BMC Psychiatry

RESEARCH

Open Access

Utilizing machine learning algorithms for predicting Anxiety-Depression Comorbidity Syndrome in Gastroenterology Inpatients (ADCS-GI)

Min Tan[1,6], Jinjin Zhao[1], Yushun Tao[1,6], Uroosa Sehar[1,6], Yan Yan[1], Qian Zou[2], Qing Liu[4], Long Xu[2], Zeyang Xia[5], Lijuan Feng[2,3*] and Jing Xiong[1,6*]

Abstract

Background Accurately diagnosing Anxiety-Depression Comorbidity Syndrome in Gastroenterology Inpatients (ADCS-GI) shows significant challenges as traditional diagnostic methods fail to meet expectations due to patient hesitance and non-psychiatric healthcare professionals' limitations. Therefore, the need for objective diagnostics highlights the potential of machine learning in identifying and treating ADCS-GI.

Methods A total of 1186 ADCS patients were recruited for this study. We conducted extensive studies for the dataset, including data quantification, equilibrium, and correlation analysis. Eight machine learning models, including Gaussian Naive Bayes (NB), Support Vector Classifier (SVC), K-Neighbors Classifier, RandomForest, XGB, CatBoost, Cascade Forest, and Decision Tree, were utilized to compare prediction efficacy, with an effort to minimize the dependency on subjective questionnaires.

Results Among eight machine learning algorithms, the Decision Tree and K-nearest neighbors models demonstrated an accuracy exceeding 81% and a sensitivity in the same range for detecting ADCS in patients. Notably, when identifying moderate and severe cases, the models achieved an accuracy above 88% and a sensitivity of 90%. Furthermore, the models trained without reliance on subjective questionnaires showed promising performance, indicating the feasibility of developing questionnaire-free early detection applications.

Conclusion Machine learning algorithms can be used to identify ADCS among gastroenterology patients. This can help facilitate the early detection and intervention of psychological disorders in gastroenterology patients' care.

Keywords Gastroenterology, Psychosomatic disorders, Anxiety, Depression, Machine learning

The work is supported by National Natural Science Foundation of China (62073309, 12426305), Guangdong Basic and Applied Basic Research Foundation (2022B1515020042), and Stable Support Project of Shenzhen (Project No. 20231120164732001). This work is also supported by Shenzhen Engineering Laboratory for Diagnosis & Treatment Key Technologies of Interventional Surgical Robots (XMHT20220104009).

*Correspondence:
Lijuan Feng
fenglj@szu.edu.cn
Jing Xiong
jing.xiong@siat.ac.cn
Full list of author information is available at the end of the article

Introduction

Approximately 50–90% of gastroenterology patients suffer from Anxiety-Depression Comorbidity Syndrome (ADCS) [1–3] because of the brain-gut link between gastrointestinal disorders and psychological distress [1, 2]. Furthermore, the psychological distress accompanying inflammatory bowel disease has been shown to exacerbate the risk of surgery, hospitalizations, and other complications. These conditions affect individuals, families, and also pose a substantial obstacle to the sustainable development of nations [4, 5], highlighting the critical need for effective mental health management in gastrointestinal patients [6]. Conventional assessment tools for ADCS, including quality of life and sleep quality questionnaires, suffer from limited predictive ability, reflecting the difficulty of early detection and intervention of ADCS in gastroenterology. Moreover, non-psychiatric healthcare professionals in gastroenterology also lacked the qualification to use specialized psychiatric questionnaires like the Hamilton Depression Rating Scale [7], which affected the accurate assessment of patients' psychological comorbidity. Self-questionnaires [8–10] completed by patients often require additional assistance from nurses, which needs extra costs and time. In addition, many patients, influenced by common misconceptions and societal stigma surrounding mental health, are reluctant to acknowledge or accurately report their psychological distress [4]. Taking our prior research as an example [11], we distributed 1,315 psychosomatic questionnaires, but 129 patients (nearly 10%) refused to participate despite the full explanation of healthcare professionals. This reluctance hampers the promotion of subjective questionnaire-based diagnostics, leading to either underdiagnosis or misdiagnosis of ADCS. Therefore, it is crucial to develop more objective, reliable, and patient-friendly diagnostic tools to overcome these challenges. Unlike traditional questionnaires, machine learning models reduce reliance on patient-reported outcomes while providing an accessible tool for non-specialist clinicians. By enabling earlier and more accurate detection, the model has the potential to reduce healthcare burden posed by untreated psychological distress in gastroenterology patients.

Recently, auto-detection methods for psychiatric disorders have emerged, incorporating complex multidimensional data such as electroencephalography (EEG) and facial image analysis. For instance, Li et al. [12] proposed an EEG-based depression recognition method using a deep learning framework. Mumtaza et al. [13] proposed an EEG-based neural network that automatically discriminated between depressed and healthy controls (HCs). Xie et al. [14] proposed a deep network diagnosis model based on facial expression and body movement video. However, The inherent complexity of EEG signals and facial images poses significant challenges in achieving high accuracy and generalization capabilities with these methods.

