2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)
Bari, Italy, October 6-9, 2019

Classification and Characterization of Children and Adolescents with Depressive Symptomatology using Machine Learning*

Kelly Malaquias[1,2], Thiago Lima[1], Renata Santana[1], Felipe Salgado[1] Maycoln Teodoro[2], Cristiane Nobre[1]
[1]Pontifical Catholic University of Minas Gerais, [2]Federal University of Minas Gerais
{hanna.malaquias, thiagohenriqueslima, renata.cris.santana, felipe.salgado72}@gmail.com,
mlmteodoro@hotmail.com, nobre@pucminas.br

Abstract— **According to the World Health Organization (WHO), there are currently in the world more than 300 millions of people living with depression symptoms. Depression is a disorder that results from a complex interaction of biological, psychological, and social factors, and it is known for having difficulties in both diagnostics and prognostics. Machine Learning techniques are increasingly and often used to classify or characterize different profiles of diseases. This paper presents a study about major depressive disorder among Brazilian's children and adolescents by using decision trees classifiers, Support Vector Machine (SVM) and Neural Network. A discussion about the identified attributes is presented, including, for instance, the great relation of suicidal thoughts with elevated symptomatology. Beyond that, the value of the evaluation method F-Measure, the weighted harmonic mean of precision and recall, was above 88% for both classes, high and low symptomatology for depression; which reached values above 95% when used Multilayer Perceptron and SMO algorithms, they are based in Neural Networks and Support Vector Machine, respectively.**

I. Introduction

The Major Depressive Disorder, according to the World Health Organization (WHO) [1], is different from the usual humor oscillations and short duration feelings in answer to life's daily challenges. This disturb might become a serious health problem, mainly when it is long-lasting with moderate or strong intensity. It could be said that it is the main cause of health problems and inability in the whole world. More than 300 millions of people are currently living with depression, a rise of more than 18% between 2005 and 2015[1]. It is the cause of great suffering to who is affected and it interferes on their professional, student, or family development.

In the worst case, the depression can lead to suicide: about 800 thousands of people die by this every year. The suicide is the second main cause of death among young people between 15 and 29 years old [1]. The WHO itself alerts, although there are effective treatments for depression, less than half of the affected people on the world (in many countries, less than 10%) actually receive those treatments. The barriers for effective care include lack of resources, lack of health professionals and the social stigma linked-up to mental disorders. Beyond that, another obstacle to effective treatment is the imprecise assessment, which consequently causes disruption to other treatments of depression.

In the whole world, according to WHO, the depression is the main cause of illness and inabilities in adolescence. Some researches [2], [3], [4] indicate that half of all people who develop mental disorders show their first symptoms until 14 years old. In other words, taking care of children and adolescents mental health might actually avoid deaths and/or sufferings throughout life.

For the American Psychiatric Association (APA) [5] one in every six people (about 16,6%) will suffer from depression in a moment of their lives, which brings us to the impressive number of about one billion people in the whole planet. The depression may occur at any moment, but on average, it appears for the first time on the period between the final stage of adolescent and 25 years old, approximately.

Despite the existence of effective psychological and pharmacological treatments for depression [1], a correct diagnosis is fundamental for it to occurs. So, researches that contribute to the diagnosis and treatment of this disturb becomes essential especially when the subject is children and adolescents. In the computational area, works have been developed to contribute to the diagnosis and the study of depression. Among them, stand out those of Machine Learning, an area of artificial intelligence focused on the algorithms and techniques developed to extract data information [6]. The literature indicates satisfactory results coming from the combination of machine learning techniques with the identification of patterns that characterize diseases, specifically those related to depression and other mental disorders [7]. This is justified because the identification of these disorders is done through the observation of characteristics that have different levels of importance, which is often not assessed by questionnaires.

Given this context, this paper has, as the main goal, to predict if a child or adolescent has depression and discover what characteristics are determinant for the prediction. Like this, beyond having a high recall number, the method must explicit the pattern used for the predictions. To achieve this, database information of 377 Brazilian's children and adolescents aged between 10 and 16 years old was used. For the knowledge extraction about the depression in this database, build decision trees methods was used, such as J48 and SimpleCart. Beyond that, methods considered in the literature as efficient in a similar context, as Support Vector

We appreciate the financial support given by Foundation for Research Support of the State of Minas Gerais/Brazil (FAPEMIG), the Brazilian National Council for Scientific and Technological Development (CNPq), the Coordination for the Improvement of Higher Education Personnel (CAPES) and PUC Minas - Brazil.

