RESEARCH ARTICLE
VOL. 46 | NO. 1 | **JANUARY-APRIL 2025** | PP 1-17

Revista Mexicana de Ingeniería Biomédica

https://dx.doi.org/10.17488/RMIB.46.1.1487

LOCATION ID: e1487

Comparison of Machine Learning Models for Identification of Depressive Patients through Motor Activity

Comparación de Modelos de Aprendizaje para la Identificación de Pacientes Depresivos por Medio de Actividad Motora

Gerardo Neftalí Rivera-Rojas[1], *Carlos Eric Galván-Tejada*[1], *Jorge Isaac Galván-Tejeda*[1], *José M. Celaya-Padilla*[1], *Erika Acosta-Cruz*[2]

[1] Universidad Autónoma de Zacatecas, Zacatecas - México

[2] Universidad Autónoma de Coahuila, Coahuila - México

ABSTRACT

The present study aims to evaluate various classification algorithms for data pertaining to subjects diagnosed with depression and non-depressive subjects. To this end, the data obtained from the "depresjon" dataset proposed by Garcia-Ceja, E., *et al* were analyzed. This dataset comprises motor activity recorded by the Actiwatch device (Cambridge Neurotechnology Ltd, England, model AW4). Predictions were made using various machine learning models, including synthetic data. Subsequently, metrics such as specificity, sensitivity, and precision were compared. The results highlight the best features of the data and the best machine learning model (using an ensemble model) for classifying potential depressive episodes in activity during the afternoon and night, with a precision of 96.6 %, sensitivity of 100 %, and specificity of 93.33 %.

KEYWORDS: data mining, data analysis, depression, machine learning and motor activity

RESUMEN

El presente estudio tiene como objetivo evaluar diversos algoritmos de clasificación de datos pertenecientes a sujetos diagnosticados con depresión y sujetos no depresivos. Para ello, se analizaron los datos obtenidos del *dataset "depresjon"* propuesto por Garcia-Ceja, E., *et al*, el cual se compone de la actividad motora captada por el dispositivo *Actiwatch (Cambridge Neurotechnology Ltd, England, model AW4).* Mediante distintos modelos de aprendizaje automático se realizaron predicciones incluyendo datos sintéticos. Posteriormente, se compararon métricas como especificidad, sensibilidad y precisión. Los resultados muestran las mejores características de los datos, así como el mejor modelo de aprendizaje automático (mediante modelo de ensamble) para realizar la clasificación de posibles episodios depresivos en la actividad durante la tarde y la noche, con una precisión del 96.6 %, una sensibilidad del 100 % y una especificidad del 93.33 %.

PALABRAS CLAVE: actividad motora, análisis de datos, aprendizaje automático, depresión, minería de datos

Corresponding author

TO: CARLOS ERIC GALVÁN-TEJADA

INSTITUTION: UNIVERSIDAD AUTÓNOMA DE ZACATECAS

ADDRESS: JARDÍN JUÁREZ #147, CENTRO HISTÓRICO, ZACATECAS, ZAC. C.P. 98000, MÉXICO.

EMAIL: ericgalvan@uaz.edu.mx

Received:

05 November 2024

Accepted:

24 February 2025

Published:

24 April 2025

INTRODUCTION

Portable electronic devices known as "wearables" have been utilized in the field of healthcare over the last decade. These devices not only accurately capture data, but have also found applications in several areas such as medicine, psychology, rehabilitation, and intervention for various psychological disorders[1]. They are particularly valuable in non-invasive research, providing greater precision with measurements ranging from seconds to entire weeks. Additionally, the implementation of these devices enables continuous monitoring of subjects under study without the need for direct observation by the experimenter, resulting in cost savings in research and minimal disruption to the daily activities of the subjects.

Major Depressive Disorder (MDD), a condition affecting approximately 280 million people worldwide[2], is characterized by persistent sadness, loss of interest in previously enjoyable activities, inability to perform daily activities for at least two weeks, decreased energy, changes in appetite, alterations in circadian rhythms, among other clinical features[3]. Due to the reduction in daily activities, various mental illnesses have been studied using electronic devices[4], and different machine learning models[5][6]. As a result, the use of different devices for recording markers has become increasingly common in current research. These markers prove useful when analyzing variables related to morbid processes, including but not limited to blood pressure[7], blood oxygenation[8], motor activity[9], and other parameters to identify different physical and mental health alterations.

Recent studies indicate that motor activity during different times of the day can strongly correlate with depression. The use of wearables has highlighted a key clinical feature: reduced movement, linked to symptoms such as drowsiness, insomnia, decreased interest in physical activity, and unwarranted fatigue. Additionally, sociodemographic characteristics have established connections between depression and risk groups. However, the diversity in parenting styles, social skills, cognitive assets, and external variables complicates psychological research, making it difficult to find well-supported, generalizable relationships for accurate differential diagnosis[10]. Thus, providing precise prognoses and treatments requires careful consideration of these diverse factors.

Accurate diagnosis of affective disorders is crucial, since these conditions can lead to suicidal ideation and irritability, endangering both the affected individuals and their caregivers[11]. Research has explored the impact on quality of life[12] and the potential comorbidities with cardiovascular, metabolic, and cancer-related diseases[13]. While MDD is treatable with medication and psychotherapy, early intervention significantly improves outcomes. Thus, certain tools are essential for the timely detection of depressive episodes. Existing tests, such as the Beck Depression Inventory (BDI)[11], the Diagnostic and Statistical Manual (DSM-V)[12], the International Classification of Diseases (ICD-10)[14], and the Montgomery-Asberg Depression Rating Scale (MADRS)[13], rely on patient-reported information and are subject to potential human error. Additionally, technical requirements, frequent updates by healthcare institutions, and the willingness of at-risk individuals can hinder timely treatment in some populations.

