2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)
Mexico City, Mexico. September 11-13, 2019

Classification of Motor Imagery Using Statistical Models

Rivas-Posada Eduardo
Division de Estudios de Posgrado e Investigacion
Tecnologico Nacional de Mexico/Instituto Tecnologico de Chihuahua
Chihuahua, Chih., Mexico
erivas@itchihuahua.edu.mx

Chacon-Murguia Mario Ignacio
Division de Estudios de Posgrado e Investigacion
Tecnologico Nacional de Mexico/Instituto Tecnologico de Chihuahua
Chihuahua, Chih., Mexico
mchacon@ieee.org

Ramirez-Quintana Juan Alberto
Division de Estudios de Posgrado e Investigacion
Tecnologico Nacional de Mexico/Instituto Tecnologico de Chihuahua
Chihuahua, Chih., Mexico
Jarq1910@gmail.com

***Abstract*— Nowadays, brain computer interfaces have become very popular in rehabilitation therapy, tools to help people with disabilities, daily application and videogames. There exist different models that use different types of brain signals. One of them is the model that employs sensorimotor rhythms, based on motor imagery. This work describes a semi-supervised method to classify motor imagery signals of the hand, foot, and a class that indicate inactivity, or relaxation. The classifier uses features obtained with statistical moments, and independent component analysis. Two classifiers were evaluated, one based on the k-means which achieved a performance of 85.16% during training, and 85.76% in validation. The other classifier tested was based on the knn. This classifier had a training performance of 90.45% and 90.64% in validation.**

Keywords—Motor imagery, k-means, knn, ICA, BCI, statistical moments.

I. Introduction

In recent years, brain computer interfaces, BCI, have been studied as alternative ways for communication, to interact with virtual environments, and to manipulate prosthesis. One of the most common methods to design BCI employs electroencephalographic signals, EEG, obtained through non-invasive methods. Some of the EEG signals used in BCI are visual evoked potentials, event-related potentials, and the sensorimotor rhythms, which include the μ wave, 7.5-12Hz, and the β wave 12-30Hz.

A motor imagery, MI, can be defined as the mental execution of a motor act, without a physical action. It is broadly accepted that a MI involves the same brain regions/functions related to the programming and preparation of such movements. According to the previous assumption, the main difference between the motor execution and the motor imagery is that in the MI case the execution is blocked at some corticospinal level [1].

In a BCI system, once the EEG signals have been registered, the following stages are applied. The EEG is preprocessed. Then features are extracted from the signals, and finally a classifier determines the type of event that is currently generated. The works related to BCI focused on improving the previously mentioned stages in order to achieve the best performance of the BCI. For example in [2], a Bayesian neural network is used to recognize multiple motor imageries. The work described in [3] presents a motor imagery detector of the left and right hands by transforming the EEG signal into time-frequency images. These new signals become the input of a convolutional neural network. Another work reported in [4], employs linear discriminant functions. A model using independent component analysis, ICA for preprocessing, and support vector machine as the classifier is described in [5]. A recent proposal to classify MI is based on non-supervised learning. This type of system is presented in [6], it also involves wavelet information, and achieves a performance of 82.14%. Other works related to BCI designs are; a method that uses *k-nearest neighbor*, [7], which presents a good performance of 94.75% to classify MI. A model that extract features from the power spectral density and employs *knn* as a classifier has a performance of 90% with a real time response of 0.0531 seconds is reported in [8].

Other interesting on-line method presented in [9], applies ICA and SVM for a 4-channel EEG data recorded for motor imagery brain computer interfaces. They recorded brain signal data from 5 subjects and the mean classification accuracy after applying ICA for a 7-second time window was 77%.

Based on the previous information, we intended to develop a semi-supervised method to classify two types of MI, hand and foot, and the class relaxation. The method employs statistical information of the EEG signal by choosing a synchronous model, where the time interval of the MI is known.

