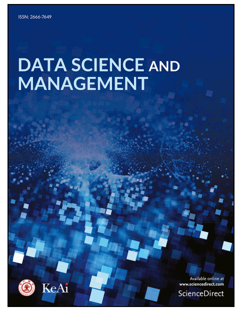
Journal Pre-proof



Predicting depression level based on human activities and feelings: A fuzzy logic-based analysis

Urmi Saha, Syed Mohammod Minhaz Hossain, Iqbal H. Sarker

PII: S2666-7649(24)00064-X

DOI: https://doi.org/10.1016/j.dsm.2024.11.003

Reference: DSM 129

To appear in: *Data Science and Management*

Received Date: 15 April 2024

Revised Date: 22 November 2024

Accepted Date: 27 November 2024

Please cite this article as: Saha, U., Minhaz Hossain, S.M., Sarker, I.H., Predicting depression level based on human activities and feelings: A fuzzy logic-based analysis, *Data Science and Management*, https://doi.org/10.1016/j.dsm.2024.11.003.

Predicting depression level based on human activities and feelings: A fuzzy logic-based analysis

Urmi Saha[a], Syed Mohammod Minhaz Hossain[b] and Iqbal H. Sarker[c]

[a] Department of CSE, Chittagong University of Engineering & Technology, Chattogram, 4349, Bangladesh

[b] Department of CSE, Premier University, Chattogram, 4000, Bangladesh

[c]Centre for Securing Digital Futures, Edith Cowan University , Perth , 6027 , WA, Australia

Corresponding Author: Syed Mohammod Minhaz Hossain, Associate Professor, Department of CSE, Premier University

Predicting depression level based on human activities and feelings: A fuzzy logic-based analysis

ARTICLE INFO

Keywords:
Depression Level Prediction
Human Activities
Human Feelings
Pearson Correlation Method
R-Squared Method
Uncertainty
Fuzzy Rules

ABSTRACT

Millions of individuals die by suicide each year, and many more suffer from severe depression. Furthermore, these deaths harm education, the economy, and healthcare worldwide. An individual's persistent feelings and activities, which include sleep disorders such as insomnia, sleeping overtime, or spending the majority of time lying down; pessimistic thinking about the future; and thoughts of committing suicide, help uncover the cause of depression. Several attempts have been made to prevent these losses and deaths. Our study proposes a simple fuzzy inference model that accurately predicts depression levels based on emotions and activities—even with incomplete data and uncertainties—and enhances mental health prediction, bridging theory and practice effectively. Psychologists and professors collaborated to create a survey to collect data for this study. The experiment was conducted using a Google survey form. This method effectively captures the ambiguity and imprecision in depression evaluation by combining linguistic elements of psychological traits. Using the Pearson correlation and R-squared methods, 15 features were chosen from 30 features, followed by five membership functions (poor, mediocre, average, decent, and good) and fuzzy rules to evaluate and create accurate forecasts of depression severity. Our proposed architecture can correctly classify depression levels based on human activities and feelings with 94% accuracy using a less sophisticated rules dictionary than previous pre-trained or hybrid models. Fuzzy logic performs better here by accurately categorizing ambiguous human emotional inputs into distinct degrees.

1. Introduction

The World Health Organization estimates that more than 264 million people worldwide suffer from depression World Health Organization (2023a). Depression is a severe illness that interferes with a person's regular functioning and mental processes, even if they are ignorant of their condition.

Among many other factors, one cause of depression is a person's constant feelings. That is, the way a person lives plays an essential role in the various types of depression. Depression affects females more than males. For example, women are 70% more likely than males to survive depression, and over 850,000 people commit suicide each year due to severe depression. Suicide is the second most significant cause of death among people aged 15–29 World Health Organization (2023b). In general, the statistics are increasing as more people are afflicted by depression in their lives, either directly or indirectly. The reason for this scenario is that a person's melancholy impacts not only themselves but also their loved ones, who may become depressed as a result. To save lives, early identification of depression and its severity is essential. The precise aspects of mental health may vary. However, typical reasons for individual emotions and continuous activities include the following:

Technological rapidness: Advancements in technology have revolutionized disease detection Das et al. (2022, 2024); Faruque et al. (2019); Ripan et al. (2021) and management, enabling early and accurate diagnoses. With the aid of technology and human abilities, the world is developing quickly, and people become so busy and materialistic that they occasionally forget to take the time to consider their mental health. Consequently, depression is a silent killer that can severely harm a person if not treated swiftly. Depression is a widespread mental health disorder that affects almost all individuals.

Lack of self-awareness: This disease is evaluated by society and individuals, and it is either deemed taboo or mocked by those diagnosed. They are unwilling to share their concerns early, escalating the condition.

Psychiatric disorders can be overcome in an era of survival with depression. Depression detection is mainly based on emotional health and measuring depression levels using intimate and personal questionnaires where giving a hundred percent privacy and not breaking any ethical law is the most challenging aspect of this study.

Most studies use social media data Alam et al. (2016), where most people express their feelings and activities. At the same time, we met counselors to learn more about depression. We decided to utilize a Google Survey Form where everyone could freely express their information without revealing their identity after collecting data from individuals and preprocessing the dataset. Then, using the Pearson Correlation and R-squared methods, 15 features were selected from among 30 features by calculating the effect of feature variables on the target feature variable. Subsequently, fuzzy rules were generated for five membership functions—poor, mediocre, average, decent, and good—-for predicting depression levels. Our study relies predominantly on fuzzy rules. Fuzzy systems are increasingly employed in a variety of practical applications. Our work involves a rule-based system in which fuzzy logic is used to express depression by simulating the interactions and relationships between variables (Shoaip et al., 2024).

Fuzzy logic principles have been effectively used in various disciplines where uncertainty and vagueness appear in multiple ways, and fuzzy logic provides better solutions

ORCID(s):

Jana et al. (2024). This experiment aims to predict depression in its early stages and, if it crosses the early stages, prevent tragic events such as suicide. The system outcome is compatible with an expert specialist's assessment of its performance Sadaf et al. (2024).

Considering the potential risks and increasing complexity of the model, a framework that can precisely detect and classify risks is desirable. It should also have no overfitting issues, require low computational cost and complexity, and apply to various operational situations. The primary motivation of this research is to help the afflicted without making them feel embarrassed or judged. Other research contributions are made by this study, which presents the use of fuzzy logic to address the crucial problem of recognizing and classifying Taneja et al. (2016), on the level of depressive status:

- Design a comprehensive, data-driven questionnaire focused on assessing mental well-being, capturing responses to key behavioral and emotional indicators.
- Collect relevant data to enable an insightful analysis while strictly adhering to ethical standards, data privacy regulations, and consent protocols, especially when handling sensitive or personal information.
- Develop a predictive model to assess depression levels by analyzing patterns in emotions and activities using fuzzy logic rules.
- Validate and test the model's accuracy and reliability on real-world data sets, ensuring that it can effectively differentiate varying depression levels.
- Provide a nuanced and accurate assessment of depression that supports healthcare providers and caregivers in identifying, diagnosing, and treating depression in both directly and indirectly afflicted persons. This tool aims to improve timely intervention, enhancing support for those affected.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 defines the methodology used in the study. Section 4 presents the evaluation results of the proposed approach. Some critical observations of our rule discovery approach, including several other application areas, are summarized in Section 5. Section 6 enhances mental health predictions, bridging theory and practice effectively. Finally, Section 7 concludes the paper and highlights future work.

2. Background and Related Work

In this section, we briefly describe the history of predicting depression and the different methods for classifying depression. We also briefly discuss related studies.

2.1. *Hybrid or deep learning model*

Ramesh et al. Ramesh and Lakshmanna (2024) proposed an Optimal Scrutiny Boosted Graph Convolutional LSTM (O-SBGC-LSTM), enhancing the SBGC-LSTM with the Eurygaster Optimization Algorithm and achieving over 98 percent accuracy in the spatial and temporal domains.

Singh et al. Singh et al. (2023) proposed an updated model using the You Look Only Once (YOLO) deep neural network method to identify melanoma lesions. This study included Deep Convolutional Neural Network (DCNN) layers, residual connections, convolutional layers, and feature concatenation. Test results showed that YOLO outperformed most classifiers regarding speed and accuracy.

Ruan et al. Ruan et al. (2024) explored using university system-gathered behavioral data to identify depression among college students. This study introduced a DeepFM network for depression diagnosis that learns linear and nonlinear relationships and recommends a modified focal loss function to address data imbalance.

2.2. *Fuzzy logic in classification*

de Souza et al. de Souza et al. (2020) suggested Fuzzy Optimal Power Flow (fuzzy OPF), which is an enhanced variant of the conventional OPF classifier that incorporates additional learning during supervised training after learning the samples' membership in an unsupervised manner.

