

dx.doi.org/10.17488/RMIB.45.1.3

E-LOCATION ID: 1397

Study of the Length of time Window in Emotion Recognition based on EEG Signals

Estudio de la Longitud de Ventana de Tiempo en el Reconocimiento de Emociones Basado en Señales EEG

Alejandro Jarillo Silva¹ , Víctor Alberto Gómez Pérez¹ , Omar Arturo Domínguez Ramírez²  .

¹Universidad de la Sierra Sur, Oaxaca - México

²Universidad Autónoma del Estado de Hidalgo, Hidalgo - México

ABSTRACT

The objective of this research is to present a comparative analysis using various lengths of time windows (TW) during emotion recognition, employing machine learning techniques and the portable wireless sensing device EPOC+. In this study, entropy will be utilized as a feature to evaluate the performance of different classifier models across various TW lengths, based on a dataset of EEG signals extracted from individuals during emotional stimulation. Two types of analyses were conducted: between-subjects and within-subjects. Performance measures such as accuracy, area under the curve, and Cohen's Kappa coefficient were compared among five supervised classifier models: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), and Decision Trees (DT). The results indicate that, in both analyses, all five models exhibit higher performance in TW ranging from 2 to 15 seconds, with the 10 seconds TW particularly standing out for between-subjects analysis and the 5-second TW for within-subjects; furthermore, TW exceeding 20 seconds are not recommended. These findings provide valuable guidance for selecting TW in EEG signal analysis when studying emotions.

KEYWORDS: electroencephalogram, emotion recognition, machine learning, time window length

RESUMEN

El objetivo de esta investigación es presentar un análisis comparativo empleando diversas longitudes de ventanas de tiempo (VT) durante el reconocimiento de emociones, utilizando técnicas de aprendizaje automático y el dispositivo de sensado inalámbrico portátil EPOC+. En este estudio, se utilizará la entropía como característica para evaluar el rendimiento de diferentes modelos clasificadores en diferentes longitudes de VT, basándose en un conjunto de datos de señales EEG extraídas de individuos durante la estimulación de emociones. Se llevaron a cabo dos tipos de análisis: entre sujetos e intra-sujetos. Se compararon las medidas de rendimiento, tales como la exactitud, el área bajo la curva y el coeficiente de Cohen's Kappa, de cinco modelos clasificadores supervisados: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) y Decision Trees (DT). Los resultados indican que, en ambos análisis, los cinco modelos presentan un mayor rendimiento en VT de 2 a 15 segundos, destacándose especialmente la VT de 10 segundos para el análisis entre los sujetos y 5 segundos intra-sujetos; además, no se recomienda utilizar VT superiores a 20 segundos. Estos hallazgos ofrecen una orientación valiosa para la elección de las VT en el análisis de señales EEG al estudiar las emociones.

PALABRAS CLAVE: aprendizaje automático, electroencefalograma, longitud de ventana de tiempo, reconocimiento de emociones

Corresponding author

TO: Omar Arturo Domínguez Ramírez

INSTITUTION: Univerdad Autónoma del Estado de Hidalgo

ADDRESS: Kilómetro 4.5 carretera Pachuca-Tulancingo en la Colonia Carboneras, Mineral de la Reforma, Hidalgo.

EMAIL: omar@uaeh.edu.mx

Received:

01 November 2023

Accepted:

31 January 2024

INTRODUCTION

Human emotions have an impact on day-to-day actions and decisions. They are important aspects of communication and human emotional intelligence, in other words, the capacity to understand and manage emotions is crucial for the success of personal interactions^[1]. There are areas of opportunity in technological innovation in the study of emotions, for example, affective computing, which has the objective of equipping machines with emotional intelligence to improve human-computer interactions (HCI)^[2]. Another area is human-robot interaction (HRI), which aims to make robots capable of interpreting and expressing emotions similar to those of human beings and thereby modulate relevant aspects of the interaction^[3]. Thus, the possible applications of an interface capable of evaluating human emotional states are numerous, ranging from medical diagnoses, rehabilitation processes, and digital commerce to new teaching methods.

In HCI, non-invasive, reliable and accessible portable sensors play an important role in the study of emotions. This is because in many work environments the use of modern technologies has increased considerably, with the objective of improving the interaction between the user and the technologies. However, many of these systems impose high demands on cognitive states, which can lead to the arousal of a person's negative emotions^[4]. One of the most effective approaches to emotion recognition is based on the use of physiological signals. Among the numerous physiological signals, it has been reported that brain signals are found to be directly related with human emotions^{[5][6][7]}. Favorable results have been reported when evaluating different classification algorithms, this during the application of diverse feature extraction techniques through electroencephalographic (EEG) signals^{[8][9][10]}.

The estimation of emotions in real-time involves processing a continuous stream of biosignals with the

lowest latency possible. Research in system development for emotional state detection is mainly focused on recognition methodology^[11]. On the other hand, the field of BCI systems using EEG signals is constantly evolving. For example, in^[12], an algorithm for attention detection during mathematical reasoning is proposed. In addition, in^[13], an analysis of EEG signals is performed using diverse classification techniques, achieving significant results in motion detection. Another relevant contribution is presented in^[14], with the introduction of a new neural network model designed for classification with a limited amount of motor imagery data. In^[15], a methodology based on EEG signals is presented to detect the level of attention in children, applying a multilayer perception neural network model. Finally, in^[16], a methodology based on motor imagery for a BCI system is presented, using convolutional neural networks. These papers highlight the diversity of approaches in current research on emotional estimation and the development of BCI systems using brain signals.

However, segmentation plays an important role in achieving real-time or continuous monitoring of emotional states (that is to say, the selection of the time window), which has received little attention and requires further research. The majority of works reported utilize different time windows (TW) as inputs for model training^{[6][7]}. The consequences of employing different TW could influence the trained model to be unsuitable for application in real-time emotion recognition because the knowledge learned from the model is related to the sampled features for subsequent detection. In addition, combining different TW in the same analysis would cause the trained model to be inconsistent due to the changing characteristics of EEG signals in temporal sequences^[17].