II. THEORETICAL CONCEPTS

A. Data base

The database used was the *BCI Competition III dataset IVa* [10] which includes the brain activity of 5 subjects acquired with 18 electrodes. Each subject executed 280 MIs for the right hand and the feet. Each MI was executed during 3.5 seconds, with a resting time of 2 seconds between a MI session and the next. The resulting signals were filtered with a band pass filter of 0.05 to 200Hz, and they were digitalized at 16 bits (0.1 μV). Finally, the signal was subsampled at 100 Hz, selecting every tenth sample.

978-1-7281-4840-3/19/$31.00 ©2019 IEEE

B. Statistical moments

When a set of data presents a strong central tendency, that is, it is observed a tendency to cluster the data based on a particular value, then the data can be characterized by a measure termed statistical moments. These moments can be used as features to classify MI of the right hand and the feet. The statistical moments are obtained as follow. Let us consider *P(x)* the frequency of occurrence of and event *x*. The probability of *x*, *p(x)* is computed by

$$p(x) = \frac{P(x)}{N} \tag{1}$$

where N is the total number of possible events. For a continuous random variable, the mean or expected value of x is determined by

$$\mu_x = \int x p(x) dx \tag{2}$$

In our case, the random variable is discrete, therefore we have

$$\mu_x = \int x \frac{P(x)}{N} dx \triangleq \frac{1}{N} \sum_k x_k P(x_k) \equiv \frac{1}{N} \sum_{i=1}^{N} x_i \tag{3}$$

with unbiased variance

$$\sigma_X^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu_x)^2 \tag{4}$$

C. Probability density of the normal distribution

The probability density function, pdf, considered in this work to model the data will be the normal distribution, defined as

$$p\left(x \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} (x-\mu)^2\right] \tag{5}$$

where $x \in \mathbb{R}^{1\times n}$, μ and σ^2, represent the data, mean and standard deviation respectively, and n is the number of samples used to generate the normal distribution.

D. Independent component analysis

Independent component analysis, ICA, is a statistical technique where random data are linearly transformed into components with maximum independence. This technique will be used to increase the discrimination power of the features used for the classifier. Let $\varsigma = [\varsigma_1, \varsigma_2, \ldots, \varsigma_m]$. It can be assumed that this vector is generated as a linear combination of m independent signals denoted by $s = [s_1, s_2, \ldots, s_m]$; that is, we have the equation system

$$\begin{aligned} \varsigma_1 &= a_{11}s_1 + a_{12}s_2 + \ldots + a_{1m}s_m \\ \varsigma_2 &= a_{21}s_1 + a_{22}s_2 + \ldots + a_{2m}s_m \\ &\vdots \\ \varsigma_m &= a_{m1}s_1 + a_{m2}s_2 + \ldots + a_{mm}s_m \end{aligned} \tag{6}$$

The coefficients a_{ij} determine a mixing matrix A, such that $\varsigma = As$. In this method, it is assumed that x has a non-Gaussian pdf.

E. K-means

K-means is a statistic non-supervised clustering method that separates data into k clusters or sets [11]. Considering the data set $\{x_1, \ldots, x_n\}$. The steps of the algorithm are the following:

1. Select k and randomly propose the means or centroids m_j. j represents the classes index.
2. Compute the Euclidean distance between the data and the means.
3. Obtain the cardinality C_j.

$$C_j = \left\{\arg_j \min\left|d_{ij}\right|\right\} \tag{7}$$

4. Update the means

$$m_j = \frac{1}{\left|C_j\right|} \sum_{i \in C_j} x(i) \tag{8}$$

5. Finally, evaluate if $m_j^t \approx m_j^{t-1}$. If it is true then finish the algorithm, otherwise return to step 2.

F. K-nearest neighbors

*k-nearest neighbor*s or *knn* is a simple method to estimate a density. The probability that a data is inside a volume *Vol* centered in a point x is given by

$$\theta = \int_{Vol(x)} p(x) dx \tag{9}$$

For a small volume $\theta \sim p(x)Vol$. Besides, the probability of θ can be approximated by the proportion of samples inside *Vol*. If k is the number of samples, out of n, that are inside *Vol* (k is a function of x) then $\theta \sim k/n$. Therefore, the estimation of the density is

$$\hat{p}(x) = \frac{k}{nVol} \tag{10}$$

Finally, if the classic counting method is applied, a decision rule will be generated $k_m > k_i$. The decision rule indicates that x is assigned to the class with the maximum number of votes considering the k nearest neighbors [12].

