Heliyon 10 (2024) e37182

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

# Screening depression among university students utilizing GHQ-12 and machine learning

Nasirul Mumenin *, A.B.M. Kabir Hossain, Md. Arafat Hossain, Partha Pratim Debnath, Mursheda Nusrat Della, Md. Mahmudul Hasan Rashed, Afzal Hossen, Md. Rubel Basar, Md. Sejan Hossain

*Bangladesh Army University of Engineering and Technology, Rajshahi, Bangladesh*

ARTICLE INFO

*Keywords:*
Mental health
Depressive disorder
Machine learning
GHQ-12
University students

ABSTRACT

The escalating incidence of depression has brought attention to the increasing concern regarding the mental well-being of university students in the current academic environment. Given the increasing mental health challenges faced by students, there is a critical need for efficient, scalable, and accurate screening methods. This study aims to address the issue by using the General Health Questionnaire-12 (GHQ-12), a well recognized tool for evaluating psychological discomfort, in combination with machine learning (ML) techniques. Firstly, for effective screening of depression, a comprehensive questionnaire has been created with the help of an expert psychiatrist. The questionnaire includes the GHQ-12, socio-demographic, and job and career-related inquiries. A total of 804 responses has been collected from various public and private universities across Bangladesh. The data has been then analyzed and preprocessed. It has been found that around 60% of the study population are suffering from depression. Lastly, 16 different ML models, including both traditional algorithms and ensemble methods has been applied to examine the data to identify trends and predictors of depression in this demographic. The models' performance has been rigorously evaluated in order to ascertain their effectiveness in precisely identifying individuals who are at risk. Among the ML models, Extremely Randomized Tree (ET) has achieved the highest accuracy of 90.26%, showcasing its classification effectiveness. A thorough investigation of the performance of the models compared, therefore clarifying their possible relevance in the early detection of depression among university students, has been presented in this paper. The findings shed light on the complex interplay among socio-demographic variables, stressors associated with one's profession, and mental well-being, which offer an original viewpoint on utilizing ML in psychological research.

## 1. Introduction

Due to the unique challenges and complexities of academic life, university students are more prone to experiencing psychological distress, making their mental well-being a significant global concern. Depression and anxiety, which are prevalent mental health

* Corresponding author.
*E-mail addresses:* nmmouno@gmail.com (N. Mumenin), kabir111192@gmail.com (A.B.M. Kabir Hossain), arafat.bauet@gmail.com (M.A. Hossain), partha@bauet.ac.bd (P.P. Debnath), nusratdella.026@gmail.com (M. Nusrat Della), mahmudulhrashed@gmail.com (M.M. Hasan Rashed), afzal-2015818726@eee.du.ac.bd (A. Hossen), rubel24434@gmail.com (M.R. Basar), sejan.me@bauet.ac.bd (M.S. Hossain).

https://doi.org/10.1016/j.heliyon.2024.e37182
Received 18 May 2024; Received in revised form 22 August 2024; Accepted 28 August 2024

Available online 2 September 2024
2405-8440/

disorders among students, significantly impair their social interactions, academic achievements, and overall well-being. To provide timely intervention and aid, it is crucial to properly and promptly recognize these disorders. Depression is a common and troubling mental condition in modern times, with negative impacts on a person's emotions, thinking, and ability to operate. Approximately 264 million people worldwide suffer from depression [1]. Studies have shown that the occurrence of depression among college students is greater than in the general population, with rates ranging from 14% to 85% [2]. A recent research revealed a higher incidence of depression among students hailing from low- and middle-income nations. Specifically, 43.7% of Indian students, 40.9% of Pakistani students, and 52.2% of Bangladeshi students reported experiencing symptoms of despair [3] [4]. Depression, a substantial mental health concern that needs careful consideration, is nonetheless often ignored and, in many societies, even considered taboo [5]. As a result, many feel uncomfortable sharing the reasons for their mental disease with anybody, even psychiatrists and counselors. Studies indicate that over 66% of young individuals refrain from discussing or seeking help for mental health concerns, such as depression [6] [7]. Chronic depression often deteriorates progressively, and its outcomes may be very perilous, such as an elevated likelihood of self-inflicted damage, dropping out or failing in education, contemplating or engaging in suicide thoughts or actions, and other risky endeavors. The occurrence of suicide thoughts among university students ranges from 6.3% to 13.8%, with depression being present in around half of these instances among American and Bangladeshi students, respectively [8] [9]. Various socio-demographic and career-related issues have been recognized as risk factors for depression among university students, as shown by the performed research. Potential drivers of depression include age, gender, place of residence, academic success, kind of university, and other important socio-demographic characteristics [10] [11]. Claims about prospective joblessness, insufficient financial aid, hardships, dissatisfaction with academics, uncertainty about one's career, and other circumstances are instances of work-related stress that lead to the onset of depression [12] [13].

The combination of psychological evaluation instruments and machine learning methods has been used to improve early diagnosis and intervention options for depression among university students. The GHQ-12 has undergone thorough validation to confirm its effectiveness in identifying psychological distress and depression in different groups, such as university students and healthcare professionals. Zhong et al. [14] assessed the efficacy of the instrument among Chinese dental healthcare professionals during the COVID-19 epidemic and verified the strength of its two-factor model, indicating high reliability and validity. This discovery affirms the effectiveness of GHQ-12 in many contexts and emphasizes its potential for mental health assessment in academic settings. To validate the GHQ-12 against the gold standard of a diagnostic interview, Baksheev et al. [15] looked into the psychometric characteristics of the test among high school students and proposed a threshold score for the diagnosis of anxiety and depression disorders. Their obtained results from ROC curve showed that the GHQ-12 identified anxiety and depression illnesses more accurately. In comparing GHQ-12 studies, Armino et al. [16] found possible risk variables for anxiety and depression as well as methodological issues for further investigations. Ozdemir and Rezaki [17] shown that university students may be screened for depression using the GHQ. They came to the conclusion that, particularly in health facilities without a psychiatrist, the GHQ is a useful instrument for detecting depression in children. By use of a population-based cohort, Lundin et al. [18] validated the GHQ-12 against standardized psychiatric interviews of depression. They discovered the GHQ-12 did very well when graded using the Likert and Corrected techniques. Performance in identifying depression disorder in the general population was adequate when graded using the Standard approach. Qin et al. [19] The mental health of nursing students was evaluated throughout their academic journey using the GHQ-12. The results showed a notable rise in psychological distress and symptoms of depression from the start to the conclusion of their studies. This research emphasizes the long-term effects of academic demands and the need for continuous support systems in educational environments. A meta-analysis focused on the pooled prevalence of depressive symptoms among medical students using various assessment tools, including GHQ [20]. It reported significant variations in depressive symptoms across different study years and regions, emphasizing the need for region-specific mental health strategies and support systems for medical students. Jiang et al. [21] conducted a study on Chinese university students using a dual-factor mental health model, combining negative and positive indicators, to analyze the students' mental health profiles. The researchers used a large set of items linked to depression and well-being for this analysis. This research has identified specific groups of mental health conditions among students, indicating the need for customized intervention approaches that are based on a detailed knowledge of the different mental health profiles of students. Anjara et al. [22] examined the feasibility of employing the GHQ-12 as a screening tool for mental health issues in a primary care environment. This research emphasized the significance of using the GHQ-12 to identify people who may have mental health problems, which may then serve as a foundation for conducting more comprehensive psychiatric assessments. Moreover, longitudinal research have yielded valuable insights into the development of psychological distress and depression in academic environments. In a longitudinal study by Sonmez et al. [23], nursing students were assessed using the GHQ-12 questionnaire. The results showed notable elevations in psychological distress and depressive symptoms during the academic years, emphasizing the need for specific mental health treatments. James et al. [24] used the GHQ-12 questionnaire to detect medical students who were prone to experiencing academic challenges. This research prospectively assessed the relationship between GHQ-12 levels and academic achievement. This methodology facilitates the prompt detection and assistance for students who may encounter academic difficulties as a result of mental health conditions.

Many researchers focused on ML models for the precise detection of psychological problems. In a study by Elovanio et al. [25], the application of ML techniques in large-scale health studies has been examined, with the random forest (RF) model being employed to assess the predictive validity of psychological scales like GHQ-12, Beck Depression Inventory, and Mental Health Index. The findings indicate that while ML can enhance the analysis, the predictive accuracy for specific outcomes like mental health service use remains challenging. A comprehensive review by Lee and Ham [26] highlights the versatility of ML methods such as logistic regression (LR), RF, support vector machines (SVM), and artificial neural networks (ANN) in diagnosing depression across various types of data.

The study emphasized the importance of selecting appropriate ML models based on the data characteristics to optimize performance metrics like accuracy, sensitivity, and specificity. Sutter et al. [27] created an ML model using many methods to forecast psychological discomfort only from ecological parameters. Eight different classification methods were used on a sample dataset. Preliminary findings indicate that an accurate and trustworthy model is feasible with further enhancements to analysis and implementation. Purwandari et al. [28] want to use ML models to predict Internet addiction disorder in young people to promote early treatment. ML was used to match the input with the respondents' answers to the GHQ-12 questionnaire to categorize the input into their general health status. Baird et al. [29] used ML methods to decipher how psychological stress among refugee children relates to digitally recorded aspects of their artwork. They discovered that the elements of children's drawings maintained by ML match previous associations discovered in clinical settings between certain drawing qualities and psychological discomfort. Bieliński et al. [30] looked at how ML algorithms handle the issue of creating a virtual mental health index. Building many ML models using a clinical dataset was evaluated according to learning time, running time during usage, and regression accuracy. Using three online forms and ML to identify the data, Mohammad and Siddiqui [31] concentrated on gathering mental health data. Every questionnaire trains a single model using hyperparameter optimization methods after RF classifiers and regressors. With the classification technique, Munir et al. [32] presented an ML model that forecasts the likelihood of depression among people. Their attention was also on pinpointing the main risk factors that lead to depression among college students. Meda et al. [33] emphasized the alarming mental health condition of university students in comparison to their classmates and other occupational groups. The study used several regression models and supervised ML algorithms to identify economic concern and cognitive and physical symptoms of depression as significant determinants of mental health outcomes. Although the models successfully accurately forecasted students' well-being, they had difficulties predicting the deterioration of symptoms. This suggests that more study is necessary, engaging persons who have personal experiences, to improve the accuracy of predictions and effectively address the mental health requirements of students.

Recent advancements in ML offer promising alternatives to traditional statistical approaches, providing the capability to analyze large datasets with high dimensionality and uncover intricate patterns that may indicate the presence of psychiatric conditions. We have utilized sixteen ML models, namely LR, Naive Bayes (NB), Decision Tree (DT), RF, K-Nearest Neighbors (KNN), SVM, Gradient Boosting Machine (GBM), Light Gradient Boosting Machine (LGBM), AdaBoost (AB), CatBoost (CB), XGBoost (XGB), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Stochastic Gradient Descent (SGD), ET, and ANN. By integrating ML models with GHQ-12 data, there is the potential to improve the accuracy and efficiency of detecting depression and anxiety among university students. This research examines the use of several machine learning algorithms to evaluate data from the GHQ-12 questionnaire, which was gathered from university students throughout the country. The aim of this study is to analyze the efficacy of these models in identifying depression and anxiety disorders, therefore enhancing mental health evaluation methods in educational environments. The objective of this project is to narrow the gap between traditional psychiatric screening techniques and cutting-edge data-driven methodologies. This will improve the early identification of mental health issues and lead to improved mental well-being for university students. The significant contributions of our research work are:

1. This research provides new insights into the patterns and prevalence of depressive symptoms in a diverse university student population. Also, a new questionnaire has been developed for screening depression among university students.
2. The integration of ML models for enhancing the potential to improve the accuracy and sensitivity of detecting depressive disorders among university students. This is crucial for implementing timely and effective interventions, thereby reducing the long-term impact of these conditions.
3. By combining psychological assessment tools with advanced data analysis techniques, this study contributes to academic literature. It sets a precedent for future research on technology and mental health care.

