Webology (ISSN: 1735-188X)
Volume 19, Number 6, 2022

Benchmarking Different Classification Techniques To Identify Depression Patterns In An Audio And Text Dataset

Sonia Jaramillo-Valbuena[1] , Cristian-Giovanny Sánchez-Pineda[2] , Sergio-Augusto Cardona-Torres[3]

[1]PhD in Engineering. Computer Engineering Dept., Universidad del Quindío, UQ. Armenia, (Colombia)

[2]Master's student in engineering. Universidad del Quindío, UQ. Armenia, (Colombia),

[3]PhD in Engineering. Computer Engineering Dept., Universidad del Quindío, UQ. Armenia, (Colombia)

Abstract

Depression is a health disorder that affects the population, regardless of their age or social status. The World Health Organization (WHO), considers it the greatest generator of incapacity worldwide. Depression increases the possibility of suicide, being the latter, the second trigger of death in people between fifteen and twenty-nine years of age. It negatively impacts different levels of the person: family, work and school and affects its ability to face daily life, aggravated preexisting medical conditions. Young or re-tired person and pregnant or postpartum women, are the groups most vulnerable to suffering from depressive disorder. In this paper, we apply two different classification techniques, namely: Bidirectional Encoder Representations from Transformers (BERT) and Support Vector Machines (SVM) in order to identify depression patterns in the Distress Analysis Interview Corpus DAIC-WOZ.

We compare the models obtained and determine their robustness, using performance metrics. The results show that the approach BERT has good performance over the SVM model, reaching an accuracy of almost 90%.

Keywords: BERT, Deep Learning, Encoder, Word embeddings.

I. INTRODUCTION

Depression is related to brain plasticity or neuroplasticity. Brain it matures according to the environment, it must adapt and evolve, based on lived experiences and behaviors. This occurs from the time you are a baby, the stimuli received are determinant to control aspects related to character. Several studies carried out in people with this disorder suggest reduced neuroplasticity [1] [2]. Aspects such as family problems, sudden changes in life, consumption of medicines, love breakups, feelings of loss of control, consumption of hallucinogens or alcoholic beverages can trigger the appearance of depression [3].

Various studies indicate that there is a genetic susceptibility to suffering from this condition [4]. In addition, that there may be a coexistence of depression with other diseases, such as, such as hypertension, thrombosis, irritable bowel syndrome, diabetes mellitus, Parkinson's, Alzheimer's, heart problems, anxiety, eating disorders, hormonal disorders and complications in cancer treatment [5] [6] [7] [8] [9].

According to the WHO, 300 million individuals suffer from depression. This mental disorder increases the risk of suicide up to 12 times. It is estimated that per annum 800,000 die by self-destruction [10]. Some of the most common symptoms are self-hatred, guilt, irritable mood, trouble falling asleep and appetite problems [11].

The use of machine learning to identify depression patterns from structured and unstructured information is known in computing as Automatic Depression Detection (ADD). This an active research topic, whose main goal is the generation of models using predictive techniques. In this regard, there are several studies where they analyze speech, text, video signals or a mix of them, for the generation of a global classifier [12] [13] [14] [15].

In this paper we apply different classification techniques to identify depression patterns in an audio and text dataset. We use the Crisp-DM methodology, which consists of the stages of identifying the business, identifying data mining goals, understanding and preprocessing data, application of modeling techniques, evaluation and deployment of the results. We use DAIC-WOZ dataset [16] and the techniques called Bidirectional Encoder Representations from Transformers (BERT) and Support Vector Machines (SVM).

The paper is organized as follows. Section 2 presents preliminaries on ADD. Section 3 describes the dataset and methods we use. Section 4 shows the relevant experiments details and selection of models. The last section we provide a summary of the findings in this work and lines of future work.

II. RELATED WORK

There are several approaches for Automatic Depression Detection. In [13] they propose a CNN-Based Speech Recognition system. The dataset they use is AVEC-2016 [17]. To obtain analytical models, the system performs a preprocessing stage, which consists of representing speech files as log-spectrograms. For the classification task they build a One-Dimensional Convolutional Neural Networks (CNN) architecture based on Ensemble. They

assess 3 different CNN-based models. Method 1 performs predictions at sample level of the M 1d-CNN and compute the average the probabilities to get the final label. Method 2, performs predictions at the level of all M classifiers. They calculate the most frequent value of number log-spectrograms to obtain the final prediction. Method 3 get the predictions at level of each of the M machines, they compute the mode of the M predictions. The model is evaluated using the F1-score metric. This metric is calculated for each class: depressive or non-depressive. Method 1 shows better performance than the other two. The results show that F1-score increases with the quantity of classifiers in the ensemble.

