 DOI: 10.3217/978-3-85125-682-6-52

sLDA classifier has been demonstrated to be more effective and more robust for small dataset than LDA [4]. In our implementation, the optimal shrinkage parameter was determined following the lemma introduced by Ledoit and Wolf.

Logistic regression (LR) is a discriminative learning classifier that directly estimates the parameters of the posterior distribution function. The maximum likelihood method is used to approximate such parameters [12]. In our implementation, the regularized logistic regression was optimized by means the liblinear [13] algorithm that supports both L1 and L2 regularization.

Support vector machine (SVM) is a classifier that uses a discriminant hyperplane to identify classes. The selected hyperplane is the ones that maximizes the distance (margin) from the nearest data points (support vectors) of each class [12]. In this study, we implemented the linear SVM and set the penalty parameter to c=1. Before selecting the value, three values (c=0.1, c=1 and c=10) were tested. No statistically significant differences were found among values.

Multilayer perceptron (MLP) is a neural network. For this analysis, we implemented a MLP that trains using a quasi-Newton algorithm which uses a backpropagation implementation of the gradient. We considered one hidden layer MLP with 20 neurons, ReLU (rectified linear unit) as activation function of the neurons and L-BFGS solver. This final setting has been defined after testing a combination of different values for the number of neurons of the hidden layer (10, 20, 40 and 80 neurons), the activation functions (ReLU, sigmoid) and the solvers (L-BFGS, Adam and RMSProp). The combination that gave the best results, in terms of classification accuracy average across subjects, was that used in this analysis.

Decision tree (DT) is a classifier which partitions the feature space until terminal nodes, each one assigned to a predicted value. Although decision trees are very easy to use for no-statisticians, they work for non-linear functions and the treatment of missing values is more satisfactory than most other model classes, we might not be able to find the best model at all. Moreover, results can be quite variable: small changes in the data can potentially lead to completely different splits (i.e. trees) [14].

Random Forest (RF) classifier is a set of decision trees merged by a probabilistic scheme. To classify an epoch, the corresponding feature vector is the input for each tree in the forest. Each tree makes a prediction and the forest chooses the prediction having the most votes over all the trees in the forest. RF can work on high-dimensional data and it can be applied to any model. Despite of its ability to returns the variable importance, it is very hard to interpret [14].

Many variant RF parameters impact on the algorithm accuracy. For each subject, we tested both the number of trees (10, 20, 50, 100, 200, 500, 1000, 2000) and the minimum number of samples required to split an internal node (from 2 to 24 in steps of two). The best set in terms of accuracy average (across subjects) was considered for the following analysis: 2000 trees and 4 samples required to split an internal node.

Validation: A 10-times cross-validation (Fig. 1) was implemented to compare classifiers and number of features. For each iteration the feature domain (epochs x number of features, at most 240 x 120) was shuffled along the first size (epochs). The first ninety and the last ten percent of the data have been the training and testing dataset, respectively.

Training dataset was the input for the feature selector (RFE-CV based on DT). It performed the feature selection ten times using the same dataset and returned the list of features sorted according to the more selected among the feature selection iterations.

The first two or ten features were considered to reduce the feature domain or all features if feature selection was not required.

Figure 1: Validation approach. For each iteration (10 in total) the steps in the hatch block were repeated. Specifically, feature domain was shuffled and divided in training and testing dataset. The former was used to train the classifier (and to select the best two or ten features, if that was the condition under investigation), the latter to test the classifier. The performance index was computed for each iteration.

The feature domain, properly reduced (only for 2 or 10 features analysis), was the input for each classifier. Each classifier was trained from the training dataset. The testing dataset, never seen before, was used to test the model and compute the performance index.

For each pair number of features-classifier (e.g. 2 features – MLP) the average of the performance index across all iterations (10 in total) has been considered the emblematic value for that pair.

Performance Measures: For each pair number of features (2, 10, 120 features) and classifier (SWLDA, sLDA, LR, SVM, MLP, DT, RF) classification accuracy was computed.

 DOI: 10.3217/978-3-85125-682-6-52

Statistical Analysis: For each pair number of features-classifier the Shapiro-Wilk test was applied to assess the normality of each performance index distribution.
To investigate the effect of the number of features as well as of the classifier and their potential interaction, classification accuracy was analysed by the repeated measure two-way analysis of variance (ANOVA). The Tukey HSD post hoc analysis was applied to assess pairwise differences. The threshold for statistical significance was set to $p<0.05$. Results are presented as mean ± standard error (SE) across subjects.