The subsequent sections of this paper are organized in the following manner: Section two describes the dataset, experimental methodology, and ML techniques utilized in this study. The third section details the results of several investigations. In conclusion, section four delineates prospective obstacles and directions for the implementation of ML in the realm of depressive disorder detection.

## 2. Methodology

This task comprises several essential stages. The comprehensive workflow is illustrated in Fig. 1. We began by consulting with a psychiatrist and creating the appropriate questionnaire for the task. Then, we have collected data from across different universities in Bangladesh. After that, the dataset is examined for missing values, and the categorical values present are converted to numeric values. The data are preprocessed and analyzed then. Following the conclusion of the preprocessing stage, ML models are developed. Subsequently, the models undergo testing using the test data after being instructed using the training data.

### *2.1. Dataset*

#### *2.1.1. Questionnaire development*

Conducting a literature evaluation was our initial objective to identify prevalent risk factors. Because these risk factors may differ internationally, we have consulted a psychiatrist, reviewed pertinent literature [32], and supplemented our findings with a few variables unique to Bangladesh. In conclusion, a survey instrument has been developed comprising thirty inquiries organized

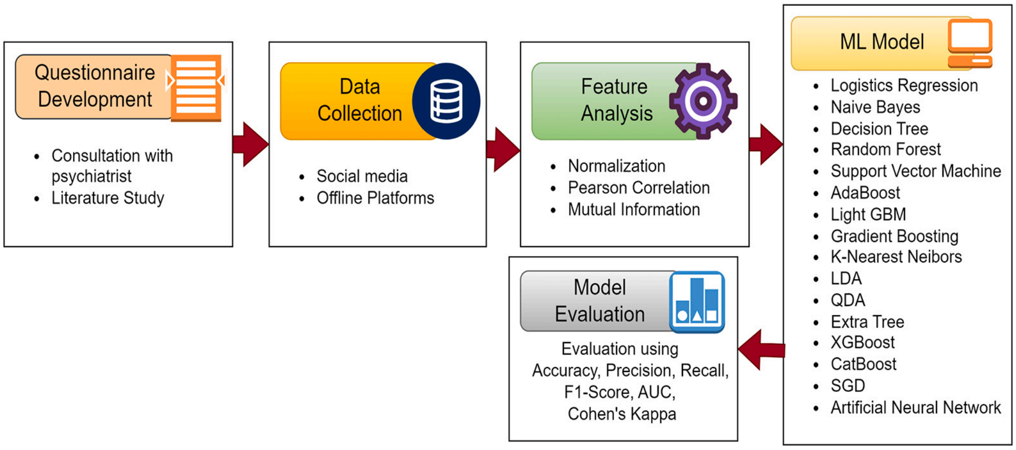


**Fig. 1.** Steps and workflow of the proposed work. The study begins with the development of the dataset and data collection. Then, feature analysis was performed, and ML models were created. Lastly, the ML models have been evaluated using various metrics.

**Table 1**
Socio-demographic Questionnaires used in this study.

| Feature | Feature Name | Question | Response Type |
|---|---|---|---|
| | SD1 | What's your gender? | Binary Choice |
| | SD2 | What's your age? | Multiple choice |
| | SD3 | What's the financial condition of your family? | Multiple choice |
| | SD4 | Where is your permanent residence? | Binary Choice |
| Socio-Demographic | SD5 | What's your current semester? | Multiple choice |
| | SD6 | Do you play any type of sports? | Binary Choice |
| | SD7 | Do you play any type of Video games? | Binary Choice |
| | SD8 | What's your current CGPA? | Multiple choice |
| | SD9 | What is the reason behind choosing your current area of study? | Multiple choice |

into three distinct sections: socio-demographic measures (Q1–Q9), career and job-seeking stress-related queries (Q10–Q21), and the GHQ-12 scale (Q22–Q30).

The participants were requested to provide fundamental personal and educational details in response to socio-demographic inquiries. Personal information-related inquiries included sex, age, financial status, residence, involvement with sports, and video games [10]. Additionally, education-related data comprises the field of study, academic year, current CGPA, and rationale for selecting the present study area [11]. Table 1 presents the socio-demographic questionnaires used in this study. Concerns about career and job-seeking stress are categorized into four distinct sub-categories: personality stress, family and surroundings environment stress, job ability stress, and employment environment stress [12]. On a Likert scale consisting of four points, responses to each question vary from three (indicating "Strongly Disagree") to zero (indicating "Strongly Agree"). Job ability-related inquiries encompass concerns regarding an applicant's capabilities, including prior work experience and skill sets. Table 2 presents this study's career and job-seeking stress-related questionnaires.

### *2.1.2. GHQ-12*

The General Health Questionnaire-12 (GHQ-12) is a concise screening instrument that has been developed to identify possible psychological disorders in both the general populace and specific healthcare environments, such as outpatient psychiatric services or primary care [34] [35]. The GHQ-12, an instrument that was initially designed by David Goldberg in the 1970s, has since gained significant popularity as a tool for diagnosing non-psychotic psychiatric conditions, including anxiety and depression [36] [37]. It requests that the respondent compare their recent psychological functioning to their typical level of functioning to assess their current state. The GHQ-12 is comprised of twelve items that assess a variety of psychological symptoms, including sleep, tension, temperament, and social functioning difficulties. It is an effective instrument for rapidly filtering large populations due to its concise nature. Utilizing its sensitivity to transient psychiatric disorders, it effectively monitors fluctuations in mental health conditions. It finds application in diverse domains such as primary care, clinical research, and community surveys. Respondents can easily complete the questionnaire, and it is also simple for professionals to grade and interpret. It has undergone validation across multiple cultures, thereby enabling its application across diverse populations. It functions as an exceptional initial screening instrument to detect individuals who might benefit from additional, more comprehensive psychological evaluation. The GHQ-12 is a highly valuable instrument for identifying individuals who may be advantageous to additional mental health assessment; thus, it promotes timely intervention and assistance. Its extensive utilization and adoption serve as evidence of its dependability and credibility as a diagnostic instrument for mental health. Table 3 presents questions in the GHQ-12 questionnaire.

**Table 2**
Career and Job Seeking related questionnaires.

| Feature | Feature Name | Question | Response Type |
|---|---|---|---|
| Career Anxiety | C1 | What is the reason behind choosing your current area of study | Multiple choice |
| | C2 | I am worried because I do not have prior work/internship experience related to the subject I am studying | Multiple choice |
| | C3 | I need to find a job right after my graduation in order to solve my family's economic problems. | Multiple choice |
| | C4 | If I don't get a job then I will lose my prestige in the eyes of the people around me (family, friends the society) | Multiple choice |
| | C5 | I am uncertain about my expertise and interests | Multiple choice |
| | C6 | I don't have enough courage and favorable circumstances for being an entrepreneur | Multiple choice |
| | C7 | I am worried because of scant job opportunities due to the potential economic recession worldwide | Multiple choice |
| | C8 | I am worried because of fierce competition created by the increasing numbers of people | Multiple choice |
| | C9 | I feel extra pressure from the people around me(family, friends relatives) to get involved in a job as soon as possible | Multiple choice |
| | C10 | I feel that our university study does not provide us with actual employment-related education | Multiple choice |
| | C11 | I think I am not confident enough to show my skills properly in the job market | Multiple choice |

**Table 3**
The questions of GHQ-12 questionnaire.

| Type | Feature | Question | Response Type |
|---|---|---|---|
| General Health Questionnaire | G1 | Have you recently been able to concentrate on whatever you're doing | MCQ |
| | G2 | Have you recently lost much sleep over worry | MCQ |
| | G3 | Have you recently felt you were playing a useful part in things | MCQ |
| | G4 | Have you recently felt capable of making decisions about things | MCQ |
| | G5 | Have you recently felt constantly under strain | MCQ |
| | G6 | Have you recently felt you couldn't overcome your difficulties | MCQ |
| | G7 | Have you recently been losing confidence in yourself | MCQ |
| | G8 | Have you recently been able to face up to your problems | MCQ |
| | G9 | Have you recently been able to enjoy your normal day-to-day activities | MCQ |
| | G10 | Have you recently been thinking of yourself as a worthless person | MCQ |
| | G11 | Have you recently been feeling reasonably happy, all things considered | MCQ |
| | G12 | Have you recently been feeling unhappy and depressed | MCQ |

A variety of methods exist for scoring the GHQ-12; however, the Likert scale (0-1-2-3) is a prevalent approach [38] [39]. In accordance with their recent experience (typically within the last few weeks), respondents are required to assess one statement from each of the twelve items comprising the GHQ-12. The inquiries have been formulated to evaluate the improvements or declines in your mental well-being. Responses are provided in four options for each topic. Conditional statements typically span a spectrum of positivity levels, including "Much more than usual," "Not at all," "No more than usual," and "Rather more than usual." The scoring system for each response option is uncomplicated, ranging from zero to three. Typically, "Much more than usual" receives a score of three, while "Not at all" gets no score. Greater distress is indicated by higher scores on this scale, which is regarded as a standardized indicator of distress. The GHQ-12 comprises a combination of negatively and positively formulated inquiries. Examples of negatively phrased inquiries include "Have you recently experienced persistent strain?" and "Have you been able to enjoy your usual daily activities recently?" Response bias, which can result from identically worded queries, is intended to be mitigated by incorporating both categories of items. A greater balance is achieved in the evaluation of an individual's mental health through the use of this variety of question formats. Consideration is given to the wording of the questions when determining the score for each item on the Likert scale (0-1-2-3) [22]. In order to maintain a consistent indicative of heightened psychological distress throughout the questionnaire, negatively formulated inquiries are assigned direct scores (0-1-2-3), while positively phrased inquiries require reverse scoring (3-2-1-0) [40] [41] [42]. The threshold value for screening disorders with GHQ-12 depends on various aspects and may differ from region to region [37]. For this study, we have set the threshold value to 16, considering all socio-demographic and financial accepts. The value has been chosen after consultation with an expert psychiatrist.

### *2.2. Data preprocessing*

Data preparation plays a crucial role in the creation of ML models. Real-world data is frequently insufficient, inconsistent, inaccurate, and lacks essential attribute values. Data preparation involves cleaning, organizing, and formatting raw data to be suitable for ML models. In this study, we have employed a variety of data preparation approaches, such as:

1. Categorical data encoding: Category data encoding is the process of converting categorical variables into numerical variables to make them usable in ML models. Categorical data comprises variables grouped into different categories, such colors, locations, or types of items. Given that the majority of ML models rely on mathematical equations, it is essential to transform categorical data into numerical data to prevent any issues. We have converted the category data in the datasets into numerical values. We utilized the LabelEncoder() function from the sklearn package for this task.

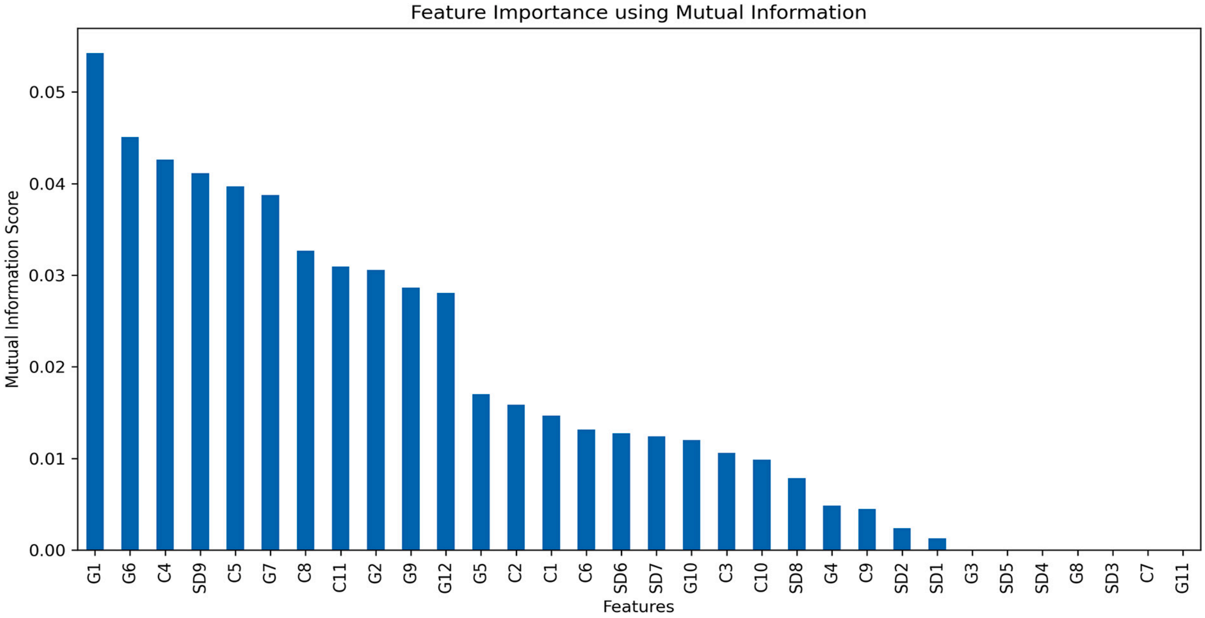


**Fig. 2.** This figure presents the feature importance graph calculated using Mutual Information. It depicts the mutual correlation among the features in the dataset.