In order to avoid this problem, analyses at different TW lengths of temporal sequences have been developed using EEG signals. For example, Lin *et al.*, report in their research a TW of 1 second to calculate the

spectrogram of an EEG, in order to investigate the relationship between emotional states and brain activities, with an accuracy of 82.29 % in their model *et al.*^[18]. On the other hand, Zheng *et al.* report a TW of 4 seconds without overlap in order to extract combined EEG features with eye tracking with the objective of carrying out emotional recognition tasks with an accuracy in their model of 71.77 %^[19]. Zhuang *et al.* used a TW of 5s for feature extraction and emotion recognition based on empirical mode decomposition with an accuracy of 69 % in their model^[20]. It is noted that different TW lengths have been used in EEG signal processing, but the appropriate length measurement for the detection of emotions is not established. Ouvan *et al.* carried out a study of the size of TW with the experiment-level batch normalization method in feature processing, in their findings they report that the best performing TW length was 2 seconds^[21]. Healey *et al.* present a study of emotions (emotion recognition) with windows of 60, 180 and 300 seconds, but they do not report the performance of each TW^[22]. Gioreski *et al.* report a laboratory study for stress detection with TW between 30 and 360 seconds, and in their findings indicate the window of 300 seconds presents better performance^[23]. However, few studies have examined the effect of wavelength on the performance of classifier models during emotion recognition.

Among the most-used parameters for measuring a classifier model's performance are accuracy (ACC), defined as the fraction of predictions that the model classifies correctly, the precision or positive predictive value (PPV) that is the percentage of correct classifications of the model within the predictions of positive emotions, completeness or sensitivity (Recall) which is defined as the proportion of emotions that were correctly identified as having a condition, true positive, over the total number of emotions that are actually positive^[24]. Other research employs the Area Under the Curve (AUC) of Receiver Operating Characteristic (ROC) in order to evaluate the

performance of classifier algorithms^{[5][25]}. Another parameter used is specificity, which measures the number of subjects who were correctly identified as having a negative emotion over the total number of subjects who actually present a negative emotion. However, for balanced studies that have on average almost the same amount of data for all categories (different emotions) the performance measures are ACC, AUC and Cohen's Kappa coefficient^{[8][10][12][26][27]}.

Another potential drawback in the study of emotions arises when variables are analyzed and reported at the group level rather than being used to evaluate the emotions in an individual. This is to say, the associations between physiological variables and emotions found through a group-level analysis may not generalize the case for evaluating emotions in an individual, as they cannot be sufficiently robust to reliably assess the emotional state at a given time for an individual^[28].

Therefore, there is a gap in defining the size of the TW that will induce the best performance from the sorting algorithm for recognizing emotions. Furthermore, it has not been clearly established if the study of data must be made on an individual or on a group level. Accordingly, the aim of this research is to evaluate classification performance in detecting emotions via EEG signals at the group and individual levels by conducting a systematic comparison of TW values below 30 seconds. The performance metrics selected for this evaluation are ACC, AUC and Cohen's Kappa coefficient.

This article is divided in the following structure: first, an Introduction is provided that establishes the contextual framework of the research. The second section addresses the Materials and Methods, detailing the proposed approach to conduct the study. Next, in the third section, the Results and Discussions are presented, highlighting the observations and analysis derived from the research. Finally, the last section covers the Conclusions, summarizing the key findings and their implications.

MATERIALS AND METHODS

In order to carry out this study of emotion recognition, the data set is employed from^[29]. For this data set, controlled experiments were designed to induce positive, negative, and neutral emotions from video clips. Participants ranged in age from 19 to 35 years (mean age 24.3). There were 8 women and 17 men.

To carry out the performance study of classification models with different TW the flow presented in Figure 1 is implemented.

The dataset contains the signals of 25 subjects from 14 electrodes of the EEG device. Every signal was broken down into alpha, beta, gamma and delta frequency bands. In this study, MATLAB 2017a libraries were used for the preprocessing of data, feature extraction and analysis of the classification algorithms. The workstation consists of a PC with an i7 processor, 8GB of RAM and 4GB of Nvidia GeForce.

Acquisition of EEG Data

The EEG sensing device used was the Emotiv EPOC+, which has 14 channels: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, plus two references: Common Mode Sense (CMS) and Driven Right Leg (DRL) in P3 and P4 (see Figure 2). This device has been widely used in different research related to emotions and the study of pathologies^{[30][31][32]}. The data obtained directly from the file of every subject were the preprocessed theta (4-8 Hz), alpha (8-12 Hz), low beta (12-16 Hz), high beta (16-25 Hz) and gamma (25-45 Hz) band signals.

Data Segmentation

The main objective of this study was to evaluate and compare the performance of different classifiers for emotion recognition with different TW shorter than 30 seconds. In Table 1 the lengths of TW considered is presented as well as the number of blocks obtained. Therefore, the format of the

features in each trial was defined as (14x5x300); 14 representing the number of electrodes, 5 representing the frequency bands and 300 being the number of features extracted from the corresponding trial. In total 10 different TW lengths were examined in order to investigate the effect on the performance of classifier models in the study of emotions employing EEG sensors at a between-subject and within-subject level.

