Depression Episodes Detection: A Neural Net and Deep Neural Net Comparison.*

Carlos H. Espino-Salinas[1 [0000-0001-8092-1333]], Carlos E. Galván-Tejada[1 [0000-0002-7635-4687]], Hamurabi Gamboa-Rosales[1 [0000-0002-9498-6602]], Huizilopoztli Luna-Garcia[1 [0000-0001-5714-7482]], and Jorge I. Galván-Tejada[1 [0000-0002-7555-5655]]

Unidad Académica de Ingeniería Eléctrica, Universidad Autónoma de Zacatecas, Jardín Juarez 147, Centro, 98000 Zacatecas, Zac, México;
`dantecore2@gmail.com, ericgalvan@uaz.edu.mx`
http://uaz.edu.mx

Abstract. Depression is a frequent mental disorder. It is estimated that it affects more than 300 million people in the world. In this investigation, a motor activity database was used, from which the readings of 55 patients (32 control patients and 23 patients with the condition) were selected, during one week in one minute intervals, obtaining a total of 385 observations (participants) and 1440 characteristics (time intervals) from which the most representative one minute intervals were extracted applying genetic algorithms that reduced the number of data to process, with this strategy it is guaranteed that the most representative genes (characteristics) in the chromosome population is included in a single machine learning model of which applied deep neural nets and neural nets with the aim of creating a comparative between the models generated and determining which model offers better performance to detecting episodes of depression. The deep neural networks obtained the best performance with 0.8086 which is equivalent to 80.86 % of precision, this deep neural network was trained with 270 of the participants which is equivalent to 70 % of the observations and was tested with 30 % Remaining data which is equal to 115 participants of which 53 were diagnosed as healthy and 40 with depression correctly. Based on these results, it can be concluded that the implementation of these models in smart devices or in some assisted diagnostic tool, it is possible to perform the automated detection of episodes of depression reliably.

Keywords: Depression · Motor activity · Deep neural nets · Depresjon dataset · Machine learning.

* The dataset was originally collected for the study of motor activity in schizophrenia and major depression https://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-3-149.

1 Introduction

Mental health encompasses a wide range of activities directly or indirectly related to the component of mental well-being included in the WHO definition of health: a state of complete physical, mental and social well-being, and not only the absence of affections or diseases. Depression is a frequent mental disorder, characterized by the presence of sadness, loss of interest or pleasure, feelings of guilt or lack of self-esteem, sleep or appetite disorders, feeling tired and lack of concentration. Depression can become chronic or recurrent and significantly impair performance at work or school and the ability to cope with daily life. In its most serious form, it can lead to suicide.If it is mild, it can be treated without the need for medication and professional psychotherapy. Around 800,000 people commit suicide each year, and suicide is the second leading cause of death among people ages 15 to 29 [16].

Depression generally begins at an early age, substantially reduces the functioning of people, is a recurring condition and has significant economic and social costs, which is why depression is at the top of the list of disabling diseases and has become a priority objective of attention worldwide [2].

Mental disorders such as depression is associated with disrupted biological rhythms caused by environmental disturbances such as seasonal change in daylight, disruption of social rhythms due, for example, to work shifts or long trips; in addition to being related to lifestyles associated with unconscious daytime rhythms with the natural cycle of daylight. The appearance of depressive symptoms is also related to physical health problems, medical side effects, life events and social factors, in addition to alcohol and substance abuse, and these factors could also cause symptoms of depression in all humans. The overall prevalence of depression in life is approximately 15%, but the incidence of episodes with a level of severity that does not meet the requirements for a depressive diagnosis is much more frequent [8].

Unfortunately, the current psychiatric evaluation method requires a great effort on the part of physicians to gather complete information about patients. Furthermore, it is highly dependent on patients' efforts to cooperate in communicating their symptoms and problems. Therefore, an objective signal-based, biological and behavioral detection mechanism is needed to improve the current diagnostic method [11].

Research papers address different approaches and certain supervised machine learning techniques that use motor activity as a source of information to create models capable of detecting episodes of depression, as well as methods that have been used for the development of classification models such as the case of genetic algorithms, which are search procedures that are based on the principle of evolution of natural selection, emulating this process using mechanisms such as the survival of the subset of more effective variables, mutation and crossing to improve combinations.

