A group of blue lines

Description automatically generated A screenshot of a graph

Description automatically generatedFig2 maybe one less good

Figure 1. Subject 7 EEG signal from the 16 second periods of stimulus with the four different frequencies.

Figure 1 shows the four EEG signals recorded from Subject 7. Although this is an extreme example, it does demonstrate several periods where the recording device clipped the signal when the results exceeded the dynamic range of the analog to digital converter. This undoubtedly has an impact on the accuracy of any algorithm attempting to detect the SSVEP so censoring of any clipped epochs of data should be considered for censoring. However, our primary goal is to compare our two neural networks to Acamp’s LR method so in most of this report we ignored the impact of the clipping since they did in their analysis without censoring any epochs. We return to this issue in depth after the main assessment.

The figure of merit used to assess the three models is accuracy. A confusion table is shown in Figure 2 for the NNovr Model with a 6 second periods obtained with 95% overlap. This model censored epochs with 4 or more consecutive saturated point from the ADC. The epoching parameters lead to over 2300 training vectors each with 168 features (frequencies from 4 to 32 Hz). Eighty percent were used for training the network and the remainder for testing.

A graph of different colored lines

Description automatically generatedA graph of different colored lines

Description automatically generatedA graph of different colored lines

Description automatically generated

A B C

Figure 3. Plot of accuracy versus overlap period for all time periods for the A) NN4Class, B) NNovr, and C) LR models.

Figure 3 shows the results the of each model run at the full set of periods and overlaps used by Acamp. The LR model is very similar to the figure found in their paper. Our NN4Class Model is similar, and the NNovr Model (B) obtains slightly higher accuracies at high overlap/long periods. Table 1 shows the mean accuracy +/- 1sd of each method compared by paired t-tests from 10 training sessions of each model. This demonstrates a significant difference between the NNovr ( 0.958+/-0.011 ) and LR Models (0.094 +/- 0.01 ) with p-value < 0.01. NN4Class exceeds the LR Model, but is not significant. Figure 4 shows a histogram from these training sessions.

Repetitions = 10

Accuracy for Period of 6 seconds with overlap 0.95

Spliting of data sets is random = True

Training proportion is 0.8

Saturation\_criterion = 4096

Accuracy and paired t-tests Note:

NN4Class: 0.9423 +/- 0.019 Change all accuracies to %

vs t-value = -4.19, p-value = 0.00234 rather than proportion

NNovr : 0.9577 +/- 0.0114

vs t-value = 6.39, p-value = 0.00013

LR : 0.9403 +/- 0.0096

vs t-value = -0.43, p-value = 0.67659

NN4Class: 0.9423 +/- 0.019

Table 1

A graph of different colored bars

Description automatically generated

Figure 4

Effect of censoring. Removing the epochs with 4 or more consecutive time points with maximum or minimum ADC output leads to improved performance of the NNovr model, representing the best performance we found with data set. With censoring the the NNovr Model accuracy increased to 97.5% and is significantly higher than the uncensored NNovr Model (p-value < 0.01.) (See Table 2).

NNovr (uncensored) versus NNovr (censored) ( 6 second periods with overlap of 95%)

NNovr (uncensored) accuracy(%): 95.8 +/- 1.2 (mean +/- sd, n = 10

NNovr (censored) accuracy(%): 97.5 +/- 1.2 (mean +/- sd, n = 10)

Unpaired t-statistic = -3.01, p-value = 0.0075

Table 2.

Despite the high accuracy obtained by our censored NNovr Model we are not convinced this could be obtained in practice. We have concerns with the assessment that Acamp employed. Training a network requires a large amount of data. Only four, 16 second periods of EEG data were obained on 11 subjects and in an attempt to “obtain more data” they devised their overlap strategy. We feel that this created an artificial correlation between epochs that probably substantially improved model performance. In an attempt to demonstrate this we trained the models on 9 of the 11 subjects and attempted to predict the remining two. The accuracy at each period/overlap dropped substantially from a random sampling of training/testing approach. The NNovr (6 second periods/95% overlap) dropped from 95.8% accuracy to 74**.**9%.

**Discussion**

**Conclusions**

More data needed to validate and compare models.

Directly compart the Acamp approach with more much more EEG data with independent eochs. (Some of the difference between the random training and biased training to predict two subjects, may be due to the generalization of a model between subjects) [Another test we could do is after epoching go into each epoch and set 15Hz bins to 0 and see how munch worse 15 Hz dose than the rest or set all to 0 and see if it gets an accuracy of over 25%.]

A larger data set would also allow a more meaningful assessment of the impact of ADC saturation. Although our censoring criterion was strict, more data could allow determination of a better censoring threshold that could reduce the frequency of eliminating epochs and improve ITR. Could saturation be partly due to dry-electrode with a variable contact and thus fluctuating voltsge

The NNovr model out performed our NN4Class model, but an advantage of the 4 Class model its predicitions are based on a meaningful probability that is available to a algorithm and could be used to modify actions based on the simple class returned. If a critical decision is to be made and the cost of a FP is high the probability could be used. (e.g. only make the decision if the Probability is>90% for a high risk response.)

ITR not accurately determined due to the way time were extracted, but given accuracy of about 97% with 6 second epochs get ITR = log2(N) + P\*log2(p)+(1-p)\*log2((1-p)/(n-1)). => 1.8 b/trial or 0.3 b/s is an upper limit. NNovr (uncensored- censoring likely loose its accuracy advantage because you would have to eliminate some trials.)

Interestingly this goes up slightly if you sacrifice accuracy for a shorter time period.

Contribution - beware of creating correlation with your preprocessing technique. Needs a formal study. Suspect that you need independent trials (huge number 100s +) Impact of saturation on epochs – when to censor. Cause of ADC saturation-?related to dry-electrode technology (lit. search)