Hence, the implementation of new technologies is necessary to enable early detection through comprehensive, harmless, and non-invasive monitoring for individuals, thereby reducing medical costs[15]. This approach offers the adaptation of treatment based on captured data as time progresses. The monitoring of motor activity in medical and psychiatric fields proves advantages in abnormal behaviors identification detecting specific periods of movement that can provide insights into recognizing behavioral patterns associated with diseases such as dementia (cita), depression[15], anxiety (cita), and schizophrenia[16]. In conditions like anxiety and depression, it has been demonstrated that affected patients tend to reduce daytime activity, increase nighttime activity, and engage in activities related to certain diagnostic criteria outlined by the DSM-V[12]. On the other hand, patients diagnosed with different types of bipolar disorder exhibit increased energy during recognizable periods accompanied by periods in which activity significantly decreases. Therefore, monitoring motor activity serves as a good indicator for differentiation between individuals with a depressive episode or bipolarity, as they show discrepancies compared to healthy individuals[17].

Motor signals have been acquired using diverse methodologies, including different types of accelerometers. It has been suggested that assorted devices of this nature are viable for behavioral analysis in sports, disease prevention, and management[18], enabling efficient and precise measurement of treatment progress. Once the information is obtained, varied methods of statistical processing and/or classification can assist in the pharmaco-psychological treatment by providing quantitative insights that complement the work of the professional team. Enrique Garcia-Ceja *et al.*[6] conducted a study where data on the motor activity of 23 patients with depression (the "condition" group) were obtained, along with sociodemographic characteristics such as age, gender, presence of melancholy, type of illness (unipolar or bipolar), patient type (outpatient or hospitalized), level of education, marital status, employment status (employed, unemployed, or subsidized by a government program), as well as scores on the depression scale (MADRS). Additionally, 32 healthy subjects (the "control" group) were included, from whom only age and gender characteristics were obtained.

Different machine learning algorithms were employed to classify depressive and non-depressive signals using the mentioned data. The study demonstrated that these models could analyze motor behavior to classify subjects. Additionally, classifying conditions based on sociodemographic characteristics significantly impacts psychological diagnosis and can guide tailored treatments. This article addresses both motor and sociodemographic characteristics to classify depressive and healthy subjects, as well as different types of depression within the "condition" group. Classification methods through machine learning algorithms will be used, involving data mining processes to minimize variables affecting predictors' performance and maximize the impact of variables highly associated with depression[18], aiming to maintain control and generate useful data for classification. The data were segmented by different time periods: early morning (00:00 - 06:00 hs), morning (06:01 - 12:00 hs), afternoon (12:01 - 18:00 hs), and night (18:01 - 24:00 hs).

The structure of the article is as follows: Materials and Methods, Results, Discussion, and Conclusion. These sections describe the processes implemented for classifying various parameters in the control and condition groups.

MATERIALS AND METHODS

Pre-processing data

The data mining process commenced with the organization and cleaning of data extracted from the dataset available at: https://datasets.simula.no//depresjon/. As depicted in, this dataset contains information regarding the motor activity (counts per minute) of patients diagnosed with some form of depression, as well as information about healthy subjects. The records were initially stored individually in separate files, each containing details such as one-minute time intervals, date of data acquisition, and motor activity. Each record was analyzed individually, and data that were deemed non-informative (exhibiting abnormal distribution, auto-correlation, or suggesting little relevance of characteristics to the explanation of the phenomenon in question) were excluded.

The records exhibited varying duration, spanning up to 20 days. Consequently, the minimum number of records per subject was considered as a threshold for extracting statistical features and subsequent training and testing of machine learning methods. The dataset was divided into 52 subsets, corresponding to each of the 4 daily periods (6 hours per period). This division involved selecting 13 continuous days of data for each subject from the aforementioned groups, ensuring that all subjects commenced and concluded their records at the same minute. The resulting dataset comprises 15 subjects from the condition group and 15 subjects from the control group, aiming to minimize potential data imbalance.

Descriptive statistics

The statistical treatment of the data began with normality tests applied to each of the 52 subsets (4 subsets for each of the 13 days), using the Shapiro-Wilk test (see Algorithm 1), revealing a non-normal distribution $p < 0.05$ of the data in most time periods analyzed in both groups. Subsequently, descriptive statistical tests were conducted to identify differences in movements between groups and generate potentially informative data. the parameters minimum value (Min), maximum value (Max), and mean (Mean) equations (2 and 3), were computed using the R function summary, while the calculation of the standard deviation (SD) was performed using the (stats) package equation (4), Additionally, the skewness (SK) and kurtosis (KURT) were calculated with the (e1071) library equations (5 and 6) were calculated for each subset Table 1. Other libraries were employed for specific purposes, such as (ggplot2) for generating graphs, (dplyr) and (tidyr) for data manipulation and preprocessing, (caret) for training and testing machine learning models, and (ape) for advanced statistical analyses.

Based on relevant research on movement and mental disorders [19], maximum values for the groups were imputed

TABLE 1. "Dataset created from the Depresjon dataset."

Periods	Number of observations	Features
Early Mornings	4680	Min, Max, Mean, SD, SK, KURT
Morning	4680	Min, Max, Mean, SD, SK, KURT
Afternoon	4680	Min, Max, Mean, SD, SK, KURT
Night	4680	Min, Max, Mean, SD, SK, KURT

using the metric of mean + 2 standard deviations. This was done to maintain expected ranges of movement in accordance with literature reports[20], thereby reducing the chances of misclassification due to data acquisition errors or potential sensor failures.

Algorithm: Shapiro-Wilk Test Formula
Input:
Sample data $X = (x_1, x_2 \dots x_n)$
Output:
Test statistic W and p-value

Steps:

1. Sort the data in ascending order: $x_{(1)} \leq x_{(2)} \ldots \leq x_{-(n)}$.
2. Calculate the coefficients a_i, b_i *and* c_i *for* $i = 1,2,\ldots n$.
3. Calculate the test statistic

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\left(\sum_{i=1}^{n} (x_{(i)} - \bar{x}\right)^2} \quad (1)$$

where $\bar{x}$ is the sample mean.