III. System Design

This section describes the design of the classifier of the MI classes termed MI-hand, MI-foot, and relaxation using semi-supervised algorithms

A. Signal Preprocessing

The signals employed in the design of the system are from the database *BCI Competition III dataset IVa* [10]. The signals selected come from the electrodes C3, Cz and C4, considering the 10-20 International system, defining the set of channels

$$Ch = \{C3,\ Cz,\ C4\} \tag{11}$$

Each one of the 5 subjects has a different number of samples for the training stage, and there is no knowledge of the exact time when the MI was generated. It is only known the time when the system indicated the start to generate the MI. There are 350 samples of each MI sampled at 100hz. As it is shown in Figure 1, the first 100 samples were discarded to yield an epoch, the next 200 samples. The features used for the classifier are obtained from the epochs. The last 50 samples are also eliminated.

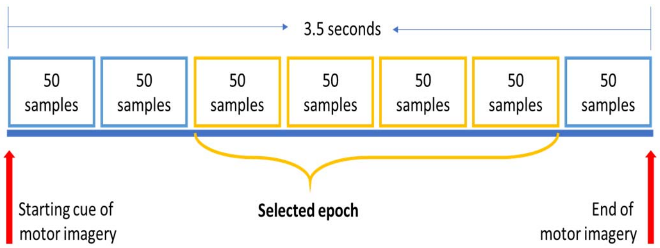


Fig. 1. Generation of an MI epoch.

The samples of the relaxation were obtained with the 100 samples after the end of a MI. The length of the epochs was determined experimentally based on what length generated better results. The number of epochs for each channel is defined as *p*. Since the classes are known, the epochs are extracted and defined from each channel

$$X_p^{Ch}(n) \qquad n=1,2,....200; \quad p\in \mathbb{N}^{1\times E_{sujeto}} \tag{12}$$

where *n* represents the number of samples of each epoch and *p* changes, E_s, from one subject to other, as shown in Figure 2.

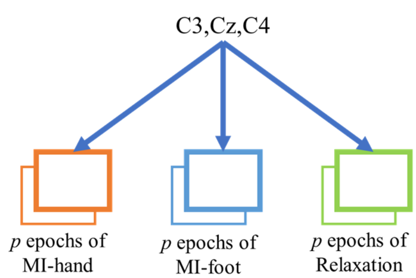


Fig. 2. Epochs by class, and form each channel.

Each epoch was normalized by

$$\overline{X}_p^{Ch}(n)=\frac{X_p^{Ch}(n)}{\left|X_p^{Ch}(n)\right|} \qquad \overline{X}_p^{Ch}\in \mathbb{R}^{3\times E_s\times 200} \tag{13}$$

B. *Feature extraction*

Once the behavior of the signal in the time domain was observed, the process to extract was determined using statistics like the histogram. In this work the normalized signal amplitude histogram evaluated in a specific rank, for all the samples of all epochs of each subject was computed by

$$hist(Ch,bin)=\frac{10}{N}\sum_{p=1}^{E_s}\sum_{n=1}^{200} g(Ch,p,bin,n) \tag{14}$$
$$bin=1,2,...,20$$

where *g(Ch,p,bin,n)* is a binary variable and it is defined in (15), *bin* is each one of the 20 intervals where the amplitude was evaluated to determine if it was inside of that rank.

$$g(Ch,p,bin,n)=\begin{cases}1 & \text{iff } T(bin)\le \overline{X}_p^{Ch}(n)<T(bin+1)\\ 0 & \text{otherwise}\end{cases} \tag{15}$$
$$T(bin)=0.1(bin-1)-1$$

here, *T(bin)* is the right limit of the interval under evaluation. Figure 3 illustrates the histograms of the subject *aa* for each one of the classes and for each one of the channels. It can be noticed that the distribution is like a Gamma distribution, with non-zero mean, which allows applying ICA to obtain the independent components.

