

<https://doi.org/10.17488/RMIB.44.4.3>

E-LOCATION ID: 1364

Feature Selection of Motor Activity in Intervals of Time with Genetics Algorithms for Depression Detection

Selección de Características de la Actividad Motora en Intervalos de Tiempo con Algoritmos Genéticos para la Detección de Depresión

Carlos H. Espino-Salinas¹ , Carlos E. Galván-Tejada¹  , Ana G. Sánchez-Reyna¹ , Huizilopoztli Luna-García¹ ,
Hamurabi Gamboa-Rosales¹ , Jorge A. Morgan-Benita¹ , José M. Celaya-Padilla¹ , Jorge I. Galván-Tejada¹ 

¹Universidad Autónoma de Zacatecas, Zacatecas - México

ABSTRACT

It is estimated that depression affects more than 300 million people in worldwide. Unfortunately, the current method of psychiatric evaluation requires a great effort on the part of clinicians to collect complete information. The aim of this paper is determine the optimal time intervals to detect depression using genetic algorithms and machine learning techniques; from motor activity readings of 55 participants during a week at one-minute intervals. The time intervals with the best performance in detecting depression in individuals were selected by applying Genetic Algorithms (GA). Methodology. 385 observations of the study participants were evaluated, obtaining an accuracy of 83.0 % with Logistic Regression (LR). Conclusion. There is a relationship between motor activity and people with depression since it is possible to detect it using machine learning techniques. However, the changes in the variables of the time intervals could be established as key factors since, at different times, they could give good or bad results because the motor activity in the patients could vary. However, the results present a first approximation for developing tools that help the opportune and objective diagnosis of depression.

KEYWORDS: artificial intelligence, depression, feature selection, genetic algorithm, motor activity

RESUMEN

Se estima que la depresión afecta a más de 300 millones de personas en el mundo. Desafortunadamente, el método de evaluación psiquiátrica actual requiere un gran esfuerzo por parte de los médicos para recopilar información completa. Objetivo. Determinar los intervalos de tiempo óptimos para detectar depresión mediante algoritmos genéticos y técnicas de aprendizaje automático, a partir de las lecturas de actividad motora de 55 sujetos durante una semana en intervalos de un minuto. Los intervalos de tiempo con mejor desempeño en la detección de depresión en individuos fueron seleccionados aplicando algoritmos genéticos. Metodología. Se evaluaron 385 observaciones de los sujetos de estudio, obteniendo una precisión del 83.0 % con Regresión Logística (LR). Conclusión. Existe una relación entre la actividad motora y las personas con depresión ya que es posible detectarla utilizando técnicas de aprendizaje automático. Sin embargo, los cambios en las variables de los intervalos de tiempo podrían establecerse como factores clave ya que en diferentes momentos podrían dar buenos o malos resultados debido a que la actividad motora en los pacientes podría llegar a variar. No obstante, los resultados presentan una primera aproximación para el desarrollo de herramientas que ayuden al diagnóstico oportuno y objetivo de la depresión.

PALABRAS CLAVE: actividad motora, algoritmos genéticos, depresión, inteligencia artificial, selección de características

Corresponding author

TO: Carlos E. Galván-Tejada

INSTITUTION: Universidad Autónoma de Zacatecas

ADDRESS: Jdn. Juárez #147, Centro Histórico, 98000
Zacatecas, Zac.

EMAIL: ericgalvan@uaz.edu.mx

Received:

19 July 2023

Accepted:

14 September 2023

INTRODUCTION

Per the World Health Organization (WHO), depression is distinguished from typical mood variations and brief emotional reactions to everyday life challenges. Particularly when it becomes recurrent and exhibits moderate or severe intensity, depression can evolve into a significant health concern ^[1]. Generally, it emerges early in life, causing a substantial decline in the overall functioning of individuals. The condition tends to recur and imposes notable economic and social burdens, making it a prominent contributor to the list of debilitating illnesses ^[2]. In its most severe form, depression can tragically result in suicide, with nearly one million people committing suicide each year. Shockingly, it stands as the second leading cause of death among individuals aged 15 to 29, as reported by WHO data. Additionally, in the past, the COVID-19 pandemic has brought forth numerous overwhelming stresses. Some evident factors include job loss, bereavement of family members, friends, or coworkers, financial instability, and social isolation, especially for individuals living alone. When required, healthcare providers must differentiate between demoralization and depression; however, access to in-person consultations with qualified mental health experts may not be easily accessible to everyone in need ^[3].

Apart from the difficulties posed by the inability to meet patients with depression in person, there are inherent challenges in conducting psychiatric assessments. Such evaluations require considerable effort from specialists to collect objective patient information. Moreover, successful assessment heavily relies on the patient willingness to cooperate and effectively communicate their symptoms and concerns ^[4]. One of the techniques employed to assess patients' depression is the Montgomery-Asberg Depression Rating Scale (MADRS), designed to gauge the current severity of ongoing depression ^[5]. Clinicians evaluate ten depression-relevant items through observations and discussions with the patient, and a cumulative score (ranging from 0 to 60) indicates the level of depression. Scores

below ten are categorized as having no depressive symptoms ^[6]. While scores above 30 indicate a severe depressive state ^[7]. Therefore, an objective detection mechanism based on biological signals is needed to improve timely diagnosis.

The extensive use of wearable devices for monitoring mental and physical health has gained significant popularity. Now, people are consistently collecting data to improve their well-being and monitor fitness advancements. Moreover, the data collected from these devices can hold considerable value from a psychiatric perspective, extending beyond evaluating the overall quality of life. It holds the potential to aid in diagnosing various mental health conditions, including depression ^[8]. Undoubtedly, motor activity reflects social patterns influenced by cyclical biological rhythms, which are regulated by the 24-hour circadian clock and interwoven with several ultradian rhythmic cycles lasting 2 to 6 hours ^[9]. Disrupted biological rhythmic patterns have been proposed as significant indicators of mood episodes ^[10]. Actigraphy serves as a non-intrusive approach to observing human rest and activity patterns. Typically, it involves using a wrist-worn device to record gravitational acceleration units ^[11].

The actigraph is among the devices frequently employed to collect motor activity data. Numerous research studies have leveraged this data to create models using artificial intelligence techniques for classifying, detecting, and monitoring the illness. García-Ceja *et al.*, ^[12] utilize machine learning to differentiate between depressed and non-depressed patients. To assess the algorithms' performance, they employ leave-one-patient-out validation. The collective results reveal that sensor data contains valuable information for determining an individual depression status. On the other hand, Zanella-Calzada *et al.*, ^[13] proposed a novel approach to distinguish depressive participants from control participants using data from their wearable device-recorded motor activity. Statistical features were extracted from the motor activity signals, which

were then utilized to train a random forest classifier. Galván-Tejada *et al.*,^[14] investigated the accelerometer signal from smart bands to identify depressive states based on patients' activity. A statistical feature extraction technique was devised, focusing on the temporal and spectral evolution of the signal. Furthermore, an intelligent feature selection method utilizing GA was incorporated to optimize the non-invasive diagnostic process efficiently. The results demonstrate the potential to distinguish between depressive states using the smart band activity signal, offering a preliminary and automated tool for almost real-time depression diagnosis at a lower computational cost to specialists.

