Research Article

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

Carlos E. Galván-Tejada (ID),[1] **Laura A. Zanella-Calzada** (ID),[1] **Hamurabi Gamboa-Rosales** (ID),[1] **Jorge I. Galván-Tejada,**[1] **Nubia M. Chávez-Lamas,**[2] **Ma. del Carmen Gracia-Cortés,**[2] **Rafael Magallanes-Quintanar,**[1] and **José M. Celaya-Padilla**[1]

[1]*Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Jardín Juarez 147, Centro, 98000 Zacatecas, Zac, Mexico*
[2]*Clínica Comunitaria de Tacoaleche, Unidad Académica de Odontología, Universidad Autónoma de Zacatecas, Jardín Juarez 147, Centro, 98000 Zacatecas, Zac, Mexico*

Correspondence should be addressed to Hamurabi Gamboa-Rosales; hamurabigr@uaz.edu.mx

Depression is a mental disorder which typically includes recurrent sadness and loss of interest in the enjoyment of the positive aspects of life, and in severe cases fatigue, causing inability to perform daily activities, leading to a progressive loss of quality of life. Monitoring depression (unipolar and bipolar patients) stats relays on traditional method reports from patients; however, bias is commonly present, given the patients' interpretation of the experiences. Nevertheless, to overcome this problem, Ecological Momentary Assessment (EMA) reports have been proposed and widely used. These reports includes data of the behaviour, feelings, and other type of activities recorded almost in real time using different types of portable devices, which nowadays include smartphones and other wearables such as smartwatches. In this study is proposed a methodology to detect depressive patients with the motion data generated by patient activity, recorded with a smartband, obtained from the "Depresjon" database. Using this signal as information source, a feature extraction approach of statistical features, in time and spectral evolution of the signal, is done. Subsequently, a clever feature selection with a genetic algorithm approach is done to reduce the amount of information required to give a fast noninvasive diagnostic. Results show that the feature extraction approach can achieve a value of 0.734 of area under the curve (AUC), and after applying feature selection approach, a model comprised by two features from the motion signal can achieve a 0.647 AUC. These results allow us to conclude that using the activity signal from a smartband, it is possible to distinguish between depressive states, providing a preliminary and automated tool to specialists for the diagnosis of depression almost in real time.

1. Introduction

The definition of health issued by the World Health Organization (WHO) says that "health is a state of complete physical, mental and social well-being and not only of disease or infirmity." More than 350 million people in the world suffer from depression, and this can become a serious health problem, especially when it is of long duration and moderate to severe intensity, and can cause great suffering and disrupt work, school, family, economic, and emotional activities, among others. In the worst case, it can lead to suicide, which is the cause of approximately 1 million deaths annually [1]. In Latin America, there is a high rate of mental health problems in the infant and youth population; about

20% of this population has disorders that require interventions of health services, but this number is underestimated due to the tendency of adolescents to hide and disguise their own problems to adults and their lack of confidence to access therapeutic structures [2]. Depression is a mental disorder characterized fundamentally by depressive mood, loss of interest, and enjoyment of the positive aspects of life and fatigue, which impoverish the quality of life and generate difficulties in the family, work, and social environment of those who suffer it.

Depression can manifest itself regardless of age, gender, socioeconomic status, and academic program and can present with primary symptoms that do not encompass mood changes and even change cognitive function, so it is not difficult for any individual to become depressed [3]. There is no doubt that studying the sociodemographic factors of age, gender, socioeconomic stratum, and family in adolescent students is relevant, due to the relationship that may exist between these and the manifestation of depression. This is why, in the global context, we find a series of studies that report high rates of depression in this population [4].

When referring to depression, we include those mood disorders with depressive symptoms, which include unipolar major depressive disorder, dysthymia, and mood disorders due to medical illness with depressive symptoms, among others. Despite the variety of alterations, by default, when speaking of depression, reference is made to the major unipolar depressive disorder. This disorder is considered the main cause of years of life lost due to disability (AVPD) according to the Global Burden of Disease Study (GBDS), conducted by the WHO [1].

One in four people suffer from one or more mood or behavioural disorders throughout their lives, and between 50% and 70% of those with a major or minor depressive episode have a predisposition to develop a new one in the next 5 years, which generates great impact on the global economy by the cost of psychotherapeutic and pharmacological management and by the avoidant personality disorder (AVPD), which for 2010 was 2.5 trillion dollars associated with major depressive disorder, with a projected increase to 6 trillion by 2030. The prognosis would improve with timely and adequate psychological, social, and pharmacological management.

The efficacy of the current treatment of depression needs to be increased since its prevalence is very high worldwide and only half of patients experience complete remission with first-line treatments (pharmacotherapy and psychotherapy) within two years [1, 5].

Within the framework of preventive actions or psychological rehabilitation, the instrumentation to obtain a standardized measure of depression is to act from a scientific framework. Therefore, it is necessary to have evaluation instruments that demonstrate the best evidence of validity and reliability to support inferences about the early detection of some symptoms of depression. One of these measures is the State-Rask Depression Inventory [6], which has advantages over some known instruments, such as the Self-Administered Depression Scale [7] and the Beck-II Depression Inventory [8]. Another of the most representative

characteristics is that it allows us to differentiate between the persons' current experience (state) and their habitual way of behaving (trait) with regard to the affective component of depression [6]. This is of great clinical value in differentiating experience in two time frames and specifically targeting one of the constituent areas of depression: affective disorders [6]. Classical methods to achieve correct monitoring of depression states (unipolar and bipolar) in patients are done by reports from patients' recall. Nevertheless, this type of monitoring is prone to bias commonly, in addition to changes in the behaviour and understanding the real world as is reported by Shiffman et al. [9]. Another type of method to overcome these problems is Ecological Momentary Assessment (EMA). This type of report includes behaviour, feelings, and other types of activities as close as possible to the moment of the experience in real-life situations [9]. One improvement to these types of reports is done by the increase of wearable devices (for instance, smartwatches and smartglasses) and smartphones, which includes different types of sensors (motion sensors, gyroscopes, and accelerometers), allowing EMA measurements to be done almost in real time, helping to monitor mental illness and give a close view to provide treatments, interventions, and increase the coverage of mental health services in the population without the need of new specific proposal devices or modifying adding sensors to the environment where the patients are living.

One device currently used to achieve mental health illness supervision is the smartphones and similar ones like smartwatches. One proposal is presented by Gravenhorst et al. [10]; they discussed how mobile phones can increase the effectiveness of mental disorders treatment by two main approaches: in one hand, the implementation of human-computer interfaces for therapy and secondly, collection of important data from patients' daily lives to be recorded by the current state and the development of their mental problems; they also discussed the advantages and drawbacks of the most promising technologies for detecting disorders like depression or bipolar disorder.

Other interesting approach is given by Firth et al.'s [11] study; they demonstrate that psychological interventions with a smartphone as a clinical tool can reduce anxiety in schizophrenia patients. Torous et al. [12], in their study, provide data on psychiatric patients and the relationship with the use and interest of utilizing mobile applications to monitor their mental health conditions. In this study, results presented show that 50% of patients from all age groups are interested and will use mobile applications to monitor their mental health condition to control their illness. Bayindir et al. [13] present a systematic review of different works that focus in the use of mobile phone sensors to detect human behavior characteristics, describing activity detection at different abstraction levels of activity and characterizing health-related activities, like physical exercise and sleeping.