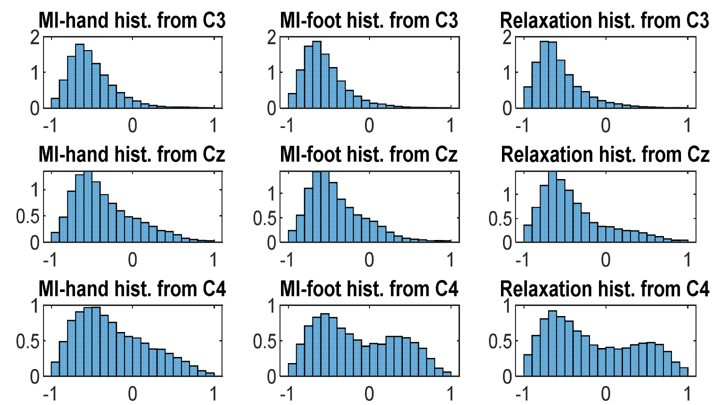


Fig. 3. Histogram for each channel and for class of the subject *aa*.

In order to model the data, and to make a better analysis, the mean and standard deviation were computed

$$\mu_{Ch} = \frac{1}{N}\sum_{p=1}^{E_{sujeto}}\sum_{n=1}^{200}\overline{X}_p^{Ch}(n) \tag{16}$$

$$\sigma_{Ch}^2=\frac{1}{N-1}\sum_{p=1}^{E_{sujeto}}\sum_{n=1}^{200}\left[\overline{X}_p^{Ch}(n)-\mu_{Ch}\right]^2 \tag{17}$$

Then, the data was simulated through a normal distribution to visualize the overlapping

$$p\left(\overline{X}_p^{Ch}\mid \mu_{Ch},\sigma_{Ch}^2\right)=\frac{1}{\sqrt{2\pi\sigma_{Ch}^2}}\exp\left[-\frac{1}{2\sigma_{Ch}^2}\left(\overline{X}_p^{Ch}-\mu_{Ch}\right)^2\right] \tag{18}$$

Figure 4 shows the normal distributions for the subject *aa*.

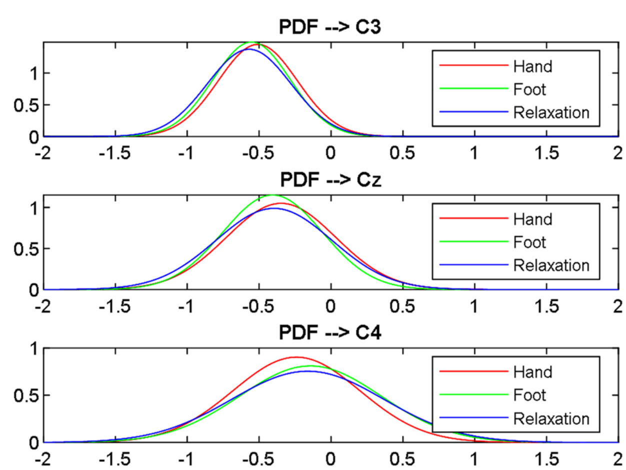


Fig. 4. Normal distributions of the data of the subject *aa* using equation (18).

The graphs in Figure 4 show that data cannot be discriminated between lasses. Therefore, the ICA process was applied to the data to provide class separation.

$$S_p^{Ch} = \overline{X}_p^{Ch} \cdot A^{-1} \qquad A \in \mathbb{R}^{3\times 3} \tag{19}$$

where A is the mixing matrix that allows the linear combination to represent the data into a new space. Thus, we expected that S_p^{Ch}, may generate class separation. Figure 5 illustrates the EEG signals after the ICA process.