Machine and Neural Networks was also used. Although those methods do not explicit the knowledge that was acquired, as wanted in this paper, they were used to evaluate the performance when compared to the J48 and SimpleCart method, which are considered simpler methods.

The remaining text is organized in the following way: in the second section the theoretical framework with the main concepts related to depression and the algorithms used in this paper are presented, followed by the works related to the addressed problem in the third section; the fourth section describes the materials and methods with a detailed description of the database used. In the fifth section the results and discussions are presented; and, finally, in the sixth section, the final considerations about this paper are revealed.

II. Theoretical Framework

Data Mining is a branch from computer science which had its beginning in the early' 80s [6] when the companies and organizations professionals started to worry about the great volume of stored data. Data Mining's concept is becoming more popular as a data discovery tool.

Data Mining is part of a bigger process denominated Knowledge Data Discovery (KDD). KDD is a set of techniques and methods for preparation and exploration of the data, interpretation of its results, and assimilation of the extracted knowledge [6]. In general, the data mining process may be seen as a set of tasks, such as classification, clustering, pattern searches, and knowledge evaluation.

To predict depression in children and adolescents, four classifiers were used: J48, an open-source Java implementation of the decision tree C4.5 algorithm [8], SimpleCart, also a Java version of the CART algorithm [9], Multilayer Perceptron and Support Vector Machine [10]. These algorithms are detailed in the following sections.

A. *Decision Tree*

A decision tree might be defined as a representation of a function based on a set of attributes inputted, which returns a unique outputted value, a decision Trees, besides having a simple representation, are also an effective classifier method. A decision tree makes a recursive subdivision of a data set until it gets a subgroup where there is only one class [8].

B. *Artificial Neural Network*

Artificial Neural Network (ANN) is a mathematical model inspired by the human brain, which reflects its "learning" and "deduction" skills. ANN is formed by a set of neurons organized in layers, where each link between the neurons may transmit a signal processed by a receptor neuron and then transmitted to the next neurons connected to it [11].

The connections between neurons have weights which vary based on the learning process [12]. In this work, the structure type Multilayer Perceptron was used, with backpropagation algorithm.

C. *Support Vector Machine*

The classification method known as Support Vector Machine [10] is based on the Statistical Learning Theory. This technique looks for the construction of a hyperplane as a decision surface, in a way to maximize the separation between classes of a problem [13].

Multiple hyperplanes can be constructed to separate the classes. Each hyperplane defines a separation margin between classes, where the instances situated in the limit are called support vectors, and the mid of the margin is our optimal separation hyperplane. It is expected that the hyperplane with larger margin may be more precise in classifying future data than the hyperplane with narrower margins.

III. Related Works

The precocious diagnosis of mental disorders is essential for life's quality promotion and its relation to machine learning techniques has been studied more and more. Specifically, in the work [14], the authors compared the performance of eight diagnostic techniques for five common mental issues in children, anxiety, Attention-deficit/hyperactivity disorder (ADHD), and Pervasive Developmental Disorders (PDD). The authors used a base with sixty instances and twenty-five attributes. The tests were made using WEKA [15] as a tool, and the classifiers with the best results were: Multilayer Perceptron, Multiclass Classifier, and LAD Tree. The data set used was small, and the authors point that, in the future, the research could be applied to a large data set with the goal to improve the precision.

In [16] the authors performed a comparison of three supervised learning models: Support Vector Machine (SVM), Logistic Regression (LR), and Naive Bayes (NB) to identify people with Major Depressive Disorder (MDD), a generalization of depression, in a way to classify people that had or not this disorder. The classifiers were applied to a set of clinical data obtained by surveys, combined with electroencephalogram of 64 individuals, being 34 of them with the disease and 30 of them not. In relation to the model performance, the obtained results were above 90% for piratically every metric evaluated, with recall up to 100% when Naive Bayes classifier was used.