Additionally to the use of applications, these devices include several embedded sensors that have been used to acquire contextual information and several niches [14, 15], including activity recognition [16] and particularly activity that helps to find mental disorders [17]. For instance,

Gruenerbl et al. [18] demonstrate that inertial sensors and GPS traces can be used as a measurement device in psychiatric diagnosis, through a methodology based on a feature extraction of physical motion levels and travel patterns, and a classification analysis using a naïve Bayes technique. Reece et al. [19] identify depressed subjects using uploaded photos of Instagram based on a random forest technique. Grünerbl et al. [20] propose the classification of depressed and maniac states in bipolar patients based on smartphone data. Maxhuni et al. [21] classify bipolar patients through audio, motor activity, and questionnaires. Berle et al. [22] propose an approach using motor activity information to reveal schizophrenia and depression patterns. Koo et al. [23] present a research of the utility of the combination of biomarkers related to different approaches, such as motor activity based on actigraphy measurements, showing that the discrimination of patients based on these biomarkers improves in the identification of depressed subjects. Averill et al. [24] examine the psychomotor change in depressive episodes based on the activity levels measured by actigraphy in order to know the response of the depression treatment, concluding that the early change in simple activity and psychomotor speed allows one to measure the treatment response in depressed patients. Garcia-Ceja et al. [15] performed an analysis of data collected through actigraphy applying machine learning to classify depressed patients, finding that the data contain information that allows determination of the depression status of a subject. Huguet et al. [25] present a review to identify self-help apps that are available for depressed people. The apps that offer cognitive behavioural therapy (CBT) or behavioural activation (BA) are evaluated since the low level of adherence to the core ingredients of the CBT and BA models causes that the utility of these apps is questionable. It was possible to conclude that the application of superior scientific, technological, and legal knowledge is required to improve the credibility of the apps for people with depression.

On the other hand, Mohr et al. [26] provide a review of sensing research related to mental health, where a layered and hierarchical model is provided for the translation of raw sensor data into markers of behaviors and states related to mental health. Finally, in the work of Guntuku et al. [27] is reviewed the study of predicting mental illness using social media, including screening surveys, public sharing on Twitter, and the membership in an online forum, concluding that automated detection methods are useful to identify depressed or individuals at risk through the monitoring of passive activity in social media.

The aim of this work is to study the signal generated by a smartbands accelerometer to detect depressive states through the activity of patients and to propose a feature extraction (using the temporal and spectral evolution of the signal), as well as a clever feature selection based on a genetic algorithm approach to minimize the data required to identify these depressive states allowing an almost real-time noninvasive diagnosis. In this type of disease, early symptomatic detection can significantly increase the development of an effective treatment and contribute to the prevention of this type of psychopathology.

One of the main advantages proposed in this work is the simplicity in the data acquisition since the device used is noninvasive, has a small size, and does not hinder daily activities, which is a benefit compared to other devices that can be interposed in day-to-day tasks, in addition to using multiple sensors for the acquisition of different types of data, which is not necessary in this approach because the same purpose is achieved with a single source of acquisition, obtaining the information required for the extraction of features that allowed the classification of depressive patients.

This paper is organized as follows: in Section 2 is detailed described the materials used for the development of this research, as well as the set of the stages of the methodology proposed. Then, Section 3 presents the results obtained. Section 4 is referred to the discussion developed based on the results previously, and finally, Section 5 shows the conclusions of this work.

2. Materials and Methods

The methodology proposed in this research consists in five main stages, shown in Figure 1. The data used for the development of this work is acquired from the "Depresjon" dataset (A). These data are initially subjected to a data preprocessing (B) step, in order to select the samples and subjects for further analysis, to normalize the data, and to eliminate missing values. Then, the feature extraction (C) is performed, obtaining temporal and frequency statistical features, which are submitted to a feature selection (D) step, using the genetic algorithm (GA) "Galgo." Finally, the set of selected features is evaluated, measuring its fitness in the classification of controls and cases, based on a random forest (RF) technique and a statistical analysis (E).

2.1. Data Description. The Depresjon dataset is a collection of data that contains the motor activity of patients monitored with an actigraph watch held on the right wrist. The actigraph watch is called "Actiwatch" (model AW4), developed by Cambridge Neurotechnology Ltd, England. The Actiwatch measures activity levels, and the sampling frequency is 32 Hz, recording movements over 0.05 g. Movements equal a corresponding voltage, which is stored as an activity count in the memory of the Actiwatch, and the number of counts is proportional to the intensity of the movement. The activity counts were recorded in intervals of one minute.

The database contains the data for the controls (absence of depression, 32 subjects) and for the cases (presence of depression, 23 subjects). The features collected for each subject were divided in two categories, actigraph data recorded over time and Montgomery Åsberg Depression Rating Scale (MADRS) scores. The data collected over time include the features "timestamp" (one minute intervals), "date" (date of measurement), and "activity" (activity measurement from the actigraph watch). In addition, MADRS scores include the features "number" (patient identifier), "day" (number of days of measurements), "gender" (1: female/2: male), "age" (age in age groups),
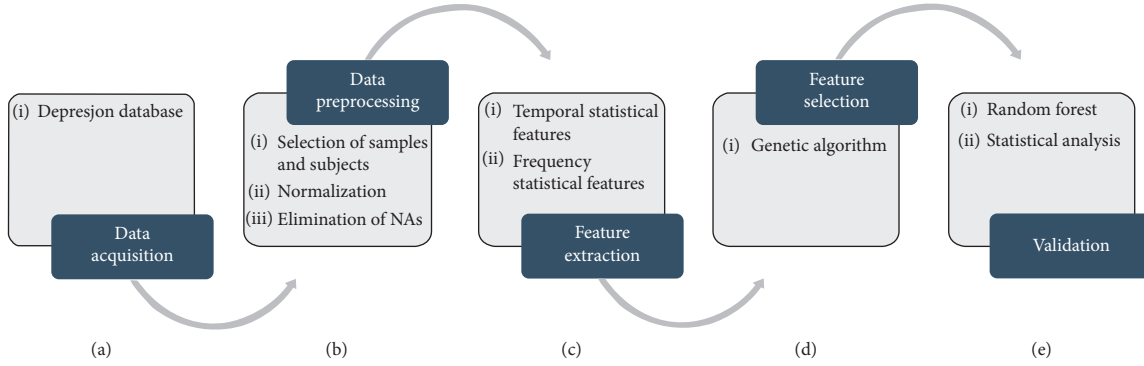
FIGURE 1: Flowchart of the methodology followed.

"afftype" (1: bipolar II, 2: unipolar depressive, and 3: bipolar I), "melanch" (1: melancholia; 2: no melancholia), "inpatient" (1: inpatient; 2: outpatient), "edu" (education grouped in years), "marriage" (1: married or cohabiting; 2: single), "work" (1: working or studying; 2: unemployed/sick leave/pension), "madrs1" (MADRS score when measurement started), and "madrs2" (MADRS when measurement stopped) [28].

For this work only the features over time were used.

2.2. Data Preprocessing. The data preprocessing consists in three main steps, the selection of samples and subjects, the normalization of the data, and the elimination of incomplete cases presented as NA (not available).