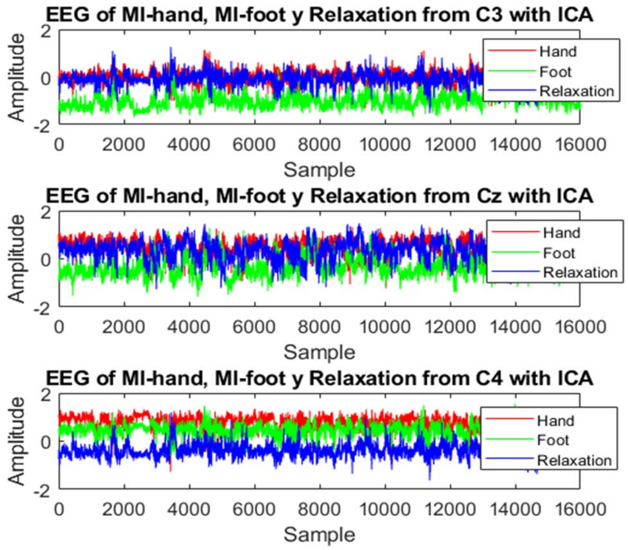


Fig. 5. EEG signal of each channel and of each epoch set of subject *aa* after ICA.

It can be observed from Figure 5 that now the data present some class separation in the channels C3 and C4. In C3 the classes MI-foot is discriminated from the classes MI-hand and relaxation. In C4 the class MI-hand is discriminated from the classes MI- foot and relaxation.

Using this transformed data, we now model the new data using the statistics

$$\begin{aligned} s\mu_{Ch} &= \frac{1}{N}\sum_{p=1}^{E_{sujeto}}\sum_{n=1}^{200} s_p^{Ch}(n) \\ s\sigma_{Ch}^2 &= \frac{1}{N-1}\sum_{p=1}^{E_{sujeto}}\sum_{n=1}^{200}\left[S_p^{Ch}(n) - s\mu_{Ch}\right]^2 \end{aligned} \tag{20}$$

With distribution

$$p\left(S_p^{Ch} \mid s\mu_{Ch}, s\sigma_{Ch}^2\right) = \frac{1}{\sqrt{2\pi\left(s\sigma_{Ch}^2\right)}} e^{-\frac{1}{2\sigma_{Ch}^2}\left(S_p^{Ch} - s\mu_{Ch}\right)^2} \tag{21}$$

The new distributions for the transformed data of the subject *aa* are shown in Figure 6.

According to the information observed in Figure 6, we can notice that the channels C3 and C4 are sufficient to discriminate among the three classes by making combinations among their distributions. From C3 we can determine the class MI-foot, and from C4 we can discriminate between MI-hand and relaxation. Figure 7a presents a plot in the feature space of the channels C3 and C4 before ICA, and Figure 7b after ICA.

C. Classification

The classification design involves the use of two clustering algorithms *k-means*, non-supervised method, and *knn*, non-parametric method.