Moreover, researchers have explored the utilization of motor activity in time series data to gather valuable insights into identifying potential cases of depression, among other applications. Frogner *et al.*,^[15] Initially employed One-Dimensional Convolutional Neural Networks (1D-CNN) to assess motor activity for depression detection. Subsequently, the study extended its scope to identify three levels of depression (no depression, mild, and severe) using the MADRS scale. The final model successfully predicts the MADRS scores of the participants.

Rodríguez-Ruiz *et al.*,^[16] introduced a series of models aimed at classifying depressive and non-depressive episodes throughout different moments of the day (day, night, and full day) based on participants motor activity levels. The Depresjon database, containing activity data from both depression patients and controls, was utilized in the study. Additionally, they proposed a Random Forest Classifier (RFC) model for multiclass classification, distinguishing schizophrenia, depression, and healthy controls using night-time activity data with an impressive 98 % accuracy in detecting all three classes. Experimental results demonstrated the model efficacy in identifying episodes of depression and schizophrenia, as well as healthy controls, surpassing prior studies that

employed computationally expensive algorithms such as CNN and Bidirectional Recurrent Neural Networks (BRNN), resulting in a noteworthy boost in accuracy.^[17]

Jakobsen *et al.*,^[18] conducted a study to investigate the potential of various machine learning algorithms in distinguishing between depressed patients and healthy controls using motor activity time series. Furthermore, their research demonstrated that machine learning capacity to reveal hidden patterns in the data aligns with the conclusions drawn from previous studies utilizing both linear and nonlinear statistical methods in motor activity analysis.

Artificial Intelligence (AI), has proven to be useful tool for detecting cases of depression. Kour *et al.*,^[19] propose that unipolar depression and bipolar depression display similar clinical symptom profiles, presenting a considerable challenge in distinguishing between the two depression types. The disruptions in motor activity offer a potential avenue to detect pathological mental states and may prove valuable in addressing this diagnostic challenge. That is why every day, there is more research studying this type of data where different types of algorithms derived from AI have been implemented, as is the case with Pacheco-González *et al.*,^[20] conducted a comparative analysis of various classification techniques, including conditional inference trees, random forest, K-Nearest Neighbor, support vector machine, and Naïve Bayes. The study aimed to predict depressive states based on patients' activity, measured using a smart band accelerometer. Conversely, Raihan *et al.*,^[21] employed a combination of motor sensor readings and demographic data along with machine learning techniques like Random Forest (RF), AdaBoost, and Artificial Neural Networks (ANN). Finally, as mentioned by Singh *et al.*,^[22] identifying disruptions in motoric activity could serve as a valuable approach to detecting pathological mental states. Thus, creating distinct motor activity database (Depresjon) containing data from patients with unipo-

lar depression, bipolar depression, and healthy individuals has demonstrated effectiveness in the timely detection of depression cases.

The objective of this research is to present an approach capable of objectively identifying episodes of depression by employing various AI algorithms in order to explore a wider range of options for detecting depression, this using a limited amount of data intelligently selected with genetics algorithms that reduced the dimensionality and redundancy of the data, it can obtain model capable of training simply with new input data and processing it to generate models adapted to different types of patients, in different environments. It also seeks to show that the behavior of each patient as well as each person is always different, depending on the situation or environment in which they find themselves, therefore, it is important to create adaptive artificial intelligence models for each participants to detect depression in a timely manner and help mental health specialist to develop a treatment for each degree of depression that can be found. Critical analysis of the current relevant literature, statement of the general aims of the work and the importance of the same. In the text of the article, the references must be numbered in the order that they appear.

MATERIALS AND METHODS

The methodology proposed in this work consists of five main stages as shown in Figure 1. Initially, the data are obtained from the Depresjon database. Next, in the second phase of the process, data pre-processing, a comprehensive explanation is given on how the dataset will be organized to extract each minute as a feature from the original data obtained in the initial step. This arrangement enables a subsequent feature selection employing GA. Lastly, to verify the significance of the chosen features, two crucial stages were undertaken: the first involved classification analysis through the implementation of AI algorithms such as Logistic Regression (LR), Artificial Neural Networks (ANN), Support Vector Machine (SVM), and Decision Tree

(DT). The second stage focused on result validation.

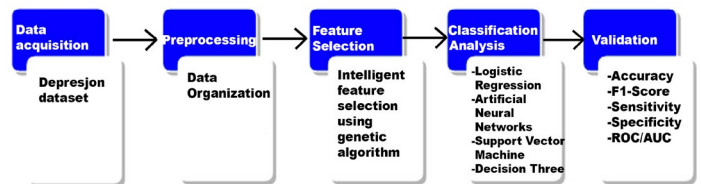


FIGURE 1. Proposed methodology for intelligent feature selection for objective detection of depressive symptoms.

Data Acquisition

The motor activity dataset comprises patient data monitored using an actigraph watch on the right wrist this measure activity using a piezoelectric accelerometer. This actigraph watch, named "Actiwatch AW4", measures activity levels at a sampling frequency of 32 Hz, recording movements above 0.05 g. These movements correspond to particular voltage (v) values, which are stored as activity counts in the Actiwatch memory. The count values directly correlate to the intensity of the movements. Continuous recording of total activity counts occurred at one-minute intervals, with the activity counts being recorded accordingly at the same one-minute intervals [1]. Information was gathered from a group of 22 psychotic patients and all used antipsychotic medications who were admitted to Hauklend University Hospital, the mean age to the first time of hospitalization was 24 +/- 9.3. The specialist diagnosed the patients using a semi structured interview based on Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [23]. Within this group, were 3 females and 19 males, with an average age of 42.6 years (ranging from 27 to 69 years). As for the healthy control group included, it included 23 hospital employees, five students, and four individual practitioner clinic [18].

Data Pre-processing

To obtain a dataset consisting of a certain number of observations capable of being processed through AI techniques, a period of a week to maintain data diversity was taken in one minute intervals. we reordered

them to identify the number of observations per participant of study as a relation between participant-day, with their respective registered data.