Seyfari et al. Seyfari and Meimandi (2024) study's experimental findings demonstrate that using their proposed approach with the K-Nearest Neighbors (KNN) classifier and a set of authorization features yielded the best feature selection results. Their proposed method, which has 1908 features, meets the accuracy criterion with a success rate of 99.02%.

Nilashi et al. Nilashi et al. (2017) used Classification and Regression Trees to create rules from medical data automatically. As a dimensionality reduction method, they also employ Principal Component Analysis to address the multi-collinearity problem in the data. This method achieved accuracy values of 0.932 and 0.941 for the Wisconsin Diagnostic Breast Cancer (WDBC) and Mammographic mass datasets, respectively.

2.3. *Traditional depression detection Machine Learning (ML) techniques*

Guntuku et al. Guntuku et al. (2017) detected depression and mental illness based on social media data in which participants were recruited to take a depression survey and share their Facebook or Twitter data and self-declared mental health status. Based on this study, the prediction of mental illness using Linear Regression and Support Vector system was based on survey responses.

Mikal et al. Mikal et al. (2017) investigated Patient Attitudes Toward the use of Social Media Data to Augment Depression Diagnosis and Treatment based on an analysis of five focus groups with Twitter users. Two groups comprised participants without a diagnosed history of depression, and three groups comprised participants with a diagnosed history of depression. The study used NLP for mood tracking under the supervision of a mental health partitioner.

Aldarwish et al. Aldarwish and Ahmad (2017) also worked on Predicting Depression Levels Using Social Media posts using a Support Vendor Machine (SVM) and the Naive Bayes classifier. The web application gathered the User-Generated Content (UGC) from the patient's Twitter and Facebook. It categorizes patients into one of four levels (Minimal, Mild, Moderate, or Extreme melancholy). Using Rapid Miner to test two classifiers (SVM and Naive Bayes classifier), they created an origin before misery model.

Jamil et al. Jamil (2017) monitored tweets for depression to detect at-risk users and recognize people who might be depressed. They proposed a mechanized framework to recognize clients in danger from their public web-based media actions. To accomplish this objective, they prepared a client-level classifier utilizing an SVM that can recognize in-danger clients with a review of 0.8750 and an accuracy of 0.7778. In addition, they prepared a tweet-level classifier that predicts whether a tweet indicates trouble. This assignment was substantially more troublesome because of the imbalanced information.

2.4. AI on depression detection and classification system

O'Dea et al. O'dea et al. (2015) detected suicidality on Twitter. They planned to inspect whether the degree of worry for a self-destruction-related post on Twitter could be resolved given the post's substance, as determined by human coders and afterward duplicated by AI.

In Almeida et al. (2017), the prediction framework depended on a gathering characterization approach that combined supervised learning, information retrieval, and feature selection methods. During the testing stage, starter tests were performed using three techniques: simple rule-based classification using a sentiment analysis library, deep learning-based classification using a Recurrent Neural Network (RNN), and topic extraction using Latent Dirichlet Allocation. The system's Information Retrieval (IR)-based resources must be enhanced. In Morales (2018), the authors utilized social media data (posting patterns, social activities, and text analysis) with machine learning (Random Forest, XGBoost) to detect depression. Effective but lacks interpretability. The authors in Wang et al. (2017) analyzed user post patterns and sentiments over time to predict depression—emphasizing temporal shifts—although they did not capture all uncertainties in depression levels. In Reece and Danforth (2017), the authors examined the text and visual data (sentiments and image features) on Instagram using Convolutional Neural Networks (CNNs). Rich data but highly black-box, limiting interpretability.

2.5. *Rules based model*

Lin et al. Lin et al. (2018) developed a standard-based technique for recognizing individual-level longitudinal sickness by coordinating information changes, rule revelations, and rule assessments. They further broadened the identified rules to create rule-based observation procedures for screening individuals with different illness severities. They identified 12 danger-prescient guidelines, including RuleFit, calculated relapse, and SVM.

2.6. *Recommendation models analyzing factors behind depression*

In study Yang et al. (2018), Yang et al. analyzed a person's feelings through a mobile app survey and recommended healthcare or other medications to improve the person's condition. However, there was not enough recommended content and no level of prediction; therefore, it was not easy to administer proper therapy most of the time.

Alshawwa et al. Alshawwa et al. (2019) developed an expert system to help psychologists diagnose depression. The proposed master framework represents considerable authority in the analysis of wretchedness with accompanying side effects: loss of energy, adjustment of hunger, sleeping more or less, anxiety, reduced focus, uncertainty, sensations of uselessness, guilt, or hopelessness, and thoughts of self-harm or suicide.

Unlike previous studies, we intend to include people from all walks of life and a generally trustworthy source to produce inclusive data collection, laying the groundwork for a comprehensive framework incorporating predictive analytics. Our questionnaire-based data, which served as the foundation for this framework, were produced to gather information on people's sentiments and activities.

3. Methodology

This section presents the proposed methodology for predicting depression levels and explains its constraints in detail, as shown in Fig. 1. The data collection and development procedures are also described in this section. Data preprocessing, feature selection, and fuzzy rules prediction are also included.

3.1. Questionnaire setup

Before building this model, relevant literature was reviewed, and discussions were held with local psychiatrists dealing with mental health. The primary objective was to determine the common causes of the problem. Once the necessary literature was reviewed, some common causes were identified: sleep disturbances, including insomnia, sleeping overtime, and crying overnight; frequent tiredness; fatigue or passing most of the time without moving; feeling back pain or experiencing migraines; difficulty in remembering problems, lapses in concentration, and losing interest in hobbies or sex; change in appetite; losing weight or weight gain; optimistic about the individual dream; unsatisfactory result in working sector; negative thinking about the future; feeling broken, unworthy, or burdensome; change in relationship status, bonding with family, social relationships, and feeling loneliness; remembering past traumas; feeling guilty; death of a loved one; copying others or feeling low watching other people; losing self-confidence; loss of self-control; drug

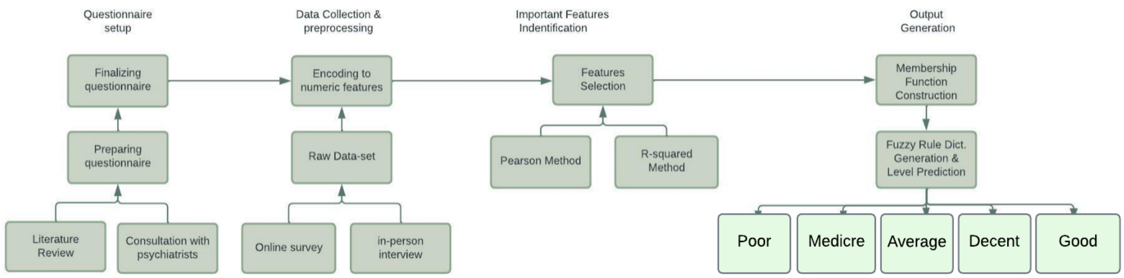


Figure 1: The Proposed System of Depression Level Prediction.

addiction, anger, and anxiety disorder; being dominated by the sub-conscious mind; thinking about the death of oneself or suicide; attempting to kill oneself, etc. Hsu and Wright (2014); Steger and Kashdan (2009). These causes can vary geographically because of differences in social structure and are verified with the help of local psychiatrists working on this issue. The questionnaire was prepared to obtain the necessary information from an individual about the causes identified as key reasons for depression among people of all ages.