In [11], the authors present a system for detecting depression automatically using Support Vector Machine (SVM) and neural networks. The features are extracted from (DAIC-WOZ 2017 [16]) database. After data comprehension and preparation stage, they use Gaussian Mixture Model (GMM) clustering [18] in resultant vectors for producing bag of words and also, Fisher vector. GMM can identify oblong clusters and performs soft classification Fisher vector is a Fisher Kernel [19] (FK)-based aggregation technique. The system considers 2-D facial landmarks and head features (Head Features, Head motion and facial expressions distance, Emotions, AUs, Gaze and Pose and Blink Rate). The models were evaluated using a validation data. They apply RMSE, MAE, kernel, cost and gamma, as metrics for the evaluation by features (Audio, Text, Head, Cyclic and fisher). In Neural Networks, they use Adam optimizer and SGD. Adam optimizer shows faster and efficient.

In the paper presented in [20] apply CNN to generate a spectrograms-based ADD system. They use DAIC-Woz dataset 2016 [16]. First of all, the system performs pre-processing and feature selection stages. It removes stage silence regions, use a frames windowing method and get features (prosodic, MFCC, LPCC, PLP, spectral, temporal). For experimentation, they apply CNN and perform a performance comparison with Hidden Markov Models, SVM and Gaussian Mixture Models. The system uses the PHQ-8 scores [21] for classifying, according to the severity of depression. The evaluation is done through cross validation and they obtain accuracy, recall, precision and F-measure metrics. The performance of the proposed system is superior to the other techniques.

The work of [22] analyses gender bias in DAIC-WOZ dataset [16] and its impact in ADD' models using audio features. They apply data redistribution and raw audio features to reduce skewness. The bias leads to overfitting or a features' unfair penalty. The system generates a DepAudioNet-base new model of [23]. They use audio-only processing and deep learning techniques. The results show better performance compared to spectrogram-based model.

Finally, in [24] proposes a speech feature fusion system based on the higher-order spectral analysis (HOSA) of bi-spectral features (BSFs) and non-linear bi-coherent features (BCFs). The system processes DAIC-WOZ dataset [16]. They use higher-order spectral analysis and their blend' features with Support Vector Machine, k-nearest Neighbor classification and convolutional neural network to obtain models. In the assess stage they

uses accuracy. The results show that convolutional neural network has better performance, followed by the k-nearest neighbor classification and SVM, respectively.

III. DATASET AND METHODS

Dataset and implementation details

In this study, we use the data from DAIC-WOZ Database [16]. It is a real-world dataset provided by University of Southern California, that has clinical interviews in different formats: audio, video and answers of questionnaires. DAIC-WOZ supports the diagnostic of depression, post-traumatic stress and anxiety. It has 189 clinical interactions. There is training, development and testing sets. The first has 107 samples, 77 Non-depressed and 30 depressed. The development set, 35 samples of which 23 are Non-depressed and 12 depressed. Finally, the testing dataset has 47 samples. For each sample, there are audio, features, features3D, pose and transcripts. Ellie, virtual pollster, is who conducts interviews between 7 and 33 minutes. The dataset takes into account levels of depression between 0 and 24, according to Patient Health Questionnaire PHQ-8 [21]. The DAIC-WOZ considers that there is depression, when the PHQ-8 score is greater than 10. From this premise, they construct a binomial class label: 0 (non-depressed) and 1 (depressed).

Regarding data understanding and data preparation, we eliminate the interviewer's sentences (leaving only the patient's) and stop words, and then, associate each interview with its corresponding class label (non-depressed or depressed). This process is repeated for each interview. During experimentation, we use different vectorizers, namely: Count Vectorizer, TF-IDF and BERT. They are in Scikit-learn, a simple and efficient library for predictive data analysis written in Python [25].

Interview classification with Bidirectional Encoder Representations from Transformers (BERT)

BERT is an ANNs-based technique for NLP pre-training. It is an unsupervised pre-trained language representation, which is bidirectional. BERT builds contextual representations. It means, that considers the context for each occurrence of a word (instead of analyzing word by word, like word2vec and GloVe does) and learns relationships between them [26]. Figure 1 shows BERT’s neural network Architecture. E1 is the vector representation of a word, T1 is the final output and the intermediate representations of the same word are called Trm. Different intermediate representations of a word have the same size.