4. Calculate the expected value E(W) and the variance Var (W) under normality for the given sample size.
5. Calculate the p-value by comparing W to the distribution of W under the null hypothesis.

The formulas used to obtain the metrics mentioned in Table 1 are presented below

$$U = \min(U_1, U_2) = R - \frac{n_1 \cdot (n_1 + 1)}{2} \quad (2)$$

where:

U_1 and U_2 are the sums of ranks for the two samples,

R is the sum of ranks for the entire sample,

n_1 and n_2 are these the sample sizes of the two groups.

$$mean = \frac{\sum_{i=1}^{n} x_i}{n} \quad (3)$$

where:

x_i represent the value of the sample

n is the total number of observations.

$$SD = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N - 1}} \quad (4)$$

where:

x_i represents the values in the sample

$\bar{x}$ is the sample mean

N is the sample size

$$SK = \frac{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^3}{(\sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2})^3} \quad (5)$$

where:

x_i represents the values in the sample

$\bar{x}$ is the sample mean

n is the sample size

$$KURT = \frac{\frac{\sum_i (x_i - \bar{x})^4}{n}}{\sigma^4} - 3 \tag{6}$$

where:

x_i represents the values in the sample

$\bar{x}$ is the sample mean

n is the sample size

σ is standard deviation

Feature selection

After organizing and obtaining metrics for each of the 52 subsets, a total of 312 features were collected. These features were then analyzed to extract relevant characteristics, achieved through multi-variable feature selection using genetic algorithms with the Galgo package[21], below is a brief description of the classification algorithms used for feature extraction through the Galgo library. Forward selection was performed using various methods: Support Vector Machines (SVM) an algorithm that maximizes a specific mathematical function with respect to a set of information, through key concepts: the separation of the hyperplane, maximizing the margin of the hyperplane, soft of the margin, and the Kernel function[22], the algorithm learn by example to assign labels to objects[23].

Neural Networks (NNET): Neural networks consist of an input layer with several nodes, internal hidden layers, and an output layer. Each node is associated with weights, activation functions such as: Linear, Tanh, ReLU, Sigmoid, to mention a few, and thresholds, connecting to the next layer until reaching the output layer where the classification results from the information in the first layer, typically, the data undergoes a training phase in which weights and thresholds are adjusted to provide a more accurate classification. On the other hand, there are various algorithms such as "Adeline, perceptron, and backpropagation" that enable the adjustment of parameters to achieve optimal classification[24].

k-Nearest Neighbors (KNN) supervised machine learning algorithm that can be used to solve both classification and regression problems, is a non-parametric classification method[25], through the calculation of distances between the data points with respect to others[26], assigning labels and iterating until well-differentiated groups are found based on distances such as Euclidean or Manhattan.

Random Forest Algorithm (RF) is a supervised classification algorithm which classifies the data by constructing a number of Classifiers (decision trees) with an aim to achieve a higher accuracy of prediction[27], the Random forest uses "Adaboost and Bootstrapping" techniques to construct multiple classifiers[28]. This algorithm has been applied in economics, medicine, commerce, and the financial sector in recent decades, as high levels of accuracy have been reported in classification tasks with large amounts of data[29]. Furthermore, this technique allows the construction of multiple classifiers that cater to specific issues, thereby minimizing errors in predictions[30].

Training and test phase

The use of machine learning algorithms previously mentioned in the analysis of human movement through wearable devices like the Actiwatch is motivated by the need to process and classify complex temporal data patterns effectively[6]. Human motor activity, as recorded by wearable sensors, often exhibits high variability due to individual differences, behavioral routines, and external influences[20]. Additionally, these data can be noisy, non-linear, and sometimes chaotic, making traditional statistical approaches insufficient for distinguishing between relevant patterns and random fluctuations.

By implementing these algorithms, we aim to improve the accuracy and reliability of classification models that distinguish between depressive and non-depressive subjects based on movement data. Furthermore, analyzing the most relevant features influencing these models can provide insights into the behavioral differences underlying these conditions. The combination of SVM, KNN, and Neural Networks allows for a comprehensive assessment of motor activity, balancing interpretability, efficiency, and performance in detecting meaningful patterns in wearable sensor data[3].

Classifications were conducted by applying the various algorithms mentioned earlier separately to identify subjects with depression and healthy subjects. This process was carried out in three stages. In the first stage, the motor activity data were partitioned into 75 % for the training phase and the remaining 25 % for the testing stage, with a k=5 cross-validation applied.

The Support Vector Machine (SVM) was implemented using the *svm()* function from the (e1071) package. By default, the kernel function used is the Radial Basis Function (RBF). The cost parameter defaults to 1, and the gamma parameter is automatically computed as equation 7:

$$\gamma = \frac{1}{number\ of\ predictors} \tag{7}$$

The Neural Network (NN) classifier was trained using the *train()* function from the (caret) package, leveraging the "nnet" method from the nnet package. The neural network consists of a single hidden layer by default, with the number of neurons set to (input features + output classes)/2. The activation function used for hidden layers is the sigmoid function, and the output layer applies the softmax function for classification tasks. The weight decay (decay) parameter, which prevents overfitting, defaults to 0, and the maximum number of iterations (maxit) defaults to 100. The optimization is performed using backpropagation with a gradient-based method.

The Random Forest (RF) classifier implemented using the "rf" method in train(), which internally calls the randomForest function from the randomForest package. By default, the number of trees (ntree) is set to 500, and the number of randomly selected predictors per split (mtry) is set to the square root of the total number of predictors for classification tasks. The model aggregates multiple decision trees and uses majority voting to improve predictive performance while reducing overfitting.

The K-Nearest Neighbors (KNN) model was implemented using the "knn" method in train(), which relies on the knn() function from the (class) package. The number of neighbors (k) is tuned automatically by (caret). The Euclidean distance metric is used to measure similarity between data points.

