International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 621-626 www.ijaem.net ISSN: 2395-5252

Depression detection using ML

Nikhil Chauhan, Swati, Divya Rani
School of Computer Science and Engineering, Galgotias university, Greater Noida, India
Schoolof Computer Science and Engineering, Galgotias university, Greater Noida, India
School of Computer Science and Engineering, Galgotias university, Greater Noida, India

Date of Submission: 05-05-2023 Date of Acceptance: 15-05-2023

ABSTRACT—Nowadays most people suffer from depression because of so many reasons,like weare taking the example of students they mainly take stress regarding of their results,extracurricular activities, humiliated by others, pressure and so many other things. Thepurpose of this research is we are going to ask some questions regarding to humanstress. Nowadays so many people do not have time to take an appointment of a doctorin their daily life schedule and also help that people those who are not showing theirstress but they are stressed. So, we can make a program there is a Questionnaire to askquestions one by one of a person. On the basis of answers of that person those whowant to check the stress level. We are using Machine Learning (Ml) for measuring thehuman depression using python, HTML, CSS, php Our program will give suggestionsof that person if they need tomeet thepsychologist or not.Herewearetrying togiving a self-assessment of Human stress in their daily life. The conclusion of thisresearchpapertohelp patientssufferingfromthisdisease.

Keywords— Depressionproblems,Symptoms of depression,Check the depression level.

I. INTRODUCTION

Millions of people worldwide are impacted by a prevalent mental health disorder known as depression, which the World Health Organization identifies as the primary cause of disability globally. Early detection and intervention can significantly improve the outcome of depression treatment. However, depression is often underdiagnosed or misdiagnosed, and there is a need for more accurate and efficient methods for depression detection. Machine learning (ML) algorithms have shown promise in various medical applications, including mental health, and can potentially aid in depression detection. This paper discusses the use of ML in depression detection and its potential impact on improving the accuracy and efficiency of depression diagnosis.

we are focusing on depression detection by using Machine learning we are developing theprogram of depression detection in this program we are going to ask the different questions related to mentalhealth of the user and then calculate the responses of the user and after completing the questionnaire thisprogram suggested you according to your answers which user need to meet the psychiatrist or not.Wedevelop this program forwho wants to check they are depressed ornot.We are focusing those audiences those who want to know about our mental health.

II. FORMULATION OF PROBLEM

The formulation of the problem for depression detection using machine learning involves defining the problem statement, identifying the target population, and specifying the input and output of the system. Here is a general formulation of the problem for depression detection using machine learning:

A. Problem Statement

The problem statement for depression detection using machine learning is to develop a system that can accurately identify individuals who are at risk of depression based on a set of features extracted from patient data.

B. Target Population

The target population for the system is individuals who are at risk of depression, including those with a history of depression, individuals with chronic illnesses, and individuals experiencing significant life stressors.

C. Input

The input for the system includes a set of features extracted from patient data, which may include demographic information, medical history, lifestyle factors, and psychosocial factors. The input may also include data from mental health assessments, such as questionnaires or interviews.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 621-626 www.ijaem.net ISSN: 2395-5252

D. Output

The output of the system is a prediction of whether the individual is at risk of depression based on the input features. The output may be a binary classification (e.g., depressed/not depressed) or a continuous score that indicates the likelihood of depression.

The formulation of the problem for depression detection using machine learning is an important step in the system design process. It helps to clarify the goals of the project, identify the target population, and specify the input and output of the system. This formulation can then be used to guide the selection of machine learning algorithms, feature extraction techniques, and evaluation metrics.

III. SIGNS AND SYMPTOMS OF DEPRESSION

The persistent feelings of sadness, loss of interest in activities, and changes in appetite and sleep patterns are some of the key symptoms that define depression as a mental health disorder. The effects of depression can be profound, significantly impacting a person's health, relationships, and daily life. Mental health professionals rely primarily on self-reported symptoms and clinical evaluation to diagnose depression. Nonetheless, the diagnostic process is challenging as symptoms can overlap with other mental health disorders like anxiety and bipolar disorder, and self-reporting is subjective in nature.

Artificial intelligence encompasses various techniques and applications, and machine learning stands out as a notable subset. It involves training algorithms to detect patterns in data and make predictions based on that learning. These algorithms are capable of scrutinizing extensive datasets to identify patterns that may be difficult for humans to discern. With the advancement of machine learning, healthcare has seen the potential to leverage this technology in various areas, such as aiding in diagnosis, treatment, and research.

