the disconnection happens after a certain number of trials, we have simulated the channel disconnection right from the first trial and hence the entire channel data is affected in our experiments. Simulating noise in a certain portion of channel data will be part of our future analysis.

V. CONCLUSION

In summary, the main goal of our experimental study was to evaluate the robustness of state-of-the-art EEG-BCI models towards noise in channel data that can commonly occur during data collection due to unexpected experimental conditions such as channel disconnections. In particular, we focused on observing the response of the models to absence of signal and presence of background noise. In addition, we also examined the relation between the model performance and SNR of noise by subjecting these models to noise with decreasing values of SNR. Our results indicate the sensitivity of these models to channel-specific noise, more so towards

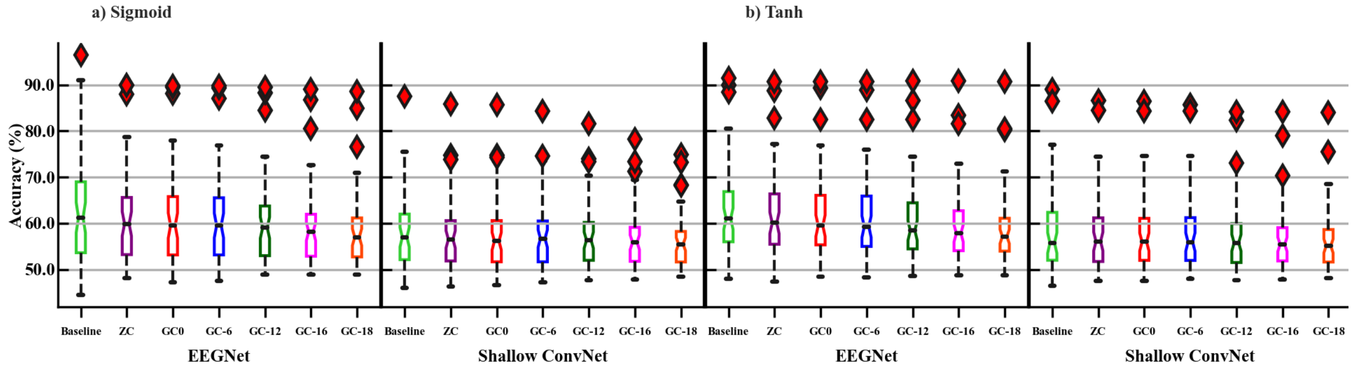

Fig. 4: Subject-specific baseline, zero channel (ZC) and Gaussian channel (GC) accuracies for EEGNet and Shallow ConvNet models using a) Sigmoid and b) Tanh activation functions. Each box represents the first and the third quartile, and the horizontal line in the middle denotes the median accuracy.

noisy signal with negative SNR. We have suggested a simple preliminary solution based on activation functions to improve the robustness of the models. Nevertheless, a more rigorous approach to address the robustness issue in deep learning based EEG-BCI models is essential. We believe that our study would evoke attention towards BCI robustness and create an awareness amongst BCI researchers of the different aspects of robustness that the BCI systems are required to satisfy in order to be fully functional in a real-world setting.

ACKNOWLEDGMENT

This work was partially supported by the RIE2020 AME Programmatic Fund, Singapore (No. A20G8b0102). The computational work for this article was partially performed on the resources provided by the National Supercomputing Centre, Singapore. We thank Dr. Tushar Chouhan for reviewing the code and for providing constructive comments on the manuscript.

REFERENCES

[1] A. Lécuyer, F. Lotte, R. B. Reilly, R. Leeb, M. Hirose, and M. Slater, “Brain-computer interfaces, virtual reality, and videogames,” *Computer*, 2008.

[2] J. N. Mak and J. R. Wolpaw, “Clinical Applications of Brain—Computer Interfaces: Current State and Future Prospects,” *IEEE Reviews in Biomedical Engineering*, 2009.

[3] K. K. Ang, C. Guan, K. S. Phua, C. Wang, I. Teh, C. W. Chen, and E. Chew, “Transcranial direct current stimulation and EEG-based motor imagery BCI for upper limb stroke rehabilitation,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2012.

[4] V. Tuyisenge, L. Trebaul, M. Bhattacharjee, B. Chanteloup-Forêt, C. Saubat-Guigui, I. Mîndruţă, S. Rheims, L. Maillard, P. Kahane, D. Taussig, and O. David, “Automatic bad channel detection in intracranial electroencephalographic recordings using ensemble machine learning,” *Clinical Neurophysiology*, 2018.

[5] T. Mulder, “Motor imagery and action observation: Cognitive tools for rehabilitation,” in *Journal of Neural Transmission*, 2007.

[6] A. Kübler, F. Nijboer, J. Mellinger, T. M. Vaughan, H. Pawelzik, G. Schalk, D. J. McFarland, N. Birbaumer, and J. R. Wolpaw, “Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface,” *Neurology*, 2005.

[7] K. K. Ang, C. Guan, K. S. G. Chua, B. T. Ang, C. Kuah, C. Wang, K. S. Phua, Y. Zheng Chin, and H. Zhang, “Clinical study of neurorehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback,” in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC’10*, 2010.

[8] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Müller, “Optimizing spatial filters for robust EEG single-trial analysis,” *IEEE Signal Processing Magazine*, 2008.

[9] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, “Filter Bank Common Spatial Pattern (FBCSP) in brain-computer interface,” in *Proceedings of the International Joint Conference on Neural Networks*, 2008.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, 2012.

[11] A. Craik, Y. He, and J. L. Contreras-Vidal, “Deep learning for electroencephalogram (EEG) classification tasks: A review,” 2019.

[12] R. T. Schirrmeister, J. T. Springenberg, L. D. J. 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,” *Human Brain Mapping*, 2017.

[13] 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*, 2018.

[14] S. M. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard, “Universal adversarial perturbations,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017.

[15] X. Zhang and D. Wu, “On the Vulnerability of CNN Classifiers in EEG-Based BCIs,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2019.

[16] S. Nakagome, T. P. Luu, Y. He, A. S. Ravindran, and J. L. Contreras-Vidal, “An empirical comparison of neural networks and machine learning algorithms for EEG gait decoding,” *Scientific Reports*, 2020.

[17] M. H. Lee, O. Y. Kwon, Y. J. Kim, H. K. Kim, Y. E. Lee, J. Williamson, S. Fazli, and S. W. Lee, “EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy,” *GigaScience*, 2019.

[18] D. A. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (ELUs),” in *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, 2016.

[19] V. Nair and G. E. Hinton, “Rectified linear units improve Restricted Boltzmann machines,” in *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, 2010.

[20] A. L. Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier nonlinearities improve neural network acoustic models,” in *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, 2013.

[21] J. Han and C. Moraga, “The influence of the sigmoid function parameters on the speed of backpropagation learning,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1995.

[22] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” 2015.

[23] Y. A. LeCun, L. Bottou, G. B. Orr, and K. R. Müller, “Efficient backprop,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012.