Another important work that analyzed the depression was [17]. The authors analyzed depression from a database extracted from an American survey, the Behavioral Risk Factor Surveillance System (BRFSS). After a process of characteristics selection, the authors worked with 73 attributes from the 450 available. For classification, the authors used the following algorithms: C4.5, Random Forest, Multilayer Perceptron, and Support Vector Machine.

In [18], the authors compared the efficiency of SVM, MLP, Hierarchical Fuzzy Signature (HFS), and Gaussian Mixture Models (GMM) algorithms in the diagnosis of patients with depression. An important point of this work was the creation of a hybrid model, made with SVM and GMM, which found the best metrics among the others classifiers.

In [19], using SVM, the authors classified patients with and without depression. The authors used a small database with only 14 instances of people with depression. Primarily, only clinical and demographic attributes were used to create the model. In this phase, they obtained 68% precision in the classification. Although analyzing the incorrectly classified instances from the "no depression" class it was observed that those had opposite characteristics, as the use of psychotropic medicines for the disease treatment, besides, multiples patients were attending to mental and psychological therapies. Some important points were discovered by the authors, like patients with and without depression did not have differences in their educational level, and seven from the fourteen patients had one or more characteristics of anxiety disorder.

From these related works, this work proposes a different approach, first because it is related to a specific study about depression among children and adolescents, and few works are dedicated strictly to this age group. Secondly, it characterizes depressive symptomatology in children and adolescents and evaluation of decision trees, Support Vector Machine, and Neural Network classifiers, other researches investigated these three classifiers, but not jointly, and even less, aiming at these specific group.

IV. Materials and Methods

A. Database Description

The database used in this work was obtained in a partnership with a Brazilian federal university. The database contains information of local children and adolescents aged between 10 and 16 years old, with 158 male instances and 219 female, along 75 attributes[1], and 377 instances with varying depressive symptoms indications. If the instance presents many symptoms, it is labeled as "high symptomatology" and otherwise as "low symptomatology".

Besides age and sex gender, this base is constituted of attributes like schooling, who lives with them, time spent with parents, use of medications, scores obtained with the Manual for the Youth Self-Report (YSR) [20], and 27 questions from Children Depression Inventory (CDI) [21]. Also, it has questions about, for instance, demographical characterization, who lives with them, if they had any psychiatric treatment, and parents education level. Other questions considered important by the mental health community [22] were also included, mainly factors like anxiety, aggressiveness, social and conduct problems.

B. Preprocessing

In order to obtain a more consistent and unbiased model, simple preprocessing strategies were adopted before effectively applying the classification algorithms. All data preprocessing was made using Python[2], using Jupyter notebook environment. Sequentially, the phases of database processing were:

- *Standardization of missing data*: Several representations were used and it was decided to transform all the missing data into a commonly used WEKA symbol, used in this work, the indication "?"
- *Removal of attributes with few information*: Two attributes were free text, indicating the remedies administered by the mother and father of the analyzed individual. In addition to open field problems such as lack of standardization and writing errors, no response was observed in 90% and 96% of the instances, respectively, which were then removed from the database.
- *Treating the data inconsistency*: In five cases, the values were filled in incorrectly, not fitting the context nor predefined response options. In two of these cases, it was possible to analyze the context and adjust the response values. In the rest, the fields were represented as missing.
- *Codification of numerical to nominal*: Every question that had its nominal codification, but was codified in a numerical way, such as the gender, codified as 1 or 2 for male and female, were converted to nominal.
- *Transforming attributes in binary*: In many cases, the variability of options in certain attributes was noticed being too big. In most of these cases, the attributes were nominal and non-ordinal, which were then transformed into binary classes.
- *Attribute discretization*: In order to adapt the obtained answers to the computational models used, a discretization was made, creating tracks, of six attributes that indicated the contact time per day of the child or adolescent with the parents.
- *Identification and manipulation of the classification attribute (class)*: Initially the database did not have any explicit attribute to patients classification. By using the CDI, an obtained attribute is the "CDI Soma" (CDI Sum), equivalent to the score of the interviewed individual on this inventory. To create the classes, the attribute was manipulated according to the following literature suggestions. Since there was an instance that did not have information about the CDI, that instance was removed from the database.