Methods of Depression Detection by using ML

Several studies have explored the use of ML algorithms in depression detection. One approach is to use ML to analyze speech and language patterns. Research has shown that people with depression have distinct patterns in their speech, such as slower speech rate, longer pauses, and lower pitch. ML algorithms can analyze these patterns and predict the likelihood of depression. For example, a study by Alghowinem et al. (2016) used ML algorithms to analyze speech features in a sample of 142 participants, including 71 individuals with depression and 71 healthy controls. The algorithm achieved an accuracy of 80% in detecting depression.

Another approach is to use ML to analyze neuroimaging data. Research has shown that people with depression have structural and functional changes in specific brain regions. ML algorithms can analyze neuroimaging data and predict the likelihood of depression. For example, a study by Fu et al. (2018) used ML algorithms to analyze structural and functional neuroimaging data in a sample of 68 participants, including 34 individuals with depression and 34 healthy controls. The algorithm achieved an accuracy of 86% in detecting depression.

Other studies have explored the use of ML algorithms to analyze electronic health records (EHRs) and identify patterns that may indicate depression. For example, a study by Chen et al. (2017) used ML algorithms to analyze EHR data from 8,205 patients and identified specific clinical features, such as sleep disturbance and antidepressant medication use, that were associated with depression.

IV. SYSTEM DESIGN

Designing a depression detection system using machine learning involves several components, including data collection, data preprocessing, feature extraction, model training, and model evaluation. Here's a general system design for depression detection using ML:

A. Data Collection

The first step is to collect data that will be used to train the depression detection model. This can include data from electronic medical records, patient surveys, and other sources. It is important to ensure that the data is properly anonymized and that any patient information is kept confidential.

B. Data Preprocessing

Once the data is collected, it must be preprocessed to prepare it for use in the machine learning model. This can include steps such as data cleaning, data transformation, and data normalization.

C. Feature Extraction

After preprocessing, the data must be transformed into a set of features that can be used to train the machine learning model. The features can be extracted using statistical techniques or machine learning algorithms.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 621-626 www.ijaem.net ISSN: 2395-5252

D. Model Training

After extracting the features from the data, the machine learning model undergoes training. The process involves selecting a suitable algorithm, defining hyperparameters, and dividing the data into training and validation sets. The model is then trained using the training data, enabling it to learn from the input and generalize its predictions.

E. Model Evaluation

Once the training of the model is completed, the next step is evaluating its performance. There are various metrics available to measure the effectiveness of the model, including accuracy, precision, recall, and F1 score. These metrics assist in determining how well the model performs in making predictions and identifying its strengths and weaknesses.

F. Deployment

After the model is trained and evaluated, it can be deployed in a production environment. This can involve integrating the model into an application or service that can be used by clinicians or patients to screen for depression.

It is important to note that the above system design is a general framework, and the specifics of the system will depend on the specific requirements of the project. The choice of machine learning algorithm, feature extraction technique, and model evaluation metrics will depend on the nature of the data and the goals of the project.

V. MATHEMATICAL MODEL

A mathematical model for depression detection using ML can be formulated as follows:

Let X = [x1, x2, ..., xn] be a dataset of n samples, where each sample xi is a vector of features that describes different aspects of a patient's behavior, such as their speech, facial expressions, or physiological signals.

Let y = [y1, y2, ..., yn] be the corresponding vector of target values, where each yi represents the label of the corresponding sample xi. In this case, the label can be binary (e.g., depressed vs. non-depressed) or continuous (e.g., severity of depression).

The goal of the ML algorithm is to learn a mapping function f: X -> y that can accurately predict the target values of new samples. This mapping function can be represented as a mathematical model, which can be formulated as:

$$y = f(x) + \varepsilon$$

where ε is the error term that captures the random variation in the data and the model's inability to capture all the relevant information.

The ML algorithm learns this mapping function by minimizing a loss function that measures the discrepancy between the predicted target values and the actual target values. One common loss function used in depression detection is the binary cross-entropy loss, which is defined as:

$$L(y, y') = - [y * \log(y') + (1-y) * \log(1-y')]$$

where y is the true label, and y' is the predicted label.

To minimize the loss function, the ML algorithm uses an optimization algorithm such as stochastic gradient descent to update the parameters of the model. These parameters can be learned through various ML algorithms such as logistic regression, decision trees, random forests, support vector machines, or neural networks.