The number of samples collected is not consistent, differing in the number of minutes recorded for each subject, so a selection of subjects and samples was made in order to present a balanced amount of data referring to controls and cases. The selection of the samples is carried out keeping only the first value of the 60 acquired data in the minutes equivalent to one hour, counting now the activity in intervals of one hour, whereas the selection of subjects depended on the amount of data resulting from the selection of samples, selecting the first four controls present in the dataset and the first five cases. This number of subjects allows balance of the number of samples for cases and controls.

Then, in the normalization, the data are adjusted in order to obtain a normal distribution, presenting a mean = 1 and a standard deviation = 0, and it is calculated with Equation (1), where z_i represents the normalized value, x_i represents the sample, μ is the mean of the total data, and σ is the standard deviation of the total data:

$$z_i = \frac{x_i - \mu}{\sigma}. \tag{1}$$

Finally, the missing data are eliminated, removing all the rows with presence of NAs, in order to avoid problems in the subsequent analysis.

2.3. Feature Extraction. The feature extraction is performed using two types of data, temporal and frequency data. The temporal data are directly used from the time-dependent

Depresjon data, which were collected from the activity of the subjects through the actigraph watch.

On the other hand, the frequency data are obtained through the calculation of the Fourier transform of the time-dependent Depresjon data.

Then, for each type of data, 14 statistical parameters are extracted, presented in Table 1, obtaining a total of 38 features.

The names of the features correspondent to the temporal data are "tKurtosis," "tSesgo," "tQ01," "tQ05," "tQ25," "tQ75," "tQ95," "tQ99," "tMedia," "tSD," "tVarianza," "tTrimMedia," "tCV," and "tICV," while the names of the features correspondent to the frequency data are "fKurtosis," "fSesgo," "fQ01," "fQ05," "fQ25," "fQ75," "fQ95," "fQ99," "fMedia," "fSD," "fVarianza," "fTrimMedia," "fCV," and "fICV."

2.4. Feature Selection. In this stage, the 38 features extracted are subjected to a feature selection based on a GA approach. GAs are a stochastic strategy that has been widely used in the analysis of data and they consist in a sequence of stages that starts with a random set of models and develops good local solutions reproducing the natural selection process using measures such as (1) higher rate of replication of the more accurate feature subsets, (2) mutation to generate different chromosomes, and (3) crossover to improve the combinations of the chromosomes.

A validation measure is calculated in combination during the selection process, testing the sets of chromosomes, ensuring that the multivariate feature selection is suitable. The aim of the GA is to minimize the score calculated by the fitness function, converging then into a solution, being therefore possible to select the most significant predictive subset of n features [29].

For this work, the genetic algorithm "Galgo" is used. Galgo is a package implemented under the R language, which is oriented to select models with high fitness and to analyze them, as well as for the reconstruction and characterization of representative summary models.

The procedure of Galgo begins with a random population of feature or gene subsets or chromosomes of a defined size (n), which are assessed through a fitness function for their ability to predict or classify the desirable

TABLE 1: Statistical features extracted from the temporal and frequency data.

Feature	Description
Mean	$\mu = 1/n\sum_{i=1}^{n} x_i$
Standard deviation	$sd = \sqrt{\sum_{i=1}^{N}(x_i - \mu)^2/(N-1)}$
Variance	$sd^2 = 1/n\sum_{i=1}^{n}(x_i - \mu)^2$
Trimmed mean	Mean with outliers trimmed
Coefficient of variation	$CV = sd/\mu$
Inverse coefficient of variation	$ICV = \mu/sd$
Kurtosis	$K = \mu/\sigma$
Skewness*	$S = (\mu - v)/\sigma$
Quantiles*: 1, 5, 25, 75, 95, 99%	$Q[i](p) = (1 - \gamma x[j] + \gamma x[j+1])$

*v represents the median value; $1 \le i \le 9$; $(j-m)/n \le p < (j-m+1)/n$; $x[j]$ represents the jth order statistic; n represents the sample size; γ is in function of j and g, where $j = \text{floor}(np + m)$ and $g = np + m - j$; and m represents a constant determined by the sample quantile type.

outcome or the dependent variable, obtaining a certain value of accuracy. The classification methods that can be used in the internal procedure of Galgo are k-nearest-neighbors, discriminant functions, nearest centroid, support vector machine, neural networks, and random forest.

The main idea of the process is to replace the first population with a new one, including variants of chromosomes that achieved a higher classification accuracy, and to repeat this procedure until a desired accuracy is reached. The progressive changes of the chromosomes are performed through a series of operators that simulate the process of natural selection, selection, mutation, and crossover.

The proportion of the solution space increases with the evolution of independent chromosome populations in partially isolated environments, known as niches, and chromosomes can migrate from one niche to another, in order to ensure the recombination of good solutions. A set of niches is called world [30].

This process is carried out in four main steps:

(i) First, the analysis is configured, specifying the input and the outcome features, as well as a series of parameters that will guide the behaviour of the process, such as the classification model, the desired accuracy, and the error estimation scheme, among others. The classification model can be selected from the implemented or can be defined by a function of the user; while the error estimation can be defined in two levels, with a training/test validation strategy using variant random splits, and in the internal training process using a k-fold cross-validation, random splits, or re-substitution error.

(ii) Then, the search of relevant multivariate models begins with a random population of chromosomes in each cycle of the procedure. The number of chromosomes developed needs to be large enough to ensure that the greatest amount of solutions was found and to achieve this, two approaches are designed to provide information of the chromosome composition, the level of convergence of the

solutions, and the evolution of the fitness values, diagnosing the stability of the populations.

(iii) A refinement and analysis of the population of the selected chromosomes is carried out, since not all the genes included in the best chromosome may be contributing in a significant way to the fitness value. Therefore, a backward selection strategy is implemented to obtain a model contained by genes that significantly contribute to the accuracy of the result.

(iv) Finally, the development of a significant statistical model is obtained from the population of the selected chromosomes. For this step, a forward selection strategy is included, and its operation is based on a stepwise inclusion adding the most frequent genes of the chromosome population.

The configuration of the analysis for this study is composed by 200 generations, five genes per chromosome, a desired accuracy of 0.99, and "nearest centroid" as classification model, and an error estimation scheme was used a cross-validation approach.

2.5. Classification Analysis. The classification analysis was carried out through a RF method, looking for the classification of subjects in two different states, depressed (labeled as "1") and not depressed (labeled as "0").

RF is a machine learning technique that presents two main approaches, classification and regression, and its performance is based on decision trees. In the classification option, RF provides estimators of a Bayes classifier, $f : R \mapsto y$, minimizing the error classification $P(Y \neq f(X))$.

Roughly, an ensemble of trees grows, constructed with random vectors that generate each of the trees, deciding the class to which the data correspond by voting, where the majority of the class votes determine the RF prediction. This process causes that the generalization error merges to a limiting value, thus improving the classification accuracy of the system [31].

Specifically, the trees are created using a subset of bootstrap samples with replacement, $L^1, \ldots, L^{n_{\text{tree}}}$ (of a training set L), known as a bagging approach, which means that one same sample can be selected several times for the classification analysis while the others samples may not be selected.

Every decision tree is independently constructed without any pruning, and each node is divided through a splitting rule using a specific number of features, m_{try}, randomly selected.