RESULTS

The statistical analysis revealed the significant effect of the classifier factor ($F=77.22$, $p<0.001$) as well as the number of the features ($F= 19.11$, $p<0.001$) on the classification accuracy and the significant interaction among factors ($F= 13,20$, $p<0.001$).
Figure 2 shows for each pair (classifier-number of features) the results, presented as average and standard error across subjects.
The post-hoc analysis, applied to the *classifier* factor, pointed out the overall superiority of the SWLDA classifier over all classifiers as well as better performance obtained by the sLDA respect to those of the LR, SVM, MLP and DT classifiers. All classifiers outperformed DT classifiers.
Moreover, better performances were globally (*number of features* factor) achieved when ten or all available features were used than those obtained for two features.
Since the number of features directly impacts on the number of physical EEG electrodes required to collect EEG data and, therefore, extract the needed features, results from interaction between classifiers and number of features were deeply analysed and reported in Fig. 3.
With equal number of features (both for 2 and 10 features), SWLDA (accuracy average= 0.78 evaluated for 2 features, accuracy average= 0.79 evaluated for 10 features) statistically outperformed all classifiers. DT classifier, instead, (accuracy average= 0.62 evaluated for 2 features, accuracy average= 0.67 evaluated for 10 features) did not reached good performance, revealing to be the worst classifier (Fig. 3 upper panel, left side).
When all available information were used to train the classifiers, no statistically significant differences emerged among SWLDA, sLDA and RF (Fig. 3 upper panel, right side).
Increasing the size of the feature domain significantly improved performances in both sLDA (0.70, 0.75, 0.80 accuracy average for 2, 10 and 120 features) and RF classifiers (0.67, 0.73, 0.80 accuracy average for 2, 10 and 120 features). Conversely, SWLDA performances (0.78, 0.79, 0.80 accuracy average for 2, 10 and all features) did not differ among them varying on feature number.