The methods of classification and detection of depression using machine learning tools are widely used for the study of different mental disorders, as well as systems capable of giving an approximate diagnosis about whether a person suffers from a mental disorder that today turns out to be a very useful tool for the prevention of this type of disease such as depression and that due to the delicacy with which these conditions are treated it is essential to know how to deal with these types of conditions. Based on this description, the present study aims to analyze, extract and process motor activity readings using deep neural net (DNN), artificial neural net (ANN) and genetic algorithms (GA) of patients with depression (condition) and without depression (control) to finally present a comparative based on precision, area under the curve and F1-Score of Two different machine learning approaches (DNN and ANN) and determine the model with the best performance.

2 Related Work

Because depression is the leading cause of disability and is worst-case scenario can lead to suicide, there are effective treatments to treat it. There are several works that have been developed in order to measure the level of depression and thus prevent it from reaching a serious state, even so, there are obstacles to effective care such as lack of resources, trained health personnel, stigmatization of mental disorders and inaccurate evaluation.Although effective treatment of depressed patients requires regular follow-up and monitoring, some research proposes new ways to treat depression.

In the work of Frogner et al [6]. A machine learning approach to detecting depression is presented using a data set with records of motor activity from a group of people with depression and a group without, that is, the condition group includes 23 unipolar and bipolar people, and the group control includes 32 people without depression. Using convolutional neural networks to classify depressed and non-depressed patients. In addition, different levels of depression were classified. Finally, a model was formed that predicts the Montgomery-Asberg depression scale scores. It could be concluded that detection of depression using activity data is definitely possible within a clinical setting.

Also García-Ceja et al [8]. presented their work on the classification of depression based on motor activity in unipolar / bipolar patients. They applied machine learning to classify depressed and non-depressed participants using techniques such as Random Forest and Deep Neural Networks. The main contribution is to better understand the association between depression and motor activity. The conclusions of this work follow that the reproducibility and comparability of the results is an important factor of high quality research. This article presents a set of data in the field of depression analysis that allows for reproducibility and comparability, which makes it unique in the field of computing and psychology.

Farholt-Jepsen et al. [5] I present an association-type study on bipolar disorder. The participants were 29 bipolar patients, from where the authors collected various actions from the patients' smartphones, such as daily use, the number of incoming calls and the number of text messages sent or received, and they were able to conclude that the objective data generated Automatically on aspects of behavioral activities collected with smartphones reflect the level of depressive and manic symptoms and discriminate between categories. Furthermore, these automatically generated objective data can be used to detect disease activity in patients with moderate to severe levels of depressive and manic symptoms.

On the other hand, Zanella-Calzada et al. [18] proposes a method to detect depressive states through the motor activity of the patients, using data from an intelligent band, applying an approach of extraction of characteristics of the activity developed by the patients, which will allow timely diagnosis and treatment, the proposals in this investigation, based on the information of the physical activity of the patients, taking into account that the depressed subjects are more likely to have motor movement than the healthy patients. The deepening of these aspects can contribute to the prevention of this psychopathology, as well as to the development of effective treatments, therefore, the main benefit presented in this study is a preliminary tool that can support the diagnosis of specialists. Know if a patient has depression based on the level of activity he has in a full day through the automatic diagnosis of the subjects obtained by sending this information to the model developed in this work, relating total motor activity with the presence or absence of depression , which according to the presented results have a significantly high precision, which allows reducing false positives and false negatives in the detection of this condition, thus improving the diagnosis of this disease.

In the research work of Rodríguez-Ruiz et al. [13] a data extraction process is performed for classified depressive episodes using data collected overnight, day, and full 24 hours. Comparison between the classification using different data collected over time gives a better picture of the disease and the behavior of the patients with the diagnosis. According to the results obtained, it can be seen that, from the selection of characteristics, the best set of characteristics selected is obtained from the data corresponding to the night period, since the best precision is calculated when classifying subjects with these characteristics using the activity levels presented during the night, which are generally related to subjects with depression.

Berle et al. [3] proposed to study the complexity of motor activity patterns in these patients by using actigraphy and concluded that motor activity was significantly reduced in both schizophrenic and depressed patients.However, schizophrenic patients differed from both depressed and control patients, demonstrating patterns of motor activity marked by less complexity and more structured behavior. These findings may indicate that abnormalities in motor activity reflect different pathophysiological mechanisms in schizophrenia compared to major depression.

Finally, Galván-Tejada et al. [7] studies the signal generated by a smart band accelerometer to detect depressive states through the activity of patients and propose an extraction of characteristics (using the temporal and spectral evolution of the signal), as well as an intelligent selection of characteristics based on a genetic algorithm approach to minimize the data necessary to identify these depressive states that allow a non-invasive diagnosis in near real time. The results obtained in this research demonstrate that the statistical characteristics extracted show that the information they contain provides a description of the main characteristics of the patient's full-day activity, which allows differentiation between depressed and non-depressed subjects, thus having a model of classification with a reduced number of features allowing a lower computational cost, which facilitates access to it, since it does not require specialized software or hardware for its implementation. Another of the main benefits demonstrated in this work is the high precision values obtained through a simple methodology that uses a single data source.