  Figure 1 shows a chart of the the interviewed individuals, considering the attribute "CDI Soma".

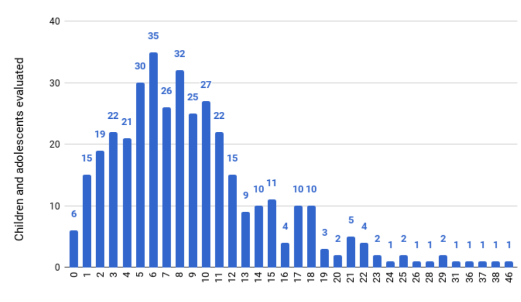

Fig. 1. CDI Sum Distribution

[1] A full description of all attributes can be found at https://goo.gl/z2wUKg

[2] Programming language of high level, interpreted and by script, in the third version of Python.

With the chart, the smaller and bigger values obtained in this database can be analyzed, respectively, are 0 to 46. The score of the CDI does not determine the existence or the absence of depression, and in fact, indicates characteristics of depression that support the evaluation made by the professional. Still, there is no unanimity about the cut-off value that determines the subdivision, since these values may vary according to the sample. The Kovacs recommendation [21] is the use of percentile 85 to indicate high symptomatology.
Figure 2 presents the number of instances which the score is bigger than the percentiles values. This way, in the considered sample, 63 individuals were classified with high symptomatology.

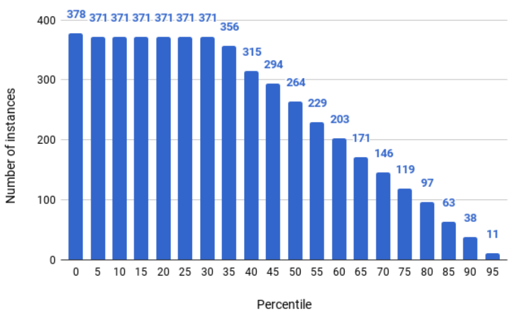

Fig. 2. Number of individuals with CDI Sum bigger than the percentile

The ratio of symptomatology classes in the database resembles real conditions, in which the number of people with an indication of the disease, even being significant, is way less than the healthy population, becoming an unbalanced quantity of people presented in these two groups.

- *Database balancing*: After the previous phases, the database had 314 and 63 individuals belonging to the classes "low" and "high", respectively. With the goal to avoid the algorithms to learn best the majoritarian class, and with that the accuracy of the minority class could be impaired, the balancing of the classes was decided. In order to characterize a profile, the use of synthetic data is not interesting, so we used a technique called Random Undersampling [23] and we made a random cut from the sample of the "low" class taking off 63 individuals, matching the number of the "high" class.
- *Train and test data set separation*: For a better validation of the generated models, it was decided to divides the database in two groups, one for training, used on the creation and evaluation of the model, and another one to test, with unevaluated data by the previous model. The division can be seen in Table I.

Considering the balanced database, 10 instances of each class were separated to test and the other 53, also of each class, were used in the construction of the model (corresponding to the Test 1 in Table I). Besides that, the samples of the "low" class were removed on the balancing process, it was also tested in the model (corresponding to Test 2 in Table I).

TABLE I

NUMBER OF INSTANCES BY CLASSES

Class	Instances before balancing	Instances after balancing	Model Building	Model Test 1	Model Test 2
HIGH	63	63	53	10	-
LOW	314	63	53	10	251
Sum	377	126	106	20	251

C. Methods

The algorithms used in this work were selected based on its wide acceptance in the literature and suitability to the proposed objectives. Among the algorithms of decision trees, we selected the J48, an implementation of the decision tree algorithm C4.5 [8] on the WEKA's environment [15]. Another method selected was SimpleCart, WEKA's version of Cart algorithm, a Classification, and Regression Tree [9], that makes and recursive binary partition on the tree building.

For the usage of the Multilayer Perceptron algorithm, the number of artificial neurons and hidden layers were adjusted, defining two hidden layers, the first one with 3 neurons and the second one with 5. In this classification algorithm, backpropagation was used, a vector that maps the mistakes which its purpose is to decrease the average of those mistakes [24].