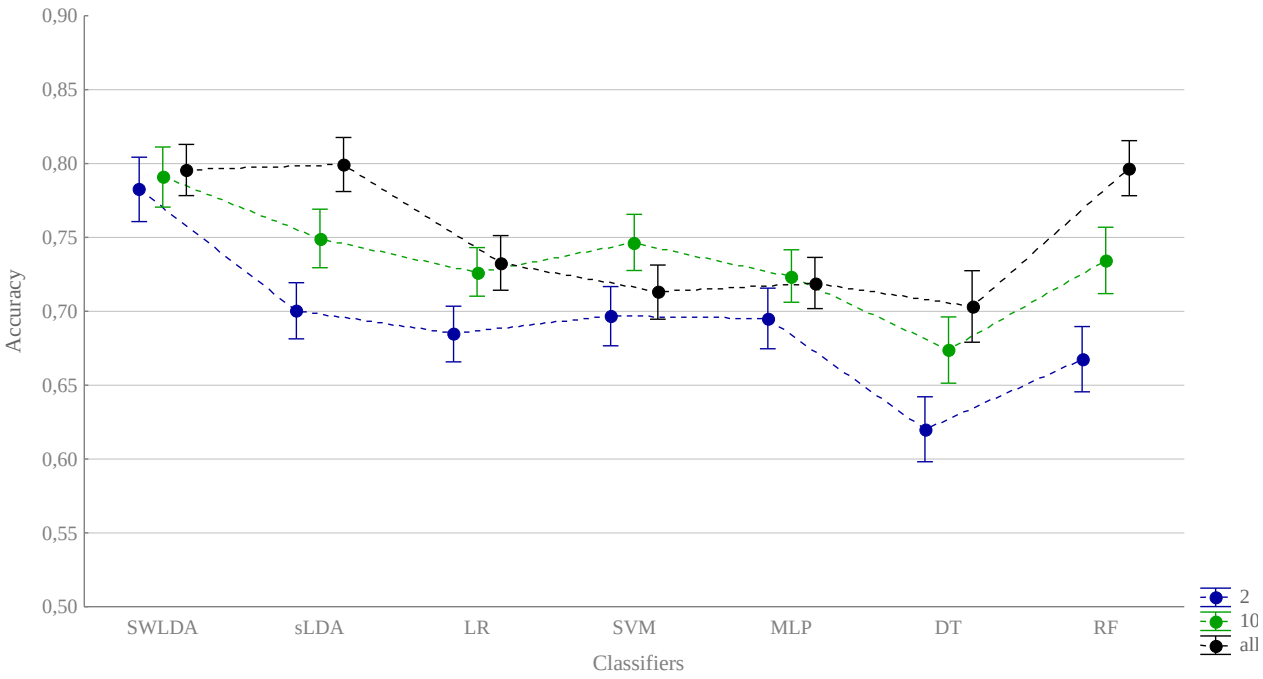

Figure 2: Classification accuracy, presented as mean ± standard error (15 stroke patients), computed for seven classifiers: stepwise linear discriminant analysis (SWLDA), shrinkage linear discriminant analysis (sLDA), logistic regression (LR), support vector machine (SVM), multilayer perceptron (ML), decision tree (DT), random forest (RF). For each classifier, accuracy was evaluated when the feature domain had been reduced to include two features (in blue), ten features (in green) and all available (black) features, i.e. no feature domain dimensionality reduction.

 DOI: 10.3217/978-3-85125-682-6-52

Figure 3: Post-hoc test results of the ANOVA interaction factor. Upper panel, left side: comparison among classifiers with the same number of features. Pairwise differences among classifiers, for 2 and 10 features, were presented in same matrix since they had returned equal results. Upper panel, right side: comparison among classifiers when all features were considered. Lower panel, left side: comparison between classifiers trained from 2 features and those trained from 10 features. Lower panel, centre: comparison between classifiers trained from 2 features and those trained from all available features. Lower panel, right side: comparison between classifiers trained from 10 features and those trained from all available features.
Matrix reading: The classifier in each column header statistically differed/ did not differ from the classifier reported in the row header. Significant/no-significant differences correspond to coloured/white boxes. Green (orange) boxes means that the classifier in the column (row) header outperformed that in the row (column) header.

The increasing trend in accuracy was observed also for LR, DT and MLP classifiers: for the first the increase from 2 to 10 or 120 features statistically improved classification accuracy, for the neural network model the trend was not supported by the statistical results. SVM seemed, instead, to be prone to overfitting (0.70, 0.75, 0.71 accuracy average for 2, 10 and 120 features).
Lastly, the cross-check between classifier and number of features revealed that even the best model of the sLDA and RF (120 features) did not significantly differ from the SWLDA based just on two or ten features (Fig. 3, lower panel, centre and right side). Therefore, even if sLDA and SWLDA are both linear and, therefore, interpretable models, the last reached comparable performance by means few features (i.e. 10 features, less than 10 EEG channels).

DISCUSSION

Identifying the optimal classification method, based on relevant features, fast and able to provide an interpretable model of the EEG reinforced pattern, is a milestone in post-stroke rehabilitation protocols supported by BCI technology. In contrast to other fields of application where optimal cursor control is pursued, in a rehabilitation context the reinforcement of the proper sensorimotor activation in terms of both topographic and spectral characteristics is the main aim.
Spectral features belonging to the sensorimotor area of the affected hemisphere, in alpha and beta bands, were extracted from EEG data of 15 stroke subjects to compare seven classifiers in terms of classification accuracy. Performance was also analysed varying on the number of features considered in the feature domain.
Stepwise linear discriminant analysis (SWLDA) revealed being the best classifier even just considering two or ten features. Considering all available features (120 features) shrinkage linear discriminant analysis (sLDA) and random forest (RF) achieved good results and comparable to those of SWLDA. Nevertheless good results, linear models (SWLDA and sLDA), resulting from the linear combination of features properly weighted, are more interpretable than RF model.
In our approach, indeed, monitoring the cursor trajectory, feedback provided to the therapist and directly related to the combination of proper features, allows to explain single trial and rehabilitative session performances.
RF belongs to the bootstrap aggregating methods based

 DOI: 10.3217/978-3-85125-682-6-52

on decision tree classifiers and, although decision tree is the simplest model because the intuitive interpretation, the structure of the RF (in our case 2000 decision trees) decreases the interpretability of the model by the clinicians. Among linear models, instead, even if SWLDA and sLDA reached the same performance, SWLDA built the model by less than 120 features: in most cases, the embedded feature selection process, starting from the empty model, did not add all predictors to the model. Moreover, the interaction among ANOVA factors did not highlight differences between SWLDA, trained with two or ten features, and both sLDA and RF trained by all features.

The linearity, the interpretability and the possibility to achieve good classification results also thank to the embedded selection process (not covered by the nature of the other classifiers) yielded SWLDA to be considered the best classification approach in the rehabilitation context. Furthermore, the possibility to monitor EEG patterns to reinforce by means few electrodes (10 features i.e. less than 10 EEG channels,) matches the use of BCI technology in clinical context.