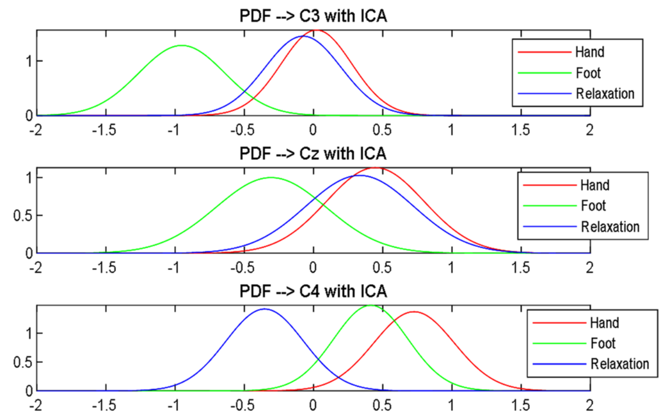
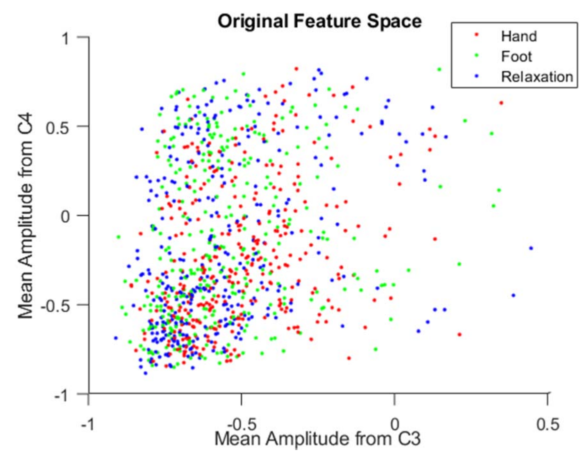
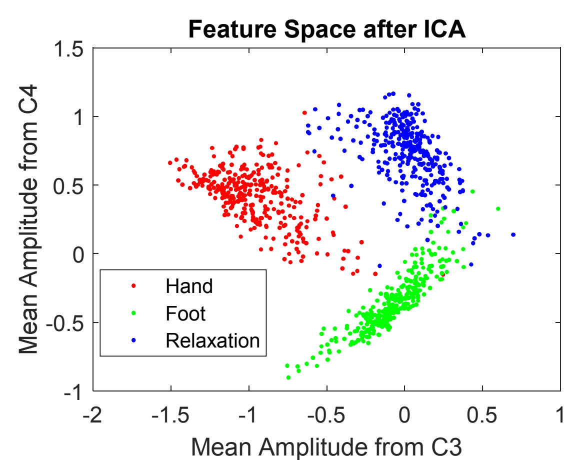


Fig. 6. New data distributions of the subject aa after ICA.



a)



b)

Fig. 7. Plots in the feature space of the channels C3 and C4, subject *aa*. a) Before ICA, b) After ICA.

The *k-means* was applied as follows.

1. The number of cluster to generate were three according to the number of classes, $J=\{MI\text{-}hand,\ MI\text{-}foot,\ relaxation\}$. The means $m_j \ j \in J$, were randomly determined.

2. The distance between the samples and the means were compute by

$$d^2_{S,m} = \left(S_p^{Ch} - m_j\right)\left(S_p^{Ch} - m_j\right)^T \tag{22}$$

3. The cardinality of the cluster j is obtained as

$$C_j = \left\{\arg_j \min \left|d_{S,m}\right|\right\} \tag{23}$$

4. Update the means

$$m_j = \frac{1}{\left|C_j\right|} \sum_{n \in C_j} S_p^{Ch}(n) \tag{24}$$

5. Finally evaluate $m_j^t \approx m_j^{t-1}$. If it is true end the algorithm, otherwise proceed to step 2.

The clustering result with the centroids of the *k-means* is illustrated in Figure 8.

If we compare the plot of Figure 8 with the plot of Figure 7, we can notice a crisp clustering of the *k-means*, which produces that certain samples are misclassified. Therefore, we decided to test another algorithm, the *knn*.

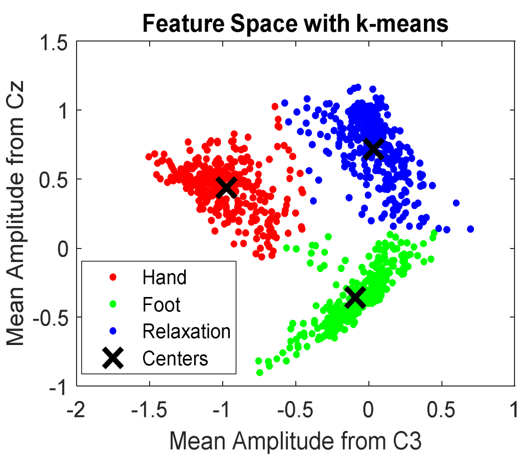


Fig. 8. Results of the *k-means* for teh data of the subject *aa*.