3 Materials and Methods

The methodology and materials used to carry out a comparison between neural nets and deep neural nets in this investigation are presented in the following subsections. Findings obtained following this approach allow defining which artificial intelligence tool is suitable for the detection of depression.

3.1 Materials

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. Total activity counts were continuously recorded in one minute intervals.The activity counts were recorded in intervals of one minute.

This dataset consists of actigraphy data collected from 23 unipolar and bipolar depressed patients (condition group). Five subjects were hospitalized during their data collection period, and 18 were outpatients. In addition, the dataset contains actigraphy data from 32 non-depressed contributors (control group), consisting of 23 hospital employees, 5 students and 4 former patients without current psychiatric symptoms.

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 Asberg 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) [8].

For this work only the features over time were used.

3.2 Data Preprocessing

Data preprocessing can often have a significant impact on generalization performance of a machine learning algorithm. Eliminating instances of noise is one of the most difficult problems in machine learning. Deleted instances typically have excessively skewed instances that have too many null entity values. These excessively deviant characteristics are also known as outliers. Also, a common approach to coping with the inability to learn from very large data sets is to select a single sample from the large data set. Data mismanagement is another problem often addressed in data preparation steps [12].

In this part, preprocessing consists of three main steps to avoid performance problems, these are: select participants' recorded motor activity for a period of one week at one-minute intervals, reorder them to identify the number of observations as a relationship between participant-day, with their respective registered data.

This means that there is a column (C) of each participant with a time range from 00:00 to 23:59 (one day), each minute with its respective record of motor activity, this column is transposed and adds to first row of an A matrix with its respective output 0 or 1 depending on dataset where the information was extracted (control / condition), successively same applies for subsequent days and participants until completing a week of information, taking as A matrix result A = [385 X 1441] including its output, as explained below.

Given columns of participant 1 on days 1,2,3,4,5,6,7 we have C_{11}, C_{12}, C_{13}, $C_{14}, C_{15}, C_{16}, C_{17}$, where each hour contains 60 minutes that multiplied by 24, which is number of hours a day has, a column of 1440 rows is obtained that is equivalent to the minutes of a day, each one with a record of motor activity given in volts (v). This is represented as follows:

$$C_{11} = \begin{array}{c} 00{:}00 \\ 00{:}01 \\ 00{:}02 \\ 00{:}03 \\ 00{:}04 \\ \vdots \\ 23{:}59 \end{array} \begin{bmatrix} 168\,v \\ 107\,v \\ 550\,v \\ 157\,v \\ 0\;\; v \\ \vdots \\ 208\,v \end{bmatrix} C_{12} = \begin{array}{c} 00{:}00 \\ 00{:}01 \\ 00{:}02 \\ 00{:}03 \\ 00{:}04 \\ \vdots \\ 23{:}59 \end{array} \begin{bmatrix} 168\,v \\ 17\,v \\ 560\,v \\ 167\,v \\ 23\;\; v \\ \vdots \\ 280\,v \end{bmatrix} C_{13} = \begin{array}{c} 00{:}00 \\ 00{:}01 \\ 00{:}02 \\ 00{:}03 \\ 00{:}04 \\ \vdots \\ 23{:}59 \end{array} \begin{bmatrix} 118\,v \\ 157\,v \\ 450\,v \\ 157\,v \\ 0\;\; v \\ \vdots \\ 305\,v \end{bmatrix} C_{14} = \begin{array}{c} 00{:}00 \\ 00{:}01 \\ 00{:}02 \\ 00{:}03 \\ 00{:}04 \\ \vdots \\ 23{:}59 \end{array} \begin{bmatrix} 148\,v \\ 157\,v \\ 120\,v \\ 457\,v \\ 121\,v \\ \vdots \\ 892\,v \end{bmatrix}$$