Focusing for each classifier on the number of features factor, if from one hand for SWLDA the increasing trend, justified by the increasing number of features (2, 10, less than 120 features), was not supported by the statistical analysis, from the other hand the trend revealed being significant for both sLDA and RF. Moreover, with equal number of features, no differences were observed between sLDA and RF, supporting, therefore, the use of RF model as a good approach to the binary classification of motor imagery tasks, as proposed in [9]. Although Steyrl et al. observed the superiority (3% in accuracy) of the RF approach to the sLDA, the characteristics of their approach, i.e. different tasks, number of channels, pipeline of EEG signal processing, should be considered in the comparison. For similar reasons, our results did no confirm results in [5]. Although we used similar spectral features, the application of the common spatial pattern filter and the lower number of the recorded EEG channels may be the reason why the multilayer perceptron and the logistic regression did not show good performances.

CONCLUSION

SWLDA classifier statistically outperformed those commonly used in SMR-BCI paradigms, achieving good performance even in case of feature domain dimensionality reduction. Monitor the brain activity by means few EEG electrodes, indeed, is the key to use BCIs in clinical realm. Linearity, interpretability and impact on the usability yielded to positively evaluating SWLDA approach in the upper limb post-stroke motor recovery protocols supported by BCI.

ACKNOWLEDGEMENTS

We thank Angelica Cottarelli for support in the preliminary analysis of EEG data. This work was partially supported by the project APOSTROPHES, Sapienza, University of Rome, Progetti di Ateneo and Promobilia Foundation (2018-H1 ref. 18076).

REFERENCES

[1] Ramos-Murguialday A et al., Brain-Machine-Interface in Chronic Stroke Rehabilitation: A Controlled Study, Ann. Neurol., vol. 74, no. 1, pp. 100–108, Jul. 2013.

[2] Pichiorri F et al., Brain-computer interface boosts motor imagery practice during stroke recovery, Ann. Neurol., vol. 77, no. 5, pp. 851–865, 2015.

[3] Lotte F, Congedo M, Lécuyer A, Lamarche F, and Arnaldi B, A review of classification algorithms for EEG-based brain-computer interfaces, J. Neural Eng., vol. 4, no. 2, pp. R1–R13, Jun. 2007.

[4] Lotte F et al., A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update, J. Neural Eng., vol. 15, no. 3, p. 031005, Jun. 2018.

[5] Bashashati H, Ward RK, Birch GE, and Bashashati A, Comparing Different Classifiers in Sensory Motor Brain Computer Interfaces, PLOS ONE, vol. 10, no. 6, p. e0129435, Jun. 2015.

[6] Colamarino E, Pichiorri F, Schettini F, Martinoia M, and Matti D and Cincotti F, Guider: A Gui for Semiautomatic, Physiologically Driven Eeg Feature Selection for a Rehabilitation Bci, Proc. 7th Graz Brain-Comput. Interface Conf. 2017 Vis. Real.

[7] Krusienski DJ et al., A comparison of classification techniques for the P300 Speller, J. Neural Eng., vol. 3, no. 4, pp. 299–305, Dec. 2006.

[8] Krusienski DJ, McFarland DJ, and Wolpaw JR, Value of amplitude, phase, and coherence features for a sensorimotor rhythm-based brain-computer interface, Brain Res. Bull., vol. 87, no. 1, pp. 130–134, 2012.

[9] Steyrl D, Scherer R, Faller J, and Müller-Putz GR, Random forests in non-invasive sensorimotor rhythm brain-computer interfaces: A practical and convenient non-linear classifier, Biomed. Tech., vol. 61, no. 1, pp. 77–86, 2016.

[10] McFarland DJ, Lefkowicz AT, and Wolpaw JR, Design and operation of an EEG-based brain-computer interface with digital signal processing technology, Behav. Res. Methods Instrum. Comput., vol. 29, no. 3, pp. 337–345, Sep. 1997.

[11] Guyon I, Weston J, Barnhill S, and Vapnik V, Gene Selection for Cancer Classification using Support Vector Machines, Mach. Learn., vol. 46, no. 1, pp. 389–422, Jan. 2002.

[12] Bishop C, Pattern Recognition and Machine Learning. New York: Springer-Verlag, 2006.

[13] 'LIBLINEAR -- A Library for Large Linear Classification'. [Online]. Available: https://www.csie.ntu.edu.tw/~cjlin/liblinear/.

[14] Breiman L, Classification and Regression Trees. Routledge, 2017.