Several classification tests were evaluated with the subject *aa* using 5 nearest neighbors and the Euclidian distance. The classification was achieved by the next condition

$$\begin{array}{lll} S_p^{Ch}(n) \in Cl_j & \text{iff} \quad k_j > k_l & \\ S_p^{Ch}(n) \notin Cl_j & \quad \text{otherwise} & \end{array} \qquad \forall l \in J \quad j \neq l \tag{25}$$

that is, $S_p^{Ch}(n)$ belongs to the cluster Cl_j if the number of votes to belong to the class j, k_j is greater than the number of votes to belong to the other classes l. The results of the *knn* are shown in the Figure 9.

In the next section, the results for all the subjects of the database are presented and analyzes.

IV. Results

This section describes the results of the proposed methodology, based on the *Ground Truths* provided in the database used. Table I shows the results employing the training data and presents a comparison between the number of samples

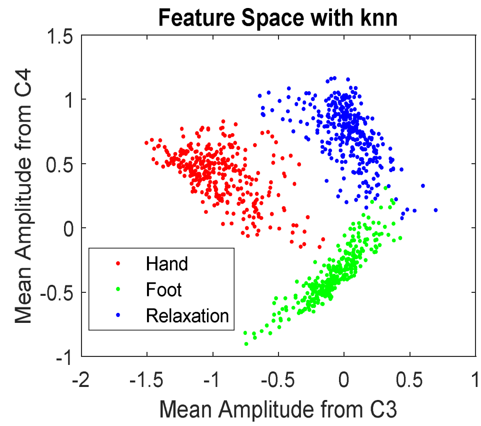


Fig. 9. Classification results with *knn* of the subject *aa* data.

by channel of each subject and the performance of the two classifiers. From Figure 9, we can observe that unlike *k-means, knn* is less crisp in its clusters, and may yield better performance.

TABLE I. Comparison between the two classifiers and the training samples.

Subject	Samples by channel	*k-means* performance	*knn* performance
aa	16,000	96.3542%	98.0208%
al	22,200	88.2132%	90.3904%
av	8,200	79.065%	86.1789%
aw	5,200	95.5128%	96.4744%
ay	2,000	66.6667%	79.1667%
Average	-	85.1624%	90.0462%

At a first view of the data in Table 1, it seems that the lower the number of samples, the lower the performance of the classifier. However, the subject *aw* presents a higher performance than the performance of the subject *aa* with less samples. Or in the case of the subject *al* who has a lower performance than the subject *aa*, and employs more samples.

To validate the results, we used the complete set of data, 28000 samples for channel. The results are shown in Table II.

TABLE II. Validation results using the complet set of data.

Subject	*k-means* performance	*knn* performance
aa	97.2619%	98.5119%
al	90.5357%	91.9048%
av	73.0952%	80.2976%
aw	95.1786%	96.6667%
ay	72.7381%	85.8333%
Average	85.7619%	90.6429%

TABLE III. COMPARISON OF THE PROPOSED METHOD WITH THE WORK REPORTED IN [7].

Classifier / **Subject**	***aa***	***al***	***av***	***aw***	***ay***	**Average**
SVM	88.09%	83.57%	82.97%	84.52%	92.14%	89.48%
knn	94.88%	92.38%	88.92%	94.76%	91.19%	94.57%
Random Forest	80.36%	75.83%	72.38%	75.24%	98.45%	78.09%
ANN (MLP)	75.83%	77.02%	76.55%	72.14%	79.52%	67.40%
C4.5	57.50%	64.88%	59.29%	60.95%	95.48%	63.55%
Naïve Bayes	63.57%	65.71%	55.83%	58.69%	53.21%	53.02%
Proposed method	98.51%	91.90%	80.29%	96.66%	85.83%	90.6429%

It can be observed that the average performance is kept, and that the knn has the best performance. In order to make a comparison with other work, the proposed methodology was compared with the work in [7]. This work used the same dataset. A comparison of the proposed method with the methods reported in [7] is presented in Table III. The proposed method turns to be better than the other methods for all subjects, except in the subjects *al*, *av*, and *ay* using the *knn* method reported in [7]. However, there are some important issues to point out. The work in [7] does not report the amount of electrode used. They only say that the *Wavelet Packed Decomposition* and the multi-scale *Principal Component Analysis* were used to extract the features. In our work, the proposed method achieved a high performance employing only 3 electrodes. Therefore, we can consider that our proposed methodology is competitive. Besides, our method deal with three classes, and the work in [7] only works with two classes.