We prepared a draft questionnaire based on the identified causes that specialized psychiatrists examined. Corrections have been made by making a few additions and exclusions to the draft. Multiple Choice Questions (MCQ) had two or more possible answers. The finalized questionnaire with 30 MCQ was as follows:

1. What is your gender?
2. How old are you?
3. What is your occupation? (Activities)
4. How many hours do you manage to sleep? (Activities)
5. Do you have sleeping disturbances at night? (Activities)
6. Addicted to social media? (Activities)
7. Addicted to drugs? (Activities)
8. What's your relationship status? (Feelings)
9. Pass most of the time laying on the bed without any body movements? (Activities)
10. Your family bonding? (Feelings)
11. Your economic status? (Feelings) Manstead (2018)
12. Nature of you? (Feelings)
13. Social relationship? (Feelings)
14. Do you feel yourself unworthy? (Feelings)
15. Do you have past traumas? (Feelings)
16. Do you cry overnight? (Activities)
17. Do you feel lonely? (Feelings)
18. Your Health? (Activities)
19. Can you eat when you are sad? (Feelings)
20. Do you feel guilt in your life? ((Activities) Hochwarter et al. (2007)
21. Do you feel concentration loss or difficulty in remembering problems? (Activities)
22. Do you feel you are a burden? (Feelings)
23. Do you have a lot of pressure in your working sector? (Activities)
24. Do you have self-control? (Feelings)
25. When do you often feel bad? (Feelings)
26. Is your subconscious mind dominating you? (Feelings)
27. Are you depressed? (Feelings)
28. Have you ever attempted suicide? (Activities and Feelings)
29. Satisfied in your working sector? (Feelings)
30. Depression type? (Target Variable)

The thirty MCQs were formulated as a questionnaire to assess reasons, local situations, and other aspects. The questions focused on family relationships, careers, financial situations, social conditions, and mental healthYang et al. (2023). All these questions were designed to consider cultural sensitivity in a sophisticated manner to extract the information needed for prediction in a classification model. To guarantee that this questionnaire collected the necessary information for our research, we conducted a pre-test with 20 participants. Once pretesting was finished, the questionnaire was improved for the final trial to produce the finalized questionnaire used in the study. Finally, the questionnaire was completed and used to collect data, and the results from 20 participants were used to validate the model. The test data were separated from the training set and are discussed in the evaluation section.

To model the various elements of depression using the diverse variables, we divided them into activities and feelings. There are a total of ten (10) activities features and fifteen (15) feelings features. However, one feature, depression type, was predicted based on both activities and feelings. Activities are occupation, hours slept, sleeping disturbances, addiction to social media, drugs, laying without movement, crying overnight, guilt in life, remembering problems, and work pressure. Feelings include relationship status, feeling unworthy, having past trauma, feeling lonely, eating when

Table 1
Raw Sample of Dataset

Gender?	How old are you?	Your occupation?	Sleeping disturbance?	Addicted to drugs?
Female	23	Student	Yes	No
Male	22	Student	Yes	No
Female	24	Student	Yes	Yes
Male	25	Employee	Yes	No

Table 2
Pre-processed Sample of Dataset.

Gender?	How old are you?	Your occupation?	Sleeping disturbance?	Addicted to drugs?
0	23	1	2	1
1	22	1	2	1
0	24	1	2	0
1	25	2	2	1

sad, feeling guilt in life, remembering problems, feeling like a burden, having self-control, often feeling bad, subconscious mind, feeling depressed, and satisfied with work. Attempted suicide stems from a suicidal thought and is considered both an activity and a feeling, along with all activities and feelings feature variables with their types.

3.2. Data collection and pre-processing

Data collection aimed to gather answers to the same questions from various sectors and age groups. The questionnaire was designed to minimize complexity yet avoid oversimplifying the process. The responses were gathered through an online survey. There were 545 data points in the training set; a raw sample of the dataset is shown in Table 1, and the variation among the datasets is exhibited in Fig. 2 and Fig. 3 based on the gender and profession, which are strong demographics of our model data to be generalizable for people of all walks of life.

To guarantee that the questionnaire collected the necessary information for our research, we conducted a pre-test with 20 participants. Once the pretesting was finished, the questionnaire was improved for the final trial to produce the finalized questionnaire used in the study. Finally, the questionnaire was completed and utilized to collect data, and the results from 20 participants were used to validate the model. The test data were separated from the training set.

For patients suffering from depression, data were collected through in-person interviews because of social stigma and difficulty in verifying or identifying an individual as a patient. Obtaining a voluntary response from people suffering from this problem is inevitably complex because of society's negative attitude towards individuals known as patients. Consequently, recognizing and collecting enough individual data required for this study voluntarily by Self-Assessment Questionnaire or other means, such as an online survey from patients, was implausible. Each interview took approximately 20 minutes to conduct. Data were collected using a Google form delivered personally to the recipients via email or social media from voluntary participants. The link to the form is attached in the Data Availability section. The form contains prefatory statements describing the purpose of the study and asking for permission to allow information to be used for research, with ethical laws being upheld. As mental health data are highly sensitive, we ensured informed consent, making participants fully aware of how their data will be used and protected. We also prevent unauthorized access by applying anonymization techniques and enforcing strong data security measures, such as encryption and secure storage.

Table 3
Feature Variables in Dataset

Key Facts	Q No.	Feature Name	Type
Activities	3	What is your occupation?	Categorical
	4	Hours you manage to sleep?	Ordinal
	5	Sleeping disturbances?	Categorical
	6	Addicted to social media?	Categorical
	7	Addicted to drugs?	Categorical
Feelings	8	Relationship status	Categorical
Activities	9	Laying without movements?	Binary
Feelings	10	Family bonding	Categorical
	11	Economic status	Ordinal
	12	Nature	Binary
	13	Social relationship	Ordinal
	14	Feeling unworthy?	Categorical
	15	Past traumas	Ordinal
Activities	16	Crying overnight?	Categorical
Feelings	17	Feeling lonely?	Categorical
Activities	18	Health	Categorical
Feelings	19	Eating when sad?	Binary
Activities	20	Guilt in life	Ordinal
	21	Memory issues	Binary
Feelings	22	Feeling burdensome?	Ordinal
Activities	23	Work pressure	Ordinal
Feelings	24	Self-control	Ordinal
	25	Often feeling bad	Categorical
	26	Subconscious mind	Ordinal
	27	Are you depressed?	Binary
Actv and Feeling	28	Suicide attempt?	Binary
Feelings	29	Satisfaction with work	Ordinal
Target	30	Depression type	Ordinal

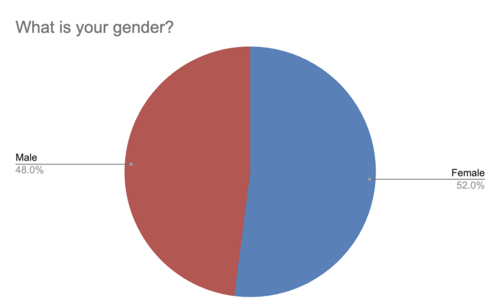


Figure 2: Gender Distribution in Training Data.

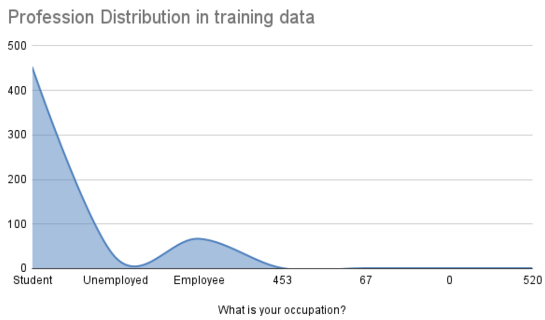


Figure 3: Profession Distribution in Training Data.

The spreadsheet constitutes each column in the dataset, and the answers to each question are replaced with categorical encoding into values that are "0," "1," "2," "3," or "4" depending on whether the question is a binary or five-point Likert scale question. The target variable titled "Depression Type" is in the dataset to predict depression levels followed by five membership functions (Poor, Mediocre, Average, Decent, and Good). For values within the column, "Depression Type" is a five-point Likert scale: "0" represents Negative, "1" represents Seasonal, "2" represents Mild, "3" represents Moderate, and "4" represents Severe. Finally, 30 variables were included in our dataset, with 29 feature variables and one target variable. The missing data were then handled using the average values within the column. A preprocessed dataset sample is shown in Table 2. Table 3 explains why different procedures were used for in-person interviews.

3.3. Key features identification

This dataset represented all category types. We used a method to determine the association between categories and quantify the impact of the variables. Features were selected using two statistical methods: the Pearson correlation approach and the R-squared method.

3.3.1. Pearson correlation method

Pearson's relationship creates a score that can change from -1 to +1. Two features with a high score (close to +1) are similar. Two uncorrelated features would have a Pearson score of almost zero. Two features that connected contrarily would have a Pearson score close to -1. The formula for Pearson's correlation is shown in Eq. (1), and the algorithm for our Dep_Fuzzy Model is shown in Algorithm 1.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

where
x_i= data value of the input feature
$\bar{x}$= average value for the input feature
y_i= data value of the target feature
$\bar{y}$= average value for the target feature
n = total number of observations for an input target feature in the dataset.

Algorithm 1 Pearson Correlation Coefficient

procedure CAL_PEARSON(X, Y)
  Input: Symptoms of Depression $X = [x_1, x_2, \dots, x_n]$ and Depression State Y
  Output: feature_target_corr_p
  Calculate the average of X:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  Calculate the average of Y:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

  Finalize the co-variance of X and Y:

$$\text{avg\_p}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

  Determine the standard deviation of X:

$$\sigma_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

  Calculate the standard deviation of Y:

$$\sigma_Y = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

  Calculate the Pearson correlation coefficient:

$$r = \frac{\text{avg\_p}(X, Y)}{\sigma_X \sigma_Y}$$

  return feature_target_corr_p
end procedure

The names of the features that have a strong relationship with the target variable according to the Pearson correlation method are listed in Table 4.