$$C_{15} = \begin{array}{c} 00{:}00 \\ 00{:}01 \\ 00{:}02 \\ 00{:}03 \\ 00{:}04 \\ \vdots \\ 23{:}59 \end{array} \begin{bmatrix} 168\,v \\ 107\,v \\ 550\,v \\ 157\,v \\ 0\;\; v \\ \vdots \\ 208\,v \end{bmatrix} C_{16} = \begin{array}{c} 00{:}00 \\ 00{:}01 \\ 00{:}02 \\ 00{:}03 \\ 00{:}04 \\ \vdots \\ 23{:}59 \end{array} \begin{bmatrix} 168\,v \\ 17\,v \\ 560\,v \\ 167\,v \\ 23\;\; v \\ \vdots \\ 280\,v \end{bmatrix} C_{17} = \begin{array}{c} 00{:}00 \\ 00{:}01 \\ 00{:}02 \\ 00{:}03 \\ 00{:}04 \\ \vdots \\ 23{:}59 \end{array} \begin{bmatrix} 118\,v \\ 157\,v \\ 450\,v \\ 157\,v \\ 0\;\; v \\ \vdots \\ 305\,v \end{bmatrix}$$

These columns are transposed to become a vector (V) where each one-minute interval corresponds to a characteristic and each vector is added to the rows of a matrix A.

$$A=\begin{array}{rccccccc}
 & 00{:}00 & 00{:}01 & 00{:}02 & 00{:}03 & 00{:}04 & \ldots & 23{:}59 \\
V_{11}^{T} = & [168\,v & 107\,v & 550\,v & 157\,v & 0\,v & \ldots & 208] \\
V_{12}^{T} = & [168\,v & 17\,v & 560\,v & 167\,v & 23\,v & \ldots & 280] \\
V_{13}^{T} = & [118\,v & 157\,v & 450\,v & 157\,v & 0\,v & \ldots & 305] \\
V_{14}^{T} = & [148\,v & 157\,v & 120\,v & 457\,v & 121\,v & \ldots & 892] \\
 & & & & \vdots & & & \\
V_{557}^{T} = & [563\,v & 124\,v & 456\,v & 321\,v & 265\,v & \ldots & 201]
\end{array}$$

For verification of empty fields, missmap function is used, found inside of Amelia Library that helps to graphically visualize empty fields in a data set that could cause problems when analyzing and processing information [9].

Finally, standardizing raw input data has a great effect on preparing the information to be suitable for training. Without this standardization, training machine learning models would have been very slow.There are many types of data normalization. It can be used to scale data in the same range of values for each input feature to minimize bias within the model from one feature to another. Data normalization can also speed up training time for each function within same scale. It is especially useful for modeling applications where the inputs are generally at very different scales. Mean and standard deviation are calculated for each characteristic [10]. The transformation occurs in equation 1.

$$z_i = \frac{x_i - \bar{x}}{\sigma} \tag{1}$$

Using statistical normalization avoids outlier problems as it handles outliers, but does not produce normalized data with the exact same scale

3.3 Feature extraction

In the extraction of characteristics obtain most relevant one minute intervals of selected set in data's preprocessing for classification of patients is in part for avoid over fitting during analysis, reduce redundancy, recover significant latent characteristics, generate greater understanding in the data generation process, as well as improve its performance, by reducing characteristics.

In this step, GALGO, which is a generic software package that uses genetic algorithms (GA) to solve optimization problems involving selection of a subset of variables, makes a schematic representation of a selection of variables of two classes (0 - control and 1 - condition), which implies combination of genes with which we are working (chromosomes that represent multivariate model) that allows us to distinguish between classes in a function of a classification method, where there are a series of models that perform this processing multiple times, with different combinations.

In this investigation, each minute is considered as a characteristic or gene for process of evolution and genetics of GA, this means that from two classes of samples (condition and control) it searches and evolves, combining genes (chromosomes that represent a multivariate model) that distinguish between classes.

The scheme for extraction of characteristics using genetic algorithms to generate a detection model for patients with depression is shown in Figure 1.