Other method reported in [9] can also be compared with our method. They generated their own database with 2 motor imageries, obtaining an average performance of 77% considering 5 test subjects, and using 4-channel EEG. In our work, we used ICA and *knn*, only 3 electrodes, recognized 3 motor imageries, and achieved a performance of 90.64%, but it works offline.

V. CONCLUSIONS

In this work, it was designed a methodology to classify two MI classes, hand and foot, and the class relaxation, of different subjects. The features used to classify the samples were based on statistics and ICA transformation. The classifiers were based on two clustering algorithms the *k-means* and the *knn*. The *knn* achieved a maximum performance of 90.6%, and the *k-means* 85.76%. Also, the proposed method achieves competitive results compared with others works reported in the literature. As future work, it is expected to include more features to improve the performance of the method.

ACKNOWLEDGMENT

This research was funded by Tecnologico Nacional de Mexico/ I.T. Chihuahua under grant 5684.16-P.

REFERENCES

[1] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[2] L. He, D. Hu, M. Wan, Y. Wen, K. M. Von Deneen, and M. Zhou, "Common Bayesian Network for Classification of EEG-Based Multiclass Motor Imagery BCI," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 46, no. 6, pp. 843–854, 2016.

[3] J. Zhang, C. Yan, and X. Gong, "Deep convolutional neural network for decoding motor imagery based brain computer interface," in *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2017, pp. 1–5.

[4] S. Qu, J. Liu, W. Chen, J. Zhang, and W. Chen, "Pattern Recognition of Motor Imagery EEG Signal in Noninvasive Brain-Computer Interface," *2018 13th IEEE Conf. Ind. Electron. Appl.*, pp. 1814–1819, 2018.

[5] A. Dey, S. Bhattacharjee, and D. Samanta, "Recognition of motor imagery left and right hand movement using EEG," in *2016 IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2016 - Proceedings*, pp. 426–430.

[6] A. M. Eldeib and S. Member, "Unsupervised Processing Methods for Motor Imagery-Based Brain- Computer Interface," *2018 IEEE 4th Middle East Conf. Biomed. Eng.*, pp. 106–111, 2018.

[7] M. Behri, A. Subasi, and S. M. Qaisar, "Comparison of Machine Learning Methods for Two Class Motor Imagery Tasks using EEG in Brain–Computer Interface," in *Advances in Science and Engineering Technology International Conferences (ASET)*, 2018, pp. 1–5.

[8] S. Bhaduri, A. Khasnobish, R. Bose, and D. N. Tibarewala, "Classification of Lower Limb Motor Imagery Using K Nearest Neighbor and Naive-Bayesian Classifier," in *3rd International Conference on Recent Advances in Information Technology (RAIT)*, 2016, pp. 499–504.

[9] I. Rejer and P. Gorski, "Independent component analysis in a motor imagery brain computer interface," in *17th IEEE International Conference on Smart Technologies, EUROCON 2017 - Conference Proceedings*, 2017, pp. 126–131.

[10] B. Blankertz *et al.*, "The BCI competition III:Validating alternative approaches to actual BCI problems," *IEEE Tansa. Neural Syst. Rehabil. Eng. 14*, vol. 14, no. 2, pp. 153–159, 2006.

[11] S. Lu *et al.*, "Clustering method of raw meal composition based on PCA and kmeans," in *37th Chinese Control Conference (CCC)*, 2018, pp. 9007–9010.

[12] S. Shukla and R. K. Chaurasiya, "Emotion Analysis Through EEG and Peripheral Physiological Signals Using KNN Classifier," in *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*, 2019, vol. 30, pp. 97–106.