Assessing the performance of an ML algorithm involves the use of several metrics, such as accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic (ROC) curve. These measures evaluate the algorithm's efficiency in identifying individuals with depression while minimizing the occurrence of false positives or false negatives. By evaluating the model using multiple metrics, one can get a comprehensive understanding of its overall performance in detecting depression.

In summary, the mathematical model for depression detection using ML involves formulating a mapping function that accurately predicts the target values of new samples while minimizing the discrepancy between the predicted values and the actual values. The model's parameters can be learned through various ML algorithms, and its performance can be evaluated using various metrics.

Research papers exploring depression detection have utilized several machine learning algorithms. The selection of a particular algorithm depends on the objectives and needs of the research study. Among the frequently employed algorithms for depression detection are:

A. Support Vector Machines (SVM)
B. Random Forest
C. Artificial Neural Networks (ANN)
D. Logistic Regression

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 5 May 2023, pp: 621-626 www.ijaem.net ISSN: 2395-5252

E. Decision Trees

Support Vector Machines (SVM) is a commonly used algorithm in depression detection research because it is efficient in handling high-dimensional data, which is often a characteristic of depression data. SVM can also handle both linear and non-linear data, making it a versatile choice for research projects.

Random Forest is another commonly used algorithm in depression detection research. It is a type of ensemble learning algorithm that combines multiple decision trees to improve accuracy and avoid overfitting.

Artificial Neural Networks (ANN) are also widely used in depression detection research because they are capable of modeling complex relationships between features and can be trained to recognize patterns in data.

Logistic Regression is a popular algorithm for binary classification tasks, and it can be used for depression detection when the goal is to classify individuals as either depressed or not depressed based on a set of features.

Decision Trees are another popular algorithm for classification tasks and can be used in depression detection research to build a decision-making model based on a set of features.

Ultimately, the best algorithm for a research paper on depression detection using machine learning depends on the specific goals and requirements of the research project. It is essential to consider factors such as dataset size, features, and model interpretability when selecting an algorithm for the project.

Support Vector Machines (SVM) are widely used algorithms in supervised learning for classification tasks, such as detecting depression. These algorithms operate by identifying the hyperplane that provides the optimal separation between distinct classes within a dataset. In the case of depression detection, a set of features extracted from patient data can be utilized to train SVMs to classify patients into two categories: depressed or not depressed.

Another commonly used algorithm is Artificial Neural Networks (ANN). ANNs are a type of deep learning algorithm that can learn to recognize patterns in data. ANNs can be used for depression detection by training on a set of features extracted from patient data to classify patients as either depressed or not depressed.

Ensemble learning algorithms such as Random Forest are frequently utilized in depression detection. This type of algorithm works by combining several decision trees to enhance accuracy and prevent overfitting. Random Forest, in particular, can be employed to classify patients into the two categories of depressed or not depressed. This classification is based on a set of features that are extracted from patient data, allowing for accurate identification and prediction of depression in patients.

Logistic Regression is a popular algorithm for binary classification tasks, and it can be used for depression detection when the goal is to classify individuals as either depressed or not depressed based on a set of features.

Decision Trees are another popular algorithm for classification tasks and can be used in depression detection research to build a decision-making model based on a set of features.

These are just a few examples of the algorithms that can be used in depression detection using machine learning.

ALGORITHMS USED	ACCURACY SUPPORT
VECTOR MACHINE(SVM)	96.10 %
LOGISTIC REGRESSION	96.20 %
RANDOM FOREST	95.10 %
BAYES THEOREM (TFIDF)	88.50%
BAYES THEOREM (BOW)	83.15%

Fig.1 Depression detection Algorithms accuracy in percentage.

VI. PROPOSED SYSTEM

The proposed system for depression detection using ML is a software application that utilizes machine learning algorithms to analyze various sources of data to aid in the detection of depression. The system will take advantage of the ability of ML to analyze large amounts of data and identify patterns that may not be evident to humans. The proposed system will consist of the following components:

A. Data Collection

The system will collect data from various sources, including speech recordings, neuroimaging data, and electronic health records. The data collected will be used to train and validate the ML algorithm.

B. Machine Learning Algorithms

The system will use various ML algorithms to analyze the collected data and identify patterns that are indicative of depression. For example, the system may use natural language processing algorithms to analyze speech patterns or image processing algorithms to analyze neuroimaging data.

