Paradigm-independent classifier

*Project for Acquisition and Analysis of Biomedical Data Course

Bexultan Alimzhanov, Vladislav Mun, Natalya Makarenko,
Madina Kudaibergenova, Madina Saparbayeva, Aida Ospanova
School of Engineering and Digital Sciences,
Nazarbayev University, Nur-Sultan, Kazakhstan
bexultan.alimzhanov@nu.edu.kz, vladislav.mun@nu.edu.kz, natalya.makarenko@nu.edu.kz,
madina.kudaibergenova@nu.edu.kz, madina.saparbayeva@nu.edu.kz, aida.ospanova@nu.edu.kz

***Abstract*—Many studies have been made on the topic of the three most common BCI paradigms - ERP, MI and SSVEP. However, only a few so far explored the possibility of combining them under one classifier. Hence, the following study is dedicated to the development and proposal of a Paradigm-independent generic classifier that successfully classifies three main EEG BCI paradigms. Several possible Machine Learning solutions are discussed in the study, the choice of which depends on the level of noise in the data. The most robust and practically feasible classifier employs a neural network, and had achieved a 71% practical classification accuracy which persisted over multiple sessions and subjects. In addition, a theoretical accuracy threshold for each of the solutions was calculated and is estimated to be around 80%, highlighting possible directions of further research.**

***Index Terms*—BCI Paradigms, Machine Learning, Neural Networks, Event-Related Potentials, Steady-State Visually Evoked Potentials, Motor Imagery, OpenBMI toolbox.**

I. Introduction

A. Motivation

Brain Computer Interfaces is a multidisciplinary approach which allows us to record, process and distinguish between varying Brain signals with the help of a computer. The acquired information can be used later for many medical or practical purposes, such as the study of psychological diseases or cases of severe disabilities.

The brain signals are observed with Electroencephalograph (EEG)-based devices. This is the cheapest non-invasive method which is comparatively efficient in detecting brain signals. EEG-based BCI systems are commonly divided into three paradigms: motor imagery (MI), event-related potential (ERP), and steady-state visually evoked potential (SSVEP) [1].

Traditionally the BCI paradigms have been analyzed and observed separately, and have achieved significant improvements. For instance, Machine Learning and Signal Processing have advanced the statistical techniques of decoding class-related EEG data [2]. However, there approaches that combine together all three paradigms are not commonly implemented. Therefore, there is a need to build a BCI system which makes MI, ERP, and SSVEP paradigms available at the same time, and provides a choice of communication modes. In other words, the idea is to build paradigm-independent classifier that decodes data from all three BCI paradigms.

Acknowledgments to Professor Siamac Fazli

B. Problem

Building a paradigm-independent BCI classifier is no easy task due to the fact that MI, ERP, and SSVEP have their own characteristics. For example, each of them have their own dimensionality after feature extraction: MI has 8 columns, SSVEP - 4, and ERP - 110 columns. We present two different solutions for that problem. Then the next problem is creating the model on training data correctly using various techniques, such NCC, LDA, SVC, KNN and Deep Learning Neural Networks.

C. Solution

Starting with the raw data, we need to preprocess the data and build the one-fits-all classifier which will able to decode input trials. As we said before, we prepared it using two methods:

1) Machine learning methods on zero padded data with the assumption of zero noise coming from the wrongly preprocessed paradigms;
2) Preprocess each EEG data with three paradigm-specific feature extractions and use concatenated input in deep learning neural network architecture.

Then finally give a correctly classified class as an output.

D. Project details

The dataset was provided by [1] where a given subject participated in MI, ERP, and SSVEP paradigms. There are 54 subjects and 2 sessions in the dataset. Then after doing some data preprocessing we've got 16200 rows for each paradigm train/test split in each session. We used predefined training and testing splits in our machine learning and deep learning evaluation pipelines, as well as session-to-session performance assessment, i.e. training one one session data split and testing on the different session.