In contrast, the gastroenterology field benefits from the extensive availability of clinical laboratory data, which is non-intrusive and easily accessible. It can facilitate mental health assessments without necessitating supplementary procedures such as EEG or facial recording. Based on this advantage, we specifically utilized machine learning algorithms like the Decision Tree and K-nearest neighbors (KNN) to analyze clinical laboratory data. Our method aligns with emerging trends in medical research, where machine learning algorithms are used to predict mental disorders across different scenarios [15–19]. For instance, Yu et al. [20] used structural magnetic resonance imaging (sMRI) and support vector machine (SVM) algorithms to predict violent behavior in male patients with schizophrenia. Priyaa et al. [21] proposed multiple machine learning methods to predict anxiety and depression in employed and unemployed individuals using questionnaire data. Leightley et al. [22] proposed machine learning methods to identify post-traumatic stress disorder (PTSD) from a UK military cohort patient. Saua et al. [23] compared the performance of different machine learning algorithms for screening anxiety and depression among seafarers. Notably, our research fills a gap in this domain, as we have not observed prior studies that specifically address the detection of depression and anxiety in gastrointestinal patients.

Based on the above analysis, this research leverages machine learning algorithms to detect ADCS in gastroenterology patients. By using the Decision Tree and K-nearest neighbors, we improve the accuracy and reliability of diagnosing different levels of ADCS conditions. The performance of these machine learning models has been extensively evaluated, which proves that our method can assist gastroenterologists in providing timely and appropriate early interventions for ADCS. In summary, our contributions are in four folds:

- A clinical dataset was collected from 1186 ADCS patients. Comprehensive data equilibrium and correlation analyses were performed to concentrate the most meaningful data.
- Eight machine learning models were utilized to compare different prediction efficacy while minimizing the dependency on subjective questionnaires.
- The performance of machine learning models was evaluated to ensure accuracy and reliability in detecting varying degrees of the ADCS-GI.

Material and methods

Dataset

Participants

This study incorporated 1186 inpatients diagnosed with digestive diseases who received treatment at the Department of Gastroenterology, Xiangya Hospital of Central South University. Informed consent was procured from all participants who met the inclusion criteria. The research design received approval from the Ethics Committee of Xiangya Hospital, Central South University.

Data details

General information and questionnaire data Inpatients' general information includes demographic data such as age, gender, marital status, and education. The self-report questionnaire with rigorous on-site quality control includes the following three parts. (1) SF-36. The 36-Item Short Form Health Survey [9] is a standardized health survey questionnaire that consists of 36 questions aimed at accurately measuring an individual's physical, psychological, and emotional health. Lower SF-36 scores indicate higher disability. (2) PSQI. The Pittsburgh sleep quality index questionnaire [10] is used for evaluating patient's sleep quality. Higher PSQI scores indicating poorer sleep quality. (3) SAS and SDS. Zung's Self-Rated Anxiety Scale (SAS) and Self-Rated Depression Scale (SDS) scores [8] are both self-administered surveys designed to measure the levels of anxiety and depression, respectively. The values of SAS and SDS are utilized as classification labels to train the Machine Learning models.

Clinical laboratory data We identified six clinical laboratory data from our preceding research [11] that exhibited both independence and high correlation. They are hemoglobin, total protein (TP), serum albumin (SA), high-density lipoprotein (HDL), aspartate aminotransferase (AST) activity, and direct bilirubin (DB) concentration.

Data quantification and analysis

Our prior work [11] has already engaged in statistical analyses, including the t-test, one-way variance, and multiple logistic regression. For the purpose of building machine learning models, further data quantification and detailed exploration of data distribution and balance will be elaborated on in this section.

Data quantification SAS and SDS scales can rate the temporal frequency of anxiety symptoms in the most recent month. A standard SAS score ≥ 50 is defined as having anxiety symptoms, which is quantized as 1. For SDS, the standard score is 53 [8]. Since SAS and SDS directly represent anxiety and depression, the quantized values of SAS and SDS are utilized as classification labels in training models.

Equilibrium analysis The density graph shown in Fig. 1a demonstrates an equilibrium in the distribution of SAS and SDS among different gender and education categories. The equilibrium distribution of the dataset affirms the reliability of subsequent machine-learning experiments.

Correlation analysis Figure 1a illustrates the findings regarding the impact of sleep and life quality (SF-36) and sleep quality (PSQI) on anxiety and depression. A negative correlation exists between SF-36/PSQI and SAS/SDS scale, suggesting that poorer sleep and life quality is associated with higher anxiety or depression scores. Since the SF-36 and PSQI scales have a significant correlation with anxiety and depression, they are utilized in conjunction with clinical laboratory data for training purposes. However, for digestive-related indexes, as shown in Fig. 1b, no discernible correlation can be observed through fundamental statistical analysis, leading us to explore a machine learning-based auto-detection method.