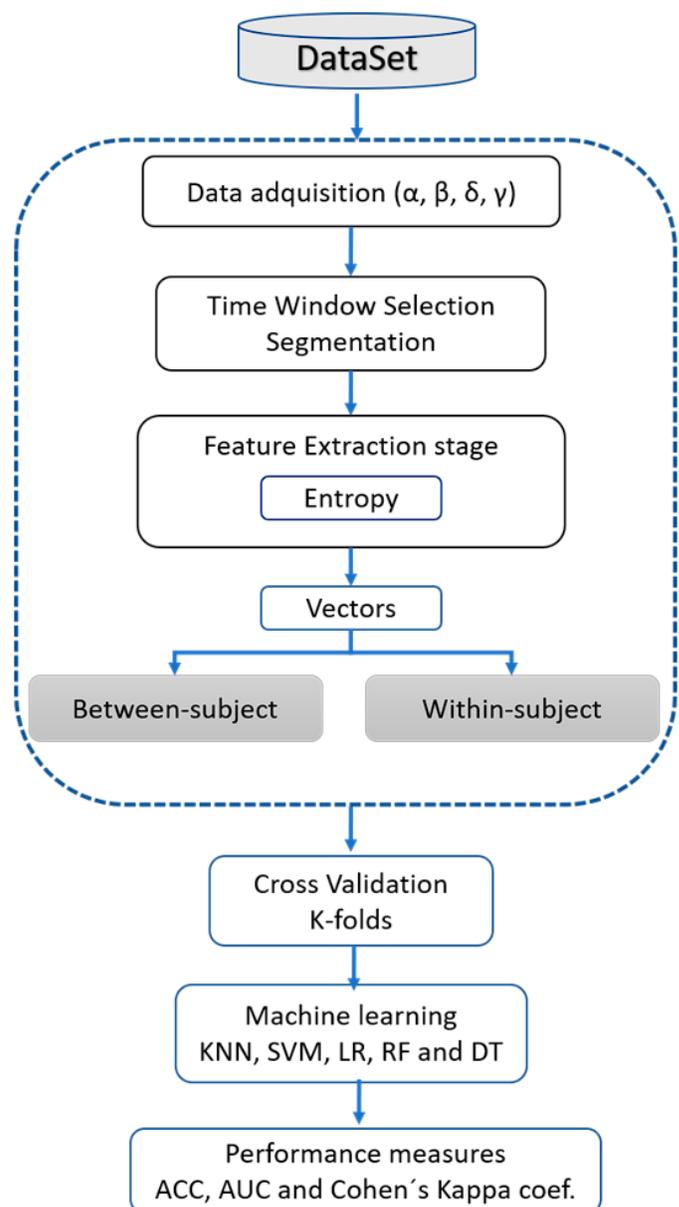


FIGURE 1. The proposed framework for emotion classification.

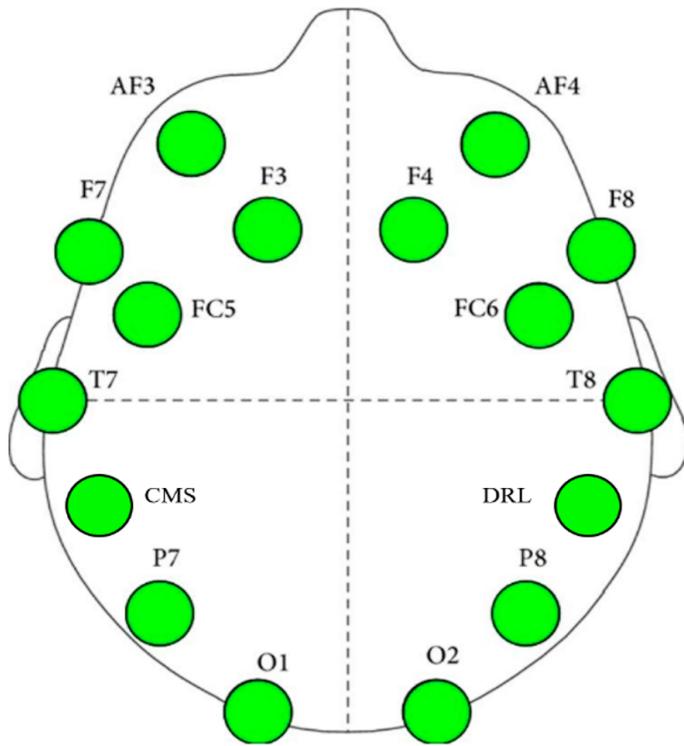


FIGURE 2. Emotiv device and electrode position system 10-20 modified.

Data Segmentation

The main objective of this study was to evaluate and compare the performance of different classifiers for emotion recognition with different TW shorter than 30 seconds. In Table 1 the lengths of TW considered is presented as well as the number of blocks obtained. Therefore, the format of the features in each trial was defined as (14x5x300); 14 representing the number of electrodes, 5 representing the frequency bands and 300 being the number of features extracted from the corresponding trial. In total 10 different TW lengths were examined in order to investigate the effect on the performance of classifier models in the study of emotions employing EEG sensors at a between-subject and within-subject level.

Feature Extraction

Since EEG signals are complex due to nonlinearity and randomness of time series data the calculation of time series entropy is incorporated^{[32][33]}. The TW is modified

separately in each analysis and the entropy is calculated as a characteristic. Several entropy functions exist; however, the Log Energy function stands out for its excellent performance in the analysis of EEG signals. This is due to its remarkable sensitivity to energy changes, lower susceptibility to high-frequency noise and rapid amplitude variations, and its ability to characterize the complexity associated with the sub-bands of EEG signals^{[34][35]}. This entropy function is based on the wavelet theory. There are different types of entropy, in this study we use Log energy. This type of entropy is based in wavelet theory where we assume a signal $x=[x_1 x_2 x_3 \dots x_n]$ and a probability distribution function $P(x_i)$ where i is the index of the signal's elements, then the entropy is defined as^[36]:

$$H(x) = \sum_{i=1}^n \log(P(x_i)^2) \quad (1)$$

under the convention that $\log(0)=0$. For this study the five frequency bands of every electrode were processed and the log energy entropy characteristic was extracted.