II. Methodology

A. Visualization

One of the steps of this project is data visualization of MI, ERP, and SSVEP. The Fig.1 illustrates P300 responses,

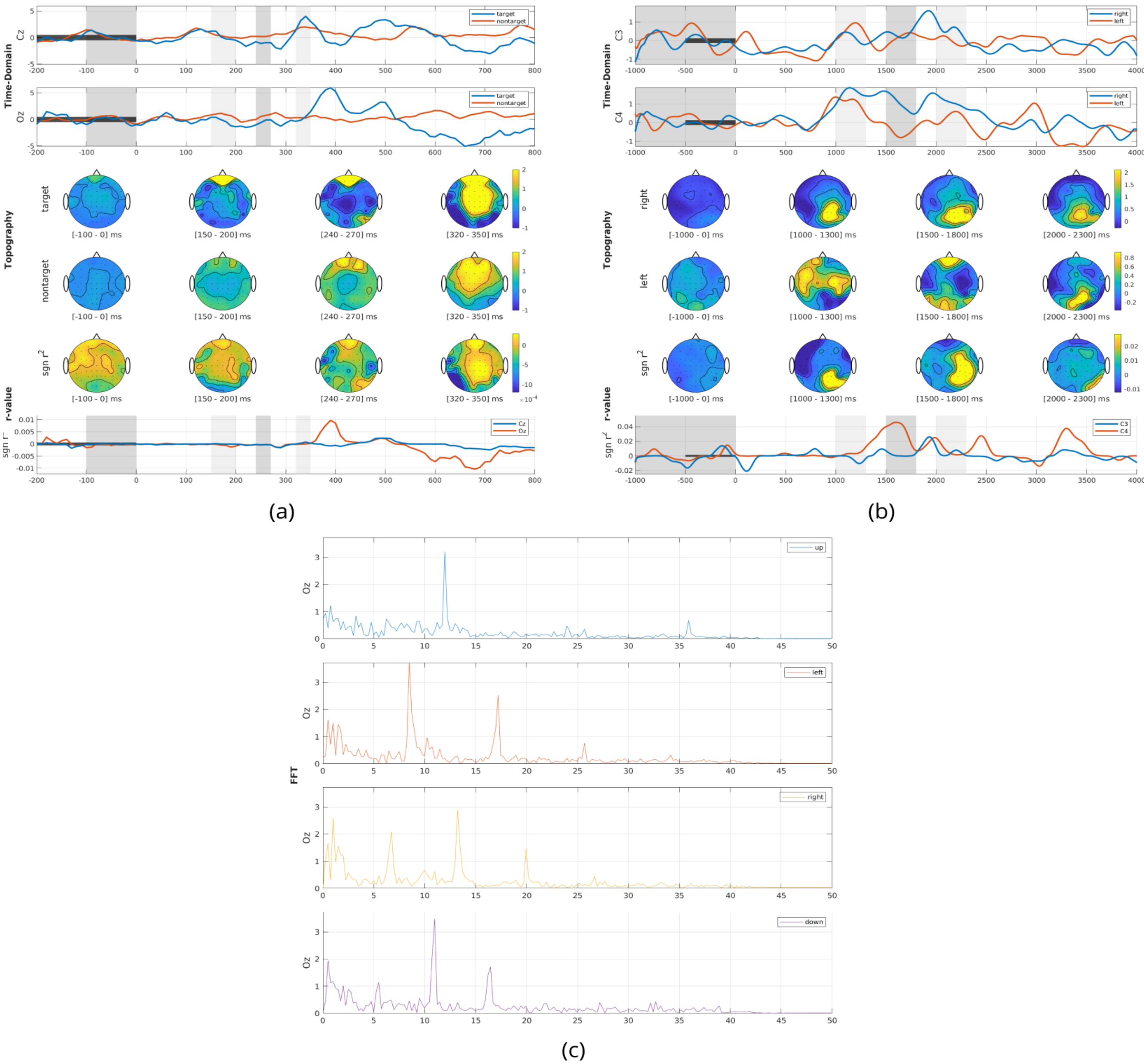

Figure 1. Visualization of P300, response (a), ERD/ERS patterns (b), and PSD (c) for ERP, MI, and SSVEP data respectively for session1 subject1

ERD/ERS, and power-spectral density (PSD) for ERP, MI, and SSVEP data respectively as shown in [1]. As an example of visualization, data from session 1 and for subject 1 were taken.

The Fig.1 (a) demonstrates visualization of P300 responses for ERP paradigm, where target and non-target trials were segmented between -200 ms and 800 ms time interval with respect to stimulus onset. According to [1] to observe ERP responses, in particular P300, the Cz and Oz electrodes should be chosen.