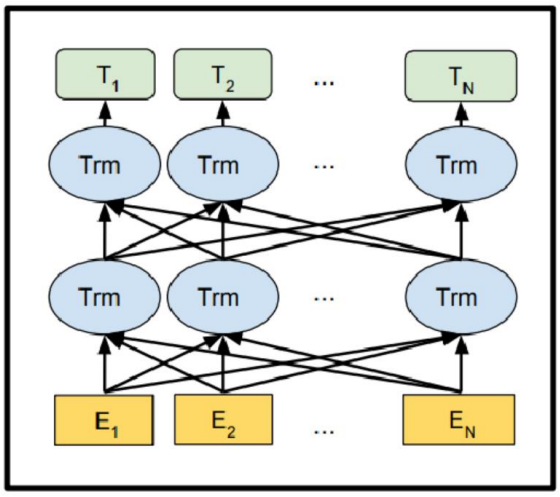

Figure 1. BERT's Neural Network Architecture, Source: [26] [27]

We use BERT to generate a binomial classification model. To do this, the system runs a script to get a list of lists (containing for each interview tokenized text, the class label and the length of the interview) and sorts by length of each interview. Data preparation considers the elimination of short sentences because they do not have a lot of meaning.

BERT text embeddings need as input the tokenized interviews. We use BERT tokenizer. Tokenization means partitioning the text into units, called tokens. Regarding the modeling stage, the batch size is 16 (after processing 16 interviews, the weights of the neural network will be updated) and we set 3 convolutional neural network layers with the kernel of 2, 3, and 4, respectively. To prevent overfitting, we use dropout regularization. It is a simple and powerful technique that randomly ignores neurons during training [28]. The first densely connected neural network is powered by the output obtained by concatenating the 3 layers of the convolutional neural network. We use a second densely connected neural network to predict the class label. The system uses the hyper parameters EMB_DIM = 200, NB_FILTERS = 100, FFN_UNITS = 256, NB_CLASSES = 2, DROPUT_RATE = 0.2 and NB_EPOCHS = 5.

Interview classification with Support Vector Machines (SVM)

The Support Vector Machine (Svm) is a supervised technique that optimally separates classes by a hyperplane or set of hyperplanes in a high-dimensional space. Support vectors are understood to be the points that define the maximum separation margin of the hyperplane that separates the classes. This technique, belonging to the family of linear classifiers, is used for classification and regression. In the event that the classes are not linearly separable, there are more than two predictor variables or a problem of non-linear separation curves, it is

necesary to use kernels. A kernel function projects the information to a higher dimensional feature space. The SVM can make use of the C hyper-parameter to tuning the regularization, see Figure 2. C=1/α, where α is the hyperparameter used in Ridge, Lasso, and ElasticNet regularizations.

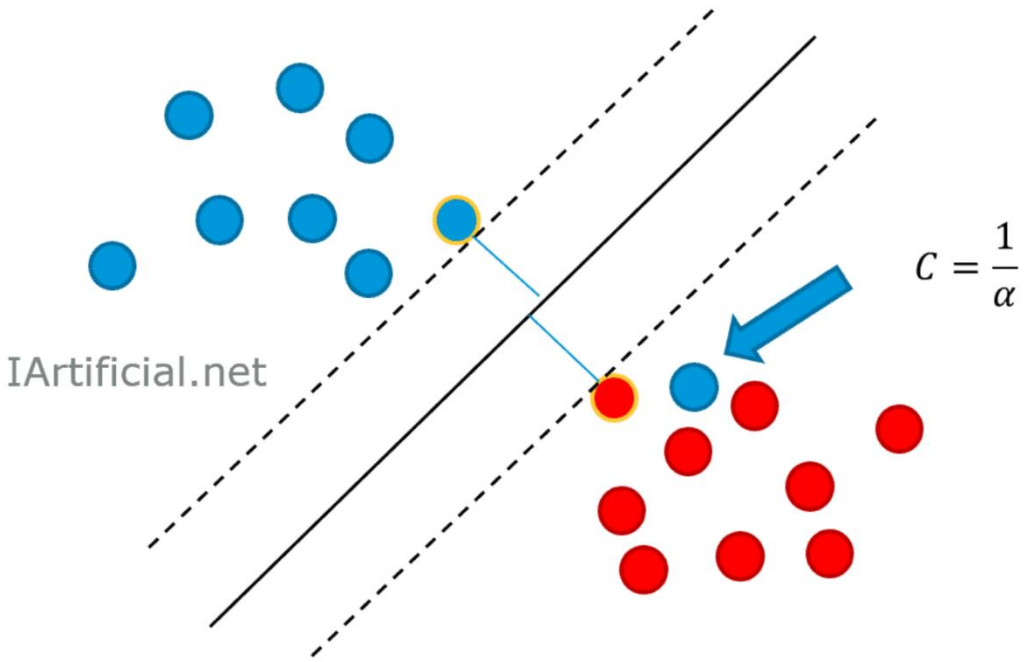

Figure 2. Regularization parameter in Support Vector Machine, Source: [29]