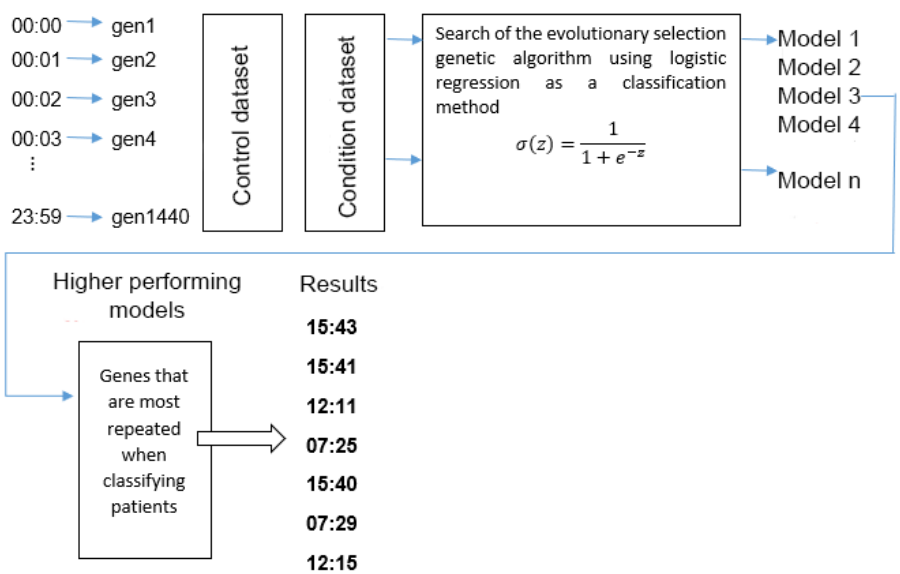


Fig. 1. Diagram process of the genetic algorithms to extraction most important characteristics for classification of patients with depression using logistic regression.

3.4 Neural nets description

As part of the analysis of comparison between artificial neural net and deep neural net, it is proposed to generate a binary classification model with two possible classes of output, participant with an episode of depression or condition (1) and a healthy or control participant (0).

Artificial neural nets are inspired by the hypothesis of functioning of the brain, they are connected to units or nodes through directed links. A link from unit i to unit j serves to propagate activation a_i from i to j. Each link has a numerical weight $w_{i,j}$ associated with it, which determines the strength and sign of connection. Like linear regression models, each unit has a dummy input $a_0 = 1$ with an associated weight, $w_{0,j}$.Each unit j first computes sum of its inputs through Equation 3.

$$in_j = \sum_{i=0}^{n} w_{i,j} a_i \tag{2}$$

Then, activation function g is applied to this sum to obtain the output, as shown in Equation 4.

$$a_j = g(in_j) = g(\sum_{i=0}^{n} w_{i,j} a_i) \tag{3}$$

Artificial neural net (ANN) are generally arranged in layers, of which each unit can receive inputs only from units found in previous layer or can be fed back, depending type of RNA. This layers that are in between in the ANN are called hidden layers. In a multilayer ANN with multiple outputs m, error vector in the output layer can be described as $y - h_w(x)$, where y is expected output and $h_w(x)$ is calculated output; however, error in hidden layers becomes complex to calculate because training data does not allow us to know what value hidden nodes should have. This problem is resolved with backpropagation, where error can be propagated from output layer to hidden layers. This process emerges directly from derivation of general error gradient.

In output layer, the rule for updating weights to minimize error is calculated with Equation 5.

$$w_i = w_i + \alpha(y - h_w(x)) \times h_w(x)(1 - h_w(x)) \times x_i \tag{4}$$

Having multiple output units, Err_k is the component kth of error vector $y - h_w$, and modified error vector is $\Delta_k = Err_k \times g'(in_k)$, so the weight update rule becomes in Equation 6.

$$w_{j,k} = w_{j,k} + \alpha \times a_j \times \Delta_k \tag{5}$$

To update connections between the input units and hidden units, it is necessary to define an analogous quantity at end of error for output nodes. This quantity is defined with the error propagated backwards, where hidden node j is responsible for a fraction of error Δ_k of each output node to which it is connected. Δ_k values are divided according to strength of the connection between hidden node and output node and are propagated backwards to provide Δ_j values of the hidden layer. Propagation rule for values Δ is calculated in Equation 7.

$$\Delta_j = g'(in_j) \sum_{k} w_{j,k} \Delta_k \tag{6}$$

Now the rule for updating weights between inputs and hidden layer becomes the same as updating rule for output layer, and is calculated with Equation 8 [14].

$$w_{i,j} = w_{i,j} + \alpha \times a_i \times \Delta_j \tag{7}$$

For this work, a deep neural nets was created with a empirically approached composed of 4 hidden layers where the first contains 6 perceptrons, second 5 perceptrons, third 4 and fourth layer 3 perceptrons each with a sigmoid activation function. First layer known as input layer is made up of the one-minute intervals extracted in genetic algorithms and an output layer with a softmax activation function, this function in mathematics field, especially in probability theory, softmax function is considered the generalization of logistics function [17]. Its mathematical expression is given by Equation 9.

$$\phi(j, T_1, .., T_n) = \phi_k^j = \frac{exp(T_j/k)}{\sum_{j=1}^{n} exp(T_j/k)} \qquad k > 0 \tag{8}$$

Softmax function has characteristics of nonlinear, monotonicity and boundedness. [15].