The ERD/ERS patterns are shown in the Fig.1 (b) in the mu rhythm band (8-12 Hz) for MI BCI paradigm. According to [1], the corresponded to motor regions C3 (left hemisphere) and C4 (right hemisphere) electrodes were chosen to specifically depict those patterns induced by left or right hand imagery tasks.

In the visualization of ERP and MI, in the first two rows there are grid plots that are shown in time (x-axis) and amplitude (y-axis) in selected channels. The topographies of brain areas on the next two rows correspond to specific time periods (also illustrated as gray areas on the grid plots). The next (fifth and six) rows are topographic and grid representations of ERP and MI paradigms for the level of significance between binary classes.

In the case of SSVEP visualization, the PSD was measured in between 1 and 25 Hz frequency band from the Oz electrode. The results are demonstrated in the Fig.1 (c), where the PSD was calculated for the four target classes. The four grid plots illustrate high amplitudes at the target frequencies corresponding to their classes [1].

B. Data Preprocessing

The initial data in the project dataset is raw, therefore, it was necessary to perform preliminary data processing before classifying it. For this, separate functions were created for each of the three paradigms in Matlab, using the OpenBMI toolbox. In addition, a main function was written that reads data from a given directory, processes it in loops that iterate over sessions and subjects, calls the corresponding function (ERP, MI or SSVEP), saves the results to files separately for training and test sessions for each subject.

The function that implements the preprocessing of the ERP data, firstly, changes the sampling rate of the given EEG signal to 100. Further, the first 100 trials were chosen from each paradigm, so that the resulting matrix with features would fit in dimensionality to the results of preprocessing in other paradigms. Next, a band-pass filter was applied to select the target frequency in the range 0.5-40Hz, and the result was segmented in the time interval [-200 800] ms, with subsequent baseline correction by subtracting the mean amplitudes in the [-200 0] ms pre-stimulus interval (the specific time interval [0 800] ms was selected). It was decided to take ERP data from the following 11 channels: 'POz', 'FC1', 'FC2', 'C1', 'Cz', 'C2', 'CP1', 'CPz', 'CP2', 'PO3', 'PO4'. Subject-dependent spatio-temporal features were extracted by calculating the mean amplitudes in 10 discriminant time intervals. And in the last step, the matrix was permuted in order [1, 3, 2] and then reshaped from 10 x 100 x 11, which led to the result of the feature matrix x with dimension 110 x 100 and matrix y with logical data (2 x 100).

The function for the MI paradigm, changing the sampling rate of the given EEG signal to the desired frequency of 100, then applies a filter for a frequency of 8-12 Hz. Further, the data was segmented in a time interval of [750 3500] ms. A Common Spatial Patterns (CSP) method with 4 patterns was applied. From the EEG epoch, subject-dependent spatio-temporal features were extracted by calculating the log variance. Thus, two matrices were obtained: x with features (8 x 100) and y with logical data (2 x 100).

For obtaining SSVEP-related features, first, the sampling rate of the raw data was changed to 100. Second, the band-pass filter was applied and it made it possible to obtain a frequency range of 0.5-40 Hz. Third, the EEG data were segmented from 0 to 4000 ms with respect to stimulus onset. Fourth, a range of channels with indices [23 : 32] was taken. Last, a general approach called multi-channel Canonical Correlation analysis (CCA) was implemented. It got following parameters: markers (1 - 'up', 2 - 'left', 3 - 'right', 4 - 'down') and reference frequencies (f1 = 12, f2 = 8.57, f3 = 6.67, and f4 = 5.45). As a result, a feature matrix of data x (4 x 100) and logical data y (4 x 100) were obtained.

Figure 2. Feature vector after data preprocessing. (Practical test)

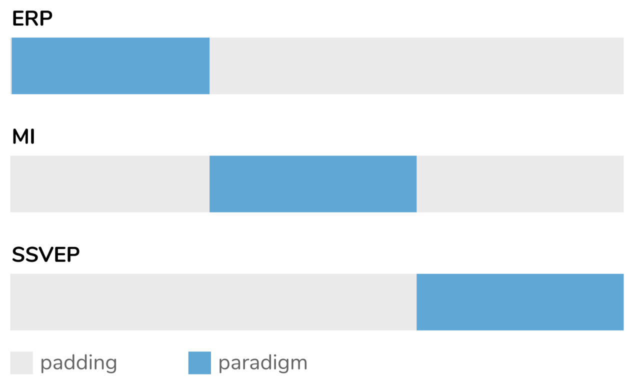

Figure 3. Padding visualization (Theoretical test)

The resulting individual paradigm-related features were then concatenated for further classification (Fig. 2).

