ISSN 1870-4069

Evaluation of Five Classifiers for Depression Episodes Detection

Susana L. Pacheco-González[1], Laura A. Zanella-Calzada[1], Carlos E. Galván-Tejada[1], Nubia M. Chávez-Lamas[2], J. Francisco Rivera-Gómez[1], Jorge I. Galván-Tejada[1]

[1] Universidad Autónoma de Zacatecas, Unidad Académica de Ingeniería Eléctrica, Mexico

[2] Universidad Autónoma de Zacatecas, Unidad Académica de Odontología, Mexico

{lzanellac, ericgalvan, jf_riverag, gatejo, nubiachavez}@uaz.edu.mx

Abstract. Depression is a mental disorder manifested through a set of psychological and physical symptoms, such as the presence of sadness, apathy, hopelessness and irritability, among others. According to the World Health Organization (WHO), depression is affecting more than 300 million people worldwide, presenting a prevalence between 3 and 21%. One of the main problems of this high prevalence is the incorrect classification of patients, since many cases are false positive and false negative diagnoses. In this work it is proposed the study of the behavior of five different classification techniques, random forest (RF), conditional inference trees (cTree), K-nearest neighbor (K-NN), support vector machine (SVM) and Naïve Bayes, to identify depressive states through the motor activity of patients contained in the Depresjon dataset. The activity of this dataset is acquired through the smart watch "Actigraph", based on actigraphy. The evaluation of these classification techniques is finally performed in terms of sensitivity, specificity, the receiver operating characteristic (ROC) curve and area under the curve (AUC), to know their performance to automatically detect depressive patients. The results shown values of sensitivity, specificity and AUC, statistically significant, specially for the RF method, which presents sensitivity = 0.8148, specificity = 0.8158 and AUC = 0.8314. Therefore, it is concluded that these classifiers are able to distinguish patients with depression from controls, based on their motor activity, allowing the development of a non-invasive diagnosis tool to support specialists in the correct diagnosis of depression.

Keywords: depression, motor activity, classification, random forest, conditional inference trees, K-nearest neighbor , support vector machine, naïve Bayes.

1 Introduction

Depression is a mental disorder which is manifested through a set of psychological and physical symptoms. Among these symptoms are present the predominance of feelings related to sadness, apathy, hopelessness and irritability, besides the deterioration in personal appearance, crying and insomnia. This disorder can also be accompanied by other emotional and / or physical illnesses, such as anxiety, alcohol and other substance abuse, systematic illnesses, eating disorders and even some personality disorders [5].

According to the World Health Organization (WHO), in 2017 the depression was present in more than 300 million people worldwide, having a current prevalence between 3 and 21 %. Mainly, this condition affects people from 15 to 45 years old and, in most countries, it occurs twice as many women as men. In addition, this condition appears as the second cause of death in people aged 15 to 29, causing about 800,000 suicides per year [11].

This disorder is regularly classified into three main categories, mild, moderate or severe, which help to make an adequate diagnosis. These categories depend on the duration with which the symptoms occur and can range from transient to persistent (the presence lasts for months or years) and, according to this classification, the disease can rebound in different ways, interfering with the work, the school, family, society, among others [8].

The diagnosis of depression also allows to receive a more appropriate treatment according to the degree to which it occurs. However, according to the literature, about 56 % of the cases are not diagnosed due to the lack of resources and trained health personnel, together with the inaccurate clinical evaluation and the mythification that is had of this disease by of patients, which is based on stigma related to mental disorders caused by lack of education.

On the other hand, the problem of false positives and false negatives in the diagnosis of depression continues to present itself in countries of all types of income in a significant way, despite the extensive study that has been focused on this disease, causing patients with depression do not receive any type of treatment while healthy patients or patients with a condition other than depression are diagnosed as positive and are subject to antidepressants [1].

Due to this diagnostic problem, different artificial intelligence (IA) tools have been implemented that can contribute to improve the detection techniques of this disorder. The main purpose of this approach is the development of tools that allow to automatically identify when a patient presents depression from certain type of information provided by the patient. To be able to carry out this task, the AI tools are based on a previous training, which consists in submitting an algorithm to an automatic learning that consists in the search of relationships of a considerable number of examples of cases similar to those that they want to classify, with the real diagnosis of those examples, being information of patients with depression specifically for this case [14].