Motor Activity Data (MAD) for each study participant is filed in "C" columns, spanning a full day from 00:00 to 23:59, with each minute corresponding to its respective MAD record. These columns are then transposed and added to the first row of an "A" matrix, accompanied by the corresponding output, where 0 indicates the control group, and 1 indicates the condition group based on the source dataset, successively. The same process is applied for the subsequent participating days until completing a week of information, resulting in a matrix $A = [385 \times 1441]$, including its output. The process is explained in more detail below.

$$\begin{aligned}
 C_{11} &= \begin{matrix} 00:00 \\ 00:01 \\ 00:02 \\ 00:03 \\ 00:04 \\ \vdots \\ 23:59 \end{matrix} \begin{bmatrix} 168 \text{ v} \\ 107 \text{ v} \\ 550 \text{ v} \\ 157 \text{ v} \\ 0 \text{ v} \\ \vdots \\ 208 \text{ v} \end{bmatrix} & C_{12} &= \begin{matrix} 00:00 \\ 00:01 \\ 00:02 \\ 00:03 \\ 00:04 \\ \vdots \\ 23:59 \end{matrix} \begin{bmatrix} 168 \text{ v} \\ 17 \text{ v} \\ 560 \text{ v} \\ 167 \text{ v} \\ 23 \text{ v} \\ \vdots \\ 280 \text{ v} \end{bmatrix} \\
 C_{13} &= \begin{matrix} 00:00 \\ 00:01 \\ 00:02 \\ 00:03 \\ 00:04 \\ \vdots \\ 23:59 \end{matrix} \begin{bmatrix} 118 \text{ v} \\ 157 \text{ v} \\ 450 \text{ v} \\ 157 \text{ v} \\ 0 \text{ v} \\ \vdots \\ 305 \text{ v} \end{bmatrix} & C_{14} &= \begin{matrix} 00:00 \\ 00:01 \\ 00:02 \\ 00:03 \\ 00:04 \\ \vdots \\ 23:59 \end{matrix} \begin{bmatrix} 148 \text{ v} \\ 157 \text{ v} \\ 120 \text{ v} \\ 457 \text{ v} \\ 121 \text{ v} \\ \vdots \\ 892 \text{ v} \end{bmatrix} \\
 C_{15} &= \begin{matrix} 00:00 \\ 00:01 \\ 00:02 \\ 00:03 \\ 00:04 \\ \vdots \\ 23:59 \end{matrix} \begin{bmatrix} 168 \text{ v} \\ 107 \text{ v} \\ 550 \text{ v} \\ 157 \text{ v} \\ 0 \text{ v} \\ \vdots \\ 208 \text{ v} \end{bmatrix} & C_{16} &= \begin{matrix} 00:00 \\ 00:01 \\ 00:02 \\ 00:03 \\ 00:04 \\ \vdots \\ 23:59 \end{matrix} \begin{bmatrix} 168 \text{ v} \\ 17 \text{ v} \\ 560 \text{ v} \\ 167 \text{ v} \\ 23 \text{ v} \\ \vdots \\ 280 \text{ v} \end{bmatrix} \\
 C_{17} &= \begin{matrix} 00:00 \\ 00:01 \\ 00:02 \\ 00:03 \\ 00:04 \\ \vdots \\ 23:59 \end{matrix} \begin{bmatrix} 118 \text{ v} \\ 157 \text{ v} \\ 450 \text{ v} \\ 157 \text{ v} \\ 0 \text{ v} \\ \vdots \\ 305 \text{ v} \end{bmatrix}
 \end{aligned} \tag{1}$$

Given the columns of the days corresponding to week 1, we would have $C_{11}, C_{12}, C_{13}, C_{14}, C_{15}, C_{16}, C_{17}$, where the first subscript indicates the week number and the sec-

ond indicates the day number with the accompanying timestamp of the motor activity record generated during that specific minute as shown in Equation 1. To obtain the total number of time intervals defined as characteristics corresponding to a day, a simple operation was carried out consisting of multiplying the 60 minutes that an hour contains by the number of hours that a day contains to obtain, as a result, a column that contains 1440 rows. These rows represent the minutes of a day with their respective MAD record, thus defining the time range to be used.

Afterward, Following the transposition, the columns are reorganized to create a vector V^T , where each one-minute interval signifies a feature. Each vector corresponds to an observation per participant, culminating in a row with 1440 columns. These vectors are then added to the matrix "A" rows. Each participant, whether depressed or not (labeled as one and zero), has their vector extracted daily, spanning 7 days (one week). Each vector is appended to matrix "A" simultaneously with the assignment of its output label. This process continues for the next participants until all 55 participants are incorporated into the new dataset configuration, as illustrated in Equation (2). This setup is utilized for feature selection to identify the most relevant minutes for depression detection.

$$\begin{aligned}
 & 00:00 \quad 00:01 \quad 00:02 \quad 00:03 \quad 00:04 \quad \dots \quad 23:59 \\
 V_{1,1}^T &= [256 \text{ v} \quad 107 \text{ v} \quad 550 \text{ v} \quad 157 \text{ v} \quad 0 \text{ v} \quad \dots \quad 208] \\
 V_{1,2}^T &= [168 \text{ v} \quad 17 \text{ v} \quad 560 \text{ v} \quad 167 \text{ v} \quad 23 \text{ v} \quad \dots \quad 280] \\
 V_{1,3}^T &= [118 \text{ v} \quad 157 \text{ v} \quad 450 \text{ v} \quad 157 \text{ v} \quad 0 \text{ v} \quad \dots \quad 305] \\
 V_{1,4}^T &= [148 \text{ v} \quad 157 \text{ v} \quad 120 \text{ v} \quad 457 \text{ v} \quad 121 \text{ v} \quad \dots \quad 892] \\
 V_{55,7}^T &= [563 \text{ v} \quad 124 \text{ v} \quad 456 \text{ v} \quad 321 \text{ v} \quad 265 \text{ v} \quad \dots \quad 201]
 \end{aligned} \tag{2}$$

Feature Selection

Bearing in mind that the data collected through a portable device known as an activewatch in intervals of one minute are considered here as characteristics for each observation, in this section, we seek to select the

minutes that provide the most significant information to generate a simple, intelligent model capable of detecting depression, considering that the number of features is much higher compared to observations, this process also aims to reduce redundancy and the amount of data to process to improve its performance.

The methodology for intelligent feature selection using Genetic Algorithms (GA) is shown in Figure 2.

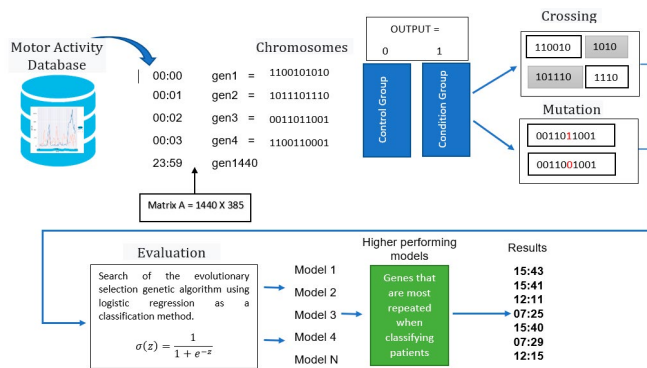


FIGURE 2. Methodology of the GA to select the most important features to classify patients with depression using Logistic Regression (LR).