For the SVM classifier training, it was necessary to adjust[3,4] three main parameters, the kernel function, the $gamma$, which is the width of the Gaussian and the parameter that narrows the margin used for the hyperplane separation, represented by C. These parameters are important to obtain good results in its execution because they are directly linked to the algorithm's training and classification stage. The kernel ($kernel = rbf$) for the SVM's, the $gamma$ ($gamma = 0.001$) and the value 12 for the parameter C were used.

The Multilayer Perceptron and SVM classifiers are considered "black box" methods because they generate models that are harder to interpret [26], which may become an obstacle to their practical application, especially when the purpose is the explanation of the classification. However, these methods present superior generalization performance compared to other classification techniques in several areas of application [27].

V. RESULTS AND DISCUSSION

In this section the results obtained from the preprocessed database are presented. Table II shows average values (in percentage) ordered by class of the evaluation metrics of the model's performance[5]. In order to define the training and validation set, the cross-validation method with 10 folds was used.

[3]To adjust the parameters from the Multilayer Perceptron and SVM, the algorithm Grid Search [25] was used. This algorithm does an exhaustive search on the parameters using specified values as parameters to the execution of the algorithm.

[4]Grid Search available at https://www.csie.ntu.edu.tw/~cjlin/libsvm/.

[5]$Precision = \frac{TP}{TP+FP}$; $Recall = \frac{TP}{TP+FN}$; $Fmeasure = \frac{2 \times Pr \times R}{Pr+R}$

TABLE II

EVALUATION METRICS OF THE MODELS, IN PERCENTAGE

Algorithm	Class	Precision	Recall	F-Measure	ROC Area
J48	HIGH	83.3	84.9	84.1	80.8
	LOW	84.6	83.0	83.8	80.9
SimpleCart	HIGH	89.1	77.4	82.8	79.7
	LOW	80.0	90.6	85.0	80.7
Multilayer Perceptron	HIGH	88.0	83.0	85.4	94.3
	LOW	83.9	88.7	86.2	94.3
SVM	HIGH	90.2	86.8	88.5	88.7
	LOW	87.3	90.6	88.9	88.7

Considering the results, the SVM got a better precision on the classes general classification. However, the methods, in general, had very similar behavior. For the diagnosis, discriminating well the individuals with and without the disease is more relevant than accuracy, and therefore, we evaluated the ROC Area, considered highly relevant for the medical community. Multilayer Perceptron yielded the best result of 94.3%, followed by SVM, J48, and Simple Cart, with 88.7%, 80.5% e 80.2%, respectively.

When it comes to the interpretable methods, those that return the rules used in the classification, J48 and Simple Cart obtained similar performances. Furthermore, as can be seen in Figures 3 and 4, the trees used different attributes in different rules.

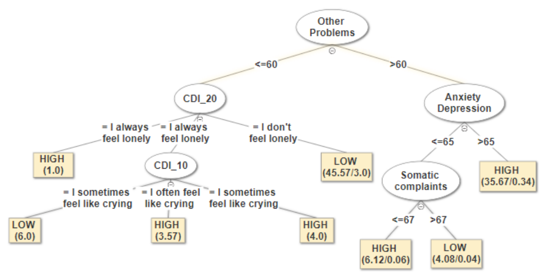

Fig. 3. Decision tree obtained with the J48 algorithm

The generated model by J48 used five distinct attributes, being them two CDI questions, signaling that, in the analyzed set, the indicators of loneliness and sadness related to crying were more relevant to determine the symptomatology. The other attributes used are scores from the YSR.

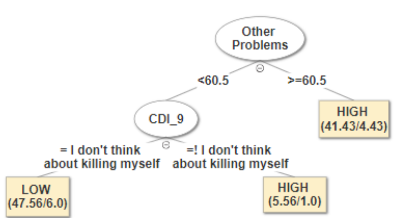

Fig. 4. Decision tree obtained by the Simple Cart algorithm

In the tree generated by the Simple Cart algorithm, only two attributes were used to classify the individuals, being one of them the score of "other problems" obtained with the YSR, which, just like with J48 it was the first attribute, the tree root, used to tell the different symptomatologies. However, in this case, the value of distinction varies from the previously used tree. The other attribute considered by this model relates directly to the existence of suicidal ideation and high symptomatology.