Based on this, the main contribution of this work is to study and compare the behavior of different types of classification techniques to predict depressive states through the activity of patients, measured with a smart band accelerometer, this evaluation in terms of sensitivity and specificity, as a receiver operating characteristic (ROC) curve and area under the curve (AUC), let to know the performance to maximize the detection of depressive or not depressive states depending on what researchers want to develop for a particular application. Additionally, this comparison allows the development of a real-time non-invasive diagnosis. Automatic depression states can increase significantly the development of an effective treatment and contribute to the prevention of this type of psychopathology.

This work is organized as follows. In section 2 is presented the materials and methods evaluated in this work, random forest (Rf), conditional inference trees (cTree), K-Nearest Neighbor (K-NN), support vector machine (SVM) and Naive Bayes. In section 3, the results obtained are shown. Finally, in section 4 these results are discussed and the conclusions are briefly described.

2 Materials and Methods

This section presents the detailed explanation of the stages proposed by this work, shown in Figure 1. Initially, a data preprocessing is applied to extract a set of data to represent the behavior of the patients. These data are subsequently subjected to five classifiers, RF, cTree, K-NN, SVM and Naive Bayes in order to fin a model that allows to the automatic classification of patients with presence of depression. Finally, the classification obtained by the different methods is validated through a set of statistical metrics, such as the ROC curve, AUC, sensitivity and specificity.

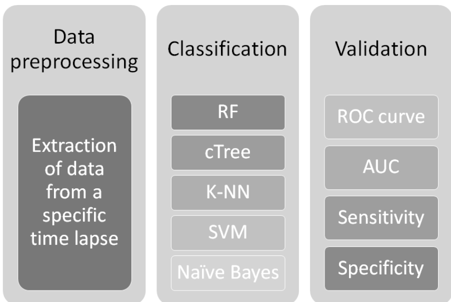

Fig. 1. Flowchart of the methodology proposed.

2.1 Data Description

For the development of this work, the Depresjon dataset is used, which is a collection of data contained by the information of the motor activity from patients with presence and absence of depression. These data have been deposited in the "control" and "condition" repositories of the dataset, being possible to access in http://datasets.simula.no/depresjon/ and / or can be directly downloaded from http://doi.org/10.5281/zenodo.1219550.

The information contained in the dataset is obtained from 55 patients, 32 controls (absence of depression, represented with 0) and 23 cases (presence of depression, represented with 1). For the acquisition of the data, the actigraph watch “Actiwatch” (model AW4, developed by Cambridge Neurotechnology Ltd, England) is used, which consists of patients carrying it on their left wrist for a series of days during the 24 hours. The Actiwatch measures the activity levels with a sampling frequency of 32 Hz, recording movements over 0.05 g. Movements are represented as voltage, which are stored as an activity count in the memory of the Actiwatch, being the number of counts proportional to the intensity of the movement. These activity counts are recorded in on minute lapses [3].

2.2 Data Preprocessing

For the preprocessing of the data, a specific time lapse was selected for each patient from the total data acquired. That is, for each observation, 60 samples are taken from a lapse of one hour, having a sampling frequency of one minute. These samples are selected within the time range of 9 pm to 5 am.

It is important to note that this time period is selected due to its clinical implication, which refers to the fact that subjects with depression tend to present a greater amount of motor activity during that time, while control subjects do not, meaning significant information for the distinction of case subjects.

2.3 Classification Methods

For this work five methods were selected to study the performance of different categories of classifiers to detect depression episodes, represented by an outcome with two possible values, 0 (controls) and 1 (cases). RF and cTree follow a decision based approach which helps to take into account the time dependence, on the other hand, K-NN and SVM is an instance based method to include all the observation as a unit, while Naive Bayes is a probabilistic approach.

Random Forest This method is a well-know classifier proposed by Breinman et al. [2], which uses two levels of randomness for the construction of the trees, beginning with a bootstrapped version of the training data, known as bagging, where a subset of the training data is used for each tree, based on the principle of replacement and the remaining data is used to estimate the error, through the out-of-bag (OOB) error. In the second level, a subset of features is randomly selected and added to each node throughout the growth of the decision trees. For each node, the best feature among a random subset of features is selected in order to reduce the label error. This technique bases its classification technique on taking the majority vote from all the decision trees. This process is recursively repeated until the trees reach a defined depth or the number of samples in a node does not exceed a threshold [15,10].