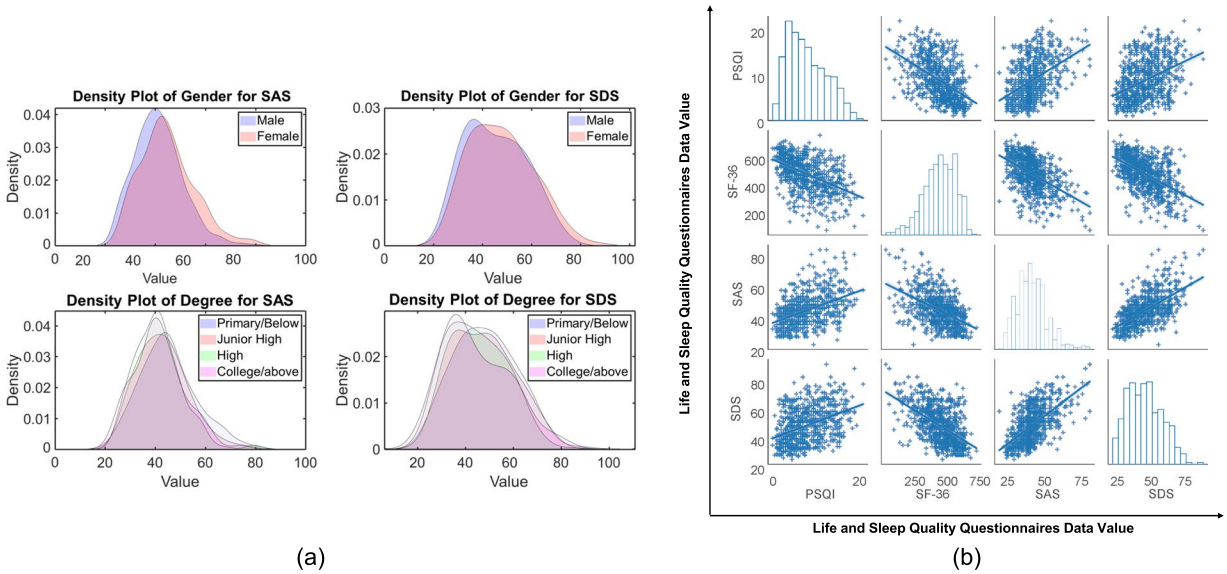
Machine learning model

This section briefly introduces the classic Decision Tree and K-nearest neighbors model.

Decision tree classifier

Decision Tree [24], serving as a straightforward method in categorizing, is very useful in statistics and data mining field. During the classifying process of a Decision Tree, the class label of a new instance is determined by a hierarchical tree which follows the divide and conquer algorithm. The algorithm for Decision Tree learning can be characterized in the following manner:

(1) Optimal attribute selection: The Attribute Selection Measures (ASM) is employed to determine the most suitable attribute for dividing the records.
(2) Dataset partition: The optimal attribute is designated as a decision node, and the dataset is partitioned into smaller subsets.
(3) Tree construction: Iteratively repeating (1) and (2) procedure for each child until one of the conditions is satisfied: all tuples belong to the same attribute value, no more attributes remain, or no more instances exist.



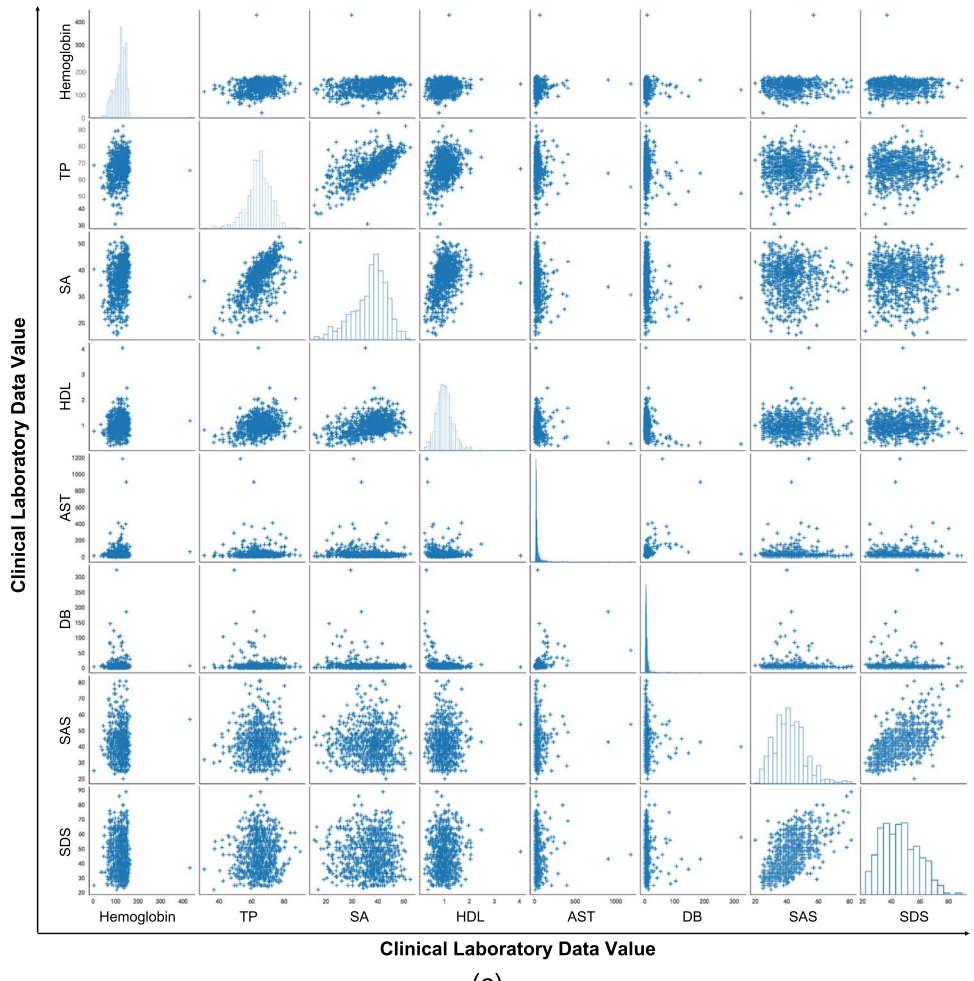


Fig. 1 Data equilibrium analysis and correlation analysis. **a** The almost overlap area of the density plot for gender and education characteristic indicates the equilibrium of the data. **b** The correlation analysis for SF-36 and PSQI. **c** The correlation analysis for clinical laboratory data

Table 1 Summary of machine learning models, characteristics, and hyperparameter settings

Model	Full name	Key characteristics	Hyperparameter settings
GaussianNB [27]	Gaussian Naive Bayes	Probabilistic model using Bayes' theorem	Not typically tunable, but priors can be adjusted.
SVC [28]	Support Vector Classifier	Finds optimal hyperplane for separation	Tune regularization *C*, kernel, and γ.
RandomForest [29]	Random Forest Classifier	Ensemble of decision trees with voting	Tune *n_estimators*, *max_depth*, and *min_samples_leaf*.
XGB [30]	XGBoost	A scalable tree boosting system	Tune learning rate *lr*, *max_depth*, and *n_estimators*.
CatBoost [31]	Categorical Boosting	Gradient boosting optimized for categorical data	Tune *depth*, *iterations*, and *lr*.
CascadeForest [32]	Cascade Forest Classifier	Deep ensemble learning with layer stacking	Tune number of estimators per layer and cascade layers.