GA draw inspiration from nature, particularly the process of natural selection. They implement a population-based search approach grounded in the fundamental principle of 'survival of the fittest.' At their core, GAs consist of several essential components: chromosome representation, selection, crossover, mutation, and fitness function computation [23].

In the context of the research, A library of genetic algorithms known as Genetic Algorithms for Multivariate Statistical Models from Large-Scale Functional Genomic Data 1.4 (GALGO) [24]. was implemented. GALGO, designed specifically for the R programming language, serves the purpose of selecting models with high fitness. This process begins by creating a random population of features with a specified size (n). These features are evaluated using a fitness function to assess their ability to classify the dependent variable, typically yielding an accuracy value.

For the specific task of depression detection, we utilized the logistic function as our classification method. In our study, the genetic algorithm iteratively explores and evolves combinations of genes (intervals of time) from a dataset comprising two classes. The goal is to identify features that effectively distinguish between these classes using the logistic regression (LR) method over 500 generations, considering 1000 possible solutions, and aiming for a fitness level of 95 %. As a result, the intervals of time obtained represent the most relevant contributors to the processed dataset. They are consistently favored within the intelligent models generated by the GA, demonstrating superior performance in classifying subjects.

Classification Analysis

Before starting the analysis of the dataset, a z-score normalization process is applied to the MADs to contain outliers as their variation in some time intervals can be significant. This normalization also speeds up the training time for each function within the same scale and is especially beneficial for modeling applications where the inputs often have varying scales. The mean and standard deviations are calculated for each feature [25], as shown in Equation 3. Then generate binary classification models since the data contains two possible classes as output, depressed (represented as 1) and not depressed (represented as 0). The techniques used to develop the models are ANN, LR, DT, and SVM. Finally, a comparison of the performance of these techniques was applied.

$$Z_i = \frac{x_i - \bar{x}}{\sigma} \quad (3)$$

The artificial neural network consists of two multi-layer perceptron (4, 2) and an output layer with a logistic activation function. On the other hand, the parameters established for the classification SVM are a linear kernel with a cost of 1.

The process for the development of the models consists in two steps, training, and testing. Therefore, the

data were randomly subsampled into two sets, one for each step. The dataset encompassed 80 % of data for training process, while the remaining 20 % was reserved for testing.

Validation

These parameters comprise accuracy, sensitivity, specificity, The Receiver Operating Characteristic Curve with the Area Under the Curve (ROC/AUC), and the F1-Score.

Accuracy (Acc) is a performance criterion that indicates the degree to which the outcome of a calculation aligns with the correct value [13]. as represented in Equation (4).

$$Accuracy(1 - Error) = \frac{TP + TN}{CP + CN} \quad (4)$$

In this context, TP represent True Positives, TN stands for True Negatives, CP denotes Truly Positive, and CN indicates Truly Negative.

The accuracy of a classifier encompasses other crucial aspects like precision, which refers to the number of correctly detected targets among all the targets detected. Additionally, there exists a relationship between the number of correctly detected targets and all known true targets, known as recall, as shown in Equations (5) and (6).

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

FP is False Positive, and FN is False Negative.

The measure that takes into account both precision and recall to evaluate the classification capability of an algorithm is referred to as the F1-Score. It is defined as

the harmonic mean of precision and recall [26], as shown in Equation (7).

$$F_1 = \frac{precision \cdot recall}{precision + recall} \quad (7)$$

Sensitivity, is the ability to accurately identify data with depressed symptoms, represents the number of condition participants correctly identified [27]. It is calculated using Equation (8).

$$Sensitivity(1 - \beta) = \frac{TP}{CP} \quad (8)$$

Specificity, referring to the ability to identify data without the condition as healthy, measures the proportion of negative samples that are accurately classified as such [28]. It is calculated using the following Equation (9).

$$Specificity(1 - \alpha) = \frac{TN}{CN} \quad (9)$$

The ROC curve is a commonly used method for evaluating machine learning models. It provides a visual representation of the classifier performance, enabling the selection of an appropriate operating point, referred to as the decision threshold, along with the AUC value [29]. The AUC can be calculated through integration, as shown in Equation (10).

$$AUC = \sum_i (1 - \beta_i \cdot \Delta\alpha) + \frac{1}{2} [\Delta(1 - \beta) \cdot \Delta\alpha] \quad (10)$$

RESULTS AND DISCUSSION

In this section, we explain the performed experiments and discuss the results.

Initially, Figure 3 shows the motor activity of depressed and non-depressed people throughout the day, extracted from the research paper entitled: Two-Dimensional Convolutional Neural Network for Depression Episodes Detection in Real Time Using Motor Activity Time Series of Depresjon Dataset [30]. As evident from the data, a significant distinction

between the two cases can be observed, with a noticeable decrease in movement observed among depressed individuals. The data was collected for each of the 55 participants under study (32 healthy and 23 depressed) for a week. Each day was treated as an individual observation for each participant, resulting in a total of 385 observations, providing enough data to develop a classification model for episodes of depression. This was achieved by multiplying the number of patients by the number of days in a week (55X 7).

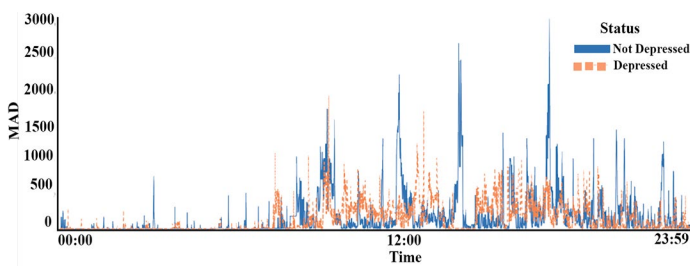


FIGURE 3. Samples collected with Actiwatch from both a non-depressed and a depressed subject of the Depresjon database [30].

The dataset comprises 385 observations, each containing 1440 minutes. Following the feature selection process, the most significant features were chosen, specifically: 15:43, 15:41, 12:11, 7:25, 15:40, 7:29, and 12:15, as depicted in Figure 4. The graph illustrates the frequency of appearance for each feature, with the highest-ranked features displayed in black.

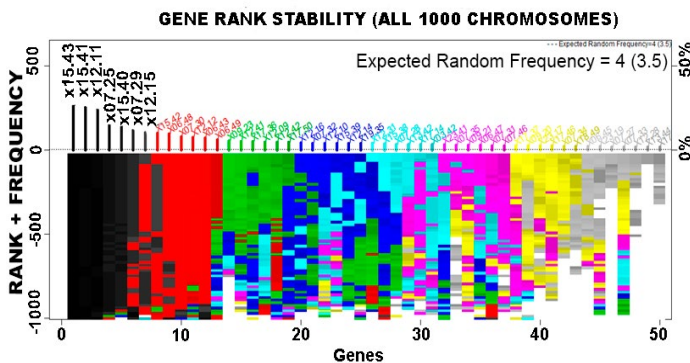


FIGURE 4. Most relevant features obtain out of the Genetic Algorithm.