The second stage involved generating synthetic random data by bootstrap method, a statistical technique used to estimate the distribution of statistics by resampling with replacement from the original data. It allows for assessing the variability of a statistic without relying on strong assumptions about the underlying population distribution[31], considering the key characteristics described, and matching the number of subjects in the generated dataset to n=30.

In the third and final stage, proportional partitioning of sociodemographic data was performed similarly to the motor activity data. Feature selection was implemented using Galgo[21]. Subsequently, various machine learning methods were applied, and the results were validated using k=5 cross-validation. Finally the Stacked Ensemble Model[32] integrates the predictions from the SVM, NN, RF, and KNN classifiers. The stacking process involves generating out-of-fold (OOF) predictions from the base models and using them as inputs for a meta-learner, which in this case was a logistic regression model (glm method) from the stats package. The goal of the stacked model is to leverage the strengths of each individual classifier and improve overall predictive accuracy.

RESULTS AND DISCUSSION

The top 10 features were extracted in order of relevance: "maximum value night 1", "maximum value afternoon 2", "maximum value night 8", "maximum value afternoon 1", "maximum value afternoon 3", "maximum value afternoon 5", "maximum value afternoon 10", "maximum value afternoon 8", "maximum value morning 4", and "maximum value morning 2". These features exhibited classification accuracy ranging from 96.6 % to 100 % for distinguishing between groups using different algorithms.

Several predictions were made using the models mentioned earlier. Finally, to assess the performance of each classification model based on metrics such as True Positives (TP) (subjects with depression correctly classified), True Negatives (TN) (control subjects correctly classified), False Positives (FP) (control subjects incorrectly classified), and False Negatives (FN) (subjects with depression incorrectly classified). Sensitivity (7), specificity (8), and accuracy (9).

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

$$Accuracy = \frac{TP}{TP + FP} \tag{10}$$

$$MCC = \frac{TN \times TP - FN \times FP}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{11}$$

The Mann-Whitney U (2) tests were conducted, contrasting the maximum values of the groups during each of the days considered in the present study. In the following tables 2 and 3, W-statistic values and the p-value for each of the different periods over the 13 days are shown. Significant differences are noted between the control and condition groups, primarily during the morning (Figure 1), afternoon, and night periods (Figure 2).

FIGURE 1. Maximum values of motor activity over 13 days (early morning and morning).

TABLE 2. Statistical results of maximum motor activity comparison between different groups during early mornings and mornings.

Early mornings	W-value	P-value	Mornings	W-value	P-value
1	160.5	0.04597	1	210	1.108e-05
2	117	0.8633	2	210	6.362e-06
3	151	0.1087	3	210	1.117e-05
4	174	0.009906	4	210	1.36e-05
5	165	0.02776	5	196	0.0002472
6	160	0.04736	6	210	1.07e-05
7	182	0.003716	7	210	8.625e-06
8	135	0.353	8	210	8.625e-06
9	143	0.2028	9	210	1.07e-05
10	155	0.07057	10	210	6.362e-06
11	210	2.83e-05	11	197	0.0001651
12	164	0.02796	12	210	1.36e-05
13	151.5	0.1072	13	210	1.07e-05

The condition group shows a tendency toward lower motor activity compared to the control group (Figures 3-4). This finding aligns with clinical criteria described in diagnostic manuals such as the DSM-5[12] and ICD-10[14], which highlight psychomotor retardation, lethargy, and reduced engagement in previously enjoyed activities as key symptoms of depression. Individuals with depression often experience diminished motivation, fatigue, and an overall reduction in voluntary movement, which could contribute to the lower motor activity observed in the experimental group.

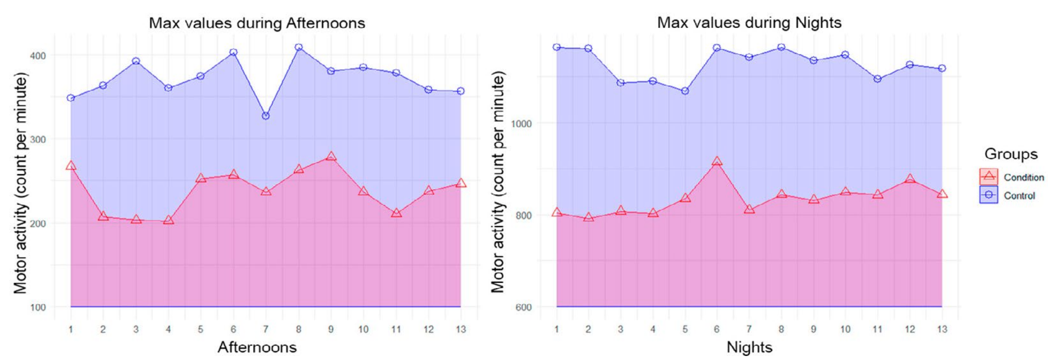

FIGURE 2. Maximum values of motor activity over 13 days (Afternoon and night).

TABLE 3. **Statistical results of maximum motor activity comparison between different groups during afternoons and nights.**

Afternoons	W-value	P-value	Nights	W-value	P-value
1	160.5	0.04597	1	210	1.108e-05
2	117	0.8633	2	210	6.362e-06
3	151	0.1087	3	210	1.117e-05
4	174	0.009906	4	210	1.36e-05
5	165	0.02776	5	196	0.0002472
6	160	0.04736	6	210	1.07e-05
7	182	0.003716	7	210	8.625e-06
8	135	0.353	8	210	8.625e-06
9	143	0.2028	9	210	1.07e-05
10	155	0.07057	10	210	6.362e-06
11	210	2.83e-05	11	197	0.0001651
12	164	0.02796	12	210	1.36e-05
13	151.5	0.1072	13	210	1.07e-05

Additionally, the Mann Whitney test was conducted with the averages of motor activity under the same conditions for both groups. The following tables 4-5 display the statistical values as well as the p-values for each of the subsets.