K-nearest neighbors classifier

K-nearest neighbors (KNN) [25, 26] identifies specific training samples that are closest to the given data instance in distance. The procedures of the KNN algorithm can be characterized as follows:

(1) Compute the distance: Computing the L2 Euclidean distance between an unknown observation and other training samples in the dataset.
(2) Identify k nearest neighbors: Identifying the *k* observations in the dataset closest to the unknown observation.
(3) Make a decision: Outputing the most common output variable among these *k* nearest neighbors as the prediction.

Comparison models

Table 1 provides an overview of the machine learning models used in this study, along with their key characteristics and the main hyperparameters. The table highlights how each model was specifically configured for optimal performance in our tasks. We employed grid search[1] as the primary optimization method to determine the best hyperparameter settings for all models. The detailed value of hyperparameter settings can be seen in Appendix 3.

Results

To evaluate anxiety and depression levels in gastrointestinal inpatients, three comprehensive experiments were conducted based on eight machine learning algorithms. Besides, five accuracy metrics were used to evaluate the model's detecting accuracy (details shown in Appendix 1: Evaluation metrics). For both anxiety and depression experiments, we used a split of 88.6% for training and 11.4% for testing, with 5-fold cross-validation on the training set. The test set maintained a 1:1 patient-to-healthy ratio for evaluation.

Auto-detection of anxiety/depression

We built the Decision Tree and K-nearest neighbors model for binary classification of anxiety and depression. Six clinical laboratory data, SF-36, and PSQI scale values were utilized to train these ML models. The auto-detection results are shown in Fig. 2 and Table 2.

Figure 2 shows that the Decision Tree and KNN classifiers significantly outperformed other machine learning models in both anxiety and depression classification tasks. For anxiety (Fig. 2a), the Decision Tree achieved the highest AUC value of 0.84, followed by Random Forest (AUC = 0.806) and SVC (AUC = 0.804), indicating its superior ability to distinguish between patients with and without anxiety. For depression (Fig. 2b), the KNN classifier achieved the highest AUC value of 0.84, surpassing all other models, highlighting its effectiveness in depression classification.

From Table 2, we can see that that the Decision Tree and KNN classifiers outperform other models across multiple metrics, providing better balance between accuracy, sensitivity, and F1-score. For anxiety classification, the Decision Tree achieves the highest accuracy (81%), precision (77%), F1-score (81%), and AUC (84%), along with a strong specificity (74%) and sensitivity (86%). This indicates that the Decision Tree provides a well-rounded performance, excelling in both identifying true positive cases (sensitivity) and reducing false positives (specificity). Similarly, the KNN classifier demonstrates competitive results, with a high sensitivity of 88%, which is critical for minimizing missed anxiety cases, although its specificity (52%) is relatively lower compared to Decision Tree and other models. This suggests that KNN may be more suited for applications prioritizing sensitivity over specificity, such as initial screenings. For depression classification, the KNN model stands out with the highest accuracy (81%), F1-score (82%), and AUC (84%), alongside balanced specificity (78%) and sensitivity (84%). This

[1] Grid search is used to find the best combination of hyperparameters by testing all possible combinations from a predefined set of values.