2. Feature Scaling: A crucial component of our study was the preparation of the data, with a particular focus on feature scaling. Feature scaling is a crucial step in machine learning models as it helps to normalize the range of independent variables. This normalization assists algorithms in converging more quickly and obtaining better performance. By scaling the attributes of our dataset, we ensure that no one feature has a dominating impact due to its magnitude, taking into account the diverse nature of the dataset. In this work, we used data normalization to achieve feature scaling. Data normalization is an essential preprocessing technique that rescales the data attributes to a predetermined range, often with the goal of improving algorithm performance and ensuring consistent results. Normalization ensures that each feature has an equal impact on distance metrics and model calculations by scaling features to a range of 0 to 1 or by adjusting features to have a mean of zero and a standard deviation of one. This helps to eliminate biases associated with the original scale of the features [43].
3. Feature Importance Calculation: When building predictive models, it is crucial to understand the importance of each information in respect to the target variable. A very effective method for evaluating the amount of relevance is to use the calculation of Mutual Information (MI) scores. Unlike simpler linear metrics, MI provides a more comprehensive measure that may include any kind of link between variables, regardless of whether it is linear or nonlinear [44]. If two variables are independent, the resulting score is zero. On the other hand, a higher score indicates a stronger relationship or connection between the variables. The Mutual Information (MI) score of a feature in relation to the target variable quantifies the extent to which understanding the feature decreases uncertainty about the target. This is particularly relevant in the context of feature selection. Mutual information scores were used to determine the characteristics in the dataset that had the most relevance to the objective variable. In our specific situation, where our dataset consisted of both linear and non-linear correlations that the MI methodology could effectively capture, our strategy demonstrated significant benefits. The characteristics were ranked based on their Mutual Information (MI) scores, which offered a clear indication of which features may potentially be more influential in predicting the target variable. To simplify the model, we may select and keep just the most useful characteristics. This will reduce the risk of overfitting and make the model easier to comprehend. Fig. 2 presents the feature importance of the features calculated using mutual information scoring.
4. Feature Correlation Analysis: An essential component of statistical analysis and data exploration is understanding the connections between variables. To achieve this goal, we adopted the Pearson correlation coefficient, a widely used method for measuring the linear relationship between two continuous variables in our dataset. The Pearson correlation coefficient (r) may take on values ranging from -1 to +1. The strength of the connection is indicated by the size of the coefficient, while the sign denotes the direction. The Pearson correlation approach was used to analyze the associations between each pair of continuous variables in the dataset [45] [46]. By finding variables that are highly correlated, one might get useful insights into hidden patterns and linkages that may indicate fundamental causal ties or interdependencies within the dataset. A strong connection between two variables may indicate the presence of redundant work. In such cases, it is possible to exclude one of the variables to reduce dimensionality without a notable loss of information. This helps to simplify our model without affecting its predictive capabilities. The findings derived from the Pearson correlation analysis significantly enhanced our understanding of the dataset, allowing us to make more informed judgments in the later modeling stages. Fig. 3 presents the feature correlation of the features calculated using Pearson correlation.

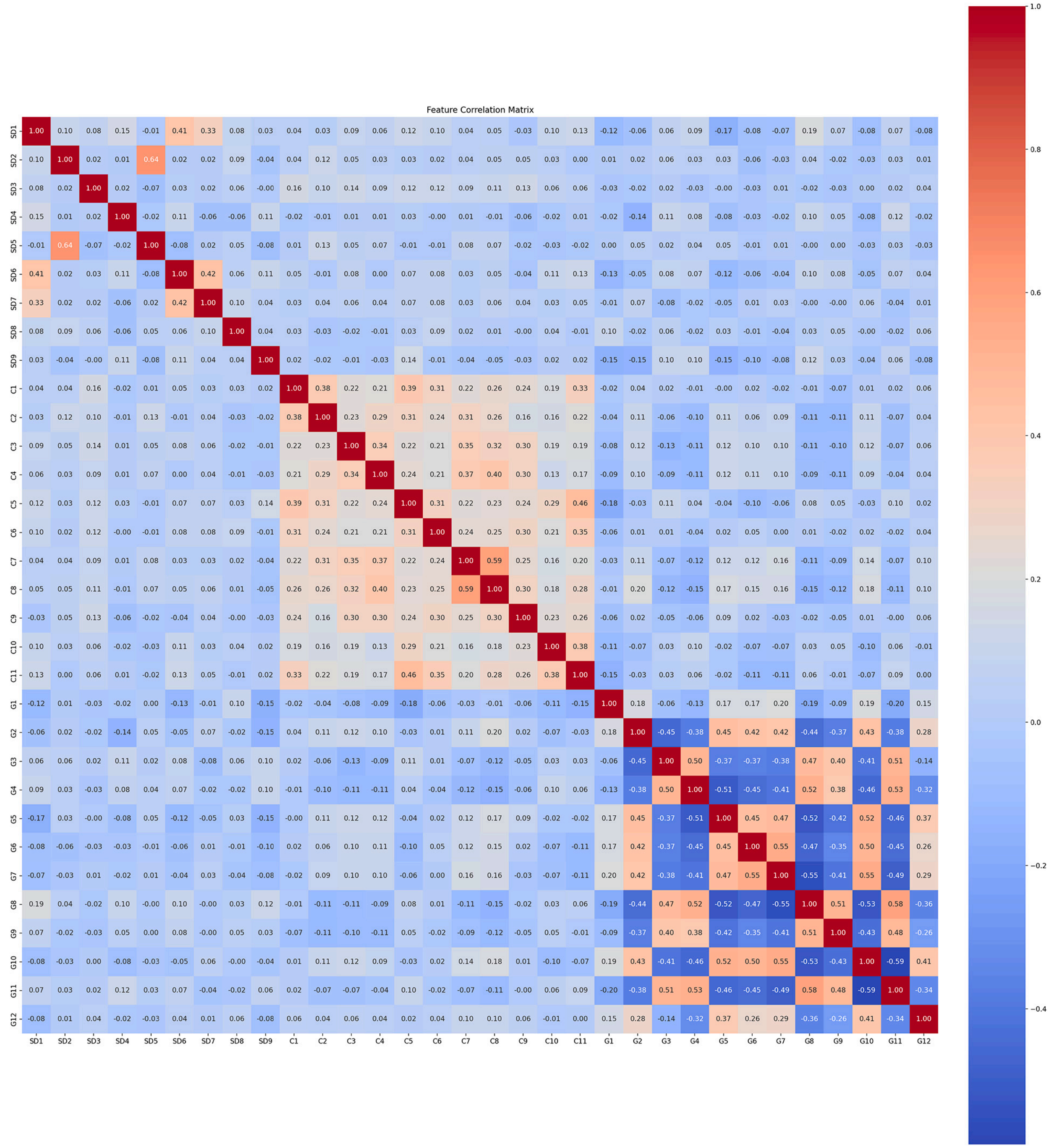


**Fig. 3.** Feature correlation heatmap generated using Pearson Correlation. The figure presents the correlation among features in the dataset.

5. Feature Reduction: The process of feature selection is crucial in determining the performance and interpretability of a model. We utilized Principal Component Analysis (PCA) for feature reduction technique. PCA is a widely used dimensionality reduction technique that aims to transform high-dimensional data into a lower-dimensional space while preserving as much of the original variance as possible [47]. PCA has emerged as a crucial technique in several domains, such as biological data analysis, where datasets often possess high dimensionality owing to the abundance of features or variables. Principal Component Analysis (PCA) is a technique that may decrease the number of dimensions in complex biomedical datasets without losing significant information. This process helps in visualizing and interpreting the data more easily and also improves computing efficiency. Principal Component Analysis (PCA) may assist in finding the most informative features or variables in the dataset by analyzing the contributions of each feature to the principal components. Principal Component Analysis (PCA) may effectively reduce the

**Table 4**
Distribution of data in response of Social and Demographic questionnaire.

| Feature | Label | Value count | Percentage |
|---|---|---|---|
| SD1 | Male (0) | 494 | 61.44 |
| | Female (1) | 310 | 38.56 |
| SD2 | 20-21 (0) | 402 | 50.00 |
| | 22-23 (1) | 317 | 39.42 |
| | 24-25 (2) | 54 | 6.72 |
| | 26-27 (3) | 22 | 2.74 |
| SD3 | Medium (0) | 492 | 61.19 |
| | Prefer not to answer (1) | 92 | 11.44 |
| | Solvent (2) | 127 | 15.8 |
| | Weak (3) | 93 | 11.57 |
| SD4 | Village (0) | 418 | 51.99 |
| | City (1) | 386 | 48.01 |
| SD5 | $1^{st}$ (0) | 150 | 18.66 |
| | $2^{nd}$ (1) | 130 | 16.17 |
| | $3^{rd}$ (2) | 103 | 12.81 |
| | $4^{th}$ (3) | 120 | 14.93 |
| | $5^{th}$ (4) | 100 | 12.44 |
| | $6^{th}$ (5) | 82 | 10.2 |
| | $7^{th}$ (6) | 81 | 10.07 |
| | $8^{th}$ (7) | 38 | 4.73 |
| SD6 | No (0) | 386 | 48.01 |
| | Yes (1) | 418 | 51.99 |
| SD7 | No (0) | 357 | 44.4 |
| | Yes (1) | 447 | 55.6 |
| SD8 | Excellent (0) | 270 | 33.58 |
| | Good (1) | 275 | 34.2 |
| | Fair (2) | 223 | 27.74 |
| | Poor (3) | 23 | 2.86 |
| | Very Poor (4) | 13 | 1.62 |
| SD9 | Financial reason (0) | 20 | 2.49 |
| | Family pressure (1) | 150 | 18.66 |
| | Friend's suggestions (2) | 30 | 3.73 |
| | Job opportunities (3) | 171 | 21.27 |
| | Personal interest (4) | 416 | 51.74 |
| | Social Influence (5) | 17 | 2.11 |

impact of noise and redundancy in biological data, resulting in higher signal-to-noise ratios and improved interpretability [48]. It can be difficult to ascertain the optimal number of features (k). Thus, we employed feature correlation and feature importance techniques. Finally, we selected 26 as the optimal number of features.