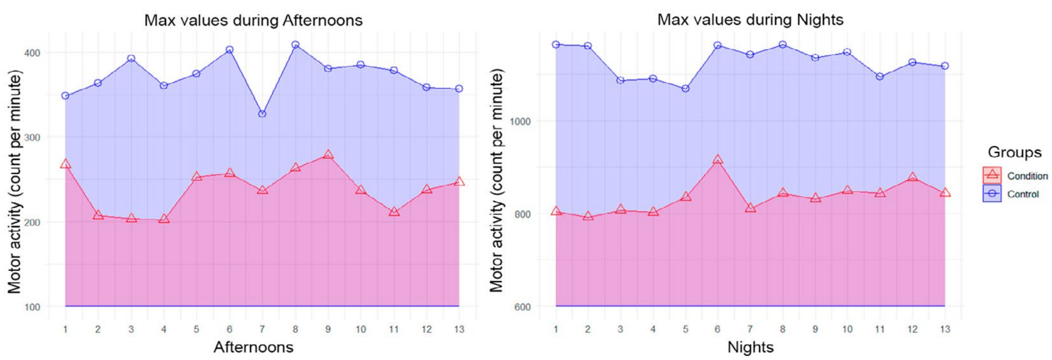

FIGURE 3. **Average values of mean motor activity over 13 days (Early morning and morning).**

TABLE 4. **Statistical results of mean motor activity comparison between different groups during early mornings and mornings.**

Early mornings	W-value	P-value	Mornings	W-value	P-value
1	130	0.4806	1	170	0.01804
2	99	0.5897	2	202	0.0002222
3	116	0.9009	3	208	8.104e-05
4	129	0.5068	4	169	0.02016
5	131	0.4552	5	152	0.1057
6	137	0.3194	6	182	0.004201
7	136	0.34	7	172	0.01438
8	99	0.5897	8	210	8.625e-06
9	103	0.7089	9	184	0.003223
10	89	0.34	10	181	0.004785
11	164	0.03436	11	131	0.4552
12	139	0.2807	12	166	0.02789
13	105	0.7715	13	174	0.01138

FIGURE 4. Average values of mean motor activity over 13 days (Afternoon and night).

TABLE 5. Statistical results of mean motor activity comparison between different groups during afternoons and nights.

Afternoons	W-value	P-value	Nights	W-value	P-value
1	146	0.171	1	179	0.006178
2	168	0.0225	2	209	6.811e-05
3	195	0.000669	3	175	0.01011
4	185	0.002817	4	156	0.07443
5	161	0.04644	5	136	0.34
6	158	0.06191	6	186	0.002457
7	152	0.1057	7	172	0.01438
8	176	0.008957	8	182	0.004201
9	147	0.1584	9	163	0.03805
10	172	0.01438	10	174	0.01138
11	179	0.006178	11	167	0.02507
12	153	0.09702	12	155	0.08143
13	157	0.06793	13	139	0.2807

We conducted within-group comparisons across different time periods. Table 6 presents the significant differences in maximum motor activity values exclusively within the control group during the early morning, morning, and night periods. Meanwhile, Table 7 displays the significant differences in average activity within the control group during the early morning, the significant differences in maximum activity during the early morning for the condition group. These findings suggest variations in activity levels across different times of the day, highlighting potential temporal patterns in motor behavior.

TABLE 6. Statistical results of the maximum motor activity comparison within the control group during early morning and morning periods.

Early mornings	W-value	P-value	Mornings	W-value	P-value	Nights	W-value	P-value
2 vs 11	64	0.02456	2 vs5	150	1.81E-02	1 vs 4	142.5	3.84E-02
3 vs 11	66	0.0312	4 vs 5	150	1.81E-02	1 vs 5	150	1.81E-02
6 vs 11	66	0.0312	5 vs 10	75	1.81E-02	4 vs 8	82.5	3.84E-02
8 vs 11	65	0.02771	-	-	-	5 vs 8	75	1.81E-02
10 vs 11	66	0.03117	-	-	-	-	-	-
11 vs 13	166	0.01509	-	-	-	-	-	-

TABLE 7. Statistical results of the mean motor activity comparison within the control group during early morning period and the maximum motor activity comparison within the condition group during early morning period.

Control group mean motor activity			Condition group maximum motor activity		
Early mornings	W-value	P-value	Early mornings	W-value	P-value
7 vs 11	64	0.04644	2 vs 7	156	0.044
8 vs 11	62	0.03805	7 vs 10	68	0.03931
11 vs 13	162	0.04206	-	-	-

Regarding the section on training, testing, and validation of different classification methods, considering the key features chosen through "forward selection features". The resulting features were as follows: "maximum value night 1", "maximum value afternoon 2", "maximum value afternoon 1", "maximum value night 8", "maximum value afternoon 3", "maximum value afternoon 5", "maximum value afternoon 10", "maximum value afternoon 8", "maximum value morning 4", and "maximum value morning 2". Below are the various performances of each individual algorithm, as well as the final model using 3 and 10 variables extracted with Galgo (Figure 5).