The splitting rule is added to the estimators calculated from the trees, represented as $\widehat{f}_1, \ldots, \widehat{f}_{n_{\text{tree}}}$. A response value is subsequently obtained from the new point, which consists in the construction of the following equation:

$$\widehat{f}(x) = \arg\max 1 \le c \le C \sum_{k=1}^{n_{\text{tree}}} 1_{\widehat{f}_k(x)=c}. \tag{2}$$

The forest is growing up to a defined number of trees, n_{tree}, and by this step, the algorithm creates tree that present two

main characteristics, high variance and low bias. The final classification decision is calculated through the arithmetic mean of the class assignment probabilities of the total number of trees. Then, an evaluation step is performed using a new set of unlabelled data input with the decision trees developed in the ensemble, giving each tree a vote for a class. The class that collects the greatest number of votes is the one selected.

Around two thirds of the total samples are usually used for the training of the trees, and they are referred as in − bag samples; then, with the remaining one third samples, referred as out − of − the − bag samples, an internal cross-validation is realized for the estimation of the model performance [32].

The estimation of this error is known as out-of-bag (OOB) error. This value measures the misclassification rate for the classification of the OOB samples. This means that a feature, X^j, is important if when breaking the relationship between X^j and Y, the error of the prediction increases, and the error of the prediction in each tree, $\widehat{f}$, is evaluated with the OOB sample using

$$\widehat{R}(\widehat{f}, \widehat{L}) = \frac{1}{|L|} \sum_{i:\, (X_i, Y_i) \varepsilon \overline{L}} 1 \, \widehat{f}(X_i) \neq Y_i. \qquad (3)$$

It is important to note that according to literature, the classification accuracy is less sensitive to n_{tree} than to m_{try}; therefore, since RF is a computationally efficient classifier that does not present problems of overfitting, n_{tree} can be a number as large as possible. On the other hand, the m_{try} parameter is usually defined by the square root of the total number of input features [33].

For the development of this study, the number of trees selected is $n_{\text{tree}} = 2000$, and the number of features at each split, m_{try}, is calculated according to the number of features as $\sqrt{p}$, with p being the number of features.

2.6. Validation. The validation stage is based on three parameters, the AUC as a single value quantity of the ROC curve, specificity, and sensitivity.

The ROC curve has been a widely used tool for the evaluation of binary classification models since it presents a series of characteristics that allow the correct interpretation of the results, such as the intuitive visual interpretation of the curve, easy comparison among multiple models, and the AUC value [34].

The calculation of the classifier's performance through the ROC curve provides a suitable operating point, called as decision threshold, for the parameterization of the classification model.

A classification problem presents two possible outputs, "correct" and "incorrect," for each class of the model. An orderly way to present this information is through a confusion matrix, a table that shows the differences between the real and the predicted classes. The values contained in a confusion matrix are the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN); besides, the value of the row totals with the truly negatives (CN) and truly positives (CP) examples, and the value of the column

totals with the predicted negative (RN) and the predicted positive (RP) examples [35].

The sensitivity is a parameter referred to as the ability to correctly identify those data with a condition, and it is calculated with the following equation:

$$\text{sensitivity}\,(1 - \beta) = \frac{\text{Tp}}{\text{Cp}}. \qquad (4)$$

On the other hand, the specificity is a parameter referred to as the ability to correctly identify those data without a condition, and it is calculated with the following equation:

$$\text{specificity}\,(1 - \alpha) = \frac{\text{Tn}}{\text{Cn}}. \qquad (5)$$

Finally, the plotted values of the sensitivity and the specificity in conjunction represent the decision threshold of the ROC curve. The AUC value of the curve can be calculated through trapezoidal integration, as shown in the following equation:

$$\text{AUC} = \sum_i \left(1 - \beta_i \cdot \Delta\alpha\right) + \frac{1}{2}\left[\Delta\,(1 - \beta) \cdot \Delta\alpha\right], \qquad (6)$$

where $\Delta\,(1 - \beta) = (1 - \beta_i) - (1 + \beta_{i-1})$ and $\Delta\alpha = \alpha_i + \alpha_{i-1}$ [35].

All the analysis is carried out in "R" (version 3.4.4), a free software environment designed for statistical computing and graphics [36]. The libraries required for this analysis are "Galgo" (version 1.2-01) [37], "pROC" (version 1.11.0) [38], "e1071" (version 1.7-0) [39], "randomForest" (version 4.6-14) [40], "caret" (version 6.0-79) [41], and "rminer" (version 1.4.2) [42].

3. Results

The results of this research are presented in this section. Through the first step of this methodology, which was the data acquisition, the number of subjects selected for the subsequent analysis was five for cases and four for controls, in order to balance the number of samples in both datasets.

Then, the feature extraction allowed collection of a series of 38 statistical features, which of the total, 14 belong to the time data and the remaining to the frequency data. It is important to remind that the frequency data were calculated through the Fourier transform of the time data.

For the third stage, a feature selection based on the GA, Galgo, is carried out, obtaining a series of graphs that allow observation of the performance of the data through the development of the different models created in the evolution of the algorithm. Figure 2 presents a graph of the frequency percentage with which each feature appeared within the different models developed, positioning each feature according to its order of appearance, from highest to lowest, where those features in black present the highest frequency and those features in gray present the lowest. According to this graph, the most significant features, according to their appearance frequency, are "tCV," "tQ99," "fCV," and "tVarianza." (Tables 2 and 3).
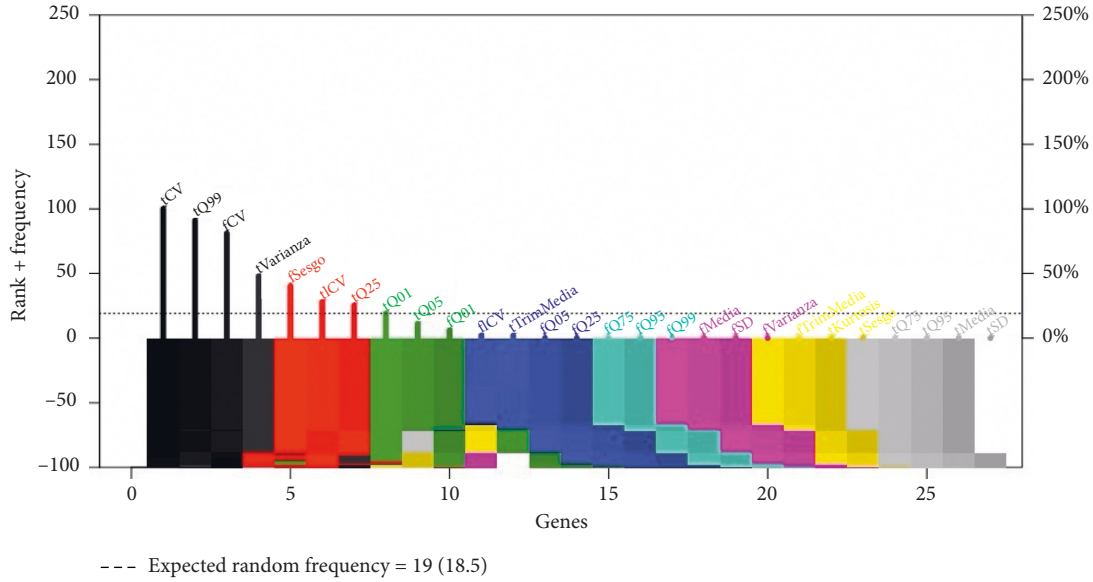
FIGURE 2: Graph of the frequency of appearance of the features within the models developed in the genetic algorithm, positioned from higher to lower rank.