C. Model Selection

1) Machine Learning Models: It was decided to compare accuracy performance across several conventional classifiers such as Linear Discriminant Analysis, Support Vector Classifier, Nearest Centroids Classifier and K-Nearest Neighbors classifier. Each of the suggested models underwent parametric grid search in order to obtain the best possible hyperparameters.

In this particular environment, data was presented as a concatenated feature vector, consisting of total number of features for ERP, MI and SSVEP paradigms. However, in order to avoid the curse of dimensionality, it was decided to select first 20 principle components of ERP data in order to reduce number of features from 110 to 20. It allowed for the total number of features for one sample of data to become 32: 20 principal components of ERP, 8 CSP filtered log-var features from MI and 4 correlation values from SSVEP. In order to avoid the problem of multi-class prediction, it was decided to padd remaining values of features for each paradigm with zeros, i.e. MI log-var features were padded from left and right with zeros in order to bring total number of features to 32. More descriptive padding can be see in Fig. 3. Target classes are concatenated into so-called "one-hot vector". Given that there are 2 classes for ERP - *target/non-target*, 2 classes for MI - *left arm/right arm* and 4 classes for SSVEP - [*up, right, left,*

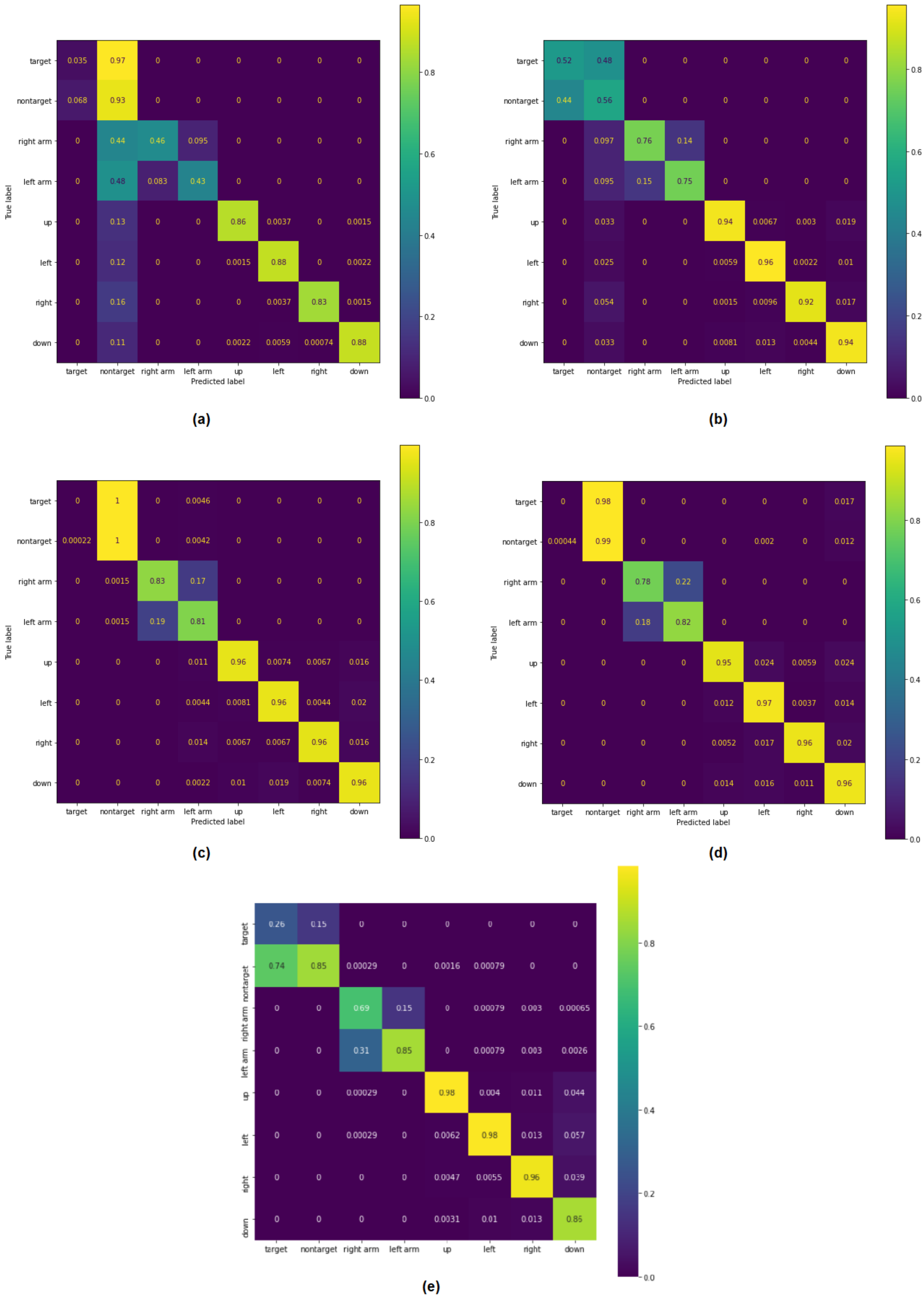

Figure 4. Estimator confusion matrices: (a) LDA, (b) NCC, (c) SVC, (d) KNN, (e) Neural Network (artificial noise).

down], our concatenated target vector consists of 8 classes. So that if data sample is ERP related, y vector will be 8 dimensional, but only first 2 values will encode ERP classes. The same practice is applied to MI and SSVEP classes (3^{rd} and 4^{th} values encode MI classes, and the rest values encode SSVEP values).

Parameterized search for all classifiers was implemented, resulting in the set of best parameters for each model as following:

- *LDA*:
  - The best shrinkage value is determined to be 0, making it unregularized LDA case;