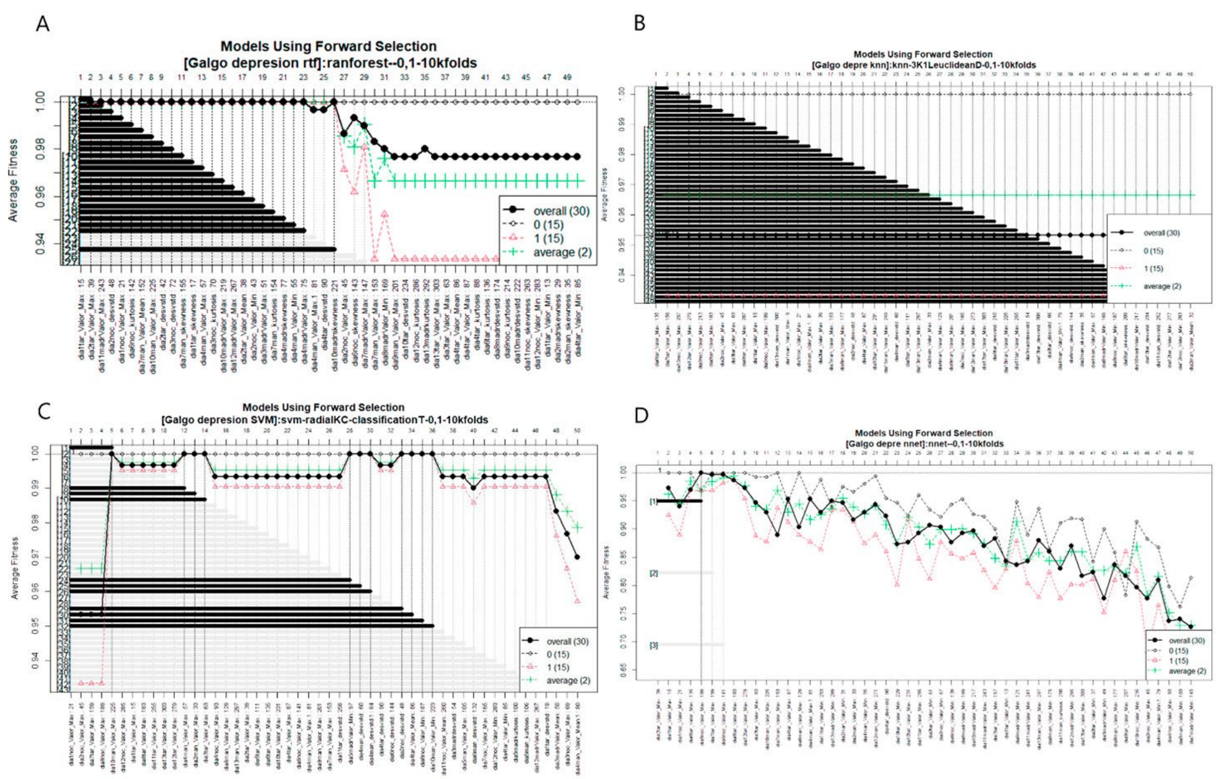

FIGURE 5. Feature selection using forward selection with the Galgo package. The graphs illustrate the average fitness of the different classification algorithms involved: A) Random Forest, B) KNN, C) SVM, and D) Neural Networks. Additionally, they show the accuracy of the models for classifying depressive and non-depressive subjects, as well as the number and names of the most relevant variables for classification. The results indicate that after selecting 10 features, the accuracy of the different algorithms begins to decline.

TABLE 8. **Evaluation metrics for different classification methods using 3 most relevant features.**

Model	Sensitivity	Specificity	Accuracy	MCC
RF	100.00 %	100.00 %	100.00 %	1
KNN	100.00 %	100.00 %	100.00 %	1
SVM	100.00 %	100.00 %	100.00 %	1
NNET	100.00 %	100.00 %	100.00 %	1
STACK	100.00 %	100.00 %	100.00 %	1
RF Synthetic data	96.66 %	100.00 %	93.30 %	0.707
KNN synthetic data	96.66 %	100.00 %	93.30 %	1
SVM synthetic data	80.20 %	77.00 %	83.00 %	0.707
NNET synthetic data	96.66 %	100.00 %	93.33 %	0.707
STACK synthetic data	96.66 %	100.00 %	93.33 %	1

The decrease in accuracy, sensitivity, and specificity is observed during the "blind" test with synthetic data n=30 as seen on table 8. While the evaluation metrics of different classification methods vary, it is evident that the final model, based on predictions from previous models, achieves an accuracy of 96.66 %.

TABLE 9. **Evaluation metrics for different classification methods using the top 10 most relevant features.**

Model	Sensitivity	Specificity	Accuracy	MCC
RF	100.00 %	100.00 %	100.00 %	1
KNN	100.00 %	100.00 %	100.00 %	1
SVM	100.00 %	100.00 %	100.00 %	1
NNET	100.00 %	100.00 %	100.00 %	1
STACK	100.00 %	100.00 %	100.00 %	1
RF Synthetic data	96.66 %	100.00 %	93.30 %	0.707
KNN synthetic data	96.66 %	100.00 %	93.30 %	1
SVM synthetic data	80.20 %	77.00 %	83.00 %	0.707
NNET synthetic data	96.66 %	100.00 %	93.33 %	0.707
STACK synthetic data	96.66 %	100.00 %	93.33 %	1

The decrease in accuracy, sensitivity, and specificity is observed during the "blind" test with synthetic data n=30 table 9. The evaluation metrics of different classification methods vary, it is noteworthy that the final model, based on predictions from previous models, achieves 100% accuracy. However, something alarming is the drastic decrease in the performance of the SVM classification model.

In this study, motor activity data from the dataset published by García-Ceja[6] were analyzed to compare different classification methods using machine learning techniques and to obtain useful data through data mining practices, also considering previous research and the findings reported by Rodríguez *et al.*[3]. In terms of descriptive statistics, significant differences ($p < 0.05$) were found between the movement of healthy individuals and those affected by depression, mainly during the day, afternoon, and night periods. Several neurobiological and psychological mechanisms may explain why the control group exhibits greater motor activity than the condition group. First, depression is associated with dysregulation of neurotransmitter systems, particularly dopaminergic and serotonergic pathways, which play a crucial role in motivation, reward processing, and motor function. Reduced dopamine levels, especially in the mesolimbic and mesocortical pathways, can lead to decreased goal-directed behavior and physical activity. Similarly, serotonin dysfunction has been linked to fatigue and reduced psychomotor speed[12], hypoactivity in the prefrontal lobe (dorsal medial prefrontal cortex [dmPFC], ventral medial prefrontal cortex [vmPFC], and dorsal lateral prefrontal cortex [dlPFC], ventral lateral prefrontal cortex [vlPFC], orbital frontal cortex [OFC])[33] regions involved in movement initiation and executive function, has been observed in individuals with depression. This reduced neural activity may contribute to slower movement, decreased exploration, and a general lack of physical engagement. Additionally, hyperactivity of the hypothalamic-pituitary-adrenal (HPA) axis, leading to chronic stress and elevated cortisol levels, has been shown to negatively affect energy levels and contribute to fatigue, further reducing motor activity[14].