Prior to the vectorization stage, we use Lancester stemmer. Regarding the generation of the predictive model with SVM, we extracted the features using Count Vectorizer and TF-IDF. The SVM implementation, in sklearn, use Stochastic Gradient Descendent, a simple approach to converge on a solution. We set the hyper parameters α =0.001, loss=hinge (soft-margin, linear Support Vector Machine) and penalty=l2 (L2 norm penalty).

IV. RESULTS OF EXPERIMENTATION

In our evaluation, we get the quality of predictive models. For this, we use the accuracy metric and cross-entropy loss. Table 1 shows a comparative analysis of BERT and SVM, for Training and test datasets. The result of the experiment shows that BERT used as a tokenizer shows better feature extractions that models such as Count Vectorizer and TF-IDF combined. The model that used SVM, even though reach a 75% of accuracy, BERT hits almost a 90% in test set.

Table 1. Model Accuracy

	Bert	**SVM**
Training set	0.7500	0.70
test set	0.8929	0.75

V. CONCLUSIONS

The use of transformer for generation of text classification models is a topic of active research. Text classification is a NLP task, which is widely used in areas such as cybersecurity, education, customer service and healthcare. In this last case, we highlight the use of deep learning for identification of patterns referring to depression in structured and unstructured data.

In this paper we apply BERT and SVM to identify depression patterns in Distress Analysis Interview Corpus DAIC-WOZ. BERT is a Bidirectional Encoder Representations from Transformers and SVM is a classical supervised learning technique. DAIC-WOZ database includes audio transcriptions of interviews from patients with and without depression.

For the analysis, we apply text cleaning techniques, being this fundamental stage for the quality of the models obtained. The results of the experimentation show that Bert outperforms SVM. BERT features such as directionality and the construction of contextual representations have a high impact on the quality obtained.

As future work, we propose the analysis of audios from spectrograms and the use of other deep learning techniques. We want to use Latent Dirichlet Allocation to get topics from DAIC-WOZ database.

REFERENCES

[1] W. Liu, T. Ge, Y. Leng, Z. Pan, J. Fan, . Yang y . Cui, The Role of Neural Plasticity in Depression: From Hippocampus to Prefrontal Cortex, Neural Plast, 2017:6871089.

[2] Y.-B. Wang, N.-N. Song, Y.-Q. Ding y L. Zhang, Neural Plasticity and Depression Treatment, IBRO Neuroscience Reports, 2022.

[3] MinSalud Colombia, "Mental health bulletin Suicidal behavior - Subdirectorate of Noncommunicable Diseases," 2018. [Online]. Available:

https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/PP/ENT/boletin-conducta-suicida.pdf. [Accessed 26 11 2022].

[4] Zalar, Blatnik, Maver, Klemenc-Ketiš y Peterlin, Family History As an Important Factor for Stratifying Participants in Genetic Studies of Major Depression, Balkan J Med Genet, 2018 Jun; 21(1): 5–12. .

[5] Takeshima, U. Ishikawa, Kudoh, Umakoshi, Yoshizawa, Ito, Hosoya, Tsutsui, Ohta y Mishima, Prevalence of Asymptomatic Venous Thromboembolism in Depressive Inpatients, Neuropsychiatr Dis Treat, vol. 16: 579–587, 2020.

[6] S. Gairing, P. Galle, J. Schattenberg, K. Kostev y C. Labenz, Portal Vein Thrombosis Is Associated with an Increased Incidence of Depression and Anxiety Disorders, J Clin Med., vol. Dec; 10(23): 5689, 2021 .

[7] L. Parkin, A. Balkwill, S. Sweetland, G. Reeves, J. Green y V. Beral, Antidepressants, Depression, and Venous Thromboembolism Risk: Large Prospective Study of UK Women, HomeJournal of the American Heart Association, Vols. %1 de %26, No. 5, 2017.

[8] K. Davidson, B. S. Jonas, K. E. Dixon y J. H. Markovitz, Do Depression Symptoms Predict Early Hypertension Incidence in Young Adults in the CARDIA Study?, Archives of Internal Medicine, vol. 160, pp. 1495-1500, May 2000.