Different artificial intelligence techniques were applied and validated with different metrics obtaining the following results. The ANN has an accuracy of 0.74, DT has 0.73, SVM obtained a performance of 0.81, and LR 0.83. Considering that logistic regression presents the best level of accuracy, it can be concluded that this technique can correctly classify a greater number of test participants than the others. To know the number of true positives and true negatives, as well as false positives and false negatives that were obtained by the different implemented algorithms, Table 1 shows the results of their confusion matrices which is often applied in machine learning to evaluate or visualize the model behavior in supervised classification scenarios [31]. Additionally, we add the Matthews Correlation Coefficient (MCC) which is a more reliable statistical index that produces a high score only if the prediction obtained good results in the four categories of the confusion matrix [32] and also the Kappa Correlation Coefficient (KCC) which is a metric to summarize the agreement between two nominal classifications, based on the same categories [33].

TABLE 1. Confusion matrix values and coefficient correlation.

ML Algorithm	Results of the algorithms with the test data set					
	FP	FN	TP	TN	MCC	KCC
SVM	5	9	27	36	0.635	0.632
LR	5	8	27	37	0.659	0.660
DT	11	11	21	34	0.411	0.411
ANN	8	10	24	35	0.523	0.523

The Results of the different validation metrics implemented to know the performance of classification models are shown in Table 2, where the values of accuracy, AUC, F1-Score, sensitivity, and specificity are specified.

TABLE 2. Validation Results with AG Features Selection.

Validation Metrics	Artificial Intelligence Algorithms			
	SVM	LR	DT	ANN
Accuracy	0.81	0.83	0.71	0.76
AUC	0.82	0.83	0.70	0.76
F1-Score	0.83	0.85	0.75	0.80
Sensitivity	0.87	0.90	0.75	0.81
Specificity	0.75	0.75	0.65	0.70

As can be seen in the results of the previous table, we can notice that the specificity values are lower in all the models generated to detect cases of depression, this is due to a class balancing problem since there are more data from healthy study participants than with a certain degree of depression, but regardless of this phenomenon and considering the limited number of observations, the algorithms used and the small number of features used, the results are favorable.

On the other hand, the ROC curves that graphically show the relationship between the sensitivity and specificity of the classification models generated when applied to a set of test data. This next image gives us a picture of the extent to which the study participants (healthy and depressed) are classified correctly.

First, in Figure 5 shows the performance obtained by the SVM algorithm to classify the participants, the model was subjected to a detailed analysis through the construction of a Receiver Operating Characteristic (ROC) curve and several key results were obtained indicating its predictive ability. The AUC was calculated to be 0.822, with a 95% confidence interval between 0.722 and 0.921. This AUC value, which is closer to 1 than 0, suggests that the model has a robust ability to discriminate between classes, indicating promising performance.

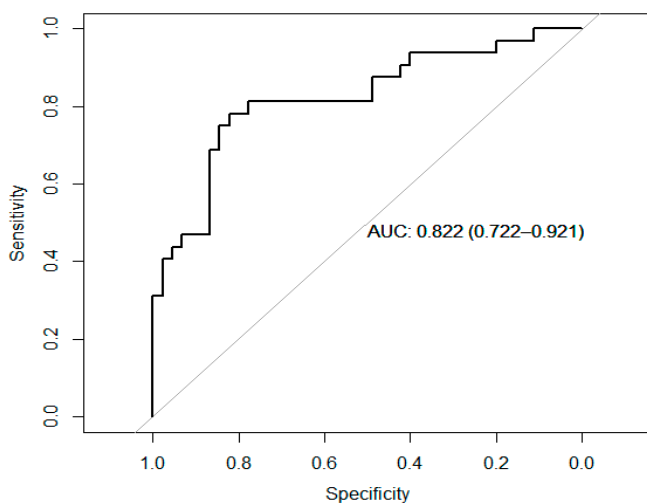


FIGURE 5. ROC curves of the SVM algorithm.

Secondly, A comprehensive analysis of the LR model was also carried out in the same classification context. Figure 6 illustrates the performance achieved by the LR algorithm in classifying the participants.

For the LR model, the AUC was calculated to be 0.838, with a 95 % confidence interval between 0.751 and 0.926. This AUC value, which is even closer to 1 than that obtained with the SVM model, reinforces the ability of the Logistic Regression model to effectively distinguish between the classes of interest. The high AUC is a strong indication of its predictive performance.

Comparing the results of both models, we observed that the Logistic Regression model obtained a slightly higher AUC (0.838) compared to the SVM model (0.822). This difference might suggest that the Logistic Regression model performs marginally better on this particular classification task. However, it is important to note that the choice between these models could depend on other factors, such as interpretability and simplicity of the model.

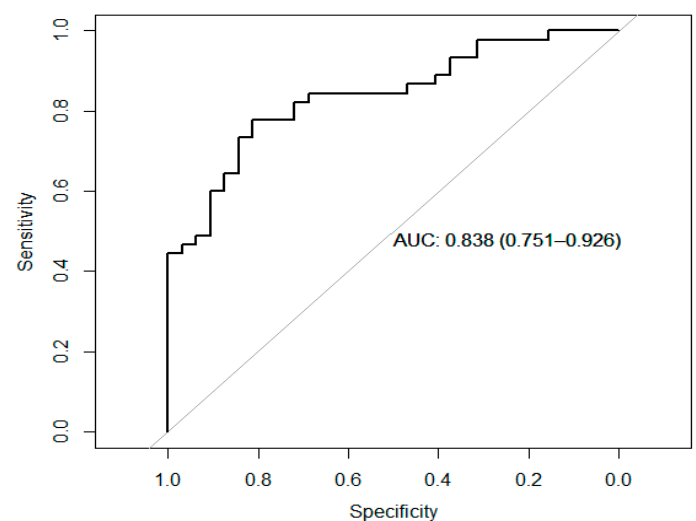


FIGURE 6. ROC curves of the LR algorithm.

Figure 7 depicts the performance attained through the DT algorithm for participant classification. Although the AUC is lower compared to the previous

models (SVM and Logistic Regression), it is still in a range that suggests some discriminative ability. However, this AUC value indicates that the Decision Tree model may have more limited predictive performance in this particular task.