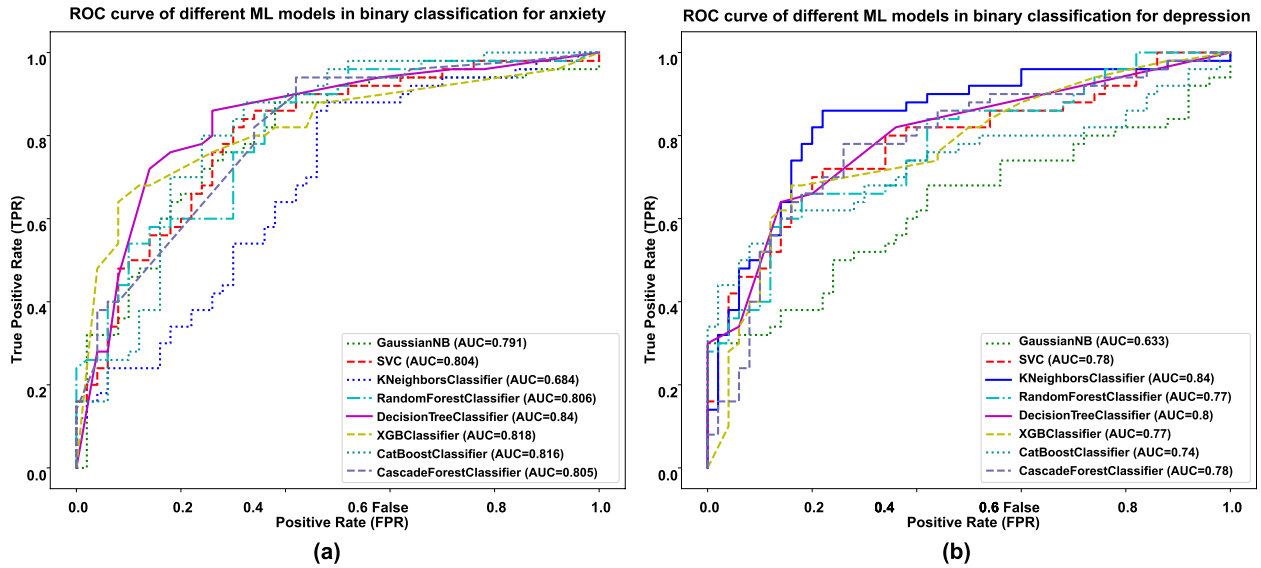


Fig. 2 Binary classification results presented by the Receiver Operator Characteristic (ROC)(ROC plots the TPR against the FPR at different thresholds, showing the trade-off between sensitivity and specificity. The closer to the upper left corner, the better the classification performance) Area Under the Curve (AUC)(AUC quantifies the ROC curve, with values closer to 1 indicating better classification performance). **a** ROC curve of different ML models for anxiety. **b** ROC curve of different ML models for depression. The AUC metric means the area under the ROC curve used to assess the performance of classification machine learning models

Table 2 Binary classification results for all machine learning models in anxiety and depression

Models	Acc	Pre	F1	AUC	Specificity	Sensitivity
Anxiety						
GaussianNB	0.73	0.69	0.75	0.79	0.64	0.82
SVC	0.74	0.72	0.75	0.8	0.70	0.78
KNeighbors	0.70	0.65	0.75	0.68	0.52	0.88
RandomForest	0.72	0.66	0.76	0.81	0.54	0.90
XGB	0.72	0.68	0.75	0.82	0.62	0.82
CatBoost	0.77	0.75	0.78	0.82	0.74	0.80
CascadeForest	0.75	0.69	0.79	0.8	0.58	**0.92**
DecisionTree	**0.81**	**0.77**	**0.81**	**0.84**	**0.74**	0.86
Depression						
GaussianNB	0.62	0.83	0.44	0.63	**0.94**	0.3
SVC	0.74	0.77	0.72	0.78	0.8	0.68
RandomForest	0.71	0.82	0.65	0.77	0.88	0.54
DecisionTree	0.75	0.82	0.72	0.8	0.86	0.64
XGB	0.76	0.81	0.74	0.77	0.84	0.68
CatBoost	0.72	**0.84**	0.66	0.74	0.9	0.54
CascadeForest	0.72	0.79	0.68	0.78	0.84	0.6
KNeighbors	**0.81**	0.79	**0.82**	**0.84**	0.78	**0.84**

robust performance indicates that KNN is particularly effective in distinguishing between patients with and without depression. In contrast, while the Decision Tree achieves a solid accuracy of 75% and an F1-score of 72%, its sensitivity (64%) is lower, suggesting it may miss more cases of depression compared to KNN.

Based on these observations, we can conclude that these two approaches are more stable and acceptable for use in clinical diagnosis for identifying anxiety and depression.

Detecting different degrees of anxiety/depression

Considering that different levels of anxiety and depression should have different treatment strategies, we evaluated the accuracy of the Decision Tree and the KNN classifier using balanced data, where data balancing aims to maintain consistent category distribution between the test set and the training set. As indicated in Table 3, all metrics of the Decision Tree and KNN improves significantly with different disorder levels. Notably, a remarkable accuracy of 90% were attained when addressing severe anxiety cases. Similarly, the K-nearest neighbors in detecting depression also yields positive results. These results may partly be explained by the fact that the mildly anxiety or depression data have high similarity with the healthy subjects. Distinguishing between depressed and healthy samples is challenging for doctors and also for machine learning algorithms. Nevertheless, if greater emphasis is placed on the sensitivity to ensure disease screenings are not missed, our machine learning algorithm achieved sensitivity values of 0.87 and 0.91 for mild anxiety or depression, respectively, which could be considered satisfactory for clinical applications.

Table 3 Detecting different degrees of anxiety and depression

Degree/Metric	Accuracy	Precision	Sensitivity	F1	AUC
Anxiety	**DecisionTree**				
mildly	0.76	0.71	0.87	0.78	0.76
moderately	0.88	0.85	0.92	0.88	0.90
severe	0.90	0.83	1.00	0.91	0.90
Depression	**KNN**				
mildly	0.65	0.65	0.91	0.76	0.80
moderately	0.79	0.74	0.90	0.81	0.86
severe	0.93	0.88	1.00	0.93	0.98

Performance without SF-36 and PSQI

Due to the early identification of anxiety and depression with only general information and clinical laboratory data is essential and practical, we built models without SF-36 and PSQI questionnaires to facilitate the development of questionnaire-free detection applications.