TABLE 1. Details of the TW information in each trial. N corresponds to the number of features.

| Parameters | | |
|---------------|-----------|-----------------------------|
| TW length (s) | TW Number | Format of Features (14x5xN) |
| 1 | 300 | (14x5x300) |
| 2 | 150 | (14x5x150) |
| 3 | 100 | (14x5x100) |
| 4 | 75 | (14x5x75) |
| 5 | 60 | (14x5x60) |
| 10 | 30 | (14x5x30) |
| 15 | 20 | (14x5x20) |
| 20 | 15 | (14x5x15) |
| 30 | 10 | (14x5x10) |

The emotional state analysis is carried out on an individual and group level. On the group level the classification models with different TW are evaluated. For the performance evaluation of the models on an individual level, statistical tests are performed to determine whether there is a statistically significant difference in classification performance between

different window lengths and between different classification algorithms.

Classification Analysis

In order to carry out the training and evaluation of the models, the k-fold cross-validation technique is used with a value of $k=5$, where k is number of folders into which the data are separated. The classifier models compared in this study were K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT) and Random Forest (RF). The use of supervised classifiers was chosen because of their ability to achieve more accurate and specific learning, as they are trained to establish direct connections between known patterns and labels. Moreover, these classifiers are widely used in the study of emotions, according to the literature^{[10][37][38][39][40][41]}. For the classification analysis the Machine Learning Toolbox 11.1 module from Matlab was used. The configuration parameters for every model are listed in Table 2.

TABLE 2. Parameter configuration in the Machine Learning Toolbox 11.1 module

| Parameter Setting | |
|-------------------|-------------------------------------------------------------------------------------------------------|
| Classifier | Parameter |
| Fine KNN | n_neighbors=1, metric=euclidean, distance=equal |
| SVM | kernel function= linear, kernel scale=automatic, box constraint level=1 multiclass method= one-vs-one |
| LR | None |
| RF | None |
| DT | Maximum number of splits=4 Split criterion= Gini's diversity index |

RESULTS AND DISCUSSION

Between-subject Study

The performance results for the models on the between-subject level are shown in Table 3. The ACC, AUC and Cohen's Kappa coefficient of each model are

presented as performance metrics. The results indicate that, regardless of the classifier model used, the TW that enhance their performances are between 2 and 15 seconds, with the 10s TWs standing out. The KNN model achieves the highest level of ACC, reaching 87.7 % in the 10 seconds TW, while the DT model exhibits the worst performance at 62.1 % in the 20 seconds TW. In terms of AUC, the best-performing model is RF with 0.93 for TW of 2 to 5 seconds, while the DT model shows the worst performance with 0.61 using a 20 seconds TW. Finally, the model with the best Cohen's Kappa coefficient is the KNN with 0.75 in the 10 seconds TW, and the DT model obtains the lowest result with 0.21 in the 30 seconds TW.

TABLE 3. Recognition results at the between-subject level (ACC, AUC and Cohen's Kappa coefficient)

| ACC-AUC-Cohen's Kappa coefficient | | | | | |
|-----------------------------------|-----------------------------------------|-----------------------------------------|---------------------------------|-----------------------------------------|---------------------------------|
| TW | KNN | SVM | LR | RF | DT |
| 1 | 83.3-0.83-0.66 | 64.1-0.69-0.28 | 63.7-0.69-0.27 | 83.6-0.92-0.67 | 63.4-0.66-0.27 |
| 2 | 83.3-0.83-0.67 | 65.1-0.71-0.30 | 64.7-0.71-0.30 | 84.3- 0.93 -0.68 | 63.4-0.66-0.27 |
| 3 | 84.8-0.85-0.70 | 65.9-0.72-0.31 | 65.2-0.72-0.30 | 84.8- 0.93 -0.69 | 62.8-0.65-0.28 |
| 4 | 85.4-0.85-0.71 | 66.1-0.72-0.32 | 65.7-0.72-0.31 | 84.4-0.92-0.69 | 63.6-0.66-0.28 |
| 5 | 86.5-0.86-0.72 | 66.9- 0.73 -0.33 | 66.5 - 0.73 -0.31 | 84.6 - 0.93 - 0.70 | 63.9 -0.66-0.25 |
| 10 | 87.7 - 0.88 - 0.75 | 67.2 - 0.73 - 0.35 | 65.8- 0.73 - 0.33 | 83.4-0.87-0.65 | 63.8- 0.67 - 0.29 |
| 15 | 86.7-0.87-0.74 | 66-0.72-0.30 | 64.1-0.72-0.31 | 80.6-0.89-0.63 | 62.9- 0.67 -0.25 |
| 20 | 84.3-0.84-0.69 | 63.2-0.7-0.29 | 64-0.7-0.29 | 78.8-0.85-0.58 | 61.2-0.62-0.27 |
| 30 | 80-0.8-0.54 | 59-0.64-0.17 | 64.4-0.69-0.23 | 73.8-0.62-0.44 | 62-0.73-0.21 |

The results indicate that, in general terms, the performance of the models tends to decrease significantly for TW greater than 20 seconds.

TABLE 4. Within-subject classifier performance results (mean of ACC, mean of AUC and mean Cohen's Kappa coefficient)