  - Least squared solution seems to be the best in the given Cross Validation scenario;
- *NCC*:
  - Metric is set to be "Manhattan" distance;
  - Shrink threshold is determined to be 0;
- *SVC*:
  - Regularization parameter is 1;
  - Radial Basis Function kernel performed best in this scenario;
  - The best gamma coefficient is set to be $\frac{1}{Numberoffeatures}$
- *KNN*:
  - 11 neighbours is found to be an optimal value;
  - Metric is determined w.r.t. manhattan distance;
  - Weighting is determined by the inverse of distance.

After hyperparameters tuning, classifiers were trained and tested for the case of subject and session independence, results of which can be seen in Table I.

However, it is a good practice not to rely on raw accuracy scores only, but analyze models with respect to distinction between different classes. One of the suitable metrics in this case is confusion matrix. Confusion matrix shows how well predicted and actual labels are aligned and what is the rate of false predictions.

Confusion matrix for each classifier can be seen in Fig. 4 (a)-(d). From analysis of confusing matrix scores it can be determined, that linear classifiers performing well in MI and SSVEP paradigms, however target prediction rate for ERP are significantly lower than any other classes. This can be explained by the substantial imbalance between target and non-target data. For the rest paradigms, Support Vector Classifier performed better by a significant margin than other classifiers.

However it is worth mentioning that zero-padding considers the assumption of 0 noise when processing EEG data of unknown paradigm. Thus, those classifiers may be considered as one-fits all on the assumption that data was preprocessed without the introduction of noise from other paradigms pre-processing pipelines. In order to combat this assumption, it was decided to utilize deep learning models, such as ANNs. The reason is because ML classifiers perform poorly without zero noise assumption, dropping accuracy below 50%.

Table I
ACCURACY SCORES.