3.3.2. R-squared method

R-squared is an integrity-of-fit measure of the fitted regression line. This measurement for our model indicates the percentage of variance in the dependent variable that the independent variables collectively explain, assessing the strength of the relationship between the model and the dependent variable on a valid 0 to 100 scale, shown in Eq. (2).

$$R^2 = \frac{\text{Variance explained by the model}}{\text{Total variance}} \tag{2}$$

Table 4
Selected Feature Variables in Dataset using Pearson Correlation Method

Feature Name	Correlation Coefficient
How old are you?	0.018
Sleeping disturbances at night?	0.095
Addicted to drugs?	0.054
What is your relationship status?	0.027
Your economic status?	0.016
Having past traumas?	0.145
Your health?	0.024
Eating when you are sad?	0.012
Feeling you are a burden?	0.106
Pressure from the working sector?	0.033
Having self-control?	0.012
Often feeling bad?	0.033
Subconscious mind?	0.078
Have you ever attempted suicide?	0.534
Are you satisfied with working?	0.073

where R denotes the coefficient of determination for one input feature to the target feature. The algorithm we used in our model is shown in Algorithm 2.

Algorithm 2 R-Squared Correlation

procedure CAL_RSQUARED($y, \hat{y}$)
  Input: Actual values $y = [y_1, y_2, \ldots, y_n]$ and predicted values $\hat{y} = [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n]$
  Output: feature_target_corr_r2
  Calculate the average of observed values:

$$avg\_r2 = \frac{1}{n}\sum_{i=1}^{n} y_i$$

  Calculate the total sum of squares:

$$\text{sums\_r2} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

  Calculate the residual sum of squares:

$$\text{pred\_r2} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

  Calculate the R-squared value:

$$R^2 = 1 - \frac{\text{sums\_r2}}{\text{pred\_r2}}$$

  return feature_target_corr_r2
end procedure

The names of the features according to the R-squared method are listed in Table 5.

Table 5
Selected Feature Variables in Dataset using R-Squared Method

Feature Name	Correlation Coefficient
Sleeping disturbances at night?	0.026
Addicted to social media?	0.027
Relationship status?	0.053
Laying without body movements?	0.059
Your family bonding?	0.021
Your economic status?	0.057
Social relationship?	0.059
Feeling unworthy?	0.742
Having past traumas?	0.029
Crying overnight?	0.059
Feeling lonely?	0.025
Your health?	0.058
Eating when you are sad?	0.016
Feeling guilt in your life?	0.028
Feeling you are a burden?	0.014
Pressure from the working sector?	0.052
Having self-control?	0.020
Often feeling bad?	0.033
Subconscious mind?	0.025
Are you satisfied with working?	0.021

3.4. Output generation

In this section, our Dep_Fuzzy Inference Model has been shown in Fig. 4. Here, we create our Fuzzification Module using the membership function, which changes the model inputs (fresh numbers) into fuzzy sets. Based on Fuzzy Inference, we then generate our output in fuzzy sets shown in Algorithms 3 and 4. Subsequently, de-fuzzification is performed, shown in Algorithm 5.

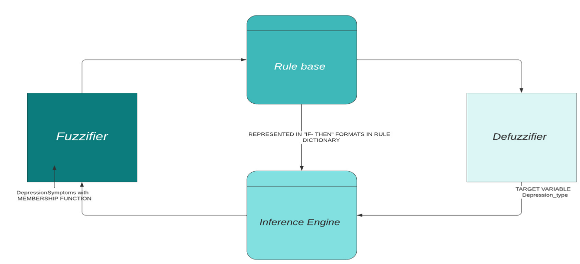


Figure 4: Dep_Fuzzy Inference Model.

3.4.1. *Fuzzifier*

Fuzzier takes all inputs determined from (TargetFeatureCorr), which will be discussed in Section 3.4.3, as fuzzy sets that are used in fuzzy inference to describe poorly specified data. In a fuzzifier, the membership function that provides each element in a fuzzy set with a degree of membership between 0 and 1 defines the fuzzy set. For instance, when predicting depression_type, an individual's

Algorithm 3 Dep_Fuzzy Inference Model

```
procedure DEP_FUZZY_INFERENCE(x_1, x_2, ..., x_15)
    Input: Crisp values for features x_1 to x_15
    Output: Predicted Depression_type
    for i = 1 to 15 do
        Fuzzifying x_i to obtain membership values for each fuzzy set
    end for
    Initialize rule_dict for Dep_ Fuzzy_Inference
    Initialize Set fuzzy output to zero
    for each fuzzy rule do
        Evaluate the degree for the rule
        Combine degrees using fuzzy conjunctive, disjunctive
        Combine the results from possible non-scattered rules
    end for
    Defuzzify the fuzzy output to obtain a crisp value
    return Predicted Depression_type
end procedure
```

mental health may fall into fuzzy groups with differing degrees of membership, such as "poor depression," "mediocre depression," "average depression," "decent depression," and "severe depression."

3.4.2. Membership function construction

We construct the membership function to define the input variables used as the algorithm input parameters. Depression_type is the feature that will be predicted based on the selected activity and feeling features shown in Table 6 for rule generation.

Table 6
Defined Input Variables in Crisp-Set

Feature Name	Input Variables
How old are you?	Input Variable x_1
Have sleeping disturbances at night?	Activity Variable x_2
Addicted to drugs?	Activity Variable x_3
Relationship status?	Feelings Variable x_4
What is your economic status?	Feelings Variable x_5
Having past traumas?	Feelings Variable x_6
Your health?	Activity Variable x_7
Eating when you are sad?	Feelings Variable x_8
Feeling you are a burden?	Feelings Variable x_9
Pressure from the working sector?	Activity Variable x_{10}
Having self-control?	Feelings Variable x_{11}
Often feeling bad?	Feelings Variable x_{12}
Subconscious mind?	Feelings Variable x_{13}
Have you ever attempted suicide?	Act and Feelings x_{14}
Are you satisfied with working?	Feelings Variable x_{15}

Membership functions permit the measurement of etymological terms and graphically address fuzzy sets. The enrollment work for a fuzzy set A on the universe of talk X is characterized as shown in Eq. (3).

$$\mu_A(x) : X \rightarrow [0, 1] \tag{3}$$

Every component of X is planned to have a worth somewhere between 0 and 1. This is called participation worth or level of enrollment. It measures the enrollment level of the component in X to the fuzzy set A. The triangular membership function is mathematically defined as Eq. (4).

$$\text{f(x,a,b,c)} = \begin{cases} 0, & x \leq a \\ \frac{\text{x - a}}{\text{b - a}}, & a \leq x \leq b, \\ \frac{\text{c - x}}{\text{c - b}}, & b \leq x \leq c \\ 0, & c \leq x \end{cases} \tag{4}$$

It can also be defined as Eq. (5).

$$\text{f(x,a,b,c)} = max\left(min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right) \tag{5}$$

where
a represents the left boundary of the fuzzy set where membership value starts to increase,
b represents the peak point of the fuzzy set where the membership value reaches the maximum,
c represents the right boundary of the fuzzy set where the membership value starts to decrease,
x is the input feature for which membership value varies.

We then used the triangular membership function on our input sets, which divided the symptoms into five stages. Let x be the input variable (Dep_Symptoms) ranging from 0 to 100. The membership functions for the fuzzy sets are defined in Eqs. (6), (7), (8), (9), and (10).

Poor

$$\mu_{\text{Poor}}(x) = \begin{cases} 1 & \text{if } x \leq 20 \\ \frac{40-x}{40-20} & \text{if } 20 < x < 40 \\ 0 & \text{if } x \geq 40 \end{cases} \tag{6}$$

Mediocre

$$\mu_{\text{Mediocre}}(x) = \begin{cases} 0 & \text{if } x \leq 20 \\ \frac{x-20}{40-20} & \text{if } 20 < x < 40 \\ \frac{60-x}{60-40} & \text{if } 40 \leq x \leq 60 \\ 0 & \text{if } x \geq 60 \end{cases} \tag{7}$$

Average

$$\mu_{\text{Average}}(x) = \begin{cases} 0 & \text{if } x \le 40 \\ \frac{x-40}{60-40} & \text{if } 40 < x < 60 \\ \frac{80-x}{80-60} & \text{if } 60 \le x \le 80 \\ 0 & \text{if } x \ge 80 \end{cases} \tag{8}$$

Decent

$$\mu_{\text{Decent}}(x) = \begin{cases} 0 & \text{if } x \le 60 \\ \frac{x-60}{80-60} & \text{if } 60 < x < 80 \\ \frac{100-x}{100-80} & \text{if } 80 \le x \le 100 \\ 0 & \text{if } x \ge 100 \end{cases} \tag{9}$$

Good

$$\mu_{\text{Good}}(x) = \begin{cases} 0 & \text{if } x \le 80 \\ \frac{x-80}{100-80} & \text{if } 80 < x < 100 \\ 1 & \text{if } x \ge 100 \end{cases} \tag{10}$$

For an ordinal input feature named "Addicted to drugs?" the membership value is determined using the triangular membership function shown in Fig. 5.