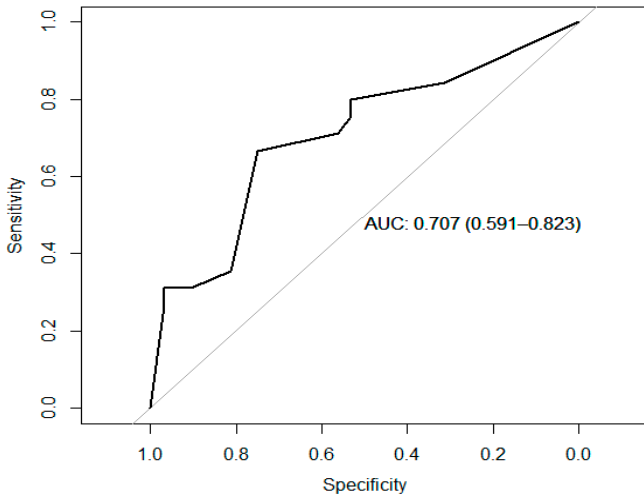


FIGURE 7. ROC curves of the DT algorithm.

In the framework of the research, the performance of an Artificial Neural Network (ANN) model as shown in Figure 8 was also evaluated in the classification task, along with SVM, Logistic Regression and Decision Tree models. When we compare the AUC of the ANN model (0.764) with the SVM (0.822) and Logistic Regression (0.838) models, we observe that ANN is in an intermediate position in terms of discrimination ability. Although it does not outperform the previous models in AUC, its performance is competitive and can be considered for applications where interpretability is not the main concern.

In summary, the results of the ROC curves of the models are highly promising. The high AUC, the balance between sensitivity and specificity, and the low false positive and false negative rate suggest that the model is an effective tool in the classification task studied. These results have important implications in the context of developing tools to support the diagnosis of depressive episodes, and could be valuable for clinical applications. However, it is important to con-

sider the limitations of the model and future areas of research, such as the optimization of classification thresholds, to maximize its usefulness in the real world.

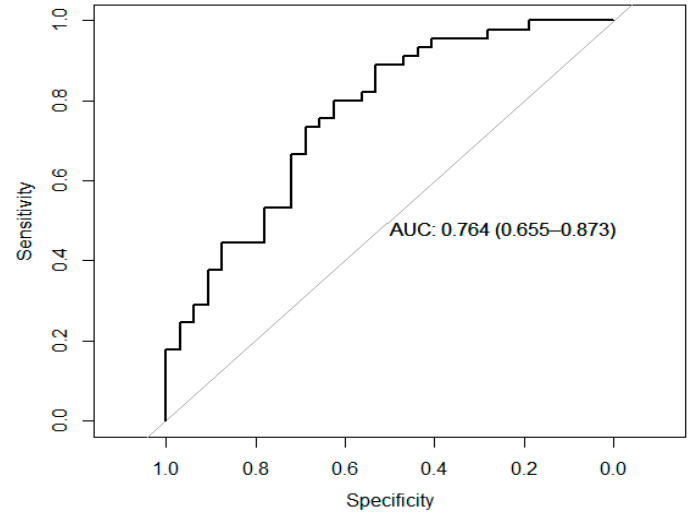


FIGURE 8. ROC curves of the ANN algorithm.

One of the assumptions that could not be omitted is where these algorithms were tested with all the features proposed in this research work; that is why we proceeded to a phase where the algorithms used the 1440 features to know their performance; the results are shown in Table 3 and Table 4.

TABLE 3. Confusion matrix values and coefficient correlation with 1440 Features.

ML Algorithm	Results of the algorithms with the test data set					
	FP	FN	TP	TN	MCC	KCC
SVM	9	16	23	29	0.358	0.352
LR	16	20	16	25	0.054	0.054
DT	10	18	22	27	0.283	0.277
ANN	10	15	22	30	0.349	0.346

TABLE 3. Validation Results with 1440 Features

Validation Metrics	Artificial Intelligence Algorithms			
	SVM	LR	DT	ANN
Accuracy	0.67	0.53	0.63	0.67
AUC	0.68	0.52	0.64	0.67
F1-Score	0.69	0.58	0.65	0.70
Sensitivity	0.76	0.60	0.72	0.75
Specificity	0.58	0.44	0.55	0.59

Results presented in the previous table are not as favorable with those presented in Table 2. This is an indicator of the importance of an intelligent feature selection phase that prevents the process of analysis and processing of the information to have better results, avoiding overfitting, reducing redundancy, and selecting the most significant data.

Although there are currently several investigations that have obtained very important results in the detection of depression, the simplicity of the process developed in this research is a very important aspect to consider since the results obtained through an algorithm such as logistic regression and reduced number of data contrast with the methodological approach used in state of the art. Table 5. shows the results obtained from the most recent research as well as the features used as source data related to the data set used in this work.

TABLE 5. Comparative performance with state of the art.

Author	Features	Technique	Acc.
Garcia-Ceja <i>et al.</i> [1]	Feature vector	SVM	0.72
Frogner <i>et al.</i> [15]	Feature Vector	1D-CNN	0.71
Jakobsen <i>et al.</i> [18]	Statistical Features	CNN	0.84
Kumar <i>et al.</i> [34]	Statistical Features	CNN	0.85
Rodríguez-Ruiz <i>et al.</i> [16]	Statistical Features	RFC	0.98
Ghate <i>et al.</i> [35]	Statistical Features	Transfer Learning	0.96
Zakariah <i>et al.</i> [36]	Statistical Features	Deep Neural Network	0.99

The table above reveals several noteworthy observations. Firstly, many previous studies employ considerably more intricate methodologies to enhance depression identification through motor activity. These approaches often involve extracting a multitude of statistical features, resulting in a substantial increase in the number of variables for analysis, processing, and the creation of classification models. This research aims to make a significant contribution by introducing

a novel feature selection method using Genetic Algorithms (GA), which has not been employed with the Depression Dataset before. The primary objective is to reduce data volume without sacrificing critical information for identifying depressive states while also mitigating data redundancy.

Additionally, the achieved accuracy provides an initial step toward the development of more efficient models in terms of time and computational cost. Lastly, it's important to note that the obtained results hold statistical significance compared to the existing literature, despite the limited amount of processed data and the unique methodology proposed in this study

The proposal offers an innovative optimization approach for future work in developing algorithms, methodologies, and tools to aid in the detection of depression.

CONCLUSIONS

Artificial intelligence algorithms such as LR, supported by an objective feature selection method such as GA, allowed the efficient generation of a model capable of detecting depression with 83.0 % accuracy using only a few time intervals at different times of the day as a data source. The time intervals that showed significant information to generate models capable of detecting depression were: 15:43, 15:41, 12:11, 7:25, 15:40, 7:29, and 12:15 of a 24 Hrs. time system. Based on these findings, it can be inferred that the methodology introduced in this paper facilitates automatic and objective depression detection through various artificial intelligence approaches, achieving noteworthy accuracy with a set of seven features derived from patients' motor activity. As a result, this preliminary development of an assisted diagnostic tool emerges, offering potential assistance in mitigating the elevated error rates associated with diagnosing this condition. However, in this research, we assume that adding other variables such as type of diet, family history, sex, age, place of birth, and habits and implementing

these algorithms in a real environment to test their efficiency and improve his learning could help in the future to improve the diagnosis of this mental illness.