	LDA	NCC	SVC	KNN
Accuracy S1 - S1	69.98%	72.7%	85.91%	84.78%
Accuracy S1 - S2	71.17%	74.96%	87.17%	86.88%

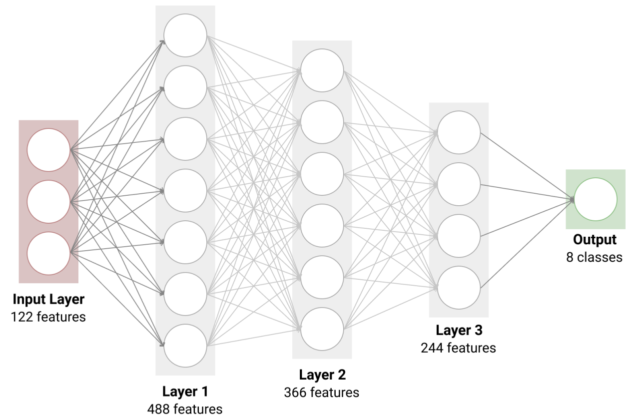

Figure 5. Neural Network structure.

2) Deep Learning Models: Many types of Deep Learning networks exist, with the most memorable being the Linear Artificial Neural Networks, Convolutional Neural Networks and Generative Adversarial Networks. CNN and GANs are often used in the image recognition and construction tasks, so those are, of course, not appropriate in the current situation. Thus, a standard densely connected network was chosen to be built.

The network consisted of three consecutive layers of varying dimensions, which can be seen in Fig. 5. In total, we had the input with a dimensionality of the concatenated feature vector, 3 hidden layers which are the multiples of this vector, and an 8-channel output layer. For the feasibility and ease of backpropagation, the hidden layers were tailored to have a Rectified Linear Unit (ReLU) output. This had ensured the partial disconnection of node outputs which are below than a 0 treshold. By doing constant backpropagation after calculating the difference between the provided and expected output, the network was able to learn the weights of the nodes and adapt to the given dataset.

Hyperparameters of 0.008 learning rate, 0.98 momentum, batch size of 200 and 50 epochs were acquired through multiple analytical trial runs. A small learning rate ensures that our updates are not too big, while the momentum further reduces the update strength as the network learns the most appropriate weights. Meanwhile, a batch size affect the loss function and ensures that we review multiple entries before deciding how to backpropagate. The remaining hyperparameter of epochs simply allows us to iterate over the same dataset multiple times. Otherwise, there would be not enough training data to approximate on.

Training and testing of the neural network consisted of two data sub-stages. Both of the stages had used the data from the first session for training, and the data from the second session for accuracy inference. During the first stage, a padded noisy data with the padding method similar to the one described in the Machine Learning section was used to estimate a theoretical accuracy on low-noise data. This provides us with an estimate of the maximum performance achievable under the conditions of reduced noise. A confusion matrix of the estimates can be seen in Fig. 4 (e).

As seen in the confusion matrix, the average inter class classification accuracy is roughly 85%, with the most confusion happening inside the original paradigm classes. That is, in theory, without a high level of noise, the classifier is able to distinguish the paradigms with a 99% rate, which is quite reasonable. The most class confusion here happens when the classifier tries to distinguish between the target and a non-target ERP response, and the right/left arm MI.

However, in real life, there is always a lot of noise, which is the original reason for employing a neural network in the first place. Numerous studies and experiments had shown that the ANN's are quite robust to a medium and even high levels of noise inside the data, and can notice intricate patterns inside the feature vector. Thus, the second stage of the classifier testing was dedicated to the study of fully concatenated noisy data (Fig. 2) acquired through a parallel processing of the three paradigms. The results were much better than the Machine Learning solutions and can be seen in Fig. 6.

The model had achieved a 71% accuracy on the training data from the first session, and a 70% success rate on the test data from the second session. This meant, that the chosen classifier was not only moderately robust to the noise acquired through the data pre-processing stage, but it was also session independent. However, it is not fully subject-independent, since each feature had relied on the CSP and CCA filtering. Upon further inspection of the confusion matrix, it can be seen that most of the misclassifications had happened between the MI and ERP paradigms. SSVEP, on the other hand, was quite strong, which leads to a conclusion that noise from generated from CCA is very random and is negligible most of the time. The accuracy of the proposed paradigm-independent classifier was combined to the previous study [2]. The comparison of the confusion matrices can be seen in Fig. 7.

The newly proposed solution had a higher average accuracy across the three paradigms, and in all of the cases, had exceeded the generally accepted threshold of 0.7. This further reinforces the idea that neural networks can perform better than the standard Machine Learning classifiers like LDA given a set of relatively noisy data.