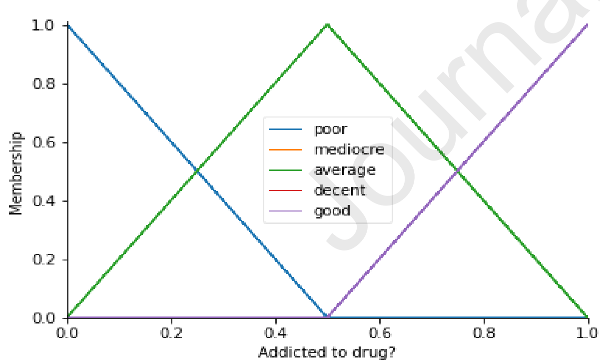


Figure 5: Membership Function for "Addicted to drugs?"

For another ordinal input feature named "Sleeping disturbances?" the membership value is presented in Fig. 6.

For a binary input feature named "Have you ever attempted suicide?" the membership value is presented in Fig. 7.

3.4.3. *Fuzzy rule base and inference engine*

Combining the input variables from TargetFeatureCorrelation, a rule_dict (Dictionary) was generated based on the input variables listed in Table 6.

Fuzzy rules based on specialists' IF-THEN standards were then built. This phenomenon works as an inference engine that recreates the human thinking measure by making fuzzy deductions from the data sources and IF-THEN

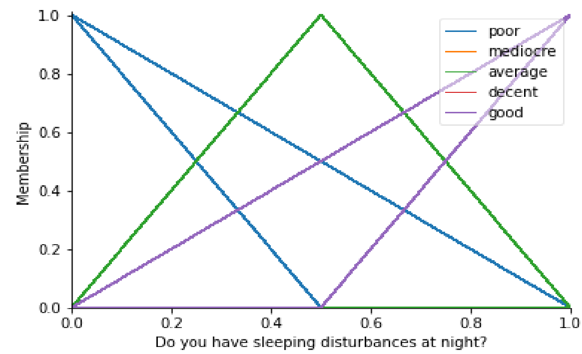


Figure 6: Membership Function for "Sleeping disturbances?"

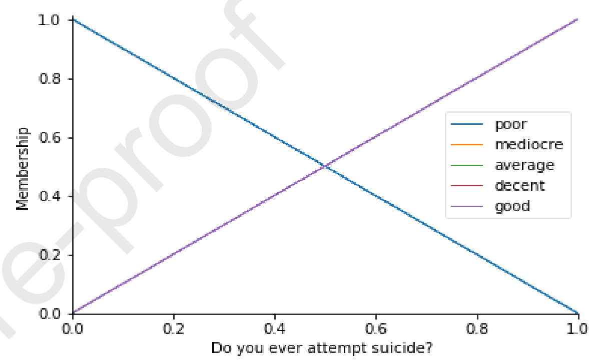


Figure 7: Membership Function for "Have you ever attempted suicide?"

Algorithm 4 Feature in rule_dict

procedure SELECTED_FEATURE($\mathcal{X}$,Thresh$_P$,Thresh$_{R^2}$)
  Input: Set of features $\mathcal{X}$, Thresh$_P$, Thresh$_{R^2}$
  Output: Selected features $\mathcal{X}_{\text{selected}}$
  Initialize $\mathcal{X}_{\text{selected}} \leftarrow \emptyset$
  for each feature $(X_i, X_j) \in \mathcal{X}$ **do**
    if $|r_{ij}| \ge \text{Thresh}_P$ **and** $R^2_{ij} \ge \text{Thresh}_{R^2}$ **then**
      Add X_i to $\mathcal{X}_{\text{selected}}$
    end if
  end for
  return $\mathcal{X}_{\text{selected}}$
end procedure

guidelines to gather decisions at a certain data point rather than allowing them to scatter.

After defining membership functions, we generated fuzzy rules to predict depression levels. Some of these are described below.

Fuzzy Rule Dictionary

For the Dep_Fuzzy model, input features with fuzzy sets defined as Poor, Mediocre, Average, Decent, and Good results in activities that impact reciprocally. If the degree of activity is ["poor"], that leads to a Depression_type of ["good"], whereas when the degree of feelings is ["poor"] (as the questionnaire denotes feelings in a meaningful, non-affirmative manner), that leads to a Depression_type of ["good"]. Three rules have been shown as hitting the target variable membership function.

Rule_dict_sample:

- If ["Do you have sleeping disturbances at night?"] is ["poor"], then ["Depression type"] is ["poor"], as shown in Fig. 8.
- If ["Do you have sleeping disturbances at night?"] is ["good"], then ["Depression type"] is ["good"], as shown in Fig. 9.
- If ["Your family bonding?"] is ["poor"], then ["Depression type"] is ["good"], as shown in Fig. 10.

The next stage is to give input and output variables numeric ranges (crisp values). An example range assigned to the defined variables is shown in Table 7.

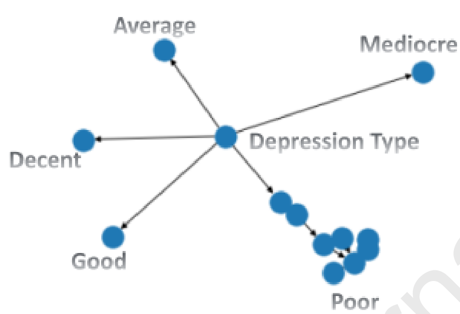


Figure 8: Dep_Fuzzy Inference Rule Generation.

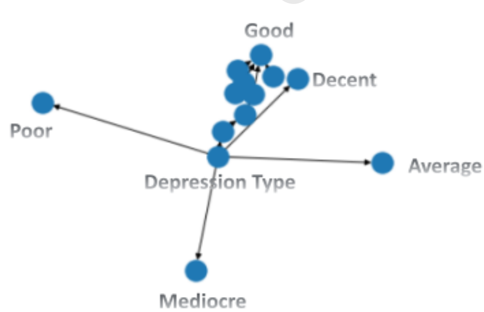


Figure 9: Dep_Fuzzy Inference Rule Generation.

3.4.4. *Defuzzifier*

The output variable is defined in Table 3 as ordinal type.

Fuzzy rule-based systems use fuzzification, inference, and composition processes to assess linguistic if-then rules. They generate fuzzy results that must usually be turned into crisp output. Therefore, we performed defuzzification, which changed the fuzzy set to a fresh value. The corresponding output was generated using the inputs shown in Fig. 11 as sample input 1, as shown in Fig. 12.

Using other sample input data shown in Fig. 13 as sample input 2, the corresponding output was generated, as shown in Fig. 14.

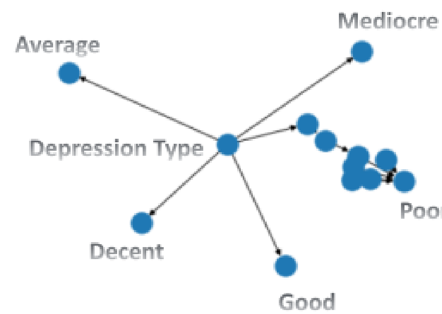


Figure 10: Dep_Fuzzy Inference Rule Generation.