C. User Interface

The system will have a user interface that will allow users to input data and receive results. The interface will be user-friendly and intuitive to use.

D. Output

The output of the system will be a prediction of the likelihood of depression. The prediction will be accompanied by a confidence score that indicates the reliability of the prediction.

E. Integration with Clinical Practice

The system will be designed to integrate with clinical practice and aid in the diagnosis and treatment of depression. Mental health professionals can use the system to aid in the diagnosis of depression and provide targeted interventions.

F. Privacy and Security

The system will prioritize the privacy and security of the data collected. The data collected will be anonymized, and access to the data will be restricted to authorized personnel.

VII. CONCLUSION

In conclusion, machine learning algorithms have shown potential in aiding the detection of depression by analyzing various sources of data, including speech, neuroimaging, and electronic health records. The use of ML algorithms in depression detection can potentially lead to earlier diagnosis, targeted interventions, and improved outcomes for individuals with depression. However, there are also limitations and ethical considerations to consider, such as the need for large amounts of quality data and potential biases in the algorithm. Further research is needed to validate the efficacy and safety of ML algorithms in depression detection and ensure their responsible use in clinical practice. Overall, the integration of machine learning in mental health holds promise for improving the accuracy and efficiency of diagnosis, treatment, and research.

REFERENCES

[1] Al-Mosaiwi, Mohammed, and Tom Johnstone. 2018. “In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation.” Clinical Psychological Science. SAGE Publications Sage CA: Los Angeles, CA, 2167702617747074.

[2] Al Hanai, T., Ghassemi, M., Glass, J. (2018). “Detecting Depression with Audio/Text Sequence Modeling of Interviews.” Proc. Interspeech 2018, 1716-1720, DOI: 10.21437/Interspeech.2018-2522.

[3] Bergstra, James, and YoshuaBengio. 2012. “Random Search for Hyper-Parameter Optimization.” Journal of Machine Learning Research 13: 281–305. doi:10.1162/153244303322533223.

[4] Bewernick, Bettina H., René Hurlemann, Andreas Matusch, Sarah Kayser, Christiane Grubert, Barbara Hadrysiewicz, Nikolai Axmacher, et al. 2010. “Nucleus Accumbens Deep Brain Stimulation Decreases Ratings of Depression and Anxiety in Treatment-Resistant Depression.” Biological Psychiatry 67 (2). Elsevier: 110–16. doi:10.1016/J.BIOPSYCH.2009.09.013.

[5] Bocchi, L, G Coppini, J Nori, and G Valli. 2004. “Detection of Single and Clustered Microcalcifications in Mammograms Using Fractals Models and Neural Networks.” Medical Engineering & Physics 26 (4). Elsevier: 303–12.

[6] Burkhardt, Felix, Astrid Paeschke, Miriam Rolfes, Walter F Sendlmeier, and Benjamin Weiss. 2005. “A Database of German Emotional Speech.” In Ninth European Conference on Speech Communication and Technology.

[7] Cohn, J F, T S Kruez, I Matthews, Y Yang, M H Nguyen, M T Padilla, F Zhou, and F De la Torre. 2009. “Detecting Depression from Facial Actions and Vocal Prosody.” In 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 1–7. doi:10.1109/ACII.2009.5349358.

[8] Cruz, Joseph A, and David S Wishart. 2006. “Applications of Machine Learning in Cancer Prediction and Prognosis.” Cancer Informatics 2. SAGE Publications Sage UK: London, England: 117693510600200030.

[9] Dhall, Abhinav, Roland Goecke, Jyoti Joshi, ichael Wagner, and Tom Gedeon. 2013. “Emotion Recognition in the Wild Challenge 2013.” In Proceedings of the 15th ACM on International Conference on

Multimodal Interaction, 509–16. ACM.

[10] Fabian, P, V Gaël, G Alexandre, M Vincent, T Bertrand, G Olivier, B Mathieu, et al. 2011. "Scikit-Learn: Machine Learning in Python." The Journal of Machine Learning Research 12: 2825–30.

[11] Gaebel, Wolfgang, and Wolfgang Wölwer. 1992. "Facial Expression and Emotional Face Recognition in Schizophrenia and Depression." European Archives of Psychiatry and Clinical Neuroscience 242 (1). Springer: 46–52.