| ACC-ACC-Cohen's Kappa coefficient | | | | | |
|-----------------------------------|------------------------|----------------------|------------------------|--------------------------|-------------------------|
| TW | KNN | SVM | LR | RF | DT |
| 1 | 82.4-0.82-0.65 | 83.48-0.90-0.67 | 81.3-0.89-0.66 | 86.95-0.938-0.74 | 78.83-0.82-0.58 |
| 2 | 82.36-0.82-0.64 | 84.5-0.91-0.69 | 81.4-0.88-0.75 | 87.7-0.94-0.75 | 80.43-0.83-0.61 |
| 3 | 83.0-0.83-0.67 | 84.6-0.91-0.70 | 81.4-0.84-0.62 | 86.9-0.83-0.74 | 81.20-0.814-0.62 |
| 4 | 83.6-0.83-0.682 | 85.4-0.91-0.682 | 82.3-0.85-0.610 | 87.26-0.938-0.745 | 80.99-0.832-0.620 |
| 5 | 85.6-0.85-0.69 | 85-0.91-0.707 | 81.1-0.85-0.629 | 87.24-0.936-0.745 | 82.2-0.832-0.636 |
| 10 | 85.8-0.86-0.715 | 83.8-0.91-0.702 | 61.9-0.62-0.209 | 85.47-0.93-0.709 | 79.17-0.8-0.583 |
| 15 | 86.1-0.86-0.704 | 81.42-0.88-0.654 | 67.6-0.69-0.35 | 82.99-0.88-0.659 | 74.14-0.748-0.483 |
| 20 | 82.7-0.82-0.664 | 79.97-0.85-0.611 | 71.4-0.75-0.458 | 81.37-0.872-0.627 | 70.89-0.718-0.423 |
| 30 | 79.5-0.79-0.642 | 72.7-0.76-0.395 | 67.7-0.71-0.368 | 72.39-0.77-0.447 | 67.88-0.679-0.358 |

Within-subject Results

Emotion recognition results within subject variability are shown in Table 4. It is observed that the KNN model achieves its best performances in temporal TW ranging from 5 to 15 seconds, with the optimal result obtained in the 10 seconds TW. On the other hand, the SVM

model shows better performances in TW from 2 to 10 seconds, with the 5 seconds TW being the most outstanding in terms of ACC, AUC, and Cohen's Kappa coefficient. The LR model performs better in TW from 1 to 5 seconds, presenting its best result in the 4 seconds TW. For the RF model, its best performance is found in TW from 1 to 10 seconds, although the 4 seconds TW stands out the most. Finally, the DT model exhibits better performance in TW from 2 to 10 seconds, achieving the best result in the 5 seconds TW.

These results demonstrate that, regardless of the model chosen from these five, the TW that promote better performance in terms of ACC, AUC, and Cohen's Kappa coefficient are between 2 and 15 seconds. Moreover, the 10-second TW appears to be the most suitable for this type of configuration.

TABLE 5. Results of the mean difference of ACC and AUC in the different within-subject TW

| Accuracy | | | | |
|---------------|----------|---------|----------|---------|
| TW length (s) | ACC | | AUC | |
| | χ^2 | P-value | χ^2 | P-value |
| 1 | 38.93 | <0.001 | 78.17 | <0.001 |
| 2 | 36.69 | <0.001 | 68.70 | <0.001 |
| 3 | 25.491 | <0.001 | 66.80 | <0.001 |
| 4 | 29.0 | <0.001 | 67.01 | <0.001 |
| 5 | 21.285 | <0.001 | 55.83 | <0.001 |
| 10 | 58.42 | <0.001 | 77.66 | <0.001 |
| 15 | 33.543 | <0.001 | 31.717 | <0.001 |
| 20 | 31.5 | <0.001 | 42.65 | <0.001 |
| 30 | 23.04 | <0.001 | 19.87 | <0.001 |

In order to identify possible significant disparities between ACC and AUC values associated with different threshold values (TW), Friedman nonparametric statistical tests were performed for repeated measurements of a single factor. This test was chosen based on its robustness to violations of normality and its lower sensitivity to outliers compared to parametric tests, such as ANOVA. The results of these tests are presented in Table 5, considering a significance level of 0.05 for the evaluation of statistical significance.

It is observed that in all TW, there are significant differences between the models. This indicates that not only does the length of the TW interfere with performance, but the chosen model also plays a role.

The Table 6 presents the results of the Friedman non-parametric hypothesis test for multiple TW sizes and Wilcoxon test for the TW pairs generated by the KNN model.

TABLE 6. KNN model results between the ACC for different TWs.

| Results | | |
|------------------------------------------|-------|---------|
| Null Hypothesis | W | P-value |
| $\mu_5 = \mu_{10} = \mu_{15} = \mu_{20}$ | 5.795 | 0.122 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4$ | 3.603 | 0.308 |
| $\mu_5 \leq \mu_3$ | 300 | <0.001 |
| $\mu_{20} \leq \mu_{30}$ | 209 | 0.016 |

When comparing the equality of the mean of ACC in TW of 5, 10, 15, and 20 seconds, it is not possible to reject the null hypothesis. In the case of testing the equality hypothesis between TW of 1, 2, 3, and 4 seconds, it is also not possible to reject the difference. However, the TW of 5 seconds presents better performance than that of 3 seconds, and that of 20 seconds presents better performance than that of 30 seconds, as there is no significant difference between the TW of 5, 10, 15, and 20 seconds it can be considered that these are the ones that present better performance in terms of ACC for emotion detection.

TABLE 7. Results of the SVM model between the ACC for different TW.

| Results | | |
|------------------------------------------|--------|---------|
| Null Hypothesis | W | P-value |
| $\mu_5 = \mu_{10} = \mu_{15} = \mu_{20}$ | 2.776 | 0.427 |
| $\mu_1 = \mu_2 = \mu_3$ | 4.0607 | 0.131 |
| $\mu_5 = \mu_4$ | 134.5 | 0.668 |
| $\mu_{20} \leq \mu_{30}$ | 173 | <0.001 |

The comparison of means of ACC of the SVM models for different TW is presented in Table 7. It is observed that the mean ACC of 20 seconds is higher than the 30 seconds, and that there is no significant statistical difference between the TW of 5, 10, 15 and 20 seconds

TABLE 8. Results of the LR model between the ACC for different TW.

| Results | | |
|-----------------------------------------|-------|---------|
| Null Hypothesis | W | P-value |
| $\mu_{15} = \mu_{10} = \mu_{30}$ | 4.151 | 0.126 |
| $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ | 8.190 | 0.052 |
| $\mu_5 \leq \mu_{10}$ | 2.0 | <0.001 |

Table 8 shows the results of the comparisons of the ACC means of the LR model between different TW. It is observed that TW less than or equal to 5 seconds present a better performance in emotion detection.