Conditional Inference Trees This method based on recursive partitioning, which is a fundamental tool to explore the structure of the data and developing decision rules for predicting a categorical or continuous outcome. The conditional distribution of statistics to measure the relationship between the input features and the output is the bases for the unbiased selection among covariates measured in different scales. In an algorithm that performs recursively partitions, a sample is developed using non-negative integer valued case weights. In cTree, two main steps are involved, (1) feature selection and (2) splitting. The step 1 consists in the selection of the covariate of strongest association with the output for splitting. Then, in step 2 a permutation test framework is developed to find the optimal binary split for the selected covariate in step 1. These steps are repeated recursively until the global null hypothesis of the independence between the output and any of the features cannot be rejected at a pre-specified threshold, which can be the p-value, to mention an example [12].

K-Nearest Neighbor This method present a non-parametric approach that has been widely used in different statistical applications since the early 1970's The basis of this algorithm consists on, from the calibration dataset, a group of K samples is found, which are nearest to unknown samples. To achieve this goal, the Euclidean distance, $||\overrightarrow{x} - \overrightarrow{y}||$, between a given set of queries and the inputs, is calculated, and the K closest input points are identified for each query [6]. From the these K samples, the output of the unknown samples are determined through the calculation of the average of the input features. The parameter K is determined based on several calculations. Therefore, as a result of this classifier, the k is a key element in the performance of the K-NN [9].

Support Vector Machine This method is based on statistical learning theory proposed by Vapniket al. [13], where the goal is to achieve structural risk minimization. A learning strategy to keep the empirical risk value fixed and minimize the confidence range is applied in this approach. Its basis is obtained from the optimal classification hyperplane with linear separability, which is mainly aimed in binary classification problems. The objective is to find a hyperplane to correctly separate the two types of outputs without errors, while the separated data points farthest from the classification surface are kept. To solve this problem and get the classifier, a constrained quadratic programming problem is constructed based on $min\frac{1}{2}||w||^2$, $sty_i(w \cdot x_i + b) \geq 1$, $i = 1, 2, ..., n$. According to the statistical learning theory, the error-free separation helps to ensure that the empirical risk is minimized, the distance between classifications is maximized and the confidence range of the generalization bound is minimized, minimizing the real risk. Therefore, SVM provides a good generalization ability.

Naïve Bayes This classifier is based on the Bayes theorem, which consists on Equation 1, where $G = G_i$ are the possible classes for a data point X and $Pr(G = G_i|X)$ represents the posterior probability based on prior knowledge combined with the observed data:

$$Pr(G = G_i|X) = \frac{Pr(X|G = G_i)Pr(G = G_i)}{Pr(X)}. \tag{1}$$

The Naïve Bayes model assumes that the inputs are conditionally independent in each class, and the classifier does not consider correlation between features, lowering the variance while the bias is increased and the overfitting is avoided on the training set. The classification of a vector X with d features is provided by taking the maximum posterior probability over all classes and applying the Bayes' theorem, as shown in Equation 2:

$$\underset{G_i \in G}{argmax} Pr(G = G_i) \prod_{j=1}^{d} Pr(X = X_j|G = G_i). \tag{2}$$

Then, all the information used to classify an arbitrary data point can be represented in a set of vectors, which is useful to carry out the classification.

2.4 Validation

For the validation stage, three parameters were measured, the AUC as a quantity value of the ROC curve, the specificity and the sensitivity.

The ROC curve is a metric widely used for the evaluation of binary classification problems, allowing to the graphic interpretation of the results based on the shape of the curve, being possible to provide its quantitative value with the AUC [7]. This parameter allows to measure the classification precision of the model through the sensitivity and the specificity.