### *2.3. Data analysis*

The Table 4 provides a distribution of values for nine distinct socio-demographic features (SD1 to SD9) within a dataset, categorizing the entries across various labels within each feature. SD1 distinguishes between Male (0) and Female (1), with males representing a larger portion at 61.44%, compared to 38.56% for females. SD2 categorizes age groups, showing a dominant group of 20-21 years at exactly 50%, followed by 22-23 years at 39.42%, with the older age groups 24-25 and 26-27 years being significantly less common at 6.72% and 2.74% respectively. SD3 describes economic status, where 'Medium' (0) is most prevalent at 61.19%. The least frequent is 'Prefer not to answer' at 11.44%, with 'Solvent' and 'Weak' being slightly more common. SD4 differentiates between Village (0) and City (1) residents, showing a fairly even split with 51.99% living in villages and 48.01% in cities. SD5 represents an ordinal scale ranging from '1st' to '8th', where lower numbers have higher frequencies, gradually decreasing as the scale increases. The most frequent is '1st' at 18.66%, with '8th' being the least at 4.73%. SD6 and SD7 feature binary choices regarding specific preferences or conditions, with SD6 evenly split between 'No' and 'Yes', and SD7 showing a slight preference for 'Yes' at 55.6% over 'No' at 44.4%. SD8 details CGPA ratings from 'Excellent' to 'Very Poor'. The majority rate their condition as 'Good' (34.2%) or 'Excellent' (33.58%), with fewer entries for 'Fair' (27.74%), and very few for 'Poor' (2.86%) and 'Very Poor' (1.62%). SD9 explains reasons for choosing a certain course of action, where 'Personal interest' dominates at 51.74%, followed by 'Job opportunities' at 21.27% and 'Family pressure' at 18.66%. Other reasons like 'Financial reason' and 'Social Influence' are much less common. Each feature in the dataset is characterized by varying degrees of distribution across the specified categories, which illustrates diverse socio-demographic patterns within the population studied. This detailed breakdown is crucial for understanding the different demographic and socioeconomic

**Table 5**
Distribution of data in response of Career and Job Seeking questionnaire. Here, 0, 1, 2, 3 indicates Strongly Agree, Agree, Disagree, and Strongly Disagree respectively.

| Feature | Label | Value count | Count Percentage |
|---|---|---|---|
| C1 | 0 | 412 | 51.24 |
| | 1 | 220 | 27.36 |
| | 2 | 125 | 15.55 |
| | 3 | 47 | 5.85 |
| C2 | 0 | 457 | 56.84 |
| | 1 | 163 | 20.27 |
| | 2 | 154 | 19.15 |
| | 3 | 30 | 3.73 |
| C3 | 0 | 316 | 39.3 |
| | 1 | 110 | 13.68 |
| | 2 | 349 | 43.41 |
| | 3 | 29 | 3.61 |
| C4 | 0 | 325 | 40.42 |
| | 1 | 146 | 18.16 |
| | 2 | 294 | 36.57 |
| | 3 | 39 | 4.85 |
| C5 | 0 | 275 | 34.2 |
| | 1 | 371 | 46.14 |
| | 2 | 85 | 10.57 |
| | 3 | 73 | 9.08 |
| C6 | 0 | 344 | 42.79 |
| | 1 | 273 | 33.96 |
| | 2 | 116 | 14.43 |
| | 3 | 71 | 8.83 |
| C7 | 0 | 405 | 50.37 |
| | 1 | 77 | 9.58 |
| | 2 | 301 | 37.44 |
| | 3 | 21 | 2.61 |
| C8 | 0 | 435 | 54.1 |
| | 1 | 88 | 10.95 |
| | 2 | 259 | 32.21 |
| | 3 | 22 | 2.74 |
| C9 | 0 | 280 | 34.83 |
| | 1 | 292 | 36.32 |
| | 2 | 172 | 21.39 |
| | 3 | 60 | 7.46 |
| C10 | 0 | 238 | 29.6 |
| | 1 | 368 | 45.77 |
| | 2 | 113 | 14.05 |
| | 3 | 85 | 10.57 |
| C11 | 0 | 331 | 41.17 |
| | 1 | 324 | 40.3 |
| | 2 | 80 | 9.95 |
| | 3 | 69 | 8.58 |

segments represented in the data. The Table 5 presents the distribution of values for eleven features (C1 to C11) in a dataset, with each feature divided into four categories (0 to 3). The data for each category is provided in terms of both count and percentage, allowing for an analysis of how values are distributed within each feature. Feature C1 shows a clear majority for category '0', making up over half of the entries. The subsequent categories decrease in frequency. Feature C2 has a dominant category '0' with 56.84%, but categories '1' and '2' also hold significant shares, indicating a slightly more even distribution than C1. Feature C3 is unique, with category '2' being the most common at 43.41%, closely followed by category '0'. Feature C4 also has a large presence of categories '0' and '2', collectively accounting for over three-quarters of the data. Feature C5 stands out with category '1' being the most prevalent, contrary to the dominance of category '0' in several other features. Feature C6 shows a more balanced distribution with category '0' leading slightly. Feature C7 exhibits a strong skew towards category '0', and much less representation in the other categories. Feature C8 follows the common pattern with category '0' at 54.1%, and the rest of the categories showing decreasing frequencies. Feature C9 has a fairly even spread between categories '0' and '1', making it one of the more balanced features. Feature C10 and C11 both show a predominant category ('1' in C10 and '0' in C11), but also display significant percentages in other categories, suggesting a reasonable

**Table 6**
Distribution of data in response of GHQ-12 questionnaire. Here, 0, 1, 2, 3 indicates Not at all, No more than usual, Rather more than usual, and Much more than usual respectively.

| Feature | Label | Value count | Count Percentage |
|---|---|---|---|
| G1 | 0 | 45 | 5.6 |
| | 1 | 160 | 19.9 |
| | 2 | 455 | 56.59 |
| | 3 | 144 | 17.91 |
| G2 | 0 | 260 | 32.34 |
| | 1 | 270 | 33.58 |
| | 2 | 188 | 23.38 |
| | 3 | 86 | 10.7 |
| G3 | 0 | 96 | 11.94 |
| | 1 | 196 | 24.38 |
| | 2 | 258 | 32.09 |
| | 3 | 254 | 31.59 |
| G4 | 0 | 109 | 13.56 |
| | 1 | 151 | 18.78 |
| | 2 | 225 | 27.99 |
| | 3 | 319 | 39.68 |
| G5 | 0 | 378 | 47.01 |
| | 1 | 207 | 25.75 |
| | 2 | 130 | 16.17 |
| | 3 | 89 | 11.07 |
| G6 | 0 | 274 | 34.08 |
| | 1 | 255 | 31.72 |
| | 2 | 182 | 22.64 |
| | 3 | 93 | 11.57 |
| G7 | 0 | 290 | 36.07 |
| | 1 | 275 | 34.2 |
| | 2 | 152 | 18.91 |
| | 3 | 87 | 10.82 |
| G8 | 0 | 121 | 15.05 |
| | 1 | 168 | 20.9 |
| | 2 | 210 | 26.12 |
| | 3 | 305 | 37.94 |
| G9 | 0 | 89 | 11.07 |
| | 1 | 164 | 20.4 |
| | 2 | 269 | 33.46 |
| | 3 | 282 | 35.07 |
| G10 | 0 | 383 | 47.64 |
| | 1 | 220 | 27.36 |
| | 2 | 114 | 14.18 |
| | 3 | 87 | 10.82 |
| G11 | 0 | 88 | 10.95 |
| | 1 | 130 | 16.17 |
| | 2 | 237 | 29.48 |
| | 3 | 349 | 43.41 |
| G12 | 0 | 599 | 74.5 |
| | 1 | 101 | 12.56 |
| | 2 | 64 | 7.96 |
| | 3 | 40 | 4.98 |

distribution. In summary, each feature has its unique distribution pattern, with some showing a significant skew towards a particular category, while others are more balanced. This variety in categorical distribution is crucial for understanding the underlying structure of the dataset and informs decisions in data processing and modeling.

Similarly, the Table 6 displays the distribution of values across twelve categorical features (G1 to G12), each divided into four categories (0 to 3). G1 is dominated by category '2' with 56.59% of entries, followed by '3', '1', and '0', suggesting a strong skew towards the middle categories. G2 displays a more balanced distribution among the categories, with '1' slightly leading at 33.58%, followed closely by '0', '2', and '3'. G3 shows a relatively uniform distribution among the higher categories ('2' and '3'), both exceeding

30%, and lower presence in '0' and '1'. In G4, category '3' is the most common at 39.68%, with the remaining categories distributed in decreasing order from '2' to '0'. For G5, category '0' holds the majority at 47.01%, with reduced frequencies through '1', '2', to '3'. G6 presents a fairly even spread across the first three categories, with '0' leading slightly over '1' and '2', and '3' being the least frequent. G7 is similar to G6 and shows a balanced distribution, with '0' slightly more common at 36.07%, and the rest of the categories fairly close in frequency. G8 is marked by a high frequency of category '3' at 37.94%, with the other categories represented less frequently, indicating a skew towards the highest category. G9 features a near uniform distribution with '3' as the most frequent at 35.07% and the others fairly balanced. G10, Like G5, is heavily skewed towards '0' at 47.64%, with other categories appearing less frequently. G11 shows a significant skew towards '3', which is the most prevalent at 43.41%, followed by '2', '1', and '0'. G12 is highly skewed with '0' representing a vast majority at 74.5%, significantly more than any other category, making it the most unbalanced feature in terms of category representation. Each feature demonstrates unique characteristics in how categories are distributed, ranging from highly skewed to fairly balanced, which is critical for understanding data dynamics and guiding data preprocessing steps in analysis or modeling efforts.

## 2.4. Machine learning models

In this investigation, we implemented ML techniques widely utilized across industries due to their straightforwardness and generalization ability. The methodologies employed in this investigation are delineated as follows:

### 2.4.1. Logistic regression

As a predictive analysis algorithm, LR is predominantly applied to problems involving binary classification [49]. It employs a logistic function, which is bounded between 0 and 1, to estimate probabilities. This feature renders it an exceptional instrument in situations where the probability of an occurrence must be predicted, such as determining the spam status of an email or the malignant or benign nature of a tumor. LR is fundamentally concerned with generating probability scores for observations and classifying them into two distinct categories via a decision threshold (typically set at 0.5).

### 2.4.2. Naive Bayes

NB is an algorithm for ML and predictive modeling that is both straightforward and potent. It operates under the assumption that the existence of a specific feature in a given class is independent of the existence of any other feature (hence the term "naive") [50]. It is founded upon Bayes' theorem. Although NB simplifies the process, the resulting models can be quite accurate, particularly when used for tasks such as sentiment analysis, spam detection, and document classification. Known for its efficiency and speed, the algorithm is well-suited for high-dimensional data and is frequently employed in text classification tasks involving sizable datasets, where the algorithm's simplicity can confer a substantial benefit.

### 2.4.3. Decision tree

A DT is an algorithm for supervised learning that is non-parametric and is utilized for both classification and regression tasks [51]. A DT, at its essence, employs a tree-like representation of choices and their potential ramifications, encompassing resource expenses, utility, and random event outcomes. It commences with a solitary node that branches into potential outcomes; each of those branches subsequently connects to additional nodes that branch off into further possibilities. The aforementioned procedure persists until it reaches a leaf node, which subsequently yields the DT's output pertaining to the corresponding input features. The routes connecting the root to the leaf signify classification rules or regression routes. At each node, a determination is executed using regression estimates or the attribute that provides the most effective separation of classes, as determined by a specific criterion. The straightforwardness of DTs facilitates their comprehension, representation, and elucidation, thereby substantially bolstering their prevalence in decision-making endeavors that demand transparency.

### 2.4.4. Random forest

Data scientists employ the RF algorithm frequently; it is among the most well-known algorithms. It is an often implemented supervised ML algorithm to address classification and regression issues. The algorithm is composed of numerous DTs, each of which processes a distinct subset of the dataset and calculates the mean in order to improve the prediction's precision. EL, which involves incorporating multiple classifiers to solve a complex problem and enhance the performance of the model, forms the foundation of the strategy. RF is an EL technique that reduces overfitting and outperforms a singular DT by aggregating the results [52].

### 2.4.5. K-nearest neighbors

KNN is an instance-based or lazy learning technique that approximates the function locally and delays all computation until function evaluation [53]. It is a versatile tool utilized for both classification and regression problems because to its simplicity and effectiveness. The K-NN algorithm operates by calculating the distances between a query and all data instances, choosing the designated number "K" of the nearest examples, and then either picking the most common label (for classification) or averaging the labels (for regression). Choosing the parameter K is crucial. A lower K number increases the impact of noise on the output, while a large value leads to increased computational costs and potential inclusion of points from other classes [54]. KNN is obvious and simple to understand; however, its computational speed decreases notably as the dataset size increases.