Table III shows the results obtained for the test base, that is, one that was not used in the model construction (relative to test 1 of Table I, where 10 instances of each class were reserved to test the model).

TABLE III

MODELS' EVALUATION ON THE TEST DATASET, VALUES IN PERCENTAGE

Algorithm	Class	Precision	Recall	F-Measure	ROC Area
J48	HIGH	83.3	100.0	90.9	83.3
	LOW	100.0	80.0	88.9	90.0
SimpleCart	HIGH	83.3	100.0	90.9	90.0
	LOW	100.0	80.0	88.9	90.0
Multilayer Perceptron	HIGH	100.0	90.0	94.7	98.0
	LOW	90.9	100.0	95.2	98.0
SVM	HIGH	100.0	100.0	100.0	100.0
	LOW	100.0	100.0	100.0	100.0

As it can be observed, SVM and MLP's ROC Area are high and very close to each other, with 100% e 98% respectively, and, therefore, exhibiting great results on the test set. Observing the recall, it can be perceived that every model, except the neural network, correctly classified every instance that was from the "high" class. This is very relevant because a good diagnosis for the high symptomatology may suggest the proper treatment faster. This means that the health professional, family, and educators can intervene in more precocious ways on the patients' treatment.

About the precision, a rate of 83.3% was observed, using the interpretable models. This is equivalent to have two instances from the "low" class, classified as "high". When isolated, these false positives are analyzed as missed algorithms in the same instances. Table IV shows the values of the attributes used by the interpretable models of those individuals.

TABLE IV

INDIVIDUALS WRONGLY CLASSIFIED

Attribute	Individual 1	Individual 2
CDI	3	6
CDI 9	I don't think about killing myself	
CDI 10	I sometimes feel like crying	
CDI 20	I don't feel lonely	
Other problems	63	67
Anxiety and depression	62	65
Somatic complaints	52	66

As we can analyze with the Table IV, the determinant factor for the instances to be classified incorrectly was the value "other problems", since that, in relation to the CDI questions, would be classified as "low". This divergence brings us to question the relation of the two inventories and their possible disagreements and interferences on the model. Lastly, we also evaluated the model quality using the instances of the "low" class removed in the balancing phase. In other words, we tested the models with the 251 instances that were removed on the class balancing (corresponds to test 2 presented in Table I). Since there are no instances of the "high" class to be evaluated, in this case, only the recall of the model was analyzed to classify the "low" class. The obtained results of recall with J48, Simple Cart, Multilayer Perceptron, and SVM, were, respectively, 76.9%, 75.3%, 99.0% and 99.0%.

With those results, the model obtained with the neural network was the best method to classify these instances, as observed. In the other hand, the interpretable methods obtained have decayed in its performance.

These results stand out the importance of evaluating different methods with the same purpose. The J48 and Simple Cart algorithm have the advantage of being easily interpretable, but the neural network and SVM algorithm, in this case, had way better values way. However, they have the disadvantage of not being easily interpreted, needing other algorithms to extract knowledge acquired with them.

VI. Final Considerations and Future Works

The goal of this work, as presented, was to study the disturb of depression in Brazilian's children and adolescents, using Decision Tree, SVM and Neural Network as classifiers.

Although there are many works in the computational area using classification algorithms to the tasks related to depression in general, few are found when it is referent to this disease in children and adolescents. When this group of people is the subject, it is hard to make other kinds of more complex evaluation that let new characteristics and data be used.

Despite the limitations found, when the metrics presented on the obtained result are observed and put in perspective that other authors and works consider those metrics above 75% as significant and satisfactory, [19] it can be said that this was a relevant work.

Every model used the "symptomatology" attribute as the final classification, this one was calculated based on the data obtained with the inventory CDI. As a future work, to apply the models in a database that uses other methods of evaluation and classification of the children and adolescents is suggested, besides to discuss another cut point that can be used for class definition, also trying to improve the algorithms looking to enhance the classifications individually. The separation of the YSR's attributes related to CDI is another possibility to analyze the behavior of each evaluation. It is also suggested, to analyze the results of SVM classifier through the separation hyperplane constructed by support vectors and should analyze which components of the feature vector affect the classification performance.