As part of future work, there is a suggestion to augment the number of experimental observations to present more robust results with a greater diversity of data, including the increase of variables other than motor activity. In addition, it is proposed to implement deep learning like convolutional neural networks in smart devices capable of monitoring the objects of study in different circumstances or recurrent neural networks to detect depression early and allows a prevention strategy.

AUTHOR CONTRIBUTIONS

C.H.E.S conceptualized the project, designed and developed the methodology, contributed to the writing of the original manuscript, and participated in the programming of the software. C.E.G.T. participated in the data curation and gathering, carried out formal analyses and participated in the development of the software for the data analyses. A.G.S.R. contributed to the writing, editing and reviewing of the manuscript and participated in the data visualization for their correct interpretation. H.L.G. analyzed and validated the results. H.G.R. obtained funding and financial resources and oversaw the project. J.A.M.B provided access to material resources and equipment, participated in the data gathering and performed experiments. J.M.C.P. supervised and guided the general research and oversaw the project. J.I.G.T. conceptualized the project. All authors reviewed and approved the final version of the manuscript.

REFERENCES

- [1] E. Garcia-Ceja, M. Riegler, P. Jakobsen, J. Tørresen, T. Nordgreen, K. J. Oedegaard, O. Bernt Fasmer, “Depresjon: A motor activity database of depression episodes in unipolar and bipolar patients,” in *MMSys '18: Proceedings of the 9th ACM Multimedia Systems Conference*, Amsterdam, Netherlands, 2018, pp. 472-477, doi: <https://doi.org/10.1145/3204949.3208125>
- [2] S. Berenzon, M. A. Lara, R. Robles, and M. E. Medina-Mora, “Depresión: Estado del conocimiento y la necesidad de políticas públicas y planes de acción en México,” *Salud Publica Mex.*, vol. 55, no. 1, pp. 74-80, Jan.-Feb. 2013, doi: <https://doi.org/10.1590/s0036-36342013000100011>
- [3] R. I. Shader, “COVID-19 and Depression,” *Clin. Ther.*, vol. 42, no. 6, pp. 962-963, Jun. 2020, doi: <https://doi.org/10.1016/j.clinthera.2020.04.010>
- [4] S. Khairuddin, S. Ahmad, A. H. Embong, N. N. W. N. Hashim, T. M. K. Altamas, S. N. S. Badaruddin, S. S. Hassan, “Classification of the Correct Quranic Letters Pronunciation of Male and Female Reciters,” *IOP Conf. Ser.: Mater. Sci.*, vol. 260, art. no. 012004, 2017, doi: <https://doi.org/10.1088/1757-899X/260/1/012004>
- [5] S. A. Montgomery and M. Asberg, “A new depression scale designed to be sensitive to change,” *Br. J. Psychiatry*, vol. 134, no. 4, pp. 382-389, Apr. 1979, doi: <https://doi.org/10.1192/bjp.134.4.382>
- [6] C. J. Hawley, T. M. Gale, and T. Sivakumaran, “Defining remission by cut off score on the MADRS: Selecting the optimal value,” *J. Affect. Disord.*, vol. 72, no. 2, pp. 177-184, Nov. 2002, doi: [https://doi.org/10.1016/s0165-0327\(01\)00451-7](https://doi.org/10.1016/s0165-0327(01)00451-7)
- [7] M. J. Müller, H. Himmerich, B. Kienzle, and A. Szegedi, “Differentiating moderate and severe depression using the Montgomery-Åsberg depression rating scale (MADRS),” *J. Affect. Disord.*, vol. 77, no. 3, pp. 255-260, Dec. 2003, doi: [https://doi.org/10.1016/s0165-0327\(02\)00120-9](https://doi.org/10.1016/s0165-0327(02)00120-9)
- [8] F. J. Penedo and J. R. Dahn, “Exercise and well-being: A review of mental and physical health benefits associated with physical activity,” *Curr. Opin. Psychiatry*, vol. 18, no. 2, pp. 189-193, Mar. 2005, doi: <https://doi.org/10.1097/00001504-200503000-00013>
- [9] C. Bourguignon and K. F. Storch, “Control of rest: Activity by a dopaminergic ultradian oscillator and the circadian clock,” *Front. Neurol.*, vol. 8, art. no. 614, Nov. 2017, doi: <https://doi.org/10.3389/fneur.2017.00614>
- [10] L. B. Alloy, T. H. Ng, M. K. Titone, and E. M. Boland, “Circadian Rhythm Dysregulation in Bipolar Spectrum Disorders,” *Curr. Psychiatry Rep.*, vol. 19, no. 4, art. no. 21, Apr. 2017, doi: <https://doi.org/10.1007/s11920-017-0772-z>
- [11] J. O. Berle, E. R. Hauge, K. J. Oedegaard, F. Holsten, O. B. Fasmer, “Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression,” *BMC Res. Notes*, vol. 3, art. no. 149, May 2010, doi: <https://doi.org/10.1186%2F1756-0500-3-149>
- [12] E. Garcia-Ceja et al., “Motor Activity Based Classification of Depression in Unipolar and Bipolar Patients,” in *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, Karlstad, Sweden, 2018, pp. 316-321, doi: <https://doi.org/10.1109/CBMS.2018.00062>
- [13] L. A. Zanella-Calzada, C. E. Galván-Tejada, N. M. Cávez-Lamas, M. C. Gracia-Cortés, R. Magallanes-Quintanar, J. M. Celaya-Padilla, J. I. Galván-Tejada, H. Gamboa-Rosales, “Feature extraction in motor activity signal: Towards a depression episodes detection in unipolar and bipolar patients,” *Diagnostics*, vol. 9, no. 1, art. no. 8, Jan. 2019, doi: <https://doi.org/10.3390/diagnostics9010008>
- [14] C. E. Galván-Tejada, L. A. Zanella-Calzada, H. Gamboa-Rosales, J. I. Galván-Tejada, N. M. Chávez-Lamas, M. C. Gracia-Cortés, R. Magallanes-Quintanar, J. M. Celaya-Padilla, “Depression Episodes Detection in Unipolar and Bipolar Patients: A Methodology with Feature Extraction and Feature Selection with Genetic Algorithms Using Activity Motion Signal as Information Source,” *Mob. Inf. Syst.*, vol. 2019, art. no. 8269695, 2019, doi: <https://doi.org/10.1155/2019/8269695>
- [15] J. I. Frogner, F. M. Noori, P. Halvorsen, S. A. Hicks, E. Garcia-Ceja, J. Torresen, M. A. Riegler, “One-dimensional convolutional neural networks on motor activity measurements in detection of depression,” in *HealthMedia '19: Proceedings of the 4th International Workshop on Multimedia for Personal Health & Health Care*, Nice, Francia, 2019, pp. 9-15, doi: <https://doi.org/10.1145/3347444.3356238>
- [16] J. G. Rodríguez-Ruiz, C. E. Galván-Tejada, L. A. Zanella-Calzada, J. M. Celaya-Padilla, et al., “Comparison of night, day and 24 h motor activity data for the classification of depressive episodes,” *Diagnostics*, vol. 10, no. 3, art. no. 162, Mar. 2020, doi: <https://doi.org/10.3390/diagnostics10030162>
- [17] J. G. Rodríguez-Ruiz, C. E. Galván-Tejada, H. Luna-García, H. Gamboa-Rosales, J. M. Celaya-Padilla, J. G. Arceo-Olague, J. I. Galván Tejada, “Classification of Depressive and Schizophrenic Episodes Using Night-Time Motor Activity Signal,” *Healthcare*, vol. 10, no. 7, art. no. 1256, Jul. 2022, doi: <https://doi.org/10.3390/healthcare10071256>
- [18] P. Jakobsen, E. Garcia-Ceja, M. Riegler, L. A. Stabell, T. Nordgreen, J. Torresen, O. B. Fasmer, K. J. Oedegaard, “Applying machine learning in motor activity time series of depressed bipolar and unipolar patients compared to healthy controls,” *PLoS One*, vol. 15, no. 8, art. no. e0231995, Ago. 2020, doi: <https://doi.org/10.1371/journal.pone.0231995>
- [19] H. Kour and M. K. Gupta, “An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM,” *Multimed. Tools Appl.*, vol. 81, no. 17, pp. 23649-23685, 2022, doi: <https://doi.org/10.1007/s11042-022-12648-y>
- [20] S. L. Pacheco-González, L. A. Zanella-Calzada, C. E. Galván-Tejada, N. M. Chávez-Lamas, J. F. Rivera-Gómez, and J. I. Galván-Tejada, “Evaluation of Five Classifiers for Depression Episodes Detection,” *Res. Comput. Sci.*, vol. 148, no. 10, pp. 129-138, 2019, doi: <https://doi.org/10.13053/rcs-148-10-11>
- [21] M. Raihan, A. K. Bairagi, and S. Rahman, “A Machine Learning Based Study to Predict Depression with Monitoring Actigraph Watch Data,” in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, India, 2021, pp. 1-5, doi: <https://doi.org/10.1109/ICCCNT51525.2021.9579614>