### 2.4.6. Support vector machine

Regression and classification problems are two applications for the powerful supervised ML method SVM. Still, classification problems are where it is mostly used. By finding the hyperplane that divides the dataset into distinct groups the best, SVM classifies data [55]. SVM seeks for the optimal hyperplane that maximizes the margin between different classes in the training set. The data points closest to the hyperplane, or support vectors, greatly influence the direction and position of the hyperplane. Support vectors are used by SVM to increase the margin between classes, hence improving classification accuracy. Widely applicable to various data formats and prediction challenges, the SVM method can manage linear and non-linear separations utilizing kernel functions [56].

### 2.4.7. Gradient boosting machine

Gradient Boosting Machine (GBM) [57] [58] is a potent ML method that enhances DTs by adding weak learners sequentially to form a robust prediction model. Every new tree aims to rectify the mistakes of the trees constructed before it. GBM employs the gradient descent approach to reduce errors in sequential models. It modifies the importance of a data point according on its prior categorization. When an observation is misclassified, its weight increases, and vice versa. This method enables the model to focus more on challenging situations for classification, leading to enhanced accuracy. GBM is versatile, applying to both regression and classification tasks, and has proven effective in addressing several real-world situations. GBM is a popular choice among data scientists for achieving excellent performance in predictive jobs due to its ability to handle different data types and produce strong predictions.

### 2.4.8. AdaBoost

AB, also known as Adaptive Boosting [59] [60], is an ensemble learning technique mainly utilized for binary classification tasks. AB's fundamental concept is amalgamating several weak classifiers to form a robust classifier. A weak classifier is a classifier that performs marginally better than random guessing. AB gives weights to each training instance and adjusts them as training advances. Classifiers are trained successively, with each new classifier emphasizing the training cases misclassified by preceding classifiers. The final prediction is generated by combining the predictions of all classifiers using a weighted majority vote or total. Boosting's adaptability stems from its emphasis on classifying challenging examples and assigning greater weight to classifiers with superior performance. AB enhances the accuracy of weak learning models, making it a potent tool for improving model performance.

### 2.4.9. XGBoost

XGBoost is short for eXtreme Gradient Boosting and is a fast and high-performance implementation of gradient boosted DTs. XGBoost is a scalable and precise version of gradient boosting machines that has been the dominant method in several ML contests. XGBoost offers parallel tree boosting, often known as GBDT or GBM, to efficiently and effectively address various data science challenges. XGBoost's main characteristics include managing missing data, using regularization to avoid overfitting, and performing both linear model solving and tree learning [61] [62]. The system is very adaptable, enabling users to establish personalized optimization goals and assessment standards, which enhances its possibilities. XGBoost's effectiveness and efficiency have established it as a favored option for data scientists and ML professionals.

### 2.4.10. Light gradient boosting machine

LGBM, akin to alternative gradient-boosting methodologies, employs an incremental model construction approach. It constructs a series of DTs, wherein each succeeding tree is designed to correct the errors committed by its antecedent [63] [64]. Assembled from the final model, this weighted sum of the individual trees. The model is denoted as follows:

$$F(x) = \sum_{i=1}^{N} f_i(x) \tag{1}$$

where NN represents the quantity of trees and $f_i(x)$ denotes the forecast of the i-th tree. The process of training entails the minimization of a loss function. In LightGBM, the objective function is composed of two components: the regularization term and the loss function. The mathematical formula has been presented in Eq. (1).

### 2.4.11. Extremely randomized tree

A member of the tree-based algorithm family, the ET algorithm, which is also referred to as Extremely Randomized Trees, is an EL method [65]. The output class is determined by the system using the mean prediction (regression) or mode of the classes (classification) of the individual trees, which are obtained after the system generates a large number of DTs during the training phase. Randomization, which increases the variance of the model in order to reduce the likelihood of overfitting, is the foundational principle that underpins the ET algorithm. ET aims to further develop this concept by incorporating arbitrary thresholds for each feature at each split, as opposed to alternative tree-based methods like RF, which employ the bagging approach and integrate randomization during feature selection at each split. When implemented on large datasets and in scenarios where the model must have low variance, it demonstrates significant effectiveness. Baseline modeling is achieved with great suitability.

### 2.4.12. CatBoost

CB is a gradient boosting algorithm that has been purposefully developed to efficiently process categorical features. The system integrates novel methodologies to attain optimal performance and resilience, specifically when confronted with situations involving diverse data types and extensive datasets [66]. CB minimizes a differentiable loss function L($y_i$,F($x_i$)), where $x_i$ is the feature vector, $y_i$

is the target variable, and $F(x_i)$ is the predicted value for the i-th instance. To minimize the loss function, CB assembles an ensemble of DTs in a sequential fashion. The model acquires knowledge of the gradient of the loss function in relation to the preceding predictions at each iteration t [67]. CB is a library for gradient boosting that has been purposefully developed to handle categorical features. It employs a gradient-boosting implementation with DTs and integrates innovative methods to efficiently manage categorical variables. A differentiable loss function is optimized by CB via gradient boosting with DTs. In order to effectively manage categorical features, the algorithm employs a number of strategies, including encoding techniques and specific treatment during tree construction.

#### *2.4.13. Linear discriminant analysis*

LDA is a supervised ML technique specifically designed for classification applications [68] [69]. It is a method employed to identify a linear combination of characteristics that most effectively distinguishes the different classes within a dataset. LDA functions by mapping the data onto a reduced-dimensional space that optimizes the distance between the different classes. It does this by identifying a group of linear discriminants that optimize the ratio of variance across classes to variance within classes. It identifies the directions in the feature space that most effectively distinguishes between the various data classes. LDA presupposes that the data follows a Gaussian distribution and that the covariance matrices of the various classes are identical. The assumption is that the data is linearly separable, allowing a linear decision boundary to effectively categorize the various classes.

#### *2.4.14. Quadratic discriminant analysis*

QDA is a method closely linked to LDA, based on the assumption that the data from each class follows a normal distribution. QDA does not assume that the covariance of each class is the same, unlike LDA [70]. The likelihood ratio test is the most suitable test for determining if a given measurement belongs to a certain class while the normalcy assumption holds true. QDA provides greater flexibility for the covariance matrix, resulting in better data fitting compared to LDA, although it requires estimating more parameters. The quantity of parameters rises substantially with QDA [71]. QDA involves having a distinct covariance matrix for each class. Having a large number of classes and a limited number of sample points might provide a challenge.

#### *2.4.15. Stochastic gradient descent*

In ML, SGD is a widely used optimization technique for training a variety of models, most notably in classification problems. Particularly well-suited for sparse and large-scale ML tasks, it is often used in natural language processing and text categorization [72]. Instead of updating the model parameters by computing the gradients based on the whole dataset as in classic gradient descent, SGD updates the parameters gradually for every training sample. For big datasets, this method is quicker and more scalable since it greatly reduces the computing load. SGD is a main approach in many ML applications as it has been shown to converge quicker when the dataset is big and the features are many [73].

#### *2.4.16. Artificial neural network*

A feed-forward Neural Network is a type of network that creates a directed graph with nodes and edges. Data is transmitted along these edges from a node to the next without forming a cycle. The ANN [74] [75] is a variant of FFN with three or more layers: an input layer, one or more hidden layers, and an output layer. Each of these layers contains many neurons or units, as defined in mathematical notation. To determine the number of hidden layers in an ANN, a hyperparameter tuning strategy is employed [76] [77]. Information is transferred from one layer to the next without taking into account previous values, and all neurons in each layer are connected, as documented in sources [78] [79].

## 3. Evaluation

### *3.1. Evaluation metrics*

Several evaluation metrics has been utilized to evaluate the effectiveness of the proposed model, i.e., *Precision*, *Recall*, $F_1$ *Score*, *Accuracy*, *Cohen's Kappa Score*, and *AUC-Score*. These have been used to evaluate the ML models. In the following equations, TP represents true positive forecasts, TF represents true negative predictions, FP represents false positive predictions, and FN represents false negative predictions.

*Precision:* Of all the classified samples, $Precision$ denotes the percentage of correctly classified samples. The mathematical formula for $Precision$ is represented in Eq. (2).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

*Recall:* $Recall$ is the ratio of correctly predicted positive observations to all observations in the actual class. It is a measure of a classifier's completeness. It can be mathematically calculated by the Eq. (3).

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

**Table 7**
Results obtained from experiments. Experimental results presents that ML models can be used to predict depression effectively.

| Model | Accuracy | Precision | Recall | F1 | AUC | Cohen's Kappa |
|---|---|---|---|---|---|---|
| LR | 77.95% | 78.02% | 77.95% | 77.93% | 85.61% | 55.88% |
| NB | 75.38% | 75.39% | 75.38% | 75.38% | 85.50% | 50.76% |
| SVM | 86.15% | 86.19% | 86.15% | 86.15% | 50.00% | 72.31% |
| DT | 71.28% | 71.63% | 71.28% | 71.18% | 71.31% | 42.60% |
| ET | **90.26%** | **90.26%** | **90.26%** | **90.26%** | **95.30%** | **80.51%** |
| RF | 86.15% | 86.16% | 86.15% | 86.15% | 93.81% | 72.31% |
| XGB | 84.10% | 84.11% | 84.10% | 84.10% | 92.02% | 68.21% |
| LGBM | 84.62% | 84.63% | 84.62% | 84.61% | 92.58% | 69.23% |
| CB | 82.56% | 82.68% | 82.56% | 82.54% | 92.08% | 65.12% |
| KNN | 74.87% | 75.88% | 74.87% | 74.64% | 81.07% | 49.79% |
| GBM | 84.10% | 84.11% | 84.10% | 84.10% | 91.68% | 68.20% |
| AB | 80.00% | 80.08% | 80.00% | 79.99% | 88.63% | 60.01% |
| QDA | 81.54% | 81.60% | 81.54% | 81.53% | 88.98% | 63.08% |
| LDA | 84.62% | 84.63% | 84.62% | 84.61% | 92.54% | 69.23% |
| SGD | 83.59% | 84.10% | 83.59% | 83.52% | 92.44% | 67.16% |
| ANN | 86.15% | 86.19% | 86.15% | 86.15% | 94.90% | 72.30% |

*Cohen's Kappa Score:* Cohen's Kappa is a statistical measure used to assess the degree of agreement between two raters or classification systems beyond what would be expected by chance. Particularly useful in the fields of psychology, education, and health sciences, Cohen's Kappa adjusts for the agreement occurring by chance, providing a more robust indication of inter-rater reliability than simple percent agreement calculation. This coefficient ranges from -1 (perfect disagreement) to +1 (perfect agreement), with 0 indicating the level of agreement that can be expected from random chance [80].

*F1 Score:* F1 Score is the harmonic mean of precision and recall. Here, recall and precision are combined into a single metric and can be measured by Eq. (4)

$$F1\ Score = \frac{2 * (Precision * Recall)}{Precision + Recall} \tag{4}$$

*Area Under the Curve:* Area Under the Curve (AUC) is a metric for measuring the effectiveness of a binary classification model. The higher the AUC, the greater the model's capacity to differentiate between positive and negative cases. It is a common metric for assessing the effectiveness of classification algorithms in ML.

*Accuracy:* Accuracy is the proportion of the correctly classified samples and the total classified samples. This is represented by Eq. (5).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

*ROC-AUC* The ROC-AUC (Receiver Operating Characteristic - Area Under the Curve) is an extensively employed metric in the assessment of binary classification models [81]. However, it is also adaptable for classification assignments involving multiple classes. The concept is to calculate the ROC-AUC for each class compared to all other classes, with the outcomes being averaged. This methodology yields a metric that quantifies the capacity of a model to differentiate among numerous classes. The ROC curve illustrates the compromise between the true positive rate and the false positive rate t different thresholds. The AUC, on the other hand, converts the ROC curve into a solitary value that signifies the probability that a classifier will assign a higher rank to a positively selected instance at random than to a negatively selected instance. The metric assesses the model's predictive accuracy regardless of the threshold for classification.