References

[1] WHO, "Depression," World Health Organization, 2017, available: http://www.who.int/mediacentre/factsheets/fs369/en/.

[2] S.-H. Wang, Y. Ding, W. Zhao, Y.-H. Huang, R. Perkins, W. Zou, and J. J. Chen, "Text mining for identifying topics in the literatures about adolescent substance use and depression," *BMC public health*, vol. 16, no. 1, p. 279, 2016.

[3] H. C. Comijs, E. van Exel, R. C. van der Mast, A. Paauw, R. O. Voshaar, and M. L. Stek, "Childhood abuse in late-life depression," *Journal of affective disorders*, vol. 147, no. 1-3, pp. 241–246, 2013.

[4] Y. Sunmoo, T. Basirah *et al.*, "Using a data mining approach to discover behavior correlates of chronic disease: a case study of depression," *Studies in health technology and informatics*, vol. 201, p. 71, 2014.

[5] APA, "Depression," American Psychiatric Association, 2017, available: https://www.psychiatry.org/.

[6] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[7] G. Gkotsis, A. Oellrich, S. Velupillai, M. Liakata, T. J. Hubbard, R. J. Dobson, and R. Dutta, "Characterisation of mental health conditions in social media using informed deep learning," *Scientific reports*, vol. 7, p. 45141, 2017.

[8] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[9] L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and regression trees (chapman y hall, eds.)," *Monterey, CA, EE. UU.: Wadsworth International Group*, 1984.

[10] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep 1995.

[11] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research," *Journal of pharmaceutical and biomedical analysis*, vol. 22, no. 5, pp. 717–727, 2000.

[12] F. Amato, A. López, E. M. Peña-Méndez, P. Vaňhara, A. Hampl, and J. Havel, "Artificial neural networks in medical diagnosis," 2013.

[13] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[14] M. Sumathi and B. Poorna, "Prediction of mental health problems among children using machine learning techniques," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 1, pp. 552–557, 2016.

[15] S. R. Garner, "Weka: The waikato environment for knowledge analysis," in *In Proc. of the New Zealand Computer Science Research Students Conference*, 1995, pp. 57–64.

[16] W. Mumtaz, S. S. A. Ali, M. A. M. Yasin, and A. S. Malik, "A machine learning framework involving eeg-based functional connectivity to diagnose major depressive disorder (mdd)," *Medical & biological engineering & computing*, vol. 56, no. 2, pp. 233–246, 2018.

[17] Y. Sunmoo, T. Basirah *et al.*, "Using a data mining approach to discover behavior correlates of chronic disease: a case study of depression," *Studies in health technology and informatics*, vol. 201, p. 71, 2014.

[18] S. Alghowinem, R. Goecke, M. Wagner, J. Epps, T. Gedeon, M. Breakspear, and G. Parker, "A comparative study of different classifiers for detecting depression from spontaneous speech," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 8022–8026.

[19] M. D. Sacchet, G. Prasad, L. C. Foland-Ross, P. M. Thompson, and I. H. Gotlib, "Support vector machine classification of major depressive disorder using diffusion-weighted neuroimaging and graph theory," *Frontiers in psychiatry*, vol. 6, p. 21, 2015.

[20] T. M. Achenbach, *Manual for the Youth Self-Report and 1991 Profile*. Dept. of Psychiatry, University of Vermont, 1991.

[21] M. Kovacs, *Children's Depression Inventory (CDI): technical manual update*. Multi-Health Systems, 1992.

[22] A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[23] G. E. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD explorations newsletter*, vol. 6, no. 1, pp. 20–29, 2004.

[24] A. Serretti and E. Smeraldi, "Neural network analysis in pharmacogenetics of mood disorders," *BMC Medical Genetics*, vol. 5, no. 1, p. 27, 2004.

[25] C.-C. Chang and C.-J. Lin, "Libsvm: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[26] G. Fung, S. Sandilya, and R. B. Rao, "Rule extraction from linear support vector machines," in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. New York, NY, USA: ACM, 2005, pp. 32–40.

[27] H. Byun and S.-W. Lee, "Applications of support vector machines for pattern recognition: A survey," in *Pattern recognition with support vector machines*. Berlin, Heidelberg: Springer, 2002, pp. 213–236.