TABLE 2: Confusion matrix of the subjects classification based on the RF approach using the complete set of features.

True	Predicted		
	Control	Case	Error
Control	819	326	0.284
Case	331	962	0.255

TABLE 3: Confusion matrix of the subjects classification based on the RF approach using the best chromosome of the GA.

True	Predicted		
	Control	Case	Error
Control	774	371	0.324
Case	373	920	0.288

Then, in Figure 3 is shown a graph of the fitness performance throughout the evolution of the 200 generations of the GA, where it is possible to observe that the average fitness reaches a stable behaviour, with a value of around 0.63.

Figure 4 presents a heat map of the best chromosome presented by the GA, contained by a model of five chromosomes, "tCV," "tQ25," "tQ99," "tICV," and "tCV."

Then, the best chromosome is subjected to a forward selection step, where for each feature added to the model, its average fitness was calculated, as shown in Figure 5. According to this graph, the model reaches its best average fitness, as well as stability, with three features, "tCV," "tQ99," and "fCV."

Finally, in Figure 6 is present a heat map of the final model obtained through backward elimination step, contained by two time features, "tCV" and "tQ99."

In the classification analysis, a RF approach is used, measuring the OOB error in order to know the accuracy classification reached through the model selected in the previous step. In Table 4 is present the confusion matrix obtained through the classification of subjects using the total set of features and the respective error values for each of the classes. The OOB error obtained was of 26.95%. In Table 4 is present the confusion matrix obtained through the classification of subjects using the best chromosome and the respective error values for each of the classes. The OOB error obtained was 30.52%. Finally, In Table 4 is present the confusion matrix obtained through the classification of subjects using the final model and the respective error values for each of the classes. The OOB error obtained was 35.97%.

For the last stage of this work, a validation step was performed, calculating the ROC curves of the models, shown in Figure 7, where in Figure 7(a) is present the ROC curve of the model contained by the total features and its respective AUC value, which obtained a sensitivity of 0.751 and a specificity of 0.717. Then, in Figure 7(b) is present the ROC curve of the model contained by the best chromosome and its respective AUC value, which obtained a sensitivity of 0.699 and a specificity of 0.694. Finally, in Figure 7(c) is present the ROC curve of the model contained by the final model and its respective AUC value, which obtained a sensitivity of 0.684 and a specificity of 0.611.

4. Discussion

In this section, the results obtained are discussed. From the total 38 statistical features extracted, a feature selection is performed based in Galgo. Initially, Galgo developed the graph shown in Figure 2, which provides the information of the frequency with which the features are part of the different chromosomes developed, ordered by rank from highest to lowest.

According to Figure 2, the four most significant features or the features that presented the highest frequency were
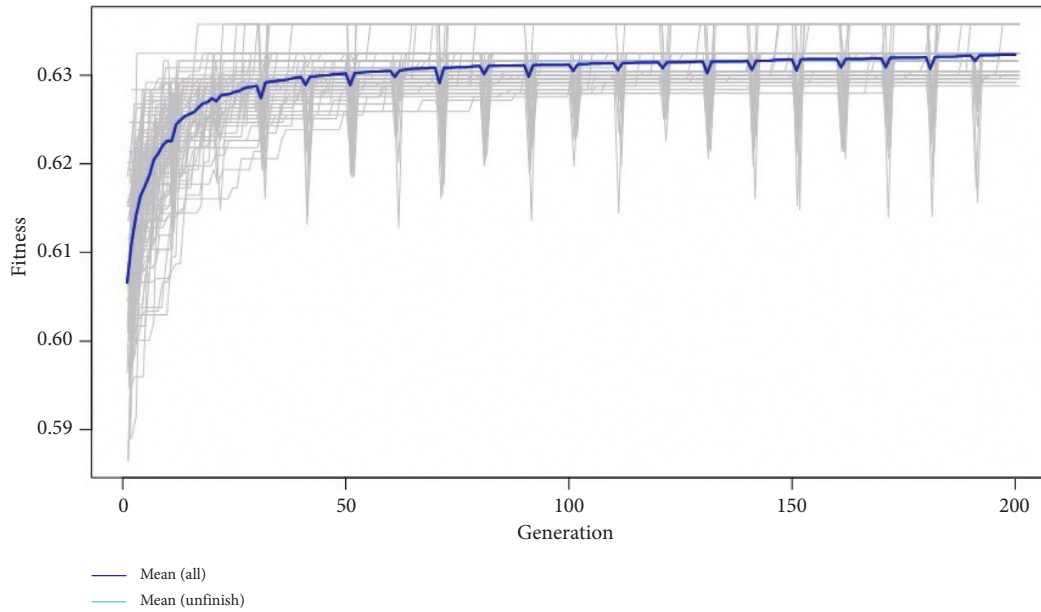
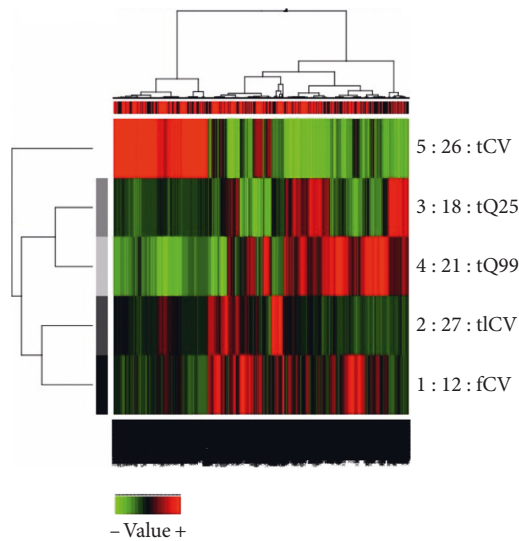FIGURE 3: Graph of fitness achieved in each generation of the genetic algorithm.

FIGURE 4: Heat map of the best chromosome calculated by the genetic algorithm.

those presented in black, of which three correspond to the temporal features and one to the frequency features, which means that temporal data are presenting more significant information than frequency data for the classification of subjects.

Then, in Figure 3 is shown a graph of the average fitness behaviour through the different generations of the GA, being possible to observe that the greatest change occurs at the beginning of the graph, within the first 50 generations, where the GA is in the process of finding the best combination of genes to obtain a chromosome suitable for classification. Subsequently, a relatively stable value is reached around generation 80, obtaining an average fitness value of around 0.63 in the last generation.

At the end of the 200 generations, the best chromosome obtained is presented in Figure 4, contained by the five features presented in the heat map, of which four correspond to the time features and the remaining corresponds to the frequency features. The first feature of the best chromosome corresponds to the frequency feature, "fCV", referred to the coefficient of variation (CV), which is related to the standard deviation and the mean value, where the higher the value of the standard deviation compared to the mean, the higher will be the CV and vice versa. This feature may imply that the frequency data could be presenting significant variations among its values between cases and controls, being able to distinguish between both classes.

The second feature is "tICV", referred to the inverse coefficient of variation (ICV), which may imply a similar meaning than the feature "fCV", where the time data can be presenting significant information in the distribution of the data that allows one to distinguish between depressed and nondepressed subjects.