These are the standard metrics of measurement that have been utilized for several studies and experiments in this work. Using these metrics, we will further analyze and discuss the proposed model.

### *3.2. Result analysis*

This section presents a detailed analysis of the performance of the employed ML models. For this experiment, the train-test split has been set to 85% and 15% respectively. The performance of various ML models are summarized in Table 7. Fig. 4(a)-4(q) presents the confusion matrix of (a) LR, (b) NB, (c) DT, (d) RF, (e) KNN, (f) SVM, (g) LDA, (h) QDA, (i) AB, (j) CB, (k) GBM, (l) LGBM, (m) XGB, (n) SGD, (o) ET, and (p) ANN model respectively. Fig. 5(a), 5(b), 5(c), and 5(d) presents the accuracy, precision, recall, and f1-score obtained from the experiment respectively. And Fig. 6 presents the combined ROC-AUC curve of the ML models. Each model is evaluated based on six criteria: Accuracy, Precision, Recall, F1 Score, AUC, and Cohen's Kappa, which are essential for understanding their effectiveness in classification tasks.

In this study, the ET model achieved the highest accuracy at 90.26%, indicating superior performance in correctly predicting both positive and negative cases. Following ET, SVM, RF, and ANN models all demonstrated strong performance with an accuracy

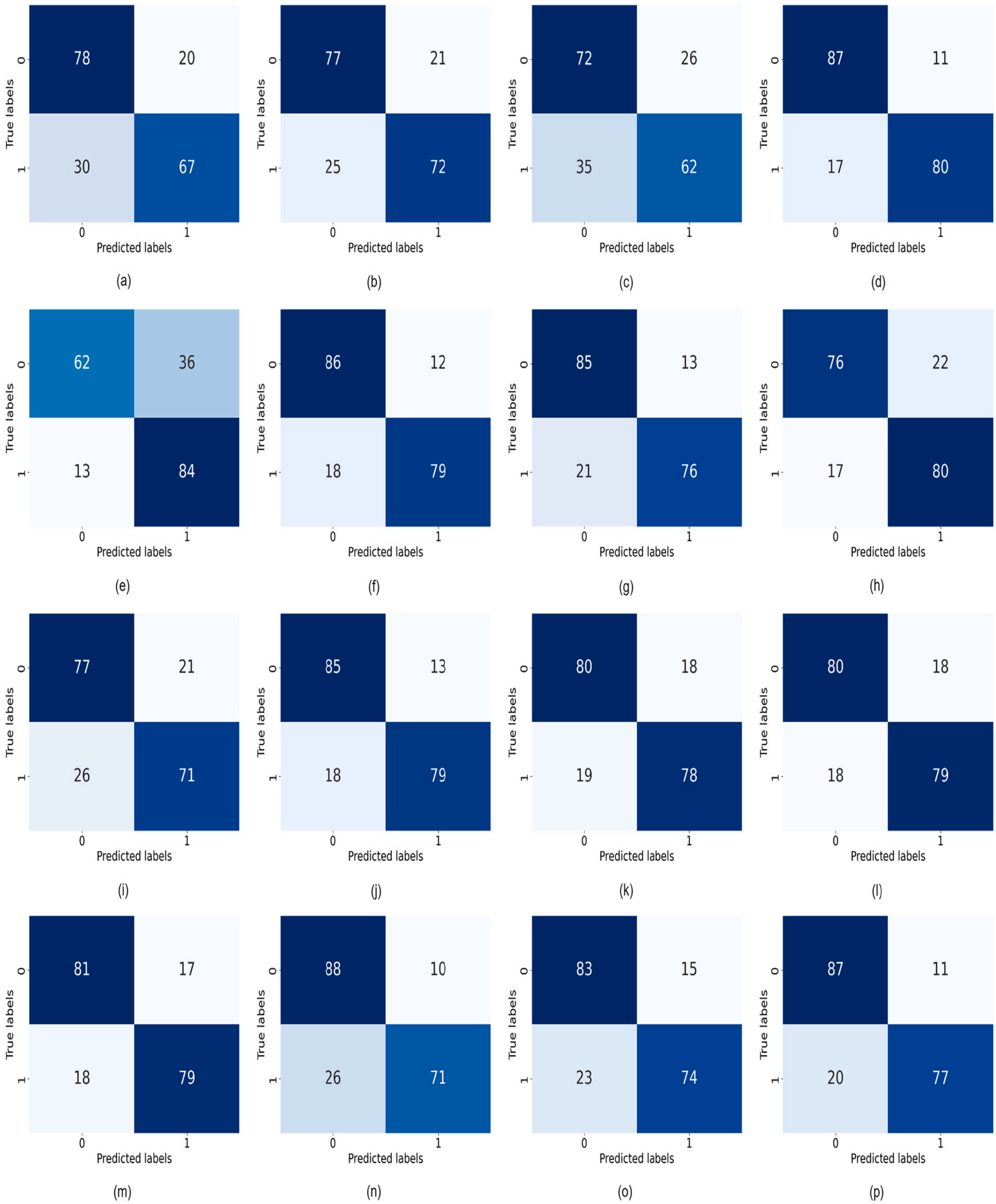


**Fig. 4.** This figure presents the confusion matrix of ML models obtained from experiments (a) LR, (b) NB, (c) DT, (d) RF, (e) KNN, (f) SVM, (g) LDA, (h) QDA, (i) AB, (j) CB, (k) GBM, (l) LGBM, (m) XGB, (n) SGD, (o) ET, and (p) ANN.

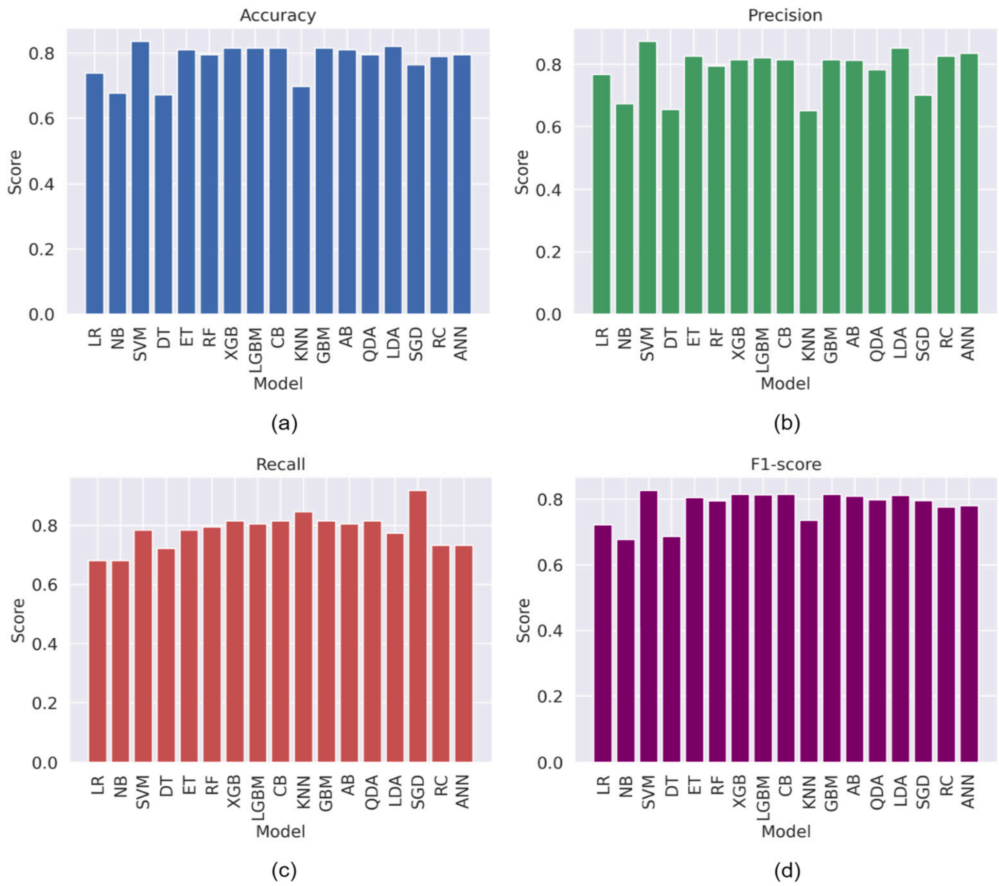


**Fig. 5.** Performance of ML models obtained from experiments. The figure is divided into four sub-parts. The figure presents comparison of (a) Accuracy, (b) Precision, (c) Recall, and (d) F1-score respectively.

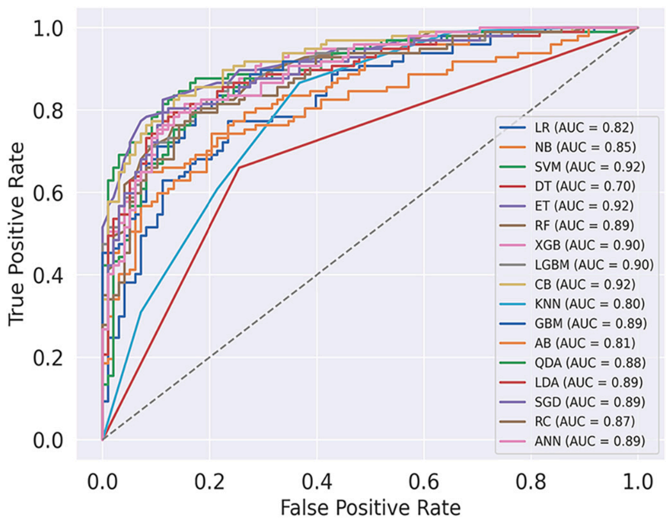


**Fig. 6.** ROC-AUC curve of ML models obtained from experiments. This graph presents the capability of the model in classification.

of 86.15%. On the other end of the spectrum, the AB model showed the lowest accuracy at 80.00%. Other models, such as LR, NB, DT, and KNN, displayed moderate accuracy, ranging from 71.28% to 77.95%. Precision, the ratio of true positive results to the total predicted positives, reflects the accuracy of positive predictions. Here, the ET model again led with the highest precision of 90.26%, underscoring its effectiveness in minimizing false positives. SVM, RF, and ANN models also performed commendably with precision values of 86.19%. In contrast, KNN exhibited the lowest precision at 75.88%. Models like LR, NB, DT, and AB showed lower precision than the leading models, indicating a higher rate of false positives in their predictions. Recall, or sensitivity, is the ratio of true positive results to the total actual positives, measuring the model's ability to identify all relevant cases. ET achieved the highest recall at 90.26%, showcasing its strength in capturing all positive instances. SVM, RF, and ANN also performed robustly with recall values of 86.15%. On the lower end, DT had the least recall at 71.28%, highlighting its limited capability to identify all positive cases. LR, NB, and KNN models exhibited lower recall than the top performers, indicating potential gaps in capturing all relevant instances. The F1 score, the harmonic mean of precision and recall, balances the two metrics to provide a comprehensive measure of the model's performance. The ET model had the highest F1 score of 90.26%, reflecting a balanced performance with high precision and recall. SVM, RF, and ANN also had high F1 scores of 86.15%, indicating their reliable and consistent performance. Conversely, DT had the lowest F1 score at 71.18%, pointing to its struggles in maintaining a balance between precision and recall. Other models, such as LR, NB, and KNN, showed moderate F1 scores compared to the top performers. AUC measures the ability of the model to distinguish between classes, with a higher AUC indicating better performance. ET achieved the highest AUC at 95.30%, highlighting its exceptional capability in differentiating between classes. RF and ANN also performed well with AUC values of 93.81% and 94.90%, respectively. SVM had the lowest AUC at 50.00%, indicating poor performance in class separation. Models like LR, NB, DT, and KNN had lower AUC compared to the leading models, reflecting less effective class distinction. Cohen's Kappa measures the agreement between predicted and actual classifications, adjusting for chance agreement. ET had the highest Cohen's Kappa at 80.51%, demonstrating strong agreement beyond chance. SVM, RF, and ANN also had high Kappa values of 72.31%, indicating reliable performance. DT had the lowest Kappa value at 42.60%, suggesting weaker agreement with actual classifications. LR, NB, and KNN models showed moderate Kappa values compared to the top performers, indicating room for improvement in agreement between predictions and actual outcomes.