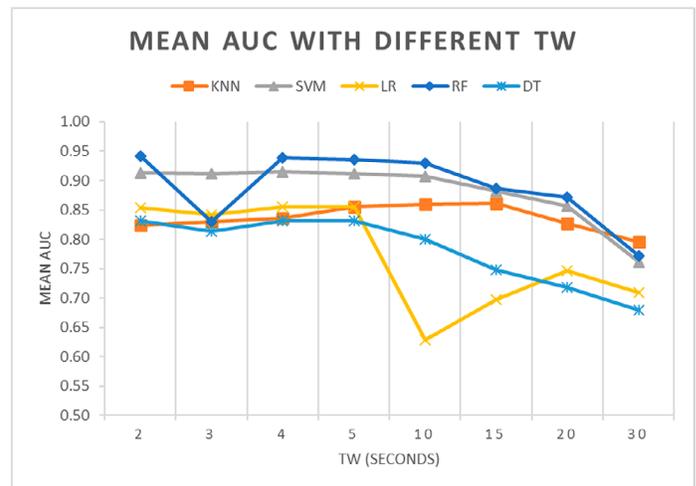


FIGURE 3. Mean AUC performance with different TW and classification models.

Figure 3 shows the performance results of the AUC averages of the five models. The best performing classifier is the SVM even though it was decreasing at 10, 15, 20 and 30 seconds TW. However, the RF model has the highest levels of AUC. The KNN model shows a growth in the AUC from TW 4 to 15 seconds. Finally, as the TW length

increases, the three classifiers tend to decrease their performance in terms of AUC.

CONCLUSIONS

In this study, the performance of three emotion classification models is evaluated, considering different TW sizes in data segmentation and two experimental setups. The first configuration involved the participation of 25 subjects in a between-subjects design. The results indicate that the window size significantly influences the performance of the classifiers. For example, the KNN model shows optimal results in TW sizes between 4 and 15 seconds, while the SVM, LR, RF and DT models excel in 4 to 10 seconds TW. In conclusion, for a between-subjects configuration, TWs of 4 to 15 seconds are recommended.

In the within-subject configuration the highest performance results are presented in TWs between 4 and 15 seconds for the KNN model with an ACC between 83.6 % and 86.12 %, respectively. However, for the SVM model the TW with the highest performance are between 2 and 10 seconds with an average AUC of 0.91. It is also observed that using the KNN model the performance results in terms of ACC and AUC do not differ significantly between the between-subject and within-subject configurations. However, for the LR and SVM models there is a significant difference when comparing the configurations: both models present better performance in the within-subjects configuration.

In general terms, it can be concluded that, in the study of emotions using EEG signals, regardless of the experimental setup or the classifier model employed, the TW that exhibit optimal performance in the classifiers, measured in terms of ACC, AUC and Cohen's Kappa coefficient, are in the range of 2 to 15 seconds. Ultimately, it is observed that, by increasing the duration of the TW above 20 seconds, all three models experience a decrease in performance. Likewise, the

use of TW equal to or longer than 20 seconds is not recommended for emotion recognition.

On the other hand, future research will focus on addressing the limitations identified in this study. This work includes: a) conducting additional analysis with a larger sample of participants, b) exploring a comparative analysis between supervised and unsupervised classification methods, and c) considering multiple features of entropy, such as Threshold Entropy and Shannon Entropy.

ACKNOWLEDGEMENTS

The authors thank the Teacher Improvement Program (PROMEP) for funding number UNSIS-CA-13. The authors thank Master Casey Hester for his support in the translation of this document.

AUTHOR CONTRIBUTIONS

A. J. S. Conceptualization (literature review, problem definition, theoretical framework selection, planning of instruments and tools), data curation, formal analysis, investigation, methodology (data analysis planning), project administration, writing original draft, writing review and editing, visualization, and resources. V. A. G. P. Conceptualization (literature review, review of alternative methodologies), methodology (methodology adjustments and peer review), supervision, validation, writing review, and editing. O. A. D. R. Conceptualization (feasibility analysis and preliminary methodological design), project administration, supervision, validation, investigation, funding acquisition, writing review, and editing.

REFERENCES

- [1] P. Salovey and J. D. Mayer, "Emotional intelligence. Imagination, cognition and personality," *Imagin. Cogn. Pers.*, vol. 9, no. 3, pp. 185-211, 1990, doi: <https://psycnet.apa.org/doi/10.2190/DUGG-P24E-52WK-6CDG>

