WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 3, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n3.49
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Rural Depression Patients Judgment Model Based on Machine Learning

Mingjiu Yang

College of Computer & Information Science, Southwest University, Chongqing, 400715, China
swuyangming9@outlook.com

ABSTRACT

Depression is a serious mental disorder. According to data released by the World Health Organization last year, there are approximately 300 million people suffering from depression worldwide. Due to the backwardness of medical care in rural areas and the low level of understanding of depression among residents, mental illness affects individuals in rural areas more seriously than in urban areas. However, with the development of artificial intelligence technology, related techniques such as machine learning have made initial gains in the field of depression determination and treatment. In this paper, based on a survey of data from Busara Center in Kenya, various machine learning methods such as K-NN, Naive Bayes methods, Support Vector Machine Classifier and Random Forest Classifier were used and evaluated for the most accurate model. The results show that Random Forest Classifier has the highest accuracy, at 0.794.

KEYWORDS

Depression, Machine learning, Prediction

1. INTRODUCTION

Depression, also known as depressive disorder, is a mental illness with a high incidence rate, good cure effect, but low treatment participation and high recurrence rate. Depression is primarily characterized by low mood, loss of interest, and lack of energy, with some early symptoms such as slowed reactions, slowed thinking, and memory loss also present, although these can vary individually. According to data released by the World Health Organization (WHO) in 2023, an estimated 3.8% of the global population suffers from depression. This includes 5% of adults, with 4% of males and 6% of females experiencing depression. Additionally, the data shows that 5.7% of adults over the age of 60 have depression. The WHO estimates that approximately 280 million people worldwide are affected by this mental health condition [1]. The impact of mental illness on individuals is more severe in areas that are relatively economically disadvantaged. Several studies have shown that the prevalence of depression is significantly higher in rural than in urban areas [2-4]. At the same time, due to factors such as inconvenient transportation conditions and limited economic income, the diagnosis and treatment of mental illnesses impose a huge financial burden on patients. Therefore, measures such as the popularization of relevant knowledge and the construction of a medical insurance system are crucial. Thanks to these initiatives, the rate of access to medical services for people with mental problem in rural areas can be increased.

The SDS (Self-Rating Depression Scale) is one of the commonly used self-rating scales that consists of 20 questions to assess the level of depression in an individual. These self-rating scales typically require participants to answer a series of questions about mood, interests, self-evaluation, and quality of life. Based on the participant's answers, a score is given which can indicate whether or not he or

she is predisposed to depression. However, in rural areas, most residents are not aware of this approach. Compared to urban areas, rural areas have remote locations, lower levels of medical care, generally lower levels of education or awareness of people than urban residents, and lagging information. In contrast, urban residents know how to use the Internet better to easily determine whether they suffer from depression or have a tendency to develop depression, so there is less advanced medical data, such as brainwave data and facial features data, for rural residents. In addition, the application of artificial intelligence (AI) technology in depression has been increasing in recent years and has achieved initial results. Currently, machine learning (ML) is popular in the field of depression [5,6]. As the core technology of AI, ML realizes the prediction of depression onset, early identification and assisted diagnosis, as well as the construction of efficacy prediction models by combining individual behaviors, clinical information, physiological signals, and other data, so as to help clinicians better formulate the diagnosis and treatment plan for patients.

In this paper, we will use machine learning methods to analyze a general population in a rural area, and classify the behavioral indicators of the population into different dimensions. An attempt is made to build a model to classify and predict the general population. In addition, this paper will compare different machine learning models to predict depression in individuals and come up with the model with the highest prediction accuracy.

This paper will be divided into four parts, the second part analyzes the existing related studies, the third part will detail the dataset used in this paper, the modeling method and details, and the experimental process, and the fourth part summarizes and discusses the experimental results.

2. LITERATURE REVIEW

Xu, L analyzed the features such as frequency of occurrence, rate of change, and intensity of action units (AU) during interviews between patients and normal people using facial data recognition techniques [7]. He found significant differences between the two and used these features to classify them by SVM. The recognition rate of depression reached 73.48% in men and 68.43% in women. Liu and her colleagues used EEG signal data from depressed patients and combined different methods to finally arrive at the best classification strategy, which was a combination of the Random Forest model and approximate entropy features [8]. Zhou and her colleagues proposed a method for constructing a domain dictionary oriented to the behavioral characteristics of depression [9]. They integrated the behavioral features refined according to the condition into the construction of the Chinese depression domain dictionary, thus making the dictionary reflect the symptomatic manifestations of depression more accurately and improving the vocabulary coverage of the domain dictionary.