In this study, LR and NB show moderate effectiveness, with LR slightly outperforming NB across most metrics. Both models exhibit relatively high AUC, indicating good performance under varied threshold settings. SVM and Ensemble Techniques such as ET, RF, and Gradient Boosting (XGB, GBM) demonstrate higher accuracy and precision. SVM, however, has a notably low AUC, suggesting limitations in its ability to distinguish between classes compared to ensemble methods, which show high AUC values, indicating superior performance in handling class separability. DT shows lower performance across all metrics, highlighting its potential overfitting and instability in handling the data variability in depression and anxiety detection. Among the ensemble methods, ET stands out with the highest scores in almost all metrics, particularly excelling in AUC, which is crucial for ensuring the robustness of the model across various decision thresholds. Models like LGBM, CB, and KNN show varied results, with LGBM and CB performing well above average. In contrast, KNN lags slightly behind, indicating its sensitivity to the dataset's noise and outliers. AB, QDA, LDA, SGD, and ANN offer competitive but varied performance metrics, with ANN showing strong potential due to high AUC and reasonable stability indicated by Cohen's Kappa. This evaluation of ML models provides insights into their suitability for implementing an effective depression and anxiety detection system in academic settings. The variation in performance across different models underscores the importance of choosing the suitable model based on specific needs and dataset characteristics.

## 4. Comparison with an existing work

In this section, a comparison has been provided with an existing work which has been conducted on similar topic. The study conducted by Meda et al. [33] analyzed the mental health of university students and found that approximately one out of five students reported severe depressive symptoms and/or suicidal ideation. A significant finding was the strong association between economic worry and depression, both at baseline and follow-up, with high-frequency economic worry tripling the odds of severe depressive symptoms. The random forest algorithm used in this study exhibited high accuracy in predicting students who maintained well-being (balanced accuracy of 0.85) but showed low accuracy for those whose symptoms worsened (balanced accuracy of 0.49). This suggests that while the algorithm was effective at identifying students who remained stable, it was less successful at predicting deterioration in mental health. In contrast, responses were gathered from 804 university students in Bangladesh in the proposed study, which identified a much higher prevalence of depression, with around 60% of the participants suffering from the condition. Sixteen different machine learning models were employed to analyze the data, and the ET model emerged as the most effective, achieving an impressive accuracy of 90.26% in identifying depression. This high accuracy underscores the potential of the ET model in accurately classifying students with depression, indicating a robust performance in detecting mental health issues. A summary of the comparison has been provided in Table 8.

## 5. Conclusion

The growing prevalence of depression among university students is a significant concern that necessitates the development of effective, scalable, and precise screening methods. This research utilized the GHQ-12 alongside various ML models to explore and identify potential predictors of depression within a diverse population of university students in Bangladesh. Our findings demonstrate that ML models can be effectively employed to analyze socio-demographic and career-related data, yielding valuable insights into the factors contributing to psychological distress among students. The rigorous evaluation of multiple ML models highlighted their potential for accurately identifying individuals at risk of depression, offering a promising approach for early detection and intervention. The ET model has achieved the highest accuracy of 90.26% among all the ML models. The other models also achieved around 75-80% of accuracy. Integrating socio-demographic variables and career-related stressors into the analysis provided a comprehensive understanding of the complex factors influencing students' mental health. This underscores the importance of considering a multifaceted approach when addressing mental health issues within academic settings.

**Table 8**
A comprehensive comparison between the proposed study and an existing work.

| Aspect | Meda et al. | Proposed Study |
|---|---|---|
| Population | University students in Italy | University students in Bangladesh |
| Sample Size | 1,388 students at baseline, 557 at follow-up | 804 students |
| Follow-Up Period | 6 months | No Follow-up |
| Key Variables Collected | Demographic information, depressive, anxiety, and obsessive-compulsive symptoms | Socio-demographic information, Job and Career-related stress |
| Questionnaires Used | Beck Depression Inventory-II (BDI-II), Beck Anxiety Inventory (BAI), Obsessive-Compulsive Inventory-Revised (OCI-R), Eating Disorder Inventory-3 (EDI-3), Eating Habits Questionnaire (EHQ) | General Health Questionnaire-12 (GHQ-12) |
| Mental Health Issues Studied | Severe depressive symptoms, anxiety, obsessive-compulsive symptoms, suicidal ideation | Depression |
| Main Findings on Prevalence | ~20% reported severe depressive symptoms and/or suicidal ideation | ~60% reported suffering from depression |
| ML Models Used | Random Forest | 16 different ML models |
| Accuracy of Best Model | Balanced accuracy: 0.85 (well-being), 0.49 (worsening symptoms) | ET model accuracy: 90.26% |
| Significant Predictors | Economic worry, cognitive and somatic symptoms of depression | Not specified |
| Data Collection Method | Online questionnaires | Online questionnaires |

Despite the promising results, this study has several limitations. Firstly, the dataset used was limited to university students in Bangladesh, which may affect the generalizability of the findings to other regions or demographics. Additionally, the self-reported nature of the GHQ-12 responses may introduce bias or inaccuracies in the data. This study also focused on limited socio-demographic and career-related variables, potentially overlooking other significant factors influencing mental health. Finally, the study's cross-sectional design limits the ability to draw causal inferences from the data. In the future, we aim to address these limitations by expanding the dataset to include a more diverse and representative sample of students from different regions and backgrounds. As longitudinal studies could provide deeper insights into the causal relationships between various factors and depression, it would be added to this research. Moreover, incorporating additional variables, such as lifestyle factors, academic performance, and social support, could enhance the predictive power of the models. Further development and refinement of ML algorithms, including exploring advanced techniques like deep learning, could also improve the accuracy and applicability of these models in real-world settings.

The results of this study contribute to the growing body of literature on the application of ML in psychological research, particularly in the context of higher education. By demonstrating the effectiveness of these models, we pave the way for future research to further refine and implement these techniques in practical settings. This research emphasizes the potential of ML to enhance our understanding of mental health challenges and to develop targeted intervention strategies that can significantly improve the well-being of university students. In conclusion, the application of ML models in conjunction with GHQ-12 provides a novel and practical approach to screening for depression among university students. The insights gained from this study advance the academic conversation on mental health in higher education and offer practical implications for developing early intervention and support systems. As universities continue to grapple with the mental health challenges students face, integrating advanced analytical techniques like ML will be crucial in promoting a healthier and more supportive academic environment.

## Research ethics

Ethical considerations were duly incorporated into each phase of the research endeavor. The informed consent form was incorporated into the questionnaire, and before the investigation of the research inquiries, the participants were required to confirm the consent form electronically. To maintain participant confidentiality, the data underwent de-identification and was assigned a unique random code. The research protocol was reviewed and approved by a Review Board of Bangladesh Army University of Engineering and Technology (BAUET) under the reference number R&P/2024-23.

## CRediT authorship contribution statement

**Nasirul Mumenin:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **A.B.M. Kabir Hossain:** Formal analysis, Data curation. **Md. Arafat Hossain:** Investigation, Data curation. **Partha Pratim Debnath:** Investigation, Data curation, Conceptualization. **Mursheda Nusrat Della:** Data curation, Conceptualization. **Md. Mahmudul Hasan Rashed:** Data curation. **Afzal Hossen:** Data curation. **Md. Rubel Basar:** Validation, Formal analysis. **Md. Sejan Hossain:** Resources, Data curation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability statement

Data will be made available on request.

## Acknowledgement

We would like to thank the Bangladesh Army University of Engineering and Technology (BAUET) for supporting us in this research. This research has been conducted under the project "Student's Mental Health Analysis and Illness Prediction Using Data Analytics and Artificial Intelligence."