The sensitivity is defined as the proportion of samples with a positive condition that are correctly classified. This metric can be calculated with Equation 3, where TP represents the true positives and FP represents the false positives [4]:

$$sensitivity = \frac{TP}{TP + FP}. \tag{3}$$

The specificity is defined as the proportion of samples with a negative condition that are correctly classified. It can be calculated with Equation 4, where TN represents the true negatives and FN represents the false negatives [4]:

$$specificity = \frac{TN}{TN + FN}. \tag{4}$$

Finally, the sensitivity and the specificity in conjunction represent the decision threshold of the ROC curve. The AUC can be calculated through trapezoidal integration, as shown in Equation 5:

$$AUC = \sum_{i} (1 - \beta_i \cdot \Delta\alpha) + \frac{1}{2}[\Delta(1 - \beta) \cdot \Delta\alpha], \tag{5}$$

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

3 Results

The results obtained from the performance of the classifiers are presented. Figure 2 shows the ROC curves calculated based on the performance of each classifier. The ROC curves (a) and (b), which correspond to RF and cTree, respectively, present very similar curves, besides being the ones that present the best behavior in conjunction with the curve (d), which corresponds to SVM. The curve (c), which corresponds to K-NN also presents an adequate behavior, although a little less than the previous curves; while the curve (e), which corresponds to Naïve Bayes, is reduced significantly compared to the others.

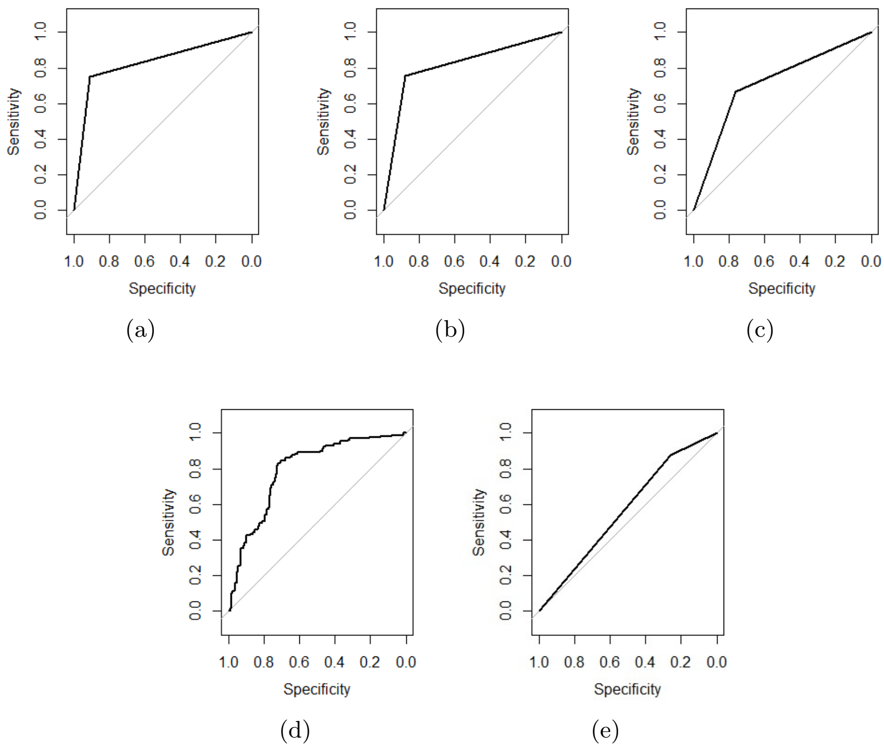

Fig. 2. ROC curves obtained using the (a) RF, (b) cTree (c) K-NN, (d) SVM and (e) Naive Bayes.

The quantitative results that justify the behavior of the ROC curves are presented in Table 1, where are shown the values obtained from the statistical metrics measured for each technique, related to the identification of patients with depression. As is possible to observe, RF presents the highest values for each metric, obtaining a sensitivity of 0.8148, specificity of 0.8158 and AUC of 0.8314. Naïve Bayes presents the lowest AUC with a value of 0.5677 and specificity with a value of 0.4606; however, its sensitivity presents a value of 0.7419, which is higher than the SVM sensitivity, since it obtained a value of 0.6364, being the lowest sensitivity.

Classifier	**Sensitivity**	**Specificity**	**AUC**
RF	0.8148	0.8158	0.8314
cTree	0.7383	0.8090	0.8183
K-NN	0.7614	0.6693	0.7153
SVM	0.6364	0.7200	0.7924
Naïve Bayes	0.7419	0.4606	0.5677

Table 1. Values obtained from the validation.