Table 7
Example of Assigned Range to Defined Variables

Variables	Min Range Value	Max Range Value
Binary	0	1
Ordinal	0	2
Categorical-1	0	3
Categorical-2	0	4

Algorithm 5 Dep_Defuzzifier

```
procedure DEP_DEFUZZIFY(FuzzySet)
    Input: Fuzzy set FuzzySet with membership functions μ_i(x) for each output membership function y
    Output: Crisp value x
    Numerator A ← 0
    Denominator B ← 0
    for each fuzzy output i in FuzzySet do
        for each value x_i within the range of fuzzy set i do
            Calculate membership value μ_i(x_i)
            Update numerator A ← A + μ_i(x_i) · x_i
            Update denominator B ← B + μ_i(x_i)
        end for
    end for
    if B = 0, then return 0 (Divide by zero Error Message)
    Calculate defuzzified value y ← A/B
    return Crisp value x = y
end procedure
```

```
classify.input["How old are you?"] = 25
classify.input["Do you have sleeping disturbances at night?"] = 1
classify.input["Addicted to drug?"] =0
classify.input["What's your relationship status?"] =2
classify.input["Your Economical status?"] =1
classify.input["Do you have past traumas?"] =1
classify.input["Your Health?"] =0
classify.input["Can you eat when you are sad?"] =0
classify.input["Do you feel yourself a burden?"] =0
classify.input["Do you have a lot of pressure on your working sector?"] =0
classify.input["Do you have self-control?"] =0
classify.input["When do you often feel bad ? "] =2
classify.input["Is your subconscious mind dominating you?"] =1
classify.input["Do you ever attempt suicide? "] =1
classify.input["Satisfied Result in your working sector?"] =0


classify.compute()
```

```
print(classify.output['Depression_type'])
rule_dict['Depression_type'].view(sim=classify)
```

```
0.4000000000000001
```

Figure 11: Sample Input 1.

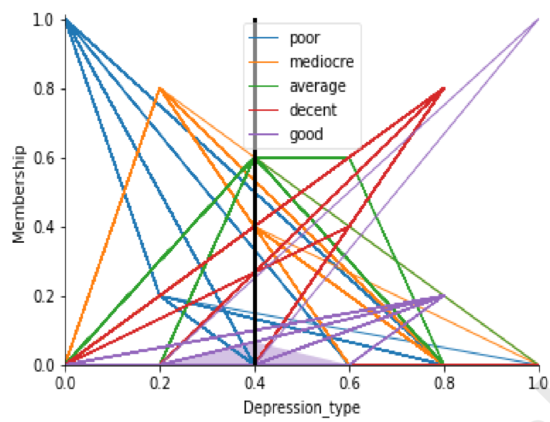


Figure 12: Defuzzified Value for Corresponding Sample Input 1.

```
classify.input["How old are you?"] = 25
classify.input["Do you have sleeping disturbances at night?"] = 1
classify.input["Addicted to drug?"] =0
classify.input["What's your relationship status?"] =5
classify.input["Your Economical status?"] =1
classify.input["Do you have past traumas?"] =0
classify.input["Your Health?"] =0
classify.input["Can you eat when you are sad?"] =0
classify.input["Do you feel yourself a burden?"] =0
classify.input["Do you have a lot of pressure on your working sector?"] =0
classify.input["Do you have self-control?"] =0
classify.input["When do you often feel bad ? "] =2
classify.input["Is your subconscious mind dominating you?"] =1
classify.input["Do you ever attempt suicide? "] =0
classify.input["Satisfied Result in your working sector?"] =1


classify.compute()
```

```
print(classify.output['Depression_type'])
rule_dict['Depression_type'].view(sim=classify)
```

```
0.15199999999999997
```

Figure 13: Sample Input 2.

4. Evaluation and Experimental Results

To evaluate the efficiency of our proposed system, we calculate the accuracy with Eqs. (11) and (12).

$$Accuracy = \frac{Number\_of\_correct\_scenarios}{Number\_of\_total\_scenarios} * 100 \quad (11)$$

For performance evaluation, we tested the model on a separate dataset of 50 participants whose data were not

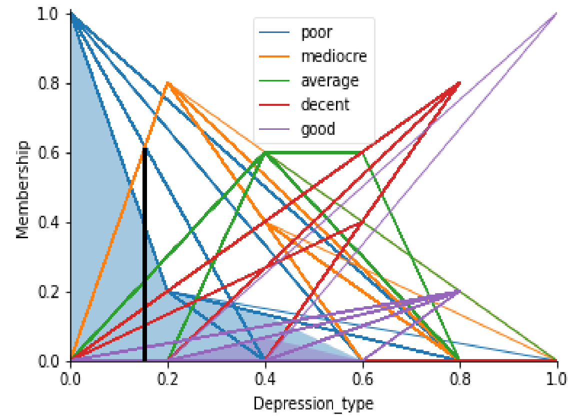


Figure 14: Defuzzified Value for Corresponding Sample Input 2.

used during training or validation. The model successfully provided accurate depression-level predictions for 47 of the 50 participants, demonstrating strong generalization and effectiveness in real-world scenarios. This result highlights the model's ability to perform reliably outside the training data, maintaining accuracy across unseen datasets.

$$Accuracy = \frac{47}{50} * 100 = 94.00\% \quad (12)$$

The complexity of depression arises from the interconnection between diverse factors that influence an individual's emotional, mental, and physical well-being. In our model, depression was influenced by various activities and feelings, each contributing in different ways. Sometimes, activities (e.g., occupation, sleeping patterns, and social media use) and feelings (e.g., relationship status, self-worth, and past traumas) are highly interdependent. The correlation coefficients are significant, as shown in Tables 4 and 5.

If ["Have sleeping disturbances at night?"] ["Mediocre"] and ["Relationship status"] ["Good"], then ["Depression_type"] leads to Poor.

However, if ["Have sleeping disturbances at night?"] ["Mediocre"] and ["Relationship status"] ["Poor"], then ["Depression_type"] leads to Average. Activities have a significant impact. At the same time:

- A greater score in Satisfied in the working sector (activities) can help the Depression_type converge to poor.

  If ["Satisfied in working sector"] ["Poor"], then ["Depression_type"] converges to poor.

- Emotional past trauma (feelings) can lead to behaviors like Suicidal thoughts, leading to attempts (activities).

  If ["Having past traumas"] ["Good"] and ["Ever attempted suicide?"], then ["Depression_type"] converges to average, decent, or good, as an example from the sample input in Figs. 11 and 13.

- If [“Have sleeping disturbances at night?”] ["Mediocre"], ["Having past traumas"] ["Good"], and ["Ever attempted suicide?"]["Good"], then ["Depression_type”] converges to good; even if the activity impact is poor, the impact of the feelings is higher.

Another aspect is that the model also includes four activity-related features and nine feelings-related features (selected features as shown in Table 6), each capable of interacting in multiple ways. For instance:

- A person addicted to social media (an activity) may also feel like a burden (a feeling), creating a loop where excessive social comparison on social media worsens their sense of worthlessness.
- Feeling guilt can emerge from multiple activities like crying overnight or engaging in harmful behaviors such as drug use.

Suicidal thoughts add another layer of complexity. The feeling of being a burden or having attempted suicide represents severe levels of depression and needs to be treated with extra caution. These thoughts are likely linked to both activities (e.g., laying without movement and drug use) and feelings (e.g., feeling worthless and having no self-control).

In the ablation test, we tested our model performance for six different scenarios: performance on only activities, only on feelings, on activities with attempting suicide features, on feelings associated with attempting suicide, and all activities and feelings without attempting suicide features, as well as all selected features (as shown in Table 6), including the activities and feelings shown in Table 8. In addition, regarding the example shown in Fig. 11, for the six scenarios, the Depression_type was

- Only activities 30.00%
- Only activities with attempting suicide features 50.062%
- Only feelings 45.00%
- Only feelings associated with attempting suicide 49.99%
- All activities and feelings without attempting suicide features 30.26%

Table 8
Ablation Test Performance for Predicting Depression Level

Ablation Test Scenarios	Performance Test (Accuracy)
Only activities	86.00%
Only activities with attempting suicide features	90.00%
Only feelings	82.00%
Only feelings associated with attempting suicide	86.00%
All activities and feelings without attempting suicide features	88.00%
All selected activities and feelings features as shown in Table 6	94.00%

In Table 9, we present a comparison with existing approaches regarding the strengths and limitations of the models and generalization. Our proposed fuzzy system significantly outperforms the other models, achieving 94% accuracy. Social media-based approaches Alshawwa et al. (2019); De Choudhury et al. (2013); Hussain et al. (2015) primarily analyze online interactions, whereas our model employs fuzzy logic to assess feelings and activities directly. Feelings and activities vary widely among individuals. Our fuzzy system was designed to be robust against this variability, allowing it to function effectively even with diverse input data. This implies that the system can still provide accurate predictions by leveraging its fuzzy rules if the user provides data reflecting high or low emotional states. Our fuzzy system demonstrates a high ability to generalize across diverse individual data, making it more reliable for real-world applications.

Table 9
Performance Evaluation

References	Key Features
SNS-based Predictive Model	
Hussain et al. (2015)	limited by variability in online behavior
Predicting via Social Media	
De Choudhury et al. (2013)	does not fully capture emotional states
Predicting using Social Media Posts	
Alshawwa et al. (2019)	lower accuracy on superficial data
Our Proposed Fuzzy System	directly analyzes personal feelings and activities

Table 10 compares the existing multimodal approaches regarding key features, interpretability, and limitations. Traditional social media-based approaches Morales (2018); Reece and Danforth (2017); Wang et al. (2017) rely on machine learning techniques that primarily analyze textual and visual elements. These models utilize metrics such as posting frequency, sentiment changes over time, and image features (e.g., brightness and facial expressions) to assess emotional and mental health states. However, their reliance on black-box models, including Random Forest, XGBoost, and CNNs, limits interpretability and can obscure the reasoning behind their predictions.