Then, there are present the features "tQ25" and "tQ99," which represent the 25 and the 99 quantiles, respectively. Quantiles are points of regular intervals of the distribution function of a random variable. Therefore, these two features may imply that, in these data intervals, the most significant information or the greatest differences between both classes are presented because taking into account that data on the amount of activity carried out as a function of time are being analyzed and that quantile data are arranged in ascending order, it is possible that by comparing the variations in the amount of activity correspondingly, that is, the greater activity of depressed patients against the higher activity of nondepressed patients, a difference is presented meaningful.

The fifth feature is "tCV," which represents the same as "fCV" but with time data. This feature may imply that the information of the physical activity of patients is presenting differences in the standard deviation and the mean between
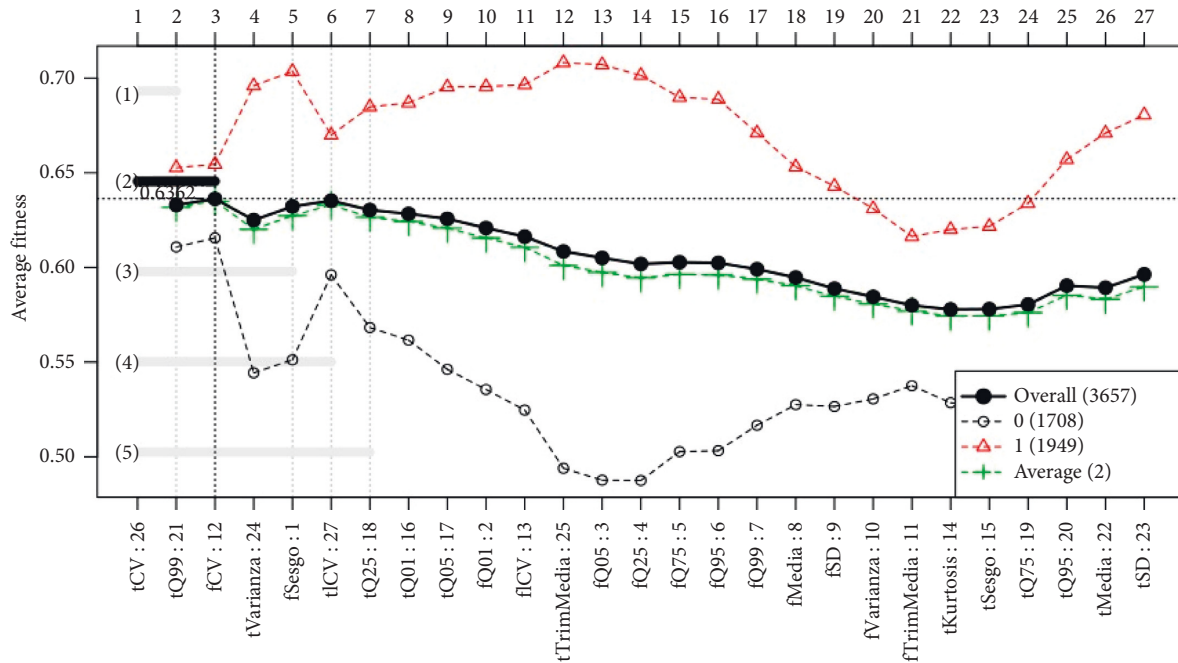
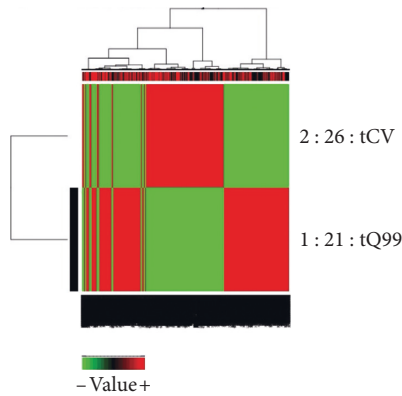FIGURE 5: Graph of average fitness obtained through a forward selection.

FIGURE 6: Heat map of the final chromosome obtained through a robust gene back elimination.

TABLE 4: Confusion matrix of the subjects classification based on the RF approach using the final model selected.

True	Predicted		
	Control	Case	Error
Control	690	455	0.397
Case	422	871	0.326

the two possible classes that provide support for the correct classification.

On the other hand, in Figure 5 is presented a graph of the behaviour of the average fitness when the features are subjected to a forward selection step. For each feature that is added to the model, the average fitness, the fitness of each class, and the total fitness are measured, in order to know the behaviour that the model has when including the information of the features and in this way to select the

adequate number of features, avoiding having nonsignificant information. According to the graph, the model achieves stability from the third feature, reaching an average fitness of 0.636.

The last step of the feature selection consisted of a robust gene back elimination (RGBE) step, in order to remove redundant information, obtaining a final model contained by two features, presented in the heat map of Figure 6. This model is dependent on the mean, standard deviation, and the 99 quantile values, which according to the previous steps, these measures provide data that allow the classification of the classes.

Then, a RF approach is used for the classification analysis, comparing the OOB error obtained through three different models: one model is contained by the total set of initial features, and it obtained an OOB error of 26.95%; the second model is contained by the features of the best chromosome obtained with GA, and it obtained an OOB error of 30.52%; and the third model is contained by the final model obtained through the RGBE step, obtaining an OOB error of 35.97%.

The OOB error values allow one to know the percentage of data that was misclassified during the construction of the decision trees that form the random forest, and as is possible to observe, this value increases when the model contains less features, presenting the lowest OOB error in the model contained by the total set of features; nevertheless, even when the final model presents an error 10% higher than the first model, the percentage that is correctly classified remains statistically significant. Besides, it is important to remark that the number of features contained in the final model is significantly smaller than that contained in the first model; therefore, the information required for the classification is much smaller, thus reducing the computational cost for the analysis of the data.
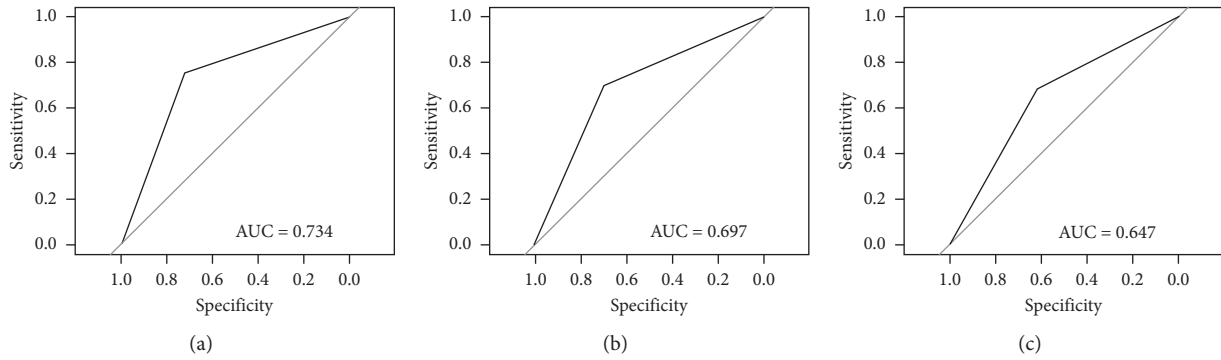
FIGURE 7: ROC curves obtained using the (a) complete features, (b) best chromosome, and (c) final chromosome.