Figure 3 displays a comparison of the models' performance with and without the SF-36 and PSQI scores. The exclusion of these features led to declines across performance metrics, such as a 10.8% reduction in recall for anxiety detection using the Decision Tree. This highlights the important role SF-36 and PSQI play in identifying anxiety-depression disorders. Despite this decline, the models achieved over 70% accuracy using only clinical laboratory data, demonstrating the potential of objective biochemical indicators in detecting anxiety and depression. Such approaches are particularly useful in scenarios

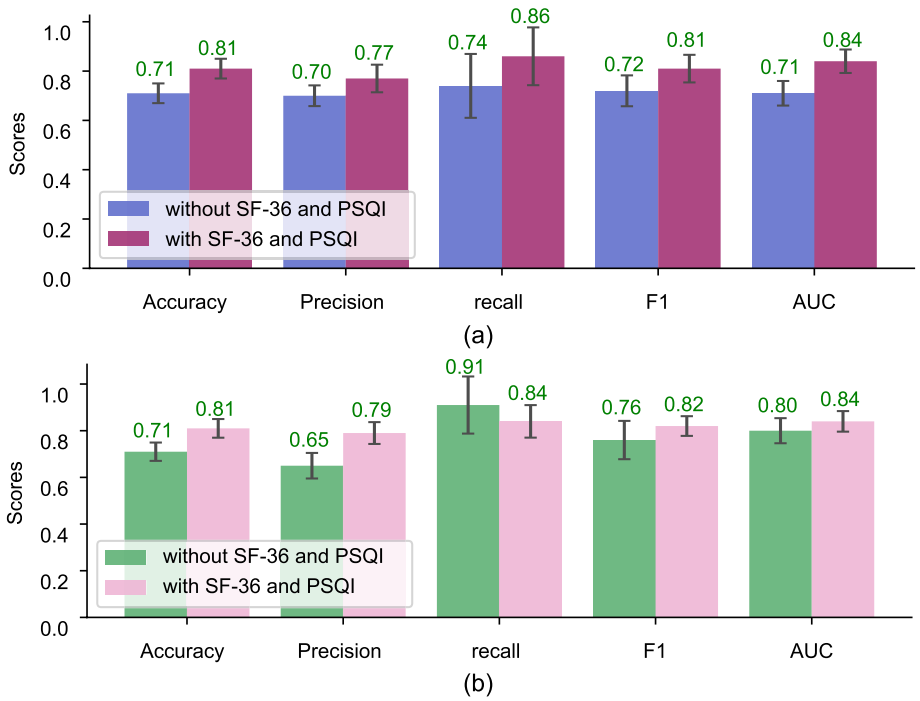


Fig. 3 Comparison of the models' performance regardless of the SF-36 and PSQI input. **a** The binary classification results of Decision Tree for anxiety. **a** The KNN binary classification results in depression

where questionnaire-based assessments are unavailable, such as time-constrained settings or patients unable to complete assessments. Future work will focus on incorporating additional laboratory features and refining algorithms to further improve questionnaire-free detection methods, enhancing their applicability in diverse clinical environments.

Discussion

Due to the higher prevalence of psychosomatic disorders in gastroenterology inpatients, this study presents a machine learning technique to identify ADCS by analyzing clinical laboratory data. To our knowledge, no prior study has directly detected the anxiety-depression status of gastroenterology patients using admittance observation data.

Our experimental study presents empirical evidence that machine learning models can extract relevant features from patients' clinical information and facilitate the initial prediction of anxiety and depression symptoms. We suggest using the Decision Tree and K-nearest neighbors model to assess the psychological condition of patients by analyzing inpatients' laboratory and questionnaire data. Compared to more complex models like Random Forest and XGBoost, Decision Tree and KNN were chosen for their higher interpretability and simplicity, which are essential for clinical decision-making. Additionally, these models demonstrated competitive performance in our study with high accuracy and AUC. Therefore, these benefits of performance, interpretability, and simplicity make them highly suitable for the diagnosis of ADCS-GI. Additionally, these automated psychological detecting methods are convenient for providing gastroenterologists with quick detection results, thus enabling doctors to identify individuals requiring special attention and prevent unnoticed psychosomatic-related diseases.

Despite this study's crucial potential application values, the following limitations must be acknowledged. Firstly, The performance of a machine learning model is heavily influenced by the size and structure of the dataset, making it difficult to generalize beyond the training data thus probabily lacking generalizability. Consequently, finding a model that simultaneously performs well on anxiety and depression has proven challenging.

Furthermore, due to the single-center origin of the dataset from Xiangya Hospital and the moderate sample size, there is a possibility of potential biases that may influence the generalizability of the findings. To minimize biases, a cross-sectional design was employed with broad disease inclusion and nurse-assisted questionnaire assessments for accuracy. Future efforts and prospective recruitment will focus on expanding the dataset and external validation to ensure broader generalizability.

Additionally, the current model does not consider the possible effects of anxiety or depression related to different gastrointestinal disorders, such as functional gastrointestinal disorders (FGIDs), including irritable bowel syndrome (IBS) and functional dyspepsia (FD), and organic gastrointestinal diseases (OGIDs). The observations of patients hospitalized in the Department of Gastroenterology revealed that psychosomatic diseases were primarily observed in two scenarios: (1) Patients with digestive tract tumors, peptic ulcers, inflammatory bowel disease, cirrhosis, and other digestive system diseases presenting anxiety and depression. (2) Patients with functional digestive disorders display symptoms of anxiety and depression despite regular examination and test results. Further research should focus on including the clinical symptoms of various gastroenterological disorders to improve the machine-learning model.

Future work will focus on expanding and diversifying the dataset. Prospective multi-center recruitment will be conducted to include patients from varied demographics, disease types, and healthcare settings. This will help mitigate biases in the current dataset and improve the model's robustness across diverse populations. Additionally, the dataset will be enriched with a broader spectrum of gastrointestinal diseases, from functional disorders like irritable bowel syndrome to organic diseases such as digestive tract tumors and inflammatory bowel disease.