[9] . Charles, . A. Bardet, A. Larive y e. al, Characterization of Depressive Symptoms Trajectories After Breast Cancer Diagnosis in Women in France, JAMA Network Open, vol. 5(4):e225118, 2022.

[10] WHO, Depression, 2022. [En línea]. Available: https://www.who.int/news-room/fact-sheets/detail/depression. [Último acceso: 27 11 2022].

[11] S. Dham, A. Sharma y A. Dhall, Depression Scale Recognition from Audio, Visual and Text Analysis, arXiv, 2017.

[12] M. Li, W. Zhang, B. Hu, J. Kang, Y. Wang y S. Lu, Automatic Assessment of Depression and Anxiety through Encoding Pupil-Wave from HCI in VR Scenes, ACM Trans. Multimedia Comput. Commun. Appl., April 2022.

[13] A. Vázquez-Romero y . Gallardo-Antolín, Automatic Detection of Depression in Speech Using Ensemble Convolutional Neural Networks, Entropy (Basel), vol. 22(6):688, 2020.

[14] S. Sardari, B. Nakisa, M. N. Rastgoo y P. Eklund, Audio Based Depression Detection Using Convolutional Autoencoder, Expert Syst. Appl., vol. 189, March 2022.

[15] B. Ay, O. Yildirim, M. Talo, U. B. Baloglu, G. Aydin, S. D. Puthankattil y U. R. Acharya, Automated Depression Detection Using Deep Representation and Sequence Learning with EEG Signals, J. Med. Syst., vol. 43, p. 1–12, July 2019.

[16] University of Southern California , Extended DAIC Database, 2019, [En línea]. Available: https://dcapswoz.ict.usc.edu/. [Último acceso: 27 11 2022].

[17] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, D. Lalanne, M. Torres Torres, S. Scherer, G. Stratou, R. Cowie y M. Pantic, AVEC 2016: Depression, Mood, and Emotion Recognition Workshop and Challenge, de Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, New York, NY, USA, 2016.

[18] E. Patel y D. S. Kushwaha, Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model, Procedia Computer Science, vol. 171, pp. 158-167, 2020.

[19] P. Figuera y P. G. Bringas, A Non-Parametric Fisher Kernel, de Hybrid Artificial Intelligent Systems: 16th International Conference, HAIS 2021, Bilbao, Spain, September 22–24, 2021, Proceedings, Berlin, 2021.

[20] N. S. Srimadhur y S. Lalitha, An End-to-End Model for Detection and Assessment of Depression Levels using Speech, Procedia Computer Science, vol. 171, pp. 12-21, 2020.

[21] Y. Wu, B. Levis, K. Riehm, N. Saadat, A. Levis, M. Azar, D. Rice, J. Boruff, P. Cuijpers, S. Gilbody, J. Ioannidis, L. Kloda, D. McMillan, S. Patten, I. Shrier, R. Ziegelstein, D. Akena, B. Arroll y L. Ayalon, Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: A systematic review and individual participant data meta-analysis, Psychol Med, Vols. %1 de %250(8), 1368–1380, 2021.

[22] A. Bailey y M. .. Plumbley, Gender Bias in Depression Detection Using Audio Features, 2021 29th European Signal Processing Conference (EUSIPCO), pp. 596-600, 2021.

[23] X. Ma, H. Yang, Q. Chen, D. Huang y Y. Wang, DepAudioNet: An Efficient Deep Model for Audio Based Depression Classification, de Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, New York, NY, USA, 2016.

[24] X. Miao, Y. Li, M. Wen, Y. Liu, I. N. Julian y H. Guo, Fusing Features of Speech for Depression Classification Based on Higher-Order Spectral Analysis, Speech Commun., vol. 143, p. 46–56, September 2022.

[25] Scikit-learn development, scikit-learn, 2022.

[26] Jacob Devlin and Ming-Wei Chang, Research Scientists, Google AI Language, Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing, 2018.

[27] J. Devlin, M.-W. Chang, K. Lee y K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, CoRR, vol. abs/1810.04805, 2018.

[28] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever y R. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research, vol. 15, p. 1929–1958, 2014.

[29] IArtificial.net, Máquinas de Vectores de Soporte, 2022. [En línea]. Available: https://www.iartificial.net/maquinas-de-vectores-de-soporte-svm/. [Último acceso: 13 12 2022].

[30] E. C. Martin, Edificios fotovoltaicos conectados a la red electrica: caracterizacion y analisis, Universidad Politecnica de Madrid, 1998.

[31] H. Dibeklioğlu, Z. Hammal, Y. Yang y J. F. Cohn, Multimodal Detection of Depression in Clinical Interviews, de Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, New York, NY, USA, 2015.