Also, from the validation that RF performs internally, a series of confusion matrices were obtained in order to explain the OOB errors obtained. In Table 4 is present the true positives ($= 962$) and true negatives ($= 819$), as well as the error value for controls ($= 0.284$) and for cases ($= 0.255$), for the model contained by the total set of features, where it is shown that even when the error is higher for controls, both classes present similar values in the classification error. Then, Table 4 shows the true positives ($= 920$), true negatives ($= 774$), error value for controls ($= 0.324$), and error value for cases ($= 0.288$), obtained through the classification using the best chromosome, where it is evident that the error values increase for both classes; nevertheless, the classification continues presenting a statistically significant aptitude even though the quantity of features contained in the model was reduced by around 86.84%. In Table 4 is present the true positives ($= 690$) and true negatives ($= 871$), and the error value for controls ($= 0.397$) and for cases ($= 0.326$), showing that, evidently, by reducing the number of features of the best chromosome, the error is increased again, especially for controls, which may represent that the activity registered by controls could be confused with the activity of cases in specific moments of time, for example, in the hours of sleep. In addition, could also be confused the time of greatest activity for both classes when in the case of controls, physical activity is not very energetic. However, this problem of confusion in the classification can be solved by increasing the number of samples in both classes, taking into account that it is important that the training of the algorithms have a balanced amount of data.

In the validation stage, the specificity and sensitivity allowed support of the previous results, obtaining higher values in the evaluation of sensitivity than in specificity, although it is important to note that the results of the validation presented significant values for the three models evaluated.

Then, the ROC curve is calculated for each of the models, as shown in Figure 7, where Figure 7(a) represents the curve obtained using the total set of features, Figure 7(b) represents the curve obtained using the best chromosome, and Figure 7(c) represents the curve obtained using the final model, obtaining AUC values of $0.734 > 0.697 > 0.647$, respectively. The AUC value is reduced by decreasing the number of features of the models; however, the difference between the AUC of the model that contains 100% of the features and the final model, which only contains 5.26% of the features, is not representative taking into account that the AUC remains statistically significant in the final model. Therefore, the ability of the final model to classify cases from controls remains significant despite the limited amount of information used, thus benefiting the computational cost necessary to carry out the classification.

Finally, in Table 5, a comparison between different techniques based on the same approach is shown, collecting data through actigraphy in order to identify depressed patients, where according to the results is possible to conclude that all works present statistical significant results; however, the complexity of the methodologies and the quantity of characteristics related to different information used on each research, as well as the information sources, are greater than those proposed in this work, since it was only necessary the extraction of a reduced set of statistical features from a database collected by a single sensor from a small set of patients, presenting as one of the main contributions the simplicity of the experimentation made for the classification of subjects with presence of depression obtaining statistically significant results, in addition to presenting a lower computational cost than that presented in the mentioned works due to the small amount of data.

5. Conclusions

In this research is proposed a methodology composed by a series of steps which mainly includes a feature selection, a classification analysis, and a validation, in order to find the relationship between a series of statistical features, based on time and frequency continuous values acquired in a specific time and the possible condition of depression.

It is important to remark that the number of subjects allows one to obtain significant results; nevertheless, this number of samples can be increased in order to mainly improve the result of the true negatives, which presents greater error than true positives. On the other hand, the extracted statistical features show that the information they contain provides a description of the main characteristics of a patient's full-day activity that allows differentiation between depressed and nondepressed subjects.

TABLE 5: Comparison with related works on identification of depression.

Work	Description
Averill et al. [24]	Examining the response of the depression treatment based on the simple activity and psychomotor speed measured through actigraphy identifying depressed episodes.
Garcia-Ceja et al. [15]	Analyzing data collected through actigraphy comparing different machine learning techniques to classify depressed subjects.
Gershon et al. [43]	Identifying activity patterns from locomotor activity collected by actigraphy extracting a series of principal components, discriminating depressive days from other states.
Koo et al. [23]	Identification of depressed patients through the combination of biomarkers related to executive dysfunctions, motor activity, and neurophysiological patterns activity, among others.

The feature selection through the GA provides a best chromosome which is subsequently reduced to a model contained by two features. These two features are statistical descriptors of temporary data that according to the validation step, despite presenting a greater error in the differentiation of cases and controls than if the whole set of features is used, the results remain statistically significant, thus allowing having a contained model with a reduced amount of features that automatically classifies depressed subjects of nondepressed subjects with significant fitness.

In addition, it is worth noting that one of the greatest advantages of the model being significantly reduced is that it is also reduced in its computational cost, making it easier to access it, since it does not require specialized software or hardware for its implementation.

Besides, one of the main benefits demonstrated in this work is the values with high precision obtained through a simple methodology using a single source of data, which in comparison with other works, where it is necessary to use more than one source for the data acquisition and a series of different techniques for the classification analysis, this approach provides simplicity and statistically significant results for less processing steps and computational cost.

Then, it is possible to conclude that the methodology implemented in this study allows one to know that evidently, there is an association between the recorded daily activity of a patient and the condition of his depressive state. Besides, the results obtained are sustained according to what is reported in the literature, where among the symptoms presented by patients with depression is the slowness of movement, poor body gesticulation, and the feeling of fatigue, tending to show lower levels of activity than subjects who do not have this condition.

Therefore, through this work is obtained a preliminary tool for the possible support in the diagnosis of the specialists to know the state of health of a patient according to his state of presence or absence of depression, based on the level of activity he has in a full day.

Data Availability

The Depresjon data used to support the findings of this study have been deposited in the "control" and "condition" repositories. This dataset can be accesed in http://datasets.simula.no/depresjon/ and/or can be directly downloaded from http://doi.org/10.5281/zenodo.1219550.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.
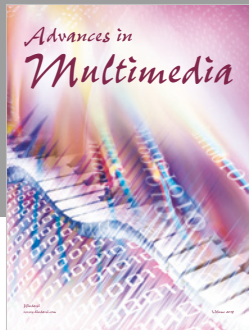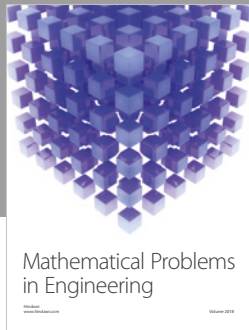
Authors' Contributions

Carlos E. Galván-Tejada and Laura A. Zanella-Calzada contributed equally to this work.

References

[1] J. Fisher, M. Cabral de Mello, V. Patel et al., "Prevalence and determinants of common perinatal mental disorders in women in low- and lower-middle-income countries: a systematic review," *Bulletin of the World Health Organization*, vol. 90, no. 2, pp. 139–149H, 2012.

[2] R. Leyva-Jiménez, A. M. Hernández-Juárez, G. Nava-Jiménez, and V. López-Gaona, "Depresión en adolescentes y funcionamiento familiar," *Revista Médica del Instituto Mexicano del Seguro Social*, vol. 45, 2007.

[3] S. Goldman, "Developmental epidemiology of depressive disorders," *Child and Adolescent Psychiatric Clinics of North America*, vol. 21, no. 2, pp. 217–235, 2012.