From a behavioral perspective, individuals with depression may experience anhedonia, a decreased ability to experience pleasure, leading to reduced engagement in activities that typically require movement or physical effort. This aligns with the behavioral inhibition system (BIS) theory, which suggests that increased sensitivity to negative stimuli in depression leads to avoidance behavior and decreased motor output[34].

The results reported in this study regarding the predictive capacity of movements recorded through accelerometers in conjunction with the application of different machine learning algorithms suggest that models like RF, KNN, and SVM are well-performance tools that can achieve accuracy above 96 % in the classification of mental illnesses, similar to what Vahia *et al.*[4], mention in their research on monitoring systems related to MDD. However, the implementation of ensemble models provides the possibility of making predictions with stronger support[35]. While this article mentions 100 % accuracy with the final ensemble model, it is recommended to reproduce the model, if possible, with a more representative sample, and conduct power tests.

CONCLUSIONS

The objectives of this study were the statistical analysis of the data provided by the dataset proposed by Garcia *et al.*[6], as well as the comparison of the performance of different machine learning models for the classification between depressive and non-depressive subjects based on the motor activity during daily activities over days. Contrasting the findings of this work with those mentioned earlier, it is concluded that motor activity is a viable parameter for identifying depressive behaviors. The significant differences between healthy and depressive subjects indicate a significant reduction in movements by those affected by this condition, a possible effect of apathy on human behavior.

Regarding the performances of the different machine learning models, although the algorithms showed performances greater than 96 % with the data extracted from the original dataset, these metrics significantly decreased when subjected to new synthetic blind data. However, in both cases, the stacked model showed an improvement in the classification of different groups, achieving an accuracy of 100 % considering 10 relevant features for the different prediction methods. Nevertheless, it is recommended to increase the sample size, as having so few data points can lead to overfitting issues.

ETHICAL STATEMENT

This publication is based on the works of Garcia-Ceja[6] The acquisition of the original dataset was part of the Introducing Mental health through Adaptive Technology (INTROMAT) project, funded by the Norwegian Research Council (259293/070).

AUTHOR CONTRIBUTION

G. N. R.-R. conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing - original draft, writing - review, and editing; C. E. G.-T. conceptualization, investigation, methodology, writing - original draft, writing - review, and editing; J. I. G.-T. supervision; J. M. C.-P. data curation, supervision; E. A. C. supervisión.

REFERENCIAS

[1] M. E. Morris and A. Aguilera, "Mobile, social, and wearable computing and the evolution of psychological practice," Prof. Psychol. Res. Pract., vol. 43, no. 6, pp. 622-626, 2012, doi: https://doi.org/10.1037/a0029041

[2] C. A. Woody et al., "A systematic review and meta-regression of the prevalence and incidence of perinatal depression," J. Affect Disord, vol. 219, pp. 86-92, 2017, doi: https://doi.org/10.1016/j.jad.2017.05.003

[3] . G. Rodríguez-Ruiz et al., "Comparison of night, day and 24 h motor activity data for the classification of depressive episodes," Diagnostics (Basel), vol. 10, no. 3, 2020, art. no. 162, doi: https://doi.org/10.3390/diagnostics10030162

[4] I. V. Vahia and D. D. Sewell, "Late-life depression: a role for accelerometer technology in diagnosis and management," Amer. J. Psychiatry, vol. 173, no. 8, pp. 763-768, 2016, doi: https://doi.org/10.1176/appi.ajp.2015.15081000

[5] A. Abd-Alrazaq et al., "Wearable artificial intelligence for anxiety and depression: scoping review," J. Med. Internet Res., vol. 25, 2023, art. no. e42672, doi: https://doi.org/10.2196/42672

[6] E. Garcia-Ceja et al., "Depresjon: a motor activity database of depression episodes in unipolar and bipolar patients," in Proc. 9th ACM Multimedia Syst. Conf., Amsterdam, Netherlands , 2018, pp. 472-477, doi: https://doi.org/10.1145/3204949.3208125

[7] A. Massaro et al., "Decisional support system with artificial intelligence oriented on health prediction using a wearable device and big data," in 2020 IEEE Int. Workshop Metrol. Ind. 4.0 & IoT, Roma, Italy, 2020, pp. 718-723, doi: https://doi.org/10.1109/MetroInd4.0IoT48571.2020.9138258

[8] D. Wang, J. Weng, Y. Zou, and K. Wu, "EmoTracer: a wearable physiological and psychological monitoring system with multi-modal sensors," in Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers, New York, NY, USA, 2022, pp. 444-449, doi: https://doi.org/10.1145/3544793.3560409

[9] P. M. Singh and P. Sathidevi, "Design and implementation of a machine learning-based technique to detect unipolar and bipolar depression using motor activity data," in Smart Trends in Computing and Communications, vol. 286, 2021, pp. 99-107, doi: https://doi.org/10.1007/978-981-16-4016-2_10

[10] C. Iribarne, V. Renner, C. Pérez, and D. L. de Guevara, "Trastornos del Ánimo y Demencia. Aspectos clínicos y estudios complementarios en el diagnóstico diferencial," Rev. Méd. Clín. Las Condes, vol. 31, no. 2, pp. 150-162, 2020, doi: https://doi.org/10.1016/j.rmclc.2020.02.001

[11] G. Jackson-Koku, "Beck Depression Inventory," Occup. Med., vol. 66, no. 2, pp. 174-175, 2016, doi: https://doi.org/10.1093/occmed/kqv087

[12] J. Morrison. DSM-5® Guía para el diagnóstico clínico. (2015). [Online]. Available: https://cdn.website-editor.net/50c6037605bc4d1e9286f706427108e6/files/uploaded/DSM-5%2520guia%2520para%2520el%2520diagnstico%2520clinico%2520-%2520James%2520Morrison.pdf.