- [2] R. W. Picard, "Affective Computing for HCI", in 8th International Conference on Human-Computer Interaction) on Human-Computer Interaction: Ergonomics and User Interfaces-Volume I - Volume I, 1999, pp. 829-833, doi: <https://dl.acm.org/doi/abs/10.5555/647943.742338>
- [3] R. Stock-Homburg, "Survey of emotions in human-robot interactions: Perspectives from robotic psychology on 20 years of research," *Int. J. Soc. Robotics*, vol. 14, no. 2, pp. 389-411, Mar. 2022, doi: <https://doi.org/10.1007/s12369-021-00778-6>
- [4] M. S. Young, K. A. Brookhuis, C. D. Wickens and P. A. Hancock, "State of science: Mental workload in ergonomics," *Ergonomics*, vol. 58, no. 1, Dec. 2014, doi: <https://doi.org/10.1080/00140139.2014.956151>
- [5] J. X. Chen, P. W. Zhang, Z. J. Mao, Y. F. Huang, D. M. Jiang, and Y. N. Zhang, "Accurate EEG-Based Emotion Recognition on Combined Features Using Deep Convolutional Neural Networks," *IEEE Access*, vol. 7, pp. 44317-44328, Jun. 2019, doi: <https://doi.org/10.1109/ACCESS.2019.2908285>
- [6] C. Qing, R. Qiao, X. Xu, and Y. Cheng, "Interpretable Emotion Recognition Using EEG Signals," *IEEE Access*, vol. 7, pp. 94160-94170, Jul. 2019, doi: <https://doi.org/10.1109/ACCESS.2019.2928691>
- [7] M. M. Duville, Y. Pérez, R. Hugues-Gudiño, N. E. Naal-Ruiz, L. M. Alonso-Valerdi, D. I. Ibarra-Zarate, "Systematic Review: Emotion Recognition Based on Electrophysiological Patterns for Emotion Regulation Detection," *Appl. Sci.*, vol. 13, no. 12, art. no. 6896, Feb. 2023, doi: <https://doi.org/10.3390/app13126896>
- [8] C. Pan, C. Shi, H. Mu, J. Li, and X. Gao, "EEG-Based Emotion Recognition Using Logistic Regression with Gaussian Kernel and Laplacian Prior and Investigation of Critical Frequency Bands," *Appl. Sci.*, vol. 9, no. 2, art. no. 1619, Apr. 2019, doi: <https://doi.org/10.3390/app10051619>
- [9] D. Wu, "Online and Offline Domain Adaptation for Reducing BCI Calibration Effort," *IEEE Trans. Hum.-Mach. Syst.*, vol. 47, no. 4, pp. 550-563, Aug. 2017, doi: <https://doi.org/10.1109/THMS.2016.2608931>
- [10] G. Li, D. Ouyang, Y. Yuan, W. Li, Z. Guo, X. Qu, P. Green, "An EEG Data Processing Approach for Emotion Recognition," *IEEE Sens. J.*, vol. 22, no. 11, pp. 10751-10763, Jun. 2022, doi: <https://doi.org/10.1109/JSEN.2022.3168572>
- [11] P. Schmidt, A. Reiss, R. Dürichen, and K. Van Laerhoven, "Wearable-Based Affect Recognition—A Review," *Sensors*, vol. 19, no. 19, art. no. 4079, Sep. 2019, doi: <https://doi.org/10.3390/s19194079>
- [12] J. J. Esqueda Elizondo, L. Jiménez Beristáin, A. Serrano Trujillo, M. Zavala Arce, et al., "Using Machine Learning Algorithms on Electroencephalographic Signals to Assess Engineering Students' Focus While Solving Math Exercises," *Rev. Mex. Ing. Biom.*, vol. 44, no. 4, pp. 23-37, Nov. 2023, doi: <https://doi.org/10.17488/RMIB.44.4.2>
- [13] F. J. Ramírez-Arias, E. E. García-Guerrero, E. Tlelo-Cuautle, et al., "Evaluation of machine learning algorithms for classification of EEG signals," *Technologies*, vol. 10, no 4, art. no. 79, Jun. 2022, doi: <https://doi.org/10.3390/technologies10040079>
- [14] M. Zheng and Y. Lin, "A deep transfer learning network with two classifiers based on sample selection for motor imagery brain-computer interface," *Biomed. Signal Process. Control*, vol. 89, art. no. 105786, Mar. 2024, doi: <https://doi.org/10.1016/j.bspc.2023.105786>
- [15] J. J. Esqueda-Elizondo, R. Juárez-Ramírez, O. R. López-Bonilla, E. E. García-Guerrero, et al., "Attention measurement of an autism spectrum disorder user using EEG signals: A case study," *Math. Comput. Appl.*, vol. 27, no. 2, art. no. 21, Mar. 2022, doi: <https://doi.org/10.3390/mca27020021>
- [16] Y. Qin, B. Li, W. Wang, X. Shi, H. Wang, and X. Wang, "ETCNet: An EEG-based motor imagery classification model combining efficient channel attention and temporal convolutional network," *Brain Res.*, vol. 1823, art. no. 148673, Nov. 2023, doi: <https://doi.org/10.1016/j.brainres.2023.148673>
- [17] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain Adaptation for EEG Emotion Recognition Based on Latent Representation Similarity," *IEEE Trans. Cogn. Dev. Syst.*, vol. 12, no. 2, pp. 344-353, Jun. 2020, doi: <https://doi.org/10.1109/TCDS.2019.2949306>
- [18] Y. -P. Lin, C.-H. Wang, T.P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, J.-H. Chen, "EEG-Based Emotion Recognition in Music Listening", *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798-1806 Jul. 2010, doi: <https://doi.org/10.1109/tbme.2010.2048568>
- [19] W.-L. Zheng, B.-N. Dong, and B.-L. Lu, "Multimodal emotion recognition using EEG and eye tracking data," in 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 2014, pp. 5040-5043, doi: <https://doi.org/10.1109/EMBC.2014.6944757>
- [20] N. Zhuang, Y. Zeng, L. Tong, C. Zhang, H. Zhang, and B. Yan, "Emotion recognition from EEG signals using multidimensional information in EMD domain," *Biomed Res. Int.*, vol. 2017, art. no. 8317357, 2017, doi: <https://doi.org/10.1155/2017/8317357>
- [21] D. Ouyang, Y. Yuan, G. Li, and Z. Guo, "The Effect of Time Window Length on EEG-Based Emotion Recognition," *Sensors*, vol. 22, no. 13, art. no. 4939, Jun. 2022, doi: <https://doi.org/10.3390/s22134939>
- [22] J. Healey, L. Nachman, S. Subramanian, J. Shahabdeen, and M. Morris, "Out of the Lab and into the Fray: Towards Modeling Emotion in Everyday Life," in 8th International Conference, Pervasive 2010, P. Floréen, A. Krüger, M. Spasojevic, Eds., Helsinki, Finland 2010, doi: https://doi.org/10.1007/978-3-642-12654-3_10
- [23] M. Gjoreski, M. Luštrek, M. Gams, and H. Gjoreski, "Monitoring stress with a wrist device using context," *J. Biomed. Inform.*, vol. 73, pp. 159-170, Sep. 2017, doi: <https://doi.org/10.1016/j.jbi.2017.08.006>
- [24] R. Alhalaseh and S. Alasasfeh, "Machine-Learning-Based Emotion Recognition System Using EEG Signals," *Computers*, vol 9, no 4, art. no. 95, 2020, doi: <https://doi.org/10.3390/computers9040095>
- [25] K. S. Kamble and J. Sengupta, "Ensemble Machine Learning-Based Affective Computing for Emotion Recognition Using Dual-Decomposed EEG Signals," *IEEE Sens. J.*, vol. 22, no. 3, pp. 2496-2507, Feb. 2022, doi: <https://doi.org/10.1109/JSEN.2021.3135953>
- [26] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring EEG Features in Cross-Subject Emotion Recognition," *Front. Neurosci.*, vol 12, art. no. 162, Mar. 2018, doi: <https://doi.org/10.3389/fnins.2018.00162>
- [27] B. Tripathi and R. K. Sharma, "EEG-Based Emotion Classification in Financial Trading Using Deep Learning: Effects of Risk Control Measures," *Sensors*, vol. 23, no. 7, art. no. 3474, Mar. 2023, doi: <https://doi.org/10.3390/s23073474>