[4] L. Mendoza, L. d. P. S. Peinado, L. A. D. Martínez, and A. Campo-Arias, "Prevalencia de sintomatología depresiva en niños y niñas escolares de bucaramanga, Colombia," *Revista Colombiana de Psiquiatr*, vol. 33, pp. 163–171, 2004.

[5] M. E. Buelna Serrano, L. Gutiérrez Herrera, and S. Ávila Saldoval, "El desarrollo de la economía de consumo en el contexto del mundo bipolar de mediados del siglo xx. una visión retrospectiva," *Análisis Económico*, vol. 30, 2015.

[6] D. Agudelo, G. Buela-Casal, and C. D. Spielberger, "Ansiedad y depresión: el problema de la diferenciación a través de los síntomas," *Salud Mental*, vol. 30, pp. 33–41, 2007.

[7] W. W. K. Zung, "A self-rating depression scale," *Archives of general psychiatry*, vol. 12, no. 1, pp. 63–70, 1965.

[8] A. T. Beck, R. A. Steer, and G. K. Brown, *Beck Depression Inventory-II*, vol. 78pp. 490–498, San Antonio, TX, USA, 1996.

[9] S. Shiffman, A. A. Stone, and M. R. Hufford, "Ecological momentary assessment," *Annual Review of Clinical Psychology*, vol. 4, no. 1, pp. 1–32, 2008.

[10] F. Gravenhorst, A. Muaremi, J. Bardram et al., "Mobile phones as medical devices in mental disorder treatment: an overview," *Personal and Ubiquitous Computing*, vol. 19, no. 2, pp. 335–353, 2015.

[11] J. Firth, J. Cotter, R. Elliott, P. French, and A. R. Yung, "A systematic review and meta-analysis of exercise interventions in schizophrenia patients," *Psychological Medicine*, vol. 45, no. 7, pp. 1343–1361, 2015.

[12] J. Torous, R. Friedman, and M. Keshavan, "Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions," *JMIR mHealth and uHealth*, vol. 2, no. 3, p. e34, 2014.

[13] L. Bayındır, "A survey of people-centric sensing studies utilizing mobile phone sensors," *Journal of Ambient Intelligence and Smart Environments*, vol. 9, no. 4, pp. 421–448, 2017.

[14] J. P. García-Vázquez, M. D. Rodríguez, Á. G. Andrade, and J. Bravo, "Supporting the strategies to improve elders' medication compliance by providing ambient aids," *Personal and Ubiquitous Computing*, vol. 15, no. 4, pp. 389–397, 2011.

[15] E. Garcia-Ceja, C. E. Galván-Tejada, and R. Brena, "Multiview stacking for activity recognition with sound and accelerometer data," *Information Fusion*, vol. 40, pp. 45–56, 2018.

[16] E. Garcia-Ceja and R. Brena, "Activity recognition using community data to complement small amounts of labeled instances," *Sensors*, vol. 16, no. 6, p. 877, 2016.

[17] E. Garcia-Ceja, V. Osmani, and O. Mayora, "Automatic stress detection in working environments from smartphones' accelerometer data: a first step," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1053–1060, 2016.

[18] A. Gruenerbl, "Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients," in *Proceedings of the 5th Augmented Human International Conference*, Kobe, Japan, March 2014.

[19] A. G. Reece and C. M. Danforth, "Instagram photos reveal predictive markers of depression," *EPJ Data Science*, vol. 6, p. 15, 2017.

[20] A. Grunerbl, A. Muaremi, V. Osmani et al., "Smartphone-based recognition of states and state changes in bipolar disorder patients," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 1, pp. 140–148, 2015.

[21] A. Maxhuni, A. Muñoz-Meléndez, V. Osmani, H. Perez, O. Mayora, and E. F. Morales, "Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients," *Pervasive and Mobile Computing*, vol. 31, pp. 50–66, 2016.

[22] J. O. Berle, E. R. Hauge, K. J. Oedegaard, F. Holsten, and O. B. Fasmer, "Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression," *BMC Research Notes*, vol. 3, no. 1, p. 149, 2010.

[23] P. C. Koo, C. Berger, G. Kronenberg et al., "Combined cognitive, psychomotor and electrophysiological biomarkers in major depressive disorder," *European Archives of Psychiatry and Clinical Neuroscience*, pp. 1–10, 2018, In press.

[24] I. R. Averill, M. Crowe, C. M. Frampton et al., "Clinical response to treatment in inpatients with depression correlates with changes in activity levels and psychomotor speed," *Australian & New Zealand Journal of Psychiatry*, vol. 52, no. 7, pp. 652–659, 2018.

[25] A. Huguet, S. Rao, P. J. McGrath et al., "A systematic review of cognitive behavioral therapy and behavioral activation apps for depression," *PLoS One*, vol. 11, no. 5, Article ID e0154248, 2016.

[26] D. C. Mohr, M. Zhang, and S. M. Schueller, "Personal sensing: understanding mental health using ubiquitous sensors and machine learning," *Annual review of clinical psychology*, vol. 13, no. 1, pp. 23–47, 2017.

[27] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "Detecting depression and mental illness on social media: an integrative review," *Current Opinion in Behavioral Sciences*, vol. 18, pp. 43–49, 2017.

[28] E. Garcia-Ceja, "Depresjon: a motor activity database of depression episodes in unipolar and bipolar patients," in *Proceedings of the 9th ACM on Multimedia Systems Conference, MMSys'18*, ACM, New York, NY, USA, 2018.

[29] D. Paul, R. Su, M. Romain, V. Sébastien, V. Pierre, and G. Isabelle, "Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier," *Computerized Medical Imaging and Graphics*, vol. 60, pp. 42–49, 2017.

[30] 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.

[31] N. Dogru and A. Subasi, "Traffic accident detection using random forest classifier," in *Proceedings of the Learning and Technology Conference*, Jinan, China, May 2018.

[32] A. Liaw and M. Wiener, *Breiman and Cutler's Random Forests for Classification And Regression, R package Version 4.6-12*, R Foundation for Statistical Computing, Vienna, Austria, 2015.

[33] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: a review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.

[34] T. Saito and M. Rehmsmeier, "Precrec: fast and accurate precision-recall and ROC curve calculations in R," *Bioinformatics*, vol. 33, no. 1, pp. 145–147, 2017.

[35] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.

[36] R Core Team, *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018.

[37] V. Trevino and F. Falciani: "Galgo: genetic algorithms for multivariate statistical models from large-scale functional genomics data, R package version 1.4", 2018.

[38] X. Robin, N. Turck, A. Hainard et al., "Proc: an open-source package for r and s+ to analyze and compare roc curves," *BMC Bioinformatics*, vol. 12, p. 77, 2011.

[39] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, *E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*, TU Wien, R package version 1.7-0, 2018.

[40] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, pp. 18–22, 2002.

[41] M. K. C. Jed Wing, *Caret: Classification and Regression Training*, 2018, R package version 6.0-79.

[42] P. Cortez, *Rminer: Data Mining Classification and Regression Methods*, 2016, R package version 1.4.2.

[43] A. Gershon, N. Ram, S. L. Johnson, A. G. Harvey, and J. M. Zeitzer, "Daily actigraphy profiles distinguish depressive and interepisode states in bipolar disorder," *Clinical Psychological Science*, vol. 4, no. 4, pp. 641–650, 2016.