3.5 Validation

Evaluation of the results obtained was carried out in the validation stage through a ROC curve-based approach. ROC curve has been widely used to measure or visualize a classifier's performance in conjunction with AUC value to select a suitable operating point, called as decision threshold [4].

Simplest way to calculate AUC is through trapezoidal integration, shown in Equation 10.

$$AUC = \sum_i (1 - \beta_i \cdot \Delta\alpha) + \frac{1}{2}[\Delta(1 - \beta) \cdot \Delta\alpha] \tag{9}$$

More significant measures can also be extracted to have certain performance criteria, such as precision, which refers degree to which the result of a calculation adjusts to correct value, shown in equation 10.

$$accuracy(1 - error) = \frac{Tp + Tn}{Cp + Cn} \tag{10}$$

Where Tp are true positives, Tn are true negatives, Cp are truly positive, and Cn are truly negative.

Accuracy of a classifier depends on the ratio between number of targets correctly detected and all detected targets, known as precision, and ratio between number of targets correctly detected and all true targets, known as recall as shown in Equations 12 and 13.

$$precision = \frac{Tp}{Tp + FP} \tag{11}$$

$$recall = \frac{Tp}{Tp + FN} \tag{12}$$

Where FP is false positive and FN is false negative.

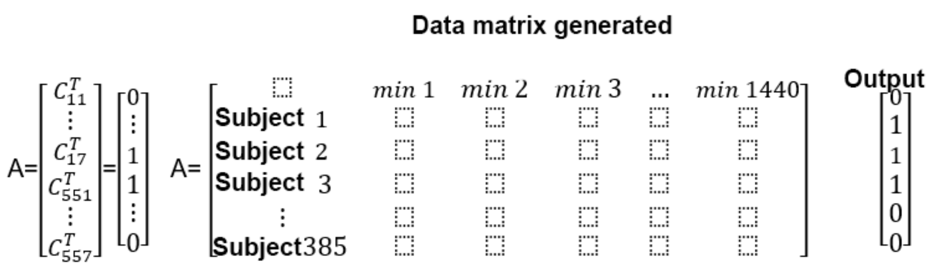
Since precision and recall are both necessary for evaluating detection capabilities of an algorithm, it is convenient to find a single measure that considers both of them. One measure that combines both of them is the F1-score, which is harmonic mean of precision and recall [1]. as show in Equation 14.

$$F_1 = \frac{precision \cdot recall}{precision + recall} \tag{13}$$

Since there are only two possible output values (healthy and depressive episode) the results are presented using a confusion matrix that shows difference between predicted and real states.

4 Experiments and Results

After data preprocessing, For each participant, both control and condition, their vector is extracted per day until completing 7 days that make up a week, it is added to matrix A at the same time that its output value is assigned and it continues with next participant until completing 55 participants that make up the database of which there is a record, which multiplied by 7 days is equivalent to 385 observations and 1440 characteristic.



This is how data set is established for this work, where a better management and understanding of information is allowed for following processes.

One of the benefits of Galgo is that it is very easy to create new methods based on the vast methods available. This can be attained by using parameters $classification.method = "user"$, and $classification.userFitnessFunc$ (a classification method based on logistic regression was specified,). However, the function specified there must follow certain rules.it must receive five parameters, *chr* which receive genes (one-minute intervals), *parent* which receive the

BigBang object, *tr* which receive training samples, *te* which receive test and *result* which receive 0 for class prediction and 1 for accuracy. Additionally, a MaxBigBangs object is configured that will store 1000 chromosomes (maxSolutions = 1000) for 500 generations with an accuracy of 95 % (goalFitness = 0.95). Other parameters define name of the saved object that is created and frequency of saving results in a file (saveFrecuency = 5). Finally, stored object has most relevant characteristics or one-minute intervals with greatest stability for a classification model are: 15:43, 15:41, 12:11, 07:25, 15:40, 7:29, 12:15. Using these characteristics instead of using all features allow provide relevant information for medical personnel about the time intervals in which a person may suffer an episode of depression. On engineering side is to perform application procedure of neural nets more efficient since it implies less time and less computational cost, this because development of a neural net with all the characteristics could imply to implement a greater number of layers and therefore it would take more processing time. Reasons for the use of GA is randomity as an essential role since both selection and the reproducibility need random procedures, also an important point to consider these algorithms as a feature extraction technique in this research work is that it always they consider a population of solutions and lastly is the ease they have to find a global optimum. It is essential to mention that GAs can be modified for different problems and in this sense they have another advantage over other methods of feature extraction.

