Tarnu Ilie-Cristian 237

Report 4

Results

To measure the performance of each model, accuracy, loss were taken into account together with sensitivity and specificity for each of the 5 classes that were predicted. The fully connected (dense) neural network and the convolutional neural network were trained for 50 epochs using 10-fold cross validation, while the recurrent neural network was trained for 50 epochs using only 1 fold.

The dense neural network had the most number of learnable parameters and had the following performance: the average accuracy when passing raw signal into the network was 50.98%, while the average accuracy for the filtered signal was 48.76%. This method had trouble differentiating between class 4 and 5 (eyes open – 5, eyes closed - 4) and between the inter ictal classes (2 and 3) in which the patient had a brain tumor but there was no seizure detected. The sensitivity for these classes varies and goes as low as 36% for class 2 and 58% for class 4. Best sensitivity was obtained by class 1 (seizure) with a value of 71%, comparable to the sensitivity when the input was filtered – 70%. When filtering the signal accuracy dropped by a little, but the sensitivity for classes 2, 4, 5 improved, while the sensitivity for class 3 dropped drastically. In terms of specificity the best performing class was class 1, about 92% for both input types.

The convolutional neural network yielded the best results out of all methods with an accuracy of 74.42% with raw signal and 48.59% when filtering the signal. In the case of raw signal input, the sensitivity of all classes but one was above 60%, with the highest being class 1 with 95%. The lowest sensitivity was for class 2 (EEG taken from the area where tumor is located): 46%. The rest of the classes had a fairly good sensitivity of around 80%. Specificity using this method was also better, all classes having a specificity of around 90%. Filtering the input signal didn't provide better metrics as the accuracy dropped significantly together with the sensitivity, especially for the seizure class: 49%. For all other classes the result is similar, a drop in accuracy.

The recurrent neural network model was the only one that wasn't trained using 10-fold cross validation, so results might differ slightly in reality or when using another data set. This model had the least number of trainable parameters, but it did yield decent results. In the case of raw signal, the average accuracy was 71.58% and for the filtered input it was 66.36%. As with the other cases the best sensitivity is for class 1 being around 90% for both input methods together with the specificity of around 98%. This method struggled the least when predicting classes 4 and 5 (eyes closed, eyes open), both classes having the same sensitivity for both input methods and that is 78%. The worst metrics for this method was the sensitivity for classes 2 and 3, which was around 50% for both input types.

Conclusion

The model with the best results was the convolutional model, using raw signal as input. This method had the highest accuracy and lowest loss of all the other methods. Even though the accuracy is a little below 75%, given the difficulty of the prediction task it's a decent accuracy.

The biggest problem for all models were the differences between class 2 and 3, and class 4 and 5 respectively, as the signals are not easily distinguishable. Oddly enough, all models performed worse when the signal was filtered as opposed to unfiltered signal. This may be because of the choice of wavelets when using discrete wavelet transform or because this filtering method is not suited for this task. The dense model had the slightest difference in accuracy when changing the input method and in my opinion is not worth pursuing any further, as any other slight tweaks to the model might not yield a better result. The recurrent model stands in between the other 2 as it had decent accuracy and the input didn’t affect the accuracy as drastically as with the convolutional network. It’s important to also take into consideration the fact that the recurrent model was trained using only one fold and that its performance might vary.

This prediction task is over all quite a difficult one as there are major differences between some classes and only minor ones between others, as shown by the metrics of each model. Using different filtering techniques and other manipulations of the time series might improve the results. Another important aspect is the fact that no artifact removal was done on the dataset and that the dataset is limited to only 500 patients, so the same EEG from one patient was treated as multiple inputs for the networks, in order to have enough data for training.