- [28] M. D. Rinderknecht, O. Lambercy, and R. Gassert, "Enhancing simulations with intra-subject variability for improved psychophysical assessments," *PLoS One*, vol. 13, no. 12, art. no. e0209839, Dec. 2018, doi: <https://doi.org/10.1371/journal.pone.0209839>
- [29] A. Jarillo-Silva, V. A. Gomez-Perez, E. A Escotto-Cordova, and O. A. Domínguez-Ramírez, "Emotion Classification form EEG signals using wearable sensors:pilot test," *ECORFAN Journal-Bolivia*, vol. 7, no. 12, pp. 1-9, Sep. 2020. [Online]. Available: https://www.ecorfan.org/bolivia/journal/vol7num12/ECORFAN_Journal_Bolivia_V7_N12.pdf
- [30] K. Kotowski, K. Stapo, J. Leski, and M. Kotas, "Validation of Emotiv EPOC+ for extracting ERP correlates of emotional face processing," *Biocybern. Biomed. Eng.*, vol. 38, no 4, pp. 773-781, 2018, doi: <https://doi.org/10.1016/j.bbe.2018.06.006>
- [31] F. Mulla, E. Eya, E. Ibrahim, A. Alhaddad, R. Qahwaji, and R. Abd-Alhameed, "Neurological assessment of music therapy on the brain using Emotiv Epoc," in *2017 Internet Technologies and Applications (ITA)*, Wrexham, UK, 2017, pp. 259-263, doi: <https://doi.org/10.1109/ITECHA.2017.8101950>
- [32] N. Browarska, A. Kawala-Sterniuk, J. Zygarlicki, M. Podpora, M. Pelc, R. Martinek, and E. J. Gorzelańczyk, "Comparison of smoothing filters influence on quality of data recorded with the emotiv epoc flex brain-computer interface headset during audio stimulation," *Brain Sci.*, vol. 11, no. 1, art. no. 98, Jan. 2021, doi: <https://doi.org/10.3390/brainsci11010098>
- [33] P. R. Patel and R. N. Annavarapu, "EEG-based human emotion recognition using entropy as a feature extraction measure," *Brain Inf.*, vol 8, art. no. 20, Oct. 2021, doi: <https://doi.org/10.1186/s40708-021-00141-5>
- [34] P. Krishnan and S. Yaacob, "Drowsiness detection using band power and log energy entropy features based on EEG signals," *Int. J. Innov. Technol. Explor. Eng.*, vol 8, no. 10, pp. 830-836, Aug. 2019, doi: <https://doi.org/10.35940/ijitee.j9025.0881019>
- [35] A. B Das and M. I. H. Bhuiyan, "Discrimination and classification of focal and non-focal EEG signals using entropy-based features in the EMD-DWT domain," *Biomed. Signal Process. Control*, vol. 29, pp. 11-21, Aug. 2016, doi: <https://doi.org/10.1016/j.bspc.2016.05.004>
- [36] R. Djemal, K. AlSharabi, S. Ibrahim, and A. Alsuwailem, "EEG-Based Computer Aided Diagnosis of Autism Spectrum Disorder Using Wavelet, Entropy, and ANN," *BioMed Res. Int.*, vol. 2017, no 1, art. no. 9816591, 2017, doi: <https://doi.org/10.1155/2017/9816591>
- [37] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, et al., "DEAP: A Database for Emotion Analysis; Using Physiological Signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18-31, 2012, doi: <https://doi.org/10.1109/T-AFFC.2011.15>
- [38] A. Bablani, D. Reddy Edla, and S. Dodia, "Classification of EEG Data using k-Nearest Neighbor approach for Concealed Information Test," *Procedia Comput. Sci.*, vol. 143, pp. 242-249, 2018, doi: <https://doi.org/10.1016/j.procs.2018.10.392>
- [39] D.-W. Chen, R. Miao, W.-Q. Yang, Y. Liang, H.-H. Chen, L. Huang, C.-J. Deng, N. Han, "A feature extraction method based on differential entropy and linear discriminant analysis for emotion recognition," *Sensors*, vol. 19, no. 7, art. no. 1631, Oct. 2019, doi: <https://doi.org/10.3390/s19071631>
- [40] S. Babeetha and S. S. Sridhar, "EEG Signal Feature Extraction Using Principal Component Analysis and Power Spectral Entropy for Multiclass Emotion Prediction," in *Fourth International Conference on Image Processing and Capsule Networks*, Bangkok, Thailand, 2023, pp. 435-448, doi: https://doi.org/10.1007/978-981-99-7093-3_29
- [41] V. Doma and M. Pirouz, "A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals," *J. Big Data*, vol. 7, art. no. 18, Mar. 2020, doi: <https://doi.org/10.1186/s40537-020-00289-7>