In contrast, our proposed fuzzy logic model was built by directly analyzing personal feelings and activities using a system of fuzzy inference rules. This approach allows for personalized and interpretable predictions as it accounts for the subjective nature of emotional states. Unlike traditional machine learning methods that may struggle to generalize owing to the variability in emotional expressions and social behaviors, our fuzzy model is designed to accommodate diverse input data. This robustness makes it well-suited for clinical settings where understanding individual emotional factors is crucial. The proposed system achieves high interpretability and personalization, offering a transparent and reliable solution that can be effectively generalized across various real-world data.

5. Discussion

This study aimed to predict depression levels based on human activities and feelings using fuzzy rules discussed in the previous sections. Necessary information on this research culled from related research has paved the way for identifying common symptoms of depression. Subsequently, to obtain participant data on these causes, the questionnaire was carefully designed to safeguard individual privacy and

Table 10
Performance Evaluation with Multimodal Models

References	Approach	Key Features	Interpretability
Morales (2018)	Utilizes social media text data with metadata applying ML techniques	Posting patterns (frequency, timing), social activity (friends, likes), and textual analysis (word frequency, sentiment)	Random Forest and XGBoost are still considered black box techniques
Wang et al. (2017)	Utilize user posts' sentiment and patterns over time to predict depression	Posting irregularity or regularity, sentiment changes over time	They emphasize the time dimension of depression, capturing sentiment dynamics does not guarantee uncertainty behind recent level depression
Reece and Danforth (2017)	Analyzes textual and visual content (e.g., Instagram posts) to estimate depression levels	Text sentiment, connection metrics, and image features (brightness, hue, facial expressions)	Still relying on uninterpretable black-box models (CNNs), used a rich dataset that combines textual and visual data
Our Proposed Fuzzy System	Directly analyzes personal feelings and activities.	Activities and Feelings	Our fuzzy logic model prioritizes interpretability and personalization using fuzzy inference rules; it is useful in clinical contexts where it is essential to understand emotional elements fully

finalized after pretesting. Participants' responses were converted into datasets after preprocessing. Using the Pearson and R-squared statistics between feature and target variables, the feature importance was calculated. Improved accuracy was achieved using only 15 necessary features out of the initial 30, as shown in Table 6. Participants were respondents aged between 18–60+ years (largely youths), which provided better opportunities to understand behavioral traits as similar age groups usually display similar behavior patterns. Local impacts unique to Bangladesh were considered to measure the risk factors feasibly to improve the model's effectiveness. Although the prediction model's performance is comparatively lower, this shortcoming could be overcome by collecting additional data. Regarding the overall performance, a good accuracy score was achieved because the questionnaire was prepared according to the required information and pretested before finalizing it, leaving no scope for unnecessary questions to remain in the questionnaire. The dataset was created using information gathered from the questionnaire, which resulted in an adequate dataset.

This study focuses on prevention by creating a model to assess depression levels, which could be a small step forward in addressing this global problem. Finally, this study demonstrates that data science plays a vital role in understanding the intricate behavioral traits of human beings to address critical psychological issues. Health officials can use the vast potential of such data to build a safer and better future. Our rule-based method can predict depression levels by using a concise set of rules representing an individual's physical and mental health conditions. The approach presented in this study is also applicable to predicting outcomes more accurately in a rule-based intelligent system.

This study was performed among Bangladeshi people, and a large portion of the dataset consisted of university-going and employed people. However, the behavior traits are generalized for Bangladesh and the general human population, except for some causes, such as cultural differences, which do not represent the world's scenario. As the causes might vary for changes in the target population at home and abroad, an increase in the size of the dataset will reveal more valuable insights that will help predict depression with a higher degree of accuracy.

The complexity of depression arises from the interconnection between diverse factors that influence an individual's emotional, mental, and physical well-being. In our model, depression was influenced by various activities and feelings, each contributing in different ways. Depression often fluctuates over time; however, when features are collected at a single point in time, they fail to capture the dynamic progression of the condition, leading to an incomplete understanding of how depression evolves. To address the challenges of capturing the temporal nature of depression in a model based on a single time point, a combination of longitudinal data collection, advanced modeling techniques, such as time series or mixed-effects models, and surrogate time-based measures can be used. While longitudinal data provide the most accurate reflection of depression's progression, alternative methods such as retrospective reporting, dynamic modeling, and synthetic data can help bridge this gap when such data is unavailable.

The model's ability to perform well on unseen data, as shown in Eq. (11), is a crucial indicator of its generalization capacity, meaning it can extend beyond the specific training set and offer reliable predictions in real-world scenarios. This is particularly important for applications in mental health, where individual variations in feelings and activities can differ greatly from person to person. Achieving a high level of accuracy across a diverse set of participants suggests that the fuzzy logic-based rules of the model are robust and adaptable to various patterns of human behavior. This study enhanced mental health predictions, bridging theory and practice effectively.

Moreover, further research, including a more varied sample representing many populations, could improve generalizability. This could involve expanding the sample size and including individuals from diverse and inclusive demographic backgrounds.

6. Theoretical and Managerial Implications

6.1. Theoretical implications

- This study enriches the theoretical understanding of using fuzzy logic to assess complex and nuanced human conditions such as depression. Due to rigid classification boundaries, traditional statistical or rule-based models typically face challenges interpreting ambiguous emotional and behavioral data. However, fuzzy logic's capacity to handle linguistic variables and uncertainty allows for more flexible and realistic modeling.

- This opens pathways for theoretical advancements in AI applications within psychology, indicating that AI models may need to incorporate human-like interpretation capacities to address complex emotional data.

- By effectively using linguistic descriptors of psychological traits (such as "pessimistic thinking" or "sleep disorders"), this study provides theoretical evidence supporting the integration of qualitative language-based inputs in predictive models. Traditional models might require quantifiable metrics, but this study shows that subjective descriptions produce accurate predictions when incorporated with fuzzy rules.

- This validation aligns with cognitive-behavioural assessment theories, supporting the view that linguistic self-descriptions reveal underlying mental states.

- The model's use of only 15 features from an initial set of 30, as shown in Table 6, without losing prediction accuracy, provides theoretical insight into the efficiency of feature selection in mental health diagnostics.

- This implication also supports research into dimensionality reduction techniques, contributing to the theoretical framework of data science by demonstrating the practicality of smaller, well-chosen feature sets in improving model interpretability and efficiency.

6.2. Managerial implications

- With its high accuracy and ability to handle ambiguous inputs, the proposed fuzzy logic-based model offers a practical tool for frontline healthcare providers.

- This model can be an auxiliary tool for general healthcare practitioners or mental health counselors, offering initial insights before more comprehensive diagnostic methods are applied.

- Such predictive tools assist mental health organizations in managing caseloads effectively, which is especially beneficial for clinics, hospitals, or social services departments facing high demand.

- As this model operates effectively without needing sophisticated machine learning infrastructures, it presents an opportunity for mental health interventions in low-resource or remote areas.

- Educational institutions and community health programs could also use such predictive tools for early interventions, integrating them into support services.

- Policymakers could leverage such predictive data to advocate for mental health funding, as the model's insights can highlight the prevalence and risk factors associated with depression.

- Additionally, such fuzzy logic-based solutions can also be applied to other domains, such as cybersecurity Sarker (2023). As an example, ambiguous Internet traffic patterns can be evaluated to detect unusual or suspicious network activity.

7. Conclusion and Future Work

This study offers a broad scope for future research. Data in this research were collected from urban youths and students in Bangladesh. With better resources and more time, the focus range of the population in this study can be increased, enabling this model to predict a larger community. A similar process can be adopted in countries where symptoms may vary slightly. In addition, using this generalized model, we hope to build another model that may work separately for men and women by identifying how symptoms may vary in predicting the level of depression. Further, different people can suffer for different reasons, such as family problems, economic crises, etc. Once the specific root causes are identified, they can be treated more accurately.

An automated online counseling chatbot can be created with the assistance of psychiatrists to recommend users' next steps based on the depression inferred by the model. We also hope to detect and predict depression in individuals using a classifier hybrid Deep Learning Model and Neural Fuzzy Inference System and control the disease by implementing prevention methods such as meditation, physical activities, and work schedules. These can significantly impact an individual's emotions and activities, reducing their severity level.