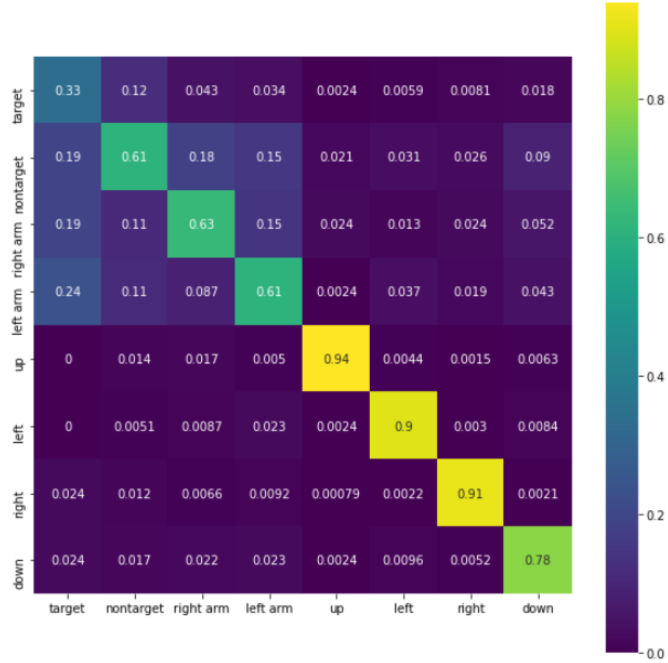

Figure 6. Neural Network confusion matrix (true noise).

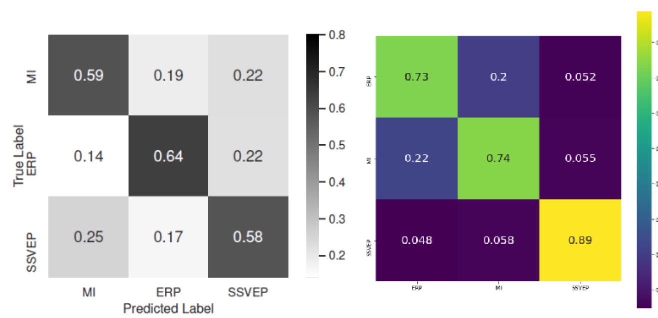

Figure 7. Paradigm accuracy comparison (Left = [2], Right = New).

Conclusion

In this study, multiple paradigm-independent classifiers for the three paradigms of MI, ERP and SSVEP were discussed. Before being used as the basis for training and testing, the data had to be pre-processed. ERP had required basic window time-slicing, while MI and SSVEP had relied on Common Spatial Patterns and Canonical Correlation Analysis respectively. Three separate functions were created in Matlab, which saved the results to files for training and test set. Several conventional classifiers such as LDA, SVC, NCC and KNN were then compared both in theoretical and practical accuracies. A standard densely connected network was also later built to for comparatory investigation purposes.

Theoretical accuracy was calculated using the low noise-padded features, while the practical accuracy was determined through direct concatenation of the parallel-processed features. All solutions had implemented parameters search with optimal parameters for each model. In the end, an eight-class classifier was produced. For low-noise data the theoretical threshold accuracy of for sessions S1 - S1 and S1 - S2 showed 69.98% and 71.17% for LDA, 72.7% and 74.96% for NCC, 85.91% and 87.17% for SVC, 84.78% and 86.88% for KNN, 85.5% and 83.2% for ANN respectively. However, practical classification with concatenated noisy data in practice had resulted in very low performance of the standard Machine Learning methods, which was lower than 50%. Only Artificial Neural Network had yielded considerable results by classifying the data with a 70% accuracy across multiple sessions. The proposed ANN had also outperformed the paradigm-classifying solution proposed in the previous studies

The confusion matrix was then used to determine metrics of analyzing models with respect to the distinction between different classes. Due to the substantial imbalance between target and non-target data, the target prediction rate for ERP is significantly lower. SSVEP had performed better than the other classifiers due to the specifics of the CCA. Future research possibilities include the study of spatio-temporal feature engineering to further reduce the noise in concatenated data to get closer to the proposed theoretical accuracy values.

References

[1] 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*, vol. 8, 01 2019. giz002.

[2] A. Li, K. Alimanov, S. Fazli, and M.-H. Lee, "Towards paradigm-independent brain computer interfaces," in *2020 8th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–6, Feb 2020.