4 Discussion and Conclusions

In this work is presented the evaluation of five well-known classifiers. These classifiers are different approaches, probabilistic, decision based, deterministic, etc. allowing the comparison in terms of sensitivity and specificity, to overcome the depression episodes automatic detection using an activity signal. this evaluation allows to conclude several specific points described below:

– The ROC curve allows to conclude that Naive Bayes is the one with worst AUC, however, one possible way to improve the behavior of this classifier is to increase the number of observations.
– Decision based approaches acquire the best AUC, however is necessary a deep study of overfitting, given that is a common issue with this type of algorithms.
– Random Forest outperforms other approaches, including other type of decision based algorithms, nevertheless is one with more computational cost when number of trees are increased.
– Random Forest outperforms in terms of sensitivity and specificity other classifiers, however, in medical diagnostics commonly is more important to increase sensitivity to find true positive states. Therefore, K-NN is one possible solution at a lower computational cost, being the second one in terms of sensitivity.

Then, it is important to remark that despite the classifiers it is possible to develop an automatic detection of depressive episodes using the motor activity signal on one hour intervals records without other type of processing. The results obtained probe that independently, there is a particular behavior on patients with depression in the body movements that can be modeled by these type of classifiers.

Therefore, this work allows the development of efficient tools for automatic diagnosis of depressive episodes to help specialists in mental illness to recognize the health status of patients based on the motor activity recorded by commonly used smart bands.

References

1. Alcalá-Rmz, V., Zanella-Calzada, L.A., Galván-Tejada, C.E., García-Hernández, A., Cruz, M., Valladares-Salgado, A., Galván-Tejada, J.I., Gamboa-Rosales, H.: Identification of diabetic patients through clinical and para-clinical features in mexico: An approach using deep neural networks. International journal of environmental research and public health 16(3), 381 (2019)
2. Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)
3. Garcia-Ceja, E., Riegler, M., Jakobsen, P., rresen, J.T., 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 on Multimedia Systems Conference. MMSys'18, ACM, New York, NY, USA (2018), http://doi.acm.org/10.1145/3204949.3208125
4. Hanley, J.A., J, M.B.: The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology (1982)
5. Instituto Nacional De Psiquiatría Ramón De La Fuente Muñiz: ADEP Ayuda para Depresión (2019), https://ayudaparadepresion.org.mx/
6. Ryoo, J., Arunachalam, M., Khanna, R., Kandemir, M.T.: Efficient k nearest neighbor algorithm implementations for throughput-oriented architectures. In: 2018 19th International Symposium on Quality Electronic Design (ISQED). pp. 144–150. IEEE (2018)
7. Saito, T., Rehmsmeier, M.: Precrec: fast and accurate precision–recall and roc curve calculations in r. Bioinformatics 33(1), 145–147 (2017)
8. Ministerio de sanidad, s.s.e.i.: Guía de Práctica Clínica sobre el Manejo de la Depresión en el Adulto. Tech. rep., Galicia (2014), http://www.guiasalud.es/GPC/GPC_534_Depresion_Adulto_Avaliat_compl.pdf
9. Thanh Noi, P., Kappas, M.: Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. Sensors 18(1), 18 (2018)
10. Vekeman, F., Piña-Garza, J.E., Cheng, W.Y., Tuttle, E., Giguère-Duval, P., Oganisian, A., Damron, J., Duh, M.S., Shen, V., Saurer, T.B., et al.: Development of a classifier to identify patients with probable lennox–gastaut syndrome in health insurance claims databases via random forest methodology. Current medical research and opinion (just-accepted), 1–1 (2019)
11. WHO: WHO — Depression (2017), https://www.who.int/mental_health/
12. Yang, J.J., Lakte, P.T.A., Kim, S.H.: Use of conditional inference trees for evaluating the effect of reclaimed asphalt pavement content and binder grade on the dynamic modulus of asphalt concrete mixtures. International Journal of Pavement Research and Technology (2018)
13. Ye, K.: Key feature recognition algorithm of network intrusion signal based on neural network and support vector machine. Symmetry 11(3), 380 (2019)
14. Zanella-Calzada, L.A., Galván-Tejada, C.E., Chávez-Lamas, N.M., 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 9(1), 8 (2019)

15. Zhu, X.e.a.: Random forest based classification of alcohol dependence patients and healthy controls using resting state mri. Neuroscience letters 676, 27–33 (2018)