## References


[1] B.J. Sadock, V.A. Sadock, P. Ruiz, Comprehensive Textbook of Psychiatry, vol. 1, Lippincott Williams & Wilkins, Philadelphia, 2000.
[2] A.K. Ibrahim, S.J. Kelly, C.E. Adams, C. Glazebrook, A systematic review of studies of depression prevalence in university students, J. Psychiatr. Res. 47 (3) (2013) 391–400.
[3] M.R. Islam, M.A. Kabir, A. Ahmed, A.R.M. Kamal, H. Wang, A. Ulhaq, Depression detection from social network data using machine learning techniques, Health Inf. Sci. Syst. 6 (2018) 1–12.
[4] A.A. Choudhury, M.R.H. Khan, N.Z. Nahim, S.R. Tulon, S. Islam, A. Chakrabarty, Predicting depression in Bangladeshi undergraduates using machine learning, in: 2019 IEEE Region 10 Symposium (TENSYMP), IEEE, 2019, pp. 789–794.
[5] D. Sarokhani, A. Delpisheh, Y. Veisani, M.T. Sarokhani, R.E. Manesh, K. Sayehmiri, et al., Prevalence of depression among university students: a systematic review and meta-analysis study, Depress. Res. Treat. 2013 (2013).
[6] L. Gao, Y. Xie, C. Jia, W. Wang, Prevalence of depression among Chinese university students: a systematic review and meta-analysis, Sci. Rep. 10 (1) (2020) 15897.
[7] S.E. Gollust, D. Eisenberg, E. Golberstein, Prevalence and correlates of self-injury among university students, J. Am. Coll. Health 56 (5) (2008) 491–498.
[8] R. Tasnim, M.S. Islam, M.S.H. Sujan, M.T. Sikder, M.N. Potenza, Suicidal ideation among Bangladeshi university students early during the covid-19 pandemic: prevalence estimates and correlates, Child. Youth Serv. Rev. 119 (2020) 105703.
[9] S. Bachmann, Epidemiology of suicide and the psychiatric perspective, Int. J. Environ. Res. Public Health 15 (7) (2018) 1425.
[10] M. Bostanci, O. Özdel, N.K. Oguzhanoglu, L. Özdel, A. Ergin, N. Ergin, F. Atesci, Depressive symptomatology among university students in Denizli, Turkey: prevalence and sociodemographic correlates, Croat. Med. J. (2005).
[11] N. Bayram, N. Bilgel, The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students, Soc. Psychiatry Psychiatr. Epidemiol. 43 (2008) 667–672.
[12] A.Y. Lim, S.-H. Lee, Y. Jeon, R. Yoo, H.-Y. Jung, Job-seeking stress, mental health problems, and the role of perceived social support in university graduates in Korea, J. Korean Med. Sci. 33 (19) (2018).
[13] D.E. Saunders, G.W. Peterson, J.P. Sampson Jr, R.C. Reardon, Relation of depression and dysfunctional career thinking to career indecision, J. Vocat. Behav. 56 (2) (2000) 288–298.
[14] X. Zhong, X. Jin, L. Yan, L. Yang, H. Long, J. Wang, H. Wang, Y. Liu, J. Pu, P. Xie, et al., Reliability and validity of general health questionnaire-12 in Chinese dental healthcare workers during the covid-19 pandemic, Front. Psychiatry 12 (2022) 792838.
[15] G.N. Baksheev, J. Robinson, E.M. Cosgrave, K. Baker, A.R. Yung, Validity of the 12-item general health questionnaire (ghq-12) in detecting depressive and anxiety disorders among high school students, Psychiatry Res. 187 (1–2) (2011) 291–296.
[16] N. Armino, V. Gouttebarge, S. Mellalieu, R. Schlebusch, J. van Wyk, S. Hendricks, Anxiety and depression in athletes assessed using the 12-item general health questionnaire (ghq-12)-a systematic scoping review, S. Afr. J. Sports Med. 33 (1) (2021) 1–13.
[17] H. Ozdemir, M. Rezaki, General health questionnaire-12 for the detection of depression, Turk Psikiyatri Derg. 18 (1) (2007) 13.
[18] A. Lundin, M. Hallgren, H. Theobald, C. Hellgren, M. Torgén, Validity of the 12-item version of the general health questionnaire in detecting depression in the general population, Publ. Health 136 (2016) 66–74.
[19] L.-L. Qin, J. Peng, M.-L. Shu, X.-Y. Liao, H.-J. Gong, B.-A. Luo, Y.-W. Chen, The fully mediating role of psychological resilience between self-efficacy and mental health: evidence from the study of college students during the covid-19 pandemic, in: Healthcare, vol. 11, MDPI, 2023, p. 420.
[20] W.B. Zatt, K. Lo, W. Tam, Pooled prevalence of depressive symptoms among medical students: an individual participant data meta-analysis, BMC Psychiatry 23 (1) (2023) 251.
[21] Y. Jiang, C. Ding, B. Shen, Latent profile analysis of mental health among Chinese university students: evidence for the dual-factor model, in: Healthcare, vol. 11, MDPI, 2023, p. 2719.
[22] S. Anjara, C. Bonetto, T. Van Bortel, C. Brayne, Using the ghq-12 to screen for mental health problems among primary care patients: psychometrics and practical considerations, Int. J. Ment. Health Syst. 14 (2020) 1–13.
[23] Y. Sonmez, M. Akdemir, A. Meydanlioglu, M.R. Aktekin, Psychological distress, depression, and anxiety in nursing students: a longitudinal study, in: Healthcare, vol. 11, MDPI, 2023, p. 636.
[24] D. James, J. Yates, E. Ferguson, Can the 12-item general health questionnaire be used to identify medical students who might 'struggle' on the medical course? A prospective study on two cohorts, BMC Med. Educ. 13 (2013) 1–8.
[25] M. Elovanio, C. Hakulinen, L. Pulkki-Råback, A.-M. Aalto, M. Virtanen, T. Partonen, J. Suvisaari, General health questionnaire (ghq-12), Beck depression inventory (bdi-6), and mental health index (mhi-5): psychometric and predictive properties in a Finnish population-based sample, Psychiatry Res. 289 (2020) 112973.
[26] K.-S. Lee, B.-J. Ham, Machine learning on early diagnosis of depression, Psychiatry Investig. 19 (8) (2022) 597.
[27] B. Sutter, R. Chiong, G.S. Budhi, S. Dhakal, Predicting psychological distress from ecological factors: a machine learning approach, in: Advances and Trends in Artificial Intelligence. Artificial Intelligence Practices: 34th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2021, Kuala Lumpur, Malaysia, July 26–29, 2021, Proceedings, Part I 34, Springer, 2021, pp. 341–352.
[28] B. Purwandari, W.S. Wibawa, N. Fitriah, M. Christia, D.R. Bintari, Internet addiction and mental health prediction using ensemble learning based on web browsing history, in: Proceedings of the 3rd International Conference on Software Engineering and Information Management, 2020, pp. 155–159.
[29] S. Baird, R. Panlilio, J. Seager, S. Smith, B. Wydick, Identifying psychological trauma among Syrian refugee children for early intervention: analyzing digitized drawings using machine learning, J. Dev. Econ. 156 (2022) 102822.
[30] A. Bieliński, I. Rojek, D. Mikołajewski, Comparison of selected machine learning algorithms in the analysis of mental health indicators, Electronics 12 (21) (2023) 4407.
[31] B.N. Shaikh Mohammad, K. Siddiqui, Random forest regressor machine learning model developed for mental health prediction based on mhi-5, phq-9 and bdi scale, in: Proceedings of the 4th International Conference on Advances in Science & Technology (ICAST2021), 2021.
[32] U.B. Munir, M.S. Kaiser, U.I. Islam, F.H. Siddiqui, Machine learning classification algorithms for predicting depression among university students in Bangladesh, in: Proceedings of the Third International Conference on Trends in Computational and Cognitive Engineering: TCCE 2021, Springer, 2022, pp. 69–80.
[33] N. Meda, S. Pardini, P. Rigobello, F. Visioli, C. Novara, Frequency and machine learning predictors of severe depressive symptoms and suicidal ideation among university students, Epidemiol. Psychiatr. Sci. 32 (2023) e42.
[34] D.P. Goldberg, General health questionnaire-12, Aust. J. Psychol. (1978).
[35] D. Goldberg, P. Williams, General Health Questionnaire (GHQ), nferNelson, Swindon, Wiltshire, UK, 2000.

[36] D.P. Goldberg, R. Gater, N. Sartorius, T.B. Ustun, M. Piccinelli, O. Gureje, C. Rutter, The validity of two versions of the ghq in the who study of mental illness in general health care, Psychol. Med. 27 (1) (1997) 191–197.
[37] D. Goldberg, T. Oldehinkel, J. Ormel, Why ghq threshold varies from one place to another, Psychol. Med. 28 (4) (1998) 915–921.
[38] M. Hankins, The factor structure of the twelve item general health questionnaire (ghq-12): the result of negative phrasing?, Clin. Pract. Epidemiol. Ment. Health 4 (2008) 1–8.
[39] A. Campbell, J. Walker, G. Farrell, Confirmatory factor analysis of the ghq-12: can I see that again?, Aust. N.Z. J. Psychiatry 37 (4) (2003) 475–483.
[40] A. Mäkikangas, T. Feldt, U. Kinnunen, A. Tolvanen, M.-L. Kinnunen, L. Pulkkinen, The factor structure and factorial invariance of the 12-item general health questionnaire (ghq-12) across time: evidence from two community-based samples, Psychol. Assess. 18 (4) (2006) 444.
[41] M.F. Rodrigo, J.G. Molina, J.-M. Losilla, J. Vives, J.M. Tomás, Method effects associated with negatively and positively worded items on the 12-item general health questionnaire (ghq-12): results from a cross-sectional survey with a representative sample of catalonian workers, BMJ Open 9 (11) (2019) e031859.
[42] R. Benoni, A. Sartorello, M. Mazzi, L. Berti, M. Sorina, E. Paiola, G. Varischi, S. Tardivo, M. Rimondini, F. Moretti, The use of 12-item general health questionnaire (ghq-12) in Ukrainian refugees: translation and validation study of the Ukrainian version, Health Qual. Life Outcomes 22 (1) (2024) 6.
[43] P.J.M. Ali, R.H. Faraj, E. Koya, P.J.M. Ali, R.H. Faraj, Data normalization and standardization: a technical report, Mach. Learn. Tech. Rep. 1 (1) (2014) 1–6.
[44] S. Liu, M. Motani, Exploring unique relevance for mutual information based feature selection, in: 2020 IEEE International Symposium on Information Theory (ISIT), IEEE, 2020, pp. 2747–2752.
[45] E.I. Obilor, E.C. Amadi, Test for significance of Pearson's correlation coefficient, Int. J. Innov. Math. Stat. Energy Policies 6 (1) (2018) 11–23.
[46] P. Sedgwick, Pearson's correlation coefficient, BMJ 345 (2012).
[47] I.T. Jolliffe, Principal Component Analysis for Special Types of Data, Springer, 2002.
[48] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philos. Trans. R. Soc. A, Math. Phys. Eng. Sci. 374 (2065) (2016) 20150202.
[49] M.P. LaValley, Logistic regression, Circulation 117 (18) (2008) 2395–2399.
[50] I. Rish, et al., An empirical study of the naive Bayes classifier, in: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, vol. 3, Citeseer, 2001, pp. 41–46.
[51] Y.-Y. Song, L. Ying, Decision tree methods: applications for classification and prediction, Shanghai Arch. Psychiatry 27 (2) (2015) 130.
[52] G. Biau, E. Scornet, A random forest guided tour, Test 25 (2016) 197–227.
[53] G. Guo, H. Wang, D. Bell, Y. Bi, K. Greer, Knn model-based approach in classification, in: On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings, Springer, 2003, pp. 986–996.
[54] S. Zhang, X. Li, M. Zong, X. Zhu, R. Wang, Efficient knn classification with different numbers of nearest neighbors, IEEE Trans. Neural Netw. Learn. Syst. 29 (5) (2017) 1774–1785.
[55] I. Steinwart, A. Christmann, Support Vector Machines, Springer Science & Business Media, 2008.
[56] S. Suthaharan, S. Suthaharan, Support vector machine, in: Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning, 2016, pp. 207–235.
[57] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, Front. Neurorobot. 7 (2013) 21.
[58] V.K. Ayyadevara, V.K. Ayyadevara, Gradient boosting machine, in: Pro Machine Learning Algorithms: A Hands-on Approach to Implementing Algorithms in Python and R, 2018, pp. 117–134.
[59] C. Ying, M. Qi-Guang, L. Jia-Chen, G. Lin, Advance and prospects of adaboost algorithm, Acta Autom. Sin. 39 (6) (2013) 745–758.
[60] T. Hastie, S. Rosset, J. Zhu, H. Zou, Multi-class adaboost, Stat. Interface 2 (3) (2009) 349–360.
[61] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[62] A. Ogunleye, Q.-G. Wang, Xgboost model for chronic kidney disease diagnosis, IEEE/ACM Trans. Comput. Biol. Bioinform. 17 (6) (2019) 2131–2140.
[63] A.A. Taha, S.J. Malebary, An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine, IEEE Access 8 (2020) 25579–25587.
[64] J. Fan, X. Ma, L. Wu, F. Zhang, X. Yu, W. Zeng, Light gradient boosting machine: an efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data, Agric. Water Manag. 225 (2019) 105758.
[65] Y.A. Alsariera, V.E. Adeyemo, A.O. Balogun, A.K. Alazzawi, Ai meta-learners and extra-trees algorithm for the detection of phishing websites, IEEE Access 8 (2020) 142532–142542.
[66] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, Catboost: unbiased boosting with categorical features, Adv. Neural Inf. Process. Syst. 31 (2018).
[67] A.V. Dorogush, V. Ershov, A. Gulin, Catboost: gradient boosting with categorical features support, arXiv preprint arXiv:1810.11363, 2018.
[68] S. Balakrishnama, A. Ganapathiraju, Linear discriminant analysis-a brief tutorial, Inst. Signal Inf. Process. 18 (1998) (1998) 1–8.
[69] A. Tharwat, T. Gaber, A. Ibrahim, A.E. Hassanien, Linear discriminant analysis: a detailed tutorial, AI Commun. 30 (2) (2017) 169–190.
[70] S. Bose, A. Pal, R. SahaRay, J. Nayak, Generalized quadratic discriminant analysis, Pattern Recognit. 48 (8) (2015) 2676–2684.
[71] Y. Qin, A review of quadratic discriminant analysis for high-dimensional data, Wiley Interdiscip. Rev.: Comput. Stat. 10 (4) (2018) e1434.
[72] L. Bottou, Large-scale machine learning with stochastic gradient descent, in: Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics, Paris France, August 22-27, 2010 Keynote, Invited and Contributed Papers, Springer, 2010, pp. 177–186.
[73] T. Zhang, Solving large scale linear prediction problems using stochastic gradient descent algorithms, in: Proceedings of the Twenty-First International Conference on Machine Learning, 2004, p. 116.
[74] M.H. Hassoun, Fundamentals of Artificial Neural Networks, MIT Press, 1995.
[75] S.-C. Wang, S.-C. Wang, Artificial neural network, in: Interdisciplinary Computing in Java Programming, 2003, pp. 81–100.
[76] Y.-c. Wu, J.-w. Feng, Development and application of artificial neural network, Wirel. Pers. Commun. 102 (2018) 1645–1656.
[77] O.I. Abiodun, A. Jantan, A.E. Omolara, K.V. Dada, N.A. Mohamed, H. Arshad, State-of-the-art in artificial neural network applications: a survey, Heliyon 4 (11) (2018).
[78] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.
[79] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[80] S. Sun, Meta-analysis of Cohen's kappa, Health Serv. Outcomes Res. Methodol. 11 (2011) 145–163.
[81] A.J. Bowers, X. Zhou, Receiver operating characteristic (roc) area under the curve (auc): a diagnostic measure for evaluating the accuracy of predictors of education outcomes, J. Educ. Stud. Placed Risk 24 (1) (2019) 20–46.