Results obtained from deep neural network that had better performance after performing the preprocessing and extraction of characteristics were validated using parameters such as sensitivity and specificity shown in DNN confusion matrix in Table 1. Where of the 385 observations, 270 equivalents 70% were used to train neural nets and 115 equivalents 30% were used as tests where, 40 test participants were classified as control or healthy (TP) and 53 with an episode of depression or condition (TN), in addition 8 participants as false positives (FP) and 14 as false negatives (FN).

Table 1. Matrix of confusion of the classification of patients based on DNN

	Control	**Condition**
Control	40	8
Condition	14	53

In Table 2. It can be seen that in each of comparative validation metrics (Accuracy, AUC and F1 Score), of neural nets implemented, DNN obtains best performance compared to ANN for detection of depression in patients, since their index is higher in each of these metrics.

Metrics	DNN	ANN
Accuracy	0.808	0.739
AUC	0.805	0.744
F1 Score	0.828	0.761

Table 2. Comparison table of the performance of neural nets implemented to detect episodes of depression.

Area under the curve (AUC) has ability to differentiate patients from each other; This means that among the 115 test patients, 40 were diagnosed as healthy and 53 with an episode of depression correctly, equivalent to 80.8 % accuracy in test data set using deep neural nets. Same type of AUC validation metric was applied to each of models generated by neural nets (ANN and DNN), as shown in Figure 2.

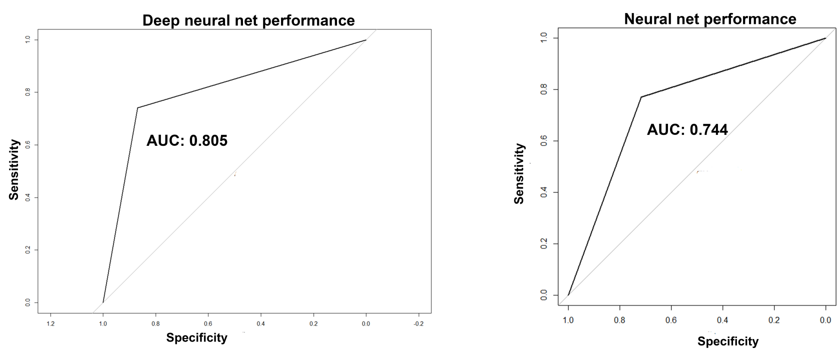


Fig. 2. Comparison of the ROC curves of different models generated by neural nets.

Although, it is known that other automatic learning techniques yield quite solid results at the time of classification, but here we intend to know behavior of neural nets at the time of processing people's motor activity, since ANN are very good when it comes to complex problems such as image classification, speech recognition, etc.

5 Conclusions and Future Work

Artificial intelligence techniques such as artificial neural networks, supported by genetic algorithms, allowed efficient generation of a model with a degree of 80 % precision, through a reduced number of characteristics that present useful information to better understand depression as a disease every more and more frequent in society.

Therefore, it is possible to conclude that the methodology proposed in this work allows automatic detection of depression through a DNN with significant precision, from a set of seven characteristics obtained from physical activity of patients. In this way, preliminary development of an assisted diagnostic tool is presented that can help reduce the high rates of error in the diagnosis of this disease, in addition to reducing its high prevalence and also offering adequate and timely treatment.

As future work, it is proposed to increase number of subjects in the experimentation to obtain a more balanced data set that generates greater diversity and obtain more robust results. Apply more complex validation methods such as Wilcoxon Signed Rank Test non-parametric test that allows obtaining a differentiation parameter between two samples proposed in this work (control and condition). In addition, it is proposed to include a features extraction stage where different methods are compared to find out which is better for contribution in detection of episodes of depression. Finally, it is also proposed to make a comparison between the approach established in this research and others approaches mentioned that are based on statistical variables of the same reference data.