Cummins and his colleagues used Mel Frequency Cepstrum Coefficient (MFCC) and resonance peak features in combination with GMM and SVM models to evaluate on a library constructed by 47 depressed patients, and the recognition correctness rate could reach 80%, confirming that speech features can be used as an effective detection index for assisting the diagnosis of depression [10]. Dogrucu, A. et al. proposed the Moodable framework, and developed the application, collected brief speech samples using the program, deployed K-nearest neighbor, SVM, and RF to the Moodable application and tested it on 335 volunteers, obtaining an F1 score of 0.766, a sensitivity of 0.75, and a specificity of 0.792 [11]. Another research shows that, a prediction model for depressed patients based on electrophysiological data by using features such as pupil size, gaze position, and gaze duration in eye movement signals was established, combined with machine learning algorithms, with higher accuracy and lower relative cost of data acquisition [12].

Unlike the traditional medical-based perspective to determine whether an ordinary person is a depressed person, the data obtained from highly sophisticated medical equipment cannot be processed because the dataset comes from a region of the country where the level of medical care is not

developed. The data behavioral indicators used in this project are mainly community surveys, including the household income and expenditure of the respondents.

3. METHODOLOGY

3.1. Dataset

The dataset used in this paper is from Kaggle, https://www.kaggle.com/datasets/diegobabativa/depression/data. The dataset is derived from a 2015 study conducted by the Busara Center in rural Siaya County near Lake Victoria in Western Kenya in the ZINDI competition data derived from the study https://zindi.africa/competitions/busara-mental-health-prediction-challenge/data. This survey included multiple characteristics including basic age, gender, marital status of the participants, household members, educational education, household economic status, and other information. In this paper, the unprocessed columns are standardized using the Z-score method. Finally, the dataset is divided into training and test sets.

3.2. Experimental Setup

In this paper, experiments were conducted using K-NN, Naive Bayes, Support Vector Machine and Random Forest methods to derive the accuracy scores of the models trained by each method. K-NN: 0.77305, Naive Bayes: 0.7766, Support Vector Machine Classifier: 0.79078, and Random Forest Classifier: 0.79078. The Random Forest Classifier, which had the best results, was subsequently tuned. The parameters obtained in this paper are shown in the following table:

Table 1. Adjusted parameters

max_depth	10
min_samples_leaf	1
min_samples_split	2
n_estimators	100

The accuracy score is 0.8447. Finally, the best parameters are used to train the Random Forest, and the model accuracy is obtained to be 0.794. The classification report is given below:

Table 2. Classification report

	precision	recall	F1-score	support
Not depressed	0.79	1.00	0.88	223
Depressed	1.00	0.02	0.03	59
Avg/total	0.84	0.79	0.71	282

The confusion matrix is as follows:

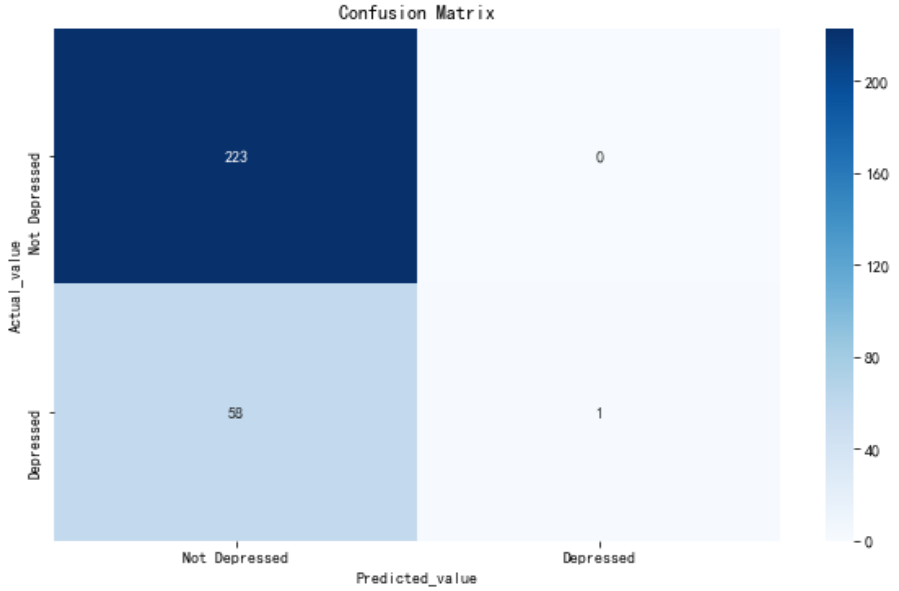

Figure 1. Confusion Matrix of Random Forest Classifier

4. CONCLUSION

Depression is a serious mood disorder characterized by persistent low mood, loss of interest and pleasure in life, loss of energy and other physical and cognitive symptoms. According to the World Health Organization, about 300 million people in the world suffer from depression. Compared with urban areas, the impact of mental illness is more severe in rural areas where the level of economic development is low, medical care is backward, and the level of awareness of depression among residents is much lower than that of urban residents. In this paper, for predicting whether an individual is suffering from depression, a machine learning approach is used. Predictive tools that may be helpful are given for poor rural areas where accurate physiological data cannot be collected. In this paper, out of the four different methods, the Random Forest Classifier has the highest accuracy, which is obtained after constant parameter adjustment of 0.794.

REFERENCES

[1] World Health Organization. (2023). Depressive disorder (depression). https://www.who.int/news-room/fact-sheets/detail/depression

[2] Rong, J., Ge, Y., Meng, N., Xie T., Ding H. Prevalence rate of depression in Chinese elderly from 2010 to 2019: a meta-analysis. Chinese Journal of Evidence-Based Medicine, 20(1), 26-31.

[3] Fu, X. et al. (2022). The mediation and interaction of depressive symptoms in activities of daily living and active aging in rural elderly: A cross-sectional survey. Frontiers in Public Health, 10, 942311.

[4] Xu, R., Liu, Y., Mu, T., Ye Y., Xu C. (2022). Determining the association between different living arrangements and depressive symptoms among over-65-year-old people: The moderating role of outdoor activities. Frontiers in Public Health, 10, 954416.

[5] Han, J., & Feng, L. (2020). Research progress on the application of artificial intelligence in the field of depression. Beijing Medicine, 42(4), 317-319+322. https://doi.org/10.15932/j.0253-9713.2020.04.013

[6] Yuan, G., Zhao, J., Zheng, D., Liu B. (2021). Study on distinguishing the severity of depression with Self-Rating Depression Scale and Beck Depression Inventory. Journal of Nervous and Mental Diseases and Mental Health, 21(12), 868-873.

[7] Xu, L. (2020). Research on depression recognition based on facial expression behavior patterns [Doctoral dissertation, Lanzhou University]. CNKI. https://doi.org/10.27204/d.cnki.glzhu.2020.000281

[8] Liu, D., Ye, J. Y., & Li, L. (2022). Feature extraction and implementation of depression based on machine learning. Experimental Technology and Management, 39(04), 153-157.

[9] Zhou, R., Zhu, G., Li, S., Duan, W., Li, J. (2024). Building domain lexicon oriented to behavioral features in depression. Big Data, 1-16.

[10] Cummins, N., Sethu, V., Epps, J., Schnieder, S., & Krajewski, J. (2015). Analysis of acoustic space variability in speech affected by depression. Speech Communication, 75, 27-49.

[11] Dogrucu, A., Perucic, A., Isaro, A., Ball, D., Toto, E., Rundensteiner, E. A., Agu, E., Davis-Martin, R., & Boudreaux, E. (2020). Moodable: On feasibility of instantaneous depression assessment using machine learning on voice samples with retrospectively harvested smartphone and social media data. Smart Health, 17, 100118.

[12] Stolicyn, A., Steele, J. D., & Seriès, P. (2022). Prediction of depression symptoms in individual subjects with face and eye movement tracking. Psychological Medicine, 52(9), 1784–1792.