To optimize the model further, automated machine learning (AutoML) pipelines will be introduced to enable dynamic online optimization of parameters and effective performance tuning. In addition, the integration of deep learning architectures will also be explored. For example, (1) combining tree-based algorithms with neural networks could leverage the strengths of interpretability and feature extraction. (2) Incorporating an attention mechanism could capture the relationships between clinical laboratory data and psychological outcomes. (3) Transfer learning is another way of utilizing pre-trained models from similar datasets, such as general mental health datasets, to enhance performance under limited data conditions.

Lastly, efforts will also be directed toward integrating the model into clinical workflows. Collaborations with clinicians will help establish a framework that emphasizes

usability and interpretability. A user-friendly interface for real-time, questionnaire-free detection will also be developed, enabling gastroenterologists to make efficient and well-informed decisions. By implementing these strategies, future research will aim to create a robust, generalizable, and practical tool to identify anxiety-depression comorbidities in gastrointestinal patients. These enhancements will ensure the model's clinical relevance and broaden its applicability in real-world settings.

Conclusions

Machine learning algorithms can be used to identify ADCS among gastroenterology patients. This can help facilitate the early detection and intervention of psychological disorders in gastroenterology patients' care.

Appendix 1

Evaluation metrics

The model's classification performance is evaluated using standard evaluation metrics including accuracy, precision, sensitivity, and F1 score. True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are utilized to calculate these metrics. The confusion matrix is comprised of these four elements.

Accuracy: The probability of correct classification:

$$accuracy = \frac{(TP + TN)}{TP + TN + FP + FN} \tag{1}$$

Precision: The probability of accurate prediction among all positive predictions:

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

Sensitivity: The proportion of positive cases that are correctly judged in the total positive samples:

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

F_1 score: F_1 score is a metric that comprehensively considers accuracy and sensitivity.

$$F_1\ score = \frac{2 * precision * sensitivity}{precision + sensitivity} \tag{4}$$

ROC curve and AUC: The Receiver Operating Characteristic curve (ROC) serves as a tool for evaluating the effectiveness of a classification model across all classification thresholds. The ROC curve uses the False Positive Rate (FPR) $= FP/(FP + TN)$ on the horizontal axis and the The True Positive Rate (TPR) on the vertical axis. The FPR known as fall-out, while TPR known as sensitivity or *sensitivity*. The Area Under the ROC Curve (AUC) is a metric for gauging the ability of a parameter to differentiate between two diagnostic groups.

Appendix 2

Importance of features

Decision tree for anxiety

Figure 4 shows the structure of the decision tree used for diagnosing anxiety. Gender appears as the feature of the first layer of the decision tree due to the algorithm's automatic feature-selection process. To mitigate potential bias, we use a variety of techniques, such as data balancing, limiting tree depth, and fairness assessment, to ensure that the model is not overly dependent on gender [33]. Moreover, splitting features closer to the root node tend to be more important than those further down the tree, such as PSQI and SF-36. Additionally, PSQI appears as a splitting condition in multiple nodes, further highlighting its importance as a key feature. Considering the changes in the Gini coefficient[2] or information gain, after splitting with the SF-36 feature, the Gini coefficient of the child nodes is significantly lower than that of the parent node. Therefore, the key features influencing the final binary classification of depression are Gender, PSQI, and SF-36. In addition, clinical laboratory data, including AST, SA, and TP, also play significant roles in identifying anxiety and depression in gastroenterology patients. For instance, elevated AST levels may indicate liver dysfunction associated with increased psychological distress [16], while changes in SA and TP levels could suggest systemic inflammation or metabolic imbalances commonly observed in patients with comorbid gastrointestinal and psychiatric conditions [17]. These findings underscore the potential utility of clinical laboratory data in enhancing diagnostic accuracy for ADCS-GI and providing a more comprehensive understanding of patient health.

[2] The Gini index quantifies the class mix in a node, with 0 indicating pure nodes and higher values indicating more mixed categories.