References

1. AlBeladi, A., Muqaibel, A.: Evaluating compressive sensing algorithms in through-the-wall radar via f1-score. International Journal of Signal and Imaging Systems Engineering **11**, 164 (01 2018). https://doi.org/10.1504/IJSISE.2018.093268
2. Berenzon, S., Lara, M.A., Robles, R., Medina-Mora, M.E.: Depresión: Estado del conocimiento y la necesidad de políticas públicas y planes de acción en méxico. Salud Publica de Mexico **55**(1) (2013). https://doi.org/10.1590/S0036-36342013000100011
3. Berle, J.O., Hauge, E.R., Oedegaard, K.J., Holsten, F., Fasmer, O.B.: Actigraphic registration of motor activity reveals a more structured behavioural pattern in schizophrenia than in major depression. BMC Research Notes (2010). https://doi.org/10.1186/1756-0500-3-149
4. Bradley, A.P.: The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recognition **30**(7), 1145 – 1159 (1997). https://doi.org/https://doi.org/10.1016/S0031-3203(96)00142-2, http://www.sciencedirect.com/science/article/pii/S0031320396001422
5. Faurholt-Jepsen, M., Vinberg, M., Frost, M., Debel, S., Margrethe Christensen, E., Bardram, J.E., Kessing, L.V.: Behavioral activities collected through smartphones and the association with illness activity in bipolar disorder. International Journal of Methods in Psychiatric Research (2016). https://doi.org/10.1002/mpr.1502
6. Frogner, J.I., Noori, F.M., Halvorsen, P., Hicks, S.A., Garcia-Ceja, E., Torresen, J., Riegler, M.A.: One-dimensional convolutional neural networks on motor activity measurements in detection of depression. In: HealthMedia 2019 - Proceedings of the 4th International Workshop on Multimedia for Personal Health and Health Care, co-located with MM 2019 (2019). https://doi.org/10.1145/3347444.3356238
7. Galván-Tejada, C.E., Zanella-Calzada, L.A., Gamboa-Rosales, H., Galván-Tejada, J.I., Chávez-Lamas, N.M., Gracia-Cortés, M.D.C., Magallanes-Quintanar, R., Celaya-Padilla, J.M.: 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. Mobile Information Systems (2019). https://doi.org/10.1155/2019/8269695
8. Garcia-Ceja, E., Riegler, M., Jakobsen, P., T\o{}rresen, J., Nordgreen, T., Oedegaard, K.J., Fasmer, O.B.: Depresjon: A Motor Activity Database of Depression Episodes in Unipolar and Bipolar Patients. In: Proceedings of the 9th ACM Multimedia Systems Conference. pp. 472–477. MMSys '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3204949.3208125, https://doi.org/10.1145/3204949.3208125
9. Honaker, J., King, G., Blackwell, M.: Amelia II: A Program for Missing Data. Journal of Statistical Software, Articles **45**(7), 1–47 (2011). https://doi.org/10.18637/jss.v045.i07, https://www.jstatsoft.org/v045/i07
10. Jayalakshmi, T., Santhakumaran, A.: Statistical Normalization and Back Propagationfor Classification. International Journal of Computer Theory and Engineering (2011). https://doi.org/10.7763/ijcte.2011.v3.288
11. Khairuddin, S., Ahmad, S., Embong, A.H., Nik Hashim, N.N.W., Altamas, T.M., Syd Badaruddin, S.N., Hassan, S.S.: Classification of the Correct Quranic Letters Pronunciation of Male and Female Reciters. In: IOP Conference Series: Materials Science and Engineering (2017). https://doi.org/10.1088/1757-899X/260/1/012004
12. Kotsiantis, S.B., Kanellopoulos, D.: Data preprocessing for supervised leaning. International Journal of ... (2006). https://doi.org/10.1080/02331931003692557
13. Rodríguez-Ruiz, J.G., Galván-Tejada, C.E., Zanella-Calzada, L.A., Celaya-Padilla, J.M., Galván-Tejada, J.I., Gamboa-Rosales, H., Luna-García, H., Magallanes-Quintanar, R., Soto-Murillo, M.A.: Comparison of night, day and 24 h motor activity data for the classification of depressive episodes. Diagnostics (2020). https://doi.org/10.3390/diagnostics10030162
14. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall Press, USA, 3rd edn. (2009)
15. Torres, R., Salas, R., Astudillo, H.: Time-based hesitant fuzzy information aggregation approach for decision-making problems. International Journal of Intelligent Systems **29**(6), 579–595 (2014). https://doi.org/10.1002/int.21658
16. WHO: OMS — Salud mental (2017), http://www.who.int/mental_health/es/
17. Yu, D.: Softmax function based intuitionistic fuzzy multi-criteria decision making and applications. Operational Research **16** (09 2015). https://doi.org/10.1007/s12351-015-0196-7
18. Zanella-Calzada, L.A., Galván-Tejada, C.E., Chávez-Lamas, N.M., del Carmen Gracia-Cortés, M., Magallanes-Quintanar, R., Celaya-Padilla, J.M., Galván-Tejada, J.I., Gamboa-Rosales, H.: Feature extraction in motor activity signal: Towards a depression episodes detection in unipolar and bipolar patients. Diagnostics (2019). https://doi.org/10.3390/diagnostics9010008