- [22] P. M. Singh and P. S. 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. Lecture Notes in Networks and Systems*, Nevada, USA, 2022, pp. 99-107, doi: https://doi.org/10.1007/978-981-16-4016-2_10
- [23] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimed. Tools Appl.*, vol. 80, no. 5, pp. 8091-8126, 2021, doi: <https://doi.org/10.1007/s11042-020-10139-6>
- [24] V. Trevino and F. Falciani, "GALGO: An R package for multivariate variable selection using genetic algorithms," *Bioinformatics*, vol. 22, no. 9, pp. 1154-1156, May 2006, doi: <https://doi.org/10.1093/bioinformatics/btl074>
- [25] T. Jayalakshmi and A. Santhakumaran, "Statistical Normalization and Back Propagation for Classification," *Int. J. Comput. Theory Eng.*, vol. 3, no. 1, pp. 89-93, 2011, doi: <https://doi.org/10.7763/IJCTE.2011.V3.288>
- [26] A. A. AlBeladi and A. H. Muqaibel, "Evaluating compressive sensing algorithms in through-the-wall radar via F1-score," *Int. J. Signal Imaging Syst. Eng.*, vol. 11, no. 3, pp. 164-171, 2018, doi: <https://doi.org/10.1504/IJSISE.2018.093268>
- [27] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. 1, no. 8476, pp. 307-310, Feb. 1986, doi: <https://doi.org/10.1016/S0140-6736%2886%2990837-8>
- [28] N. Banaei, J. Moshfegh, A. Mohseni-Kabir, J. M. Houghton, Y. Sun, and B. Kim, "Machine learning algorithms enhance the specificity of cancer biomarker detection using SERS-based immunoassays in microfluidic chips," *RSC Adv.*, vol. 9, no. 4, pp. 1859-1868, 2019, doi: <https://doi.org/10.1039/C8RA08930B>
- [29] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145-1159, Jul. 1997, doi: [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
- [30] C. H. Espino-Salinas, C. E. Galván-Tejada, H. Luna-García, H. Gamboa-Rosales, J. M. Celaya-Padilla, L. A. Zanella-Calzada, and J. I. Galván-Tejada, "Two-Dimensional Convolutional Neural Network for Depression Episodes Detection in Real Time Using Motor Activity Time Series of Depression Dataset," *Bioengineering*, vol. 9, no. 9, art. no. 458, Sep. 2022, doi: <https://doi.org/10.3390/bioengineering9090458>
- [31] O. Caelen, "A Bayesian interpretation of the confusion matrix," *Ann. Math. Artif. Intell.*, vol. 81, pp. 429-450, Sep. 2017, doi: <https://doi.org/10.1007/s10472-017-9564-8>
- [32] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, art. no. 6, Jan. 2020, doi: <https://doi.org/10.1186/s12864-019-6413-7>
- [33] D. Chicco, M. J. Warrens, and G. Jurman, "The Matthews Correlation Coefficient (MCC) is More Informative Than Cohen's Kappa and Brier Score in Binary Classification Assessment," *IEEE Access*, vol. 9, pp. 78368-78381, 2021, doi: <https://doi.org/10.1109/ACCESS.2021.3084050>
- [34] A. Kumar, S. R. Sangwan, A. Arora, and V. G. Menon, "Depress-DCNF: A deep convolutional neuro-fuzzy model for detection of depression episodes using IoMT," *Appl. Soft Comput.*, vol. 122, art. no. 108863, 2022, doi: <https://doi.org/10.1016/j.asoc.2022.108863>
- [35] R. Ghate, N. Kalnad, R. Walambe, and K. Kotecha, "Transfer Learning for Real-time Deployment of a Screening Tool for Depression Detection Using Actigraphy," in *UKSim-AMSS 25th International Conference on Modelling & Simulation*, Cambridge, United Kingdom, 2023, pp. 1-5, doi: <https://doi.org/10.48550/arXiv.2303.07847>
- [36] M. Zakariah and Y. A. Alotaibi, "Unipolar and Bipolar Depression Detection and Classification Based on Actigraphic Registration of Motor Activity Using Machine Learning and Uniform Manifold Approximation and Projection Methods," *Diagnostics*, vol. 13, no. 14, art. no. 2323, Jul. 2023, doi: <https://doi.org/10.3390/diagnostics13142323>