In summary, this study enhances the role of fuzzy logic in accurately predicting depression, offering insights into mental health modeling and practical tools for efficient assessment and resource allocation.

We intend to collect more data and construct a modified rule-based model using our rule discovery process. Furthermore, conducting a user survey to assess the method's usability at the application level could be a future project. In addition, further research, including a more varied sample representing many populations, could improve generalizability. This could involve expanding the sample size and including individuals from diverse and inclusive demographic backgrounds.

Funding Declaration

Not applicable.

Declaration of Competing Interest

The authors declare no conflicts of interest related to this study.

Availability of Data and Materials

All the data and materials have been included in the submission. The data are available at https://github.com/urmi1504046/Depression-Level-Prediction

References

World Health Organization, Depressive disorder (depression), https://www.who.int/news-room/fact-sheets/detail/depression, 2023a. Accessed on: 19 November 2023.

World Health Organization, Suicide, https://www.who.int/news-room/fact-sheets/detail/suicide, 2023b. Accessed on: 19 November 2023.

S. Das, O. Hasan, A. Chowdhury, S. M. Aslam, S. M. Minhaz Hossain, An automatic detection of heart block from ecg images using yolov4, in: International Conference on Hybrid Intelligent Systems, volume 647, Springer, 2022, pp. 981–990. doi:10.1007/978-3-031-27409-1_90.

T. Das, S. Mobassirin, S. M. M. Hossain, A. Das, A. Sen, K. M. A. Kamal, K. Deb, Patient questionnaires based parkinson's disease classification using artificial neural network, Annals of Data Science 11 (2024) 1821–1864.

M. F. Faruque, A. Asaduzzaman, S. M. M. Hossain, M. H. Furhad, I. H. Sarker, Predicting diabetes mellitus and analysing risk-factors correlation, EAI Endorsed Transactions on Pervasive Health and Technology 5 (2019) e7–e7.

R. C. Ripan, I. H. Sarker, S. M. M. Hossain, M. M. Anwar, R. Nowrozy, M. M. Hoque, M. H. Furhad, A data-driven heart disease prediction model through k-means clustering-based anomaly detection, SN Computer Science 2 (2021) 112.

K. T. Alam, S. M. M. Hossain, M. S. Arefin, Developing a framework for analyzing social networks to identify human behaviors, in: 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), IEEE, 2016, pp. 1–4. doi:10.1109/ICECTE.2016.7879589.

N. Shoaip, S. El-Sappagh, T. Abuhmed, M. Elmogy, A dynamic fuzzy rule-based inference system using fuzzy inference with semantic reasoning, Scientific Reports 14 (2024) 4275.

S. Jana, A. Patel, J. Mahanta, Deriving fuzzy topological relations from incomplete observations, Journal of Geographical Systems 26 (2024) 117–147.

M. Sadaf, Z. Iqbal, Z. Anwar, U. Noor, M. Imran, T. R. Gadekallu, A novel framework for detection and prevention of denial of service attacks on autonomous vehicles using fuzzy logic, Vehicular Communications 46 (2024) 100741.

S. Taneja, B. Suri, H. Narwal, A. Jain, A. Kathuria, S. Gupta, A new approach for data classification using fuzzy logic, in: 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence), IEEE, 2016, pp. 22–27. doi:10.1109/CONFLUENCE.2016.7508041.

B. Ramesh, K. Lakshmanna, A novel early detection and prevention of coronary heart disease framework using hybrid deep learning model and neural fuzzy inference system, IEEE Access 12 (2024) 26683–26695.

S. K. Singh, V. Abolghasemi, M. H. Anisi, Fuzzy logic with deep learning for detection of skin cancer, Applied Sciences 13 (2023).

Z. Ruan, P. Yang, J. Huang, K. Yang, Y. Lv, Z. Zhang, Automatic depression detection among higher education students based on deepfm, IEEE Transactions on Instrumentation and Measurement 73 (2024) 1–10.

R. W. R. de Souza, J. V. C. de Oliveira, L. A. Passos, W. Ding, J. P. Papa, V. H. C. de Albuquerque, A novel approach for optimum-path forest classification using fuzzy logic, IEEE Transactions on Fuzzy Systems 28 (2020) 3076–3086.

Y. Seyfari, A. Meimandi, A new approach to android malware detection using fuzzy logic-based simulated annealing and feature selection, Multimedia Tools and Applications 83 (2024) 10525–10549.

M. Nilashi, O. Ibrahim, H. Ahmadi, L. Shahmoradi, A knowledge-based system for breast cancer classification using fuzzy logic method, Telematics and Informatics 34 (2017) 133–144.

S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, J. C. Eichstaedt, Detecting depression and mental illness on social media: an integrative review, Current Opinion in Behavioral Sciences 18 (2017) 43–49.

J. Mikal, S. Hurst, M. Conway, Investigating patient attitudes towards the use of social media data to augment depression diagnosis and treatment: a qualitative study, in: Proceedings of the fourth workshop on computational linguistics and clinical psychology—from linguistic signal to clinical reality, 2017, pp. 41–47. doi:10.18653/v1/W17-3105.

M. M. Aldarwish, H. F. Ahmad, Predicting depression levels using social media posts, in: 2017 IEEE 13th international Symposium on Autonomous decentralized system (ISADS), IEEE, 2017, pp. 277–280. doi:10.1109/ISADS.2017.41.

Z. Jamil, Monitoring tweets for depression to detect at-risk users, Ph.D. thesis, Université d'Ottawa/University of Ottawa, 2017. doi:10.20381/ruor-20310.

B. O'dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, H. Christensen, Detecting suicidality on twitter, Internet Interventions 2 (2015) 183–188.

H. Almeida, A. Briand, M.-J. Meurs, Detecting early risk of depression from social media user-generated content, in: CLEF (Working Notes), 2017.

M. R. Morales, Multimodal depression detection: An investigation of features and fusion techniques for automated systems, City University of New York, 2018.

C.-C. Wang, M.-Y. Day, C.-C. Chen, J.-W. Liou, Temporal and sentimental analysis of a real case of fake reviews in taiwan, in: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 729–736. URL: https://doi.org/10.1145/3110025.3116206. doi:10.1145/3110025.3116206.

A. G. Reece, C. M. Danforth, Instagram photos reveal predictive markers of depression, EPJ Data Science 6 (2017) 15.

Y. Lin, S. Huang, G. E. Simon, S. Liu, Data-based decision rules to personalize depression follow-up, Scientific reports 8 (2018) 5064.

S. Yang, P. Zhou, K. Duan, M. S. Hossain, M. F. Alhamid, emhealth: towards emotion health through depression prediction and intelligent health recommender system, Mobile Networks and Applications 23 (2018) 216–226.

I. A. Alshawwa, M. Elkahlout, H. Q. El-Mashharawi, S. S. Abu-Naser, An expert system for depression diagnosis, International Journal of Academic Health and Medical Research (IJAHMR) 3 (2019) 20–27.

Y.-C. Hsu, C. L. Wright, The association between participation in social activity and depressive symptoms in institutionalized elders in taiwan, Geriatric nursing 35 (2014) 31–36.

M. F. Steger, T. B. Kashdan, Depression and everyday social activity, belonging, and well-being., Journal of counseling psychology 56 (2009) 289–300.

A. S. Manstead, The psychology of social class: How socioeconomic status impacts thought, feelings, and behaviour, British Journal of Social Psychology 57 (2018) 267–291.

W. A. Hochwarter, P. L. Perrewé, J. A. Meurs, C. Kacmar, The interactive effects of work-induced guilt and ability to manage resources on job and life satisfaction., Journal of occupational health psychology 12 (2007) 125.

L. Yang, J. Yuan, H. Sun, Y. Zhao, J. Yu, Y. Li, Influencing factors of depressive symptoms among undergraduates: A systematic review and meta-analysis, Plos one 18 (2023) e0279050.

M. De Choudhury, M. Gamon, S. Counts, E. Horvitz, Predicting depression via social media, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 7(1), 2013, pp. 128–137. doi:10.1609/icwsm.v7i1.14432.

J. Hussain, M. Ali, H. S. M. Bilal, M. Afzal, H. F. Ahmad, O. Banos, S. Lee, Sns based predictive model for depression, in: International Conference on Smart Homes and Health Telematics, volume 9102, Springer, 2015, pp. 349–354. doi:10.1007/978-3-319-19312-0_34.

I. H. Sarker, Multi-aspects ai-based modeling and adversarial learning for cybersecurity intelligence and robustness: A comprehensive overview, Security and Privacy 6 (2023) e295.

Declaration of Interest Statement

No conflict of interest is declared related to this paper.