[13] J. Davidson et al., "The Montgomery-Åsberg Depression Scale: reliability and validity," Acta Psychiatr. Scand., vol. 73, no. 5, pp. 544-548,1986, doi: https://doi.org/10.1016/j.csbj.2024.07.005

[14] World Health Organization, 2015, "International statistical classification of diseases and related health problems," distributed by World Health Organization, https://iris.who.int/handle/10665/246208

[15] S. Saeb et al., "The relationship between clinical, momentary, and sensor-based assessment of depression," in 2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), Istanbul, Turquía, 2015, doi: https://doi.org/10.4108/icst.pervasive-health.2015.259034

[16] Z. Y. Wee et al., "Actigraphy studies and clinical and biobehavioural correlates in schizophrenia: a systematic review," J. Neural. Transm., vol. 126, no. 5, pp. 531-558, 2019, doi: https://doi.org/10.1007/s00702-019-01993-2

[17] J. O. Berle et al., "Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression," BMC Res. Notes, vol. 3, no. 1, 2010, art. no. 149, doi: https://doi.org/10.1186/1756-0500-3-149

[18] R. Ramakrishnan et al., "Accelerometer measured physical activity and the incidence of cardiovascular disease: Evidence from the UK Biobank cohort study," PLoS Med., vol. 18, no. 1, 2021, art. no. e1003487, doi: https://doi.org/10.1371/journal.pmed.1003487

[19] B. Helgadóttir, Y. Forsell, and Ö. Ekblom, "Physical activity patterns of people affected by depressive and anxiety disorders as measured by accelerometers: a cross-sectional study," PLoS One, vol. 10, no. 1, 2015, art. no. e0115894, doi: https://doi.org/10.1371/journal.pone.0115894

[20] A. C. Cote et al., "Evaluation of wearable technology in dementia: a systematic review and meta-analysis," Front. Med., vol. 7, 2021, art. no. 501104, doi: https://doi.org/10.3389/fmed.2020.501104

[21] V. Trevino and F. Falciani, "GALGO: an R package for multivariate variable selection using genetic algorithms," Bioinformatics, vol. 22, no. 9, pp. 1154-1156, 2006, doi: https://doi.org/10.1093/bioinformatics/btl074

[22] W. S. Noble, "What is a support vector machine?," Nat. Biotechnol., vol. 24, no. 12, pp. 1565-1567, 2006, doi: https://doi.org/10.1038/nbt1206-1565

[23] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in Proc. COLT92 5th. Annu. Workshop Comput. Learn. Theory, Pittsburgh, PA, USA, 1992, pp. 144-152, doi: https://doi.org/10.1145/130385.130401

[24] A. Krogh, "What are artificial neural networks?," Nat. Biotechnol., vol. 26, no. 2, pp. 195-1977, 2008, doi: https://doi.org/10.1038/nbt1386

[25] G. Guo et al., "KNN model-based approach in classification," in On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE, Sicily, Italy, vol. 2888, 2003, pp. 986-996, doi: https://doi.org/10.1007/978-3-540-39964-3_62

[26] B. Mahesh et al., "Machine learning algorithms-a review," Int. J. Sci. Res., vol. 9, no. 1, pp. 381-386, 2020, doi: https://doi.org/10.21275/ART20203995

[27] M. Gashler, C. Giraud-Carrier, and T. Martinez, "Decision tree ensemble: Small heterogeneous is better than large homogeneous," in 2008 7th Int. Conf. Mach. Learn. Appl., San Diego, CA, USA, 2008, pp. 900-905, doi: https://doi.org/10.1109/ICMLA.2008.154

[28] A. B. Shaik and S. Srinivasan, "A brief survey on random forest ensembles in classification model," in Proc. 2018 Int. Conf. Innov. Comput. Commun., vol. 56, Nov. 2018, pp. 253-260, doi: https://doi.org/10.1007/978-981-13-2354-6_27

[29] L. Pietruczuk, L. Rutkowski, M. Jaworski, and P. Duda, "How to adjust an ensemble size in stream data mining?," Inf. Sci., vol. 381, pp. 46-54, 2017, doi: https://doi.org/10.1016/j.ins.2016.10.028

[30] S. Chetlur et al., Dec. 2014, "ccuDNN: Efficient primitives for deep learning," 2014, arXiv:1410.0759, doi: https://doi.org/10.48550/arXiv.1410.0759

[31] A. C. Davison and D. V. Hinkley, Bootstrap methods and their application (Cambridge Series in Statistical and Probabilistic Mathematics). New York, NY, USA: Cambridge University press, 1997, doi: https://doi.org/10.1017/CBO9780511802843

[32] Y. Xie et al., "Stacking ensemble learning models for daily runoff prediction using 1D and 2D CNNs," Expert Syst. Appl., vol. 217, 2023, art. no. 119469, doi: https://doi.org/10.1016/j.eswa.2022.119469

[33] I. M. Lin et al., "Prefrontal Lobe and Posterior Cingulate Cortex Activations in Patients with Major Depressive Disorder by Using Standardized Weighted Low-Resolution Electromagnetic Tomography," J. Pers. Med., vol. 11, no. 11, 2021, art. no. 1054, doi: https://doi.org/10.3390/jpm11111054

[34] S. Wang, F. Leri and S. J. Rizvi, "Anhedonia as a central factor in depression: Neural mechanisms revealed from preclinical to clinical evidence," Prog. Neuro-Psychopharmacol. Biol. Psychiatry vol. 110, 2021, art. no. 110289, doi: https://doi.org/10.1016/j.pnpbp.2021.110289

[35] J. Chen, A. Zeb, Y. A. Nanehkaran, and D. Zhang, "Stacking ensemble model of deep learning for plant disease recognition," J. Ambient Intell. Human. Comput., vol. 14, pp. 12359-12372, 2023, doi: https://doi.org/10.1007/s12652-022-04334-6