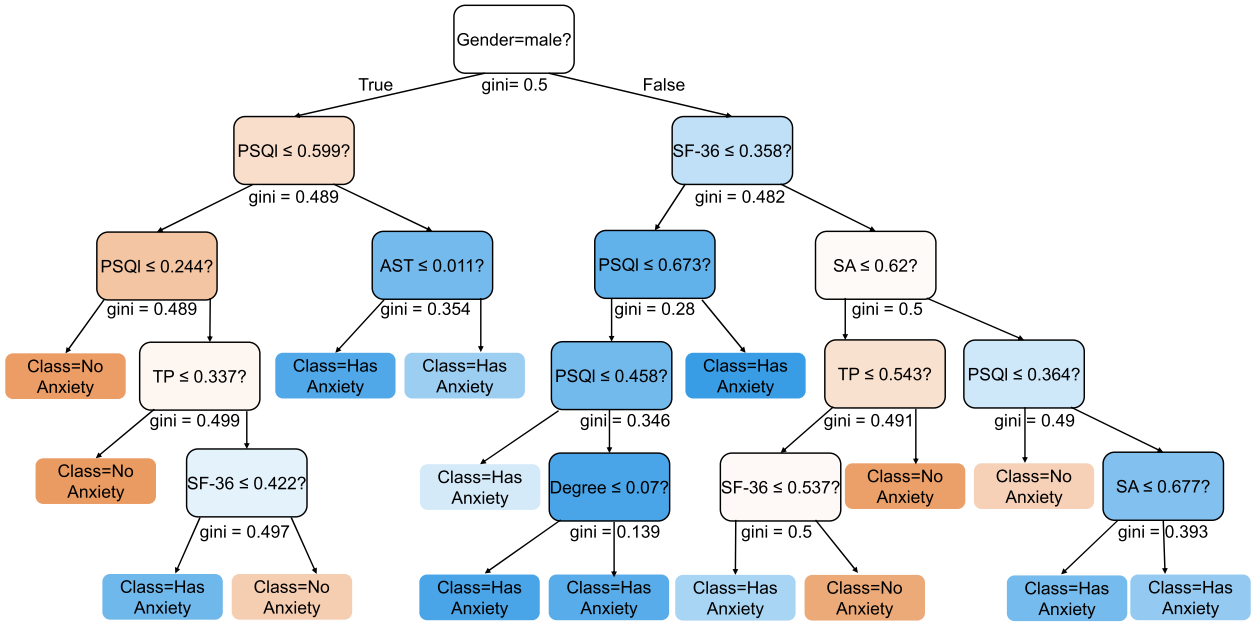


Fig. 4 Visualization of the Decision Tree model in anxiety task

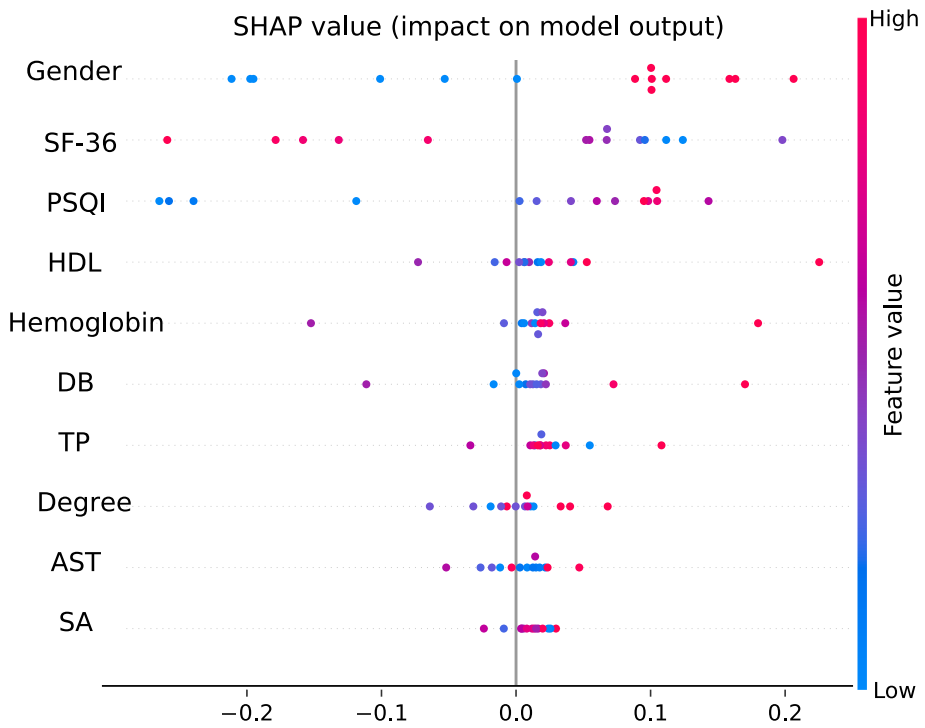


Fig. 5 Importance of features in depression task by SHAP

K-nearest Neighbors Classifier for depression

Figure 5 illustrates the importance of input features in the classification task for depression. The vertical axis ranks features using SHapley Additive exPlanations (SHAP) [34], a unified framework for interpreting machine learning models by assigning each feature an important value based on Shapley values from cooperative game theory. This method provides consistent explanations of model predictions by attributing the contribution of each feature to the output. The results indicate that features such as Gender, SF-36, and PSQI are the top contributors. In contrast, physiological indicators with a significant impact on depression classification include HDL, Hemoglobin, and DB.

Appendix 3

Model parameters

The grid search was utilized to find the optimum parameters for all machine learning models. This technique is an exhaustive search for specific parameter values and is implemented in scikit-learn [35], which requires a dictionary of hyperparameters to evaluate. Exhaustive parameter candidates were generated from GridSearchCV of scikit-learn, which allows for exploring multiple parameter grids. Eventually, the optimal parameters for the eight machine models are as follows:

Model parameters for anxiety

- GaussianNB: var smoothing=1e-9
- SVC: C=9.03, coef0=0.2, degree=1, gamma=0.1, kernel='sigmoid', maxiter=100000, probability=True
- KNeighborsClassifier: leafsize=1, njobs=-1, nneighbors=20, weights='distance', algorithm='auto', p=2
- RandomForestClassifier: criterion='gini', maxdepth = 3, maxfeatures=4, nestimators=30, bootstrap=False
- DecisionTreeClassifier: classweight='balanced', maxdepth =5, maxfeatures=8,minsamplesleaf=8, minsamplessplit =4, splitter='random', criterion='gini', minweightfractionleaf=0.0
- XGBClassifier: maxdepth=12, learningrate=0.025, minchildweight=26, gamma=1.3, subsample=1, colsamplebytree=1
- CatBoostClassifier: depth=4, earlystoppingrounds = 1000, iterations=1200, l2leafreg=9, learningrate=0.05, lossfunction='MultiClass', tasktype='GPU'
- CascadeForestClassifier: nestimators=5, criterion = 'entropy', maxdepth=3, maxlayers=3, njobs=-1, ntrees = 4, minsamplessplit=2, usepredictor=True

Model parameters for depression

- GaussianNB: var smoothing=1e-9
- SVC: C=9.03, coef0=0.8, degree=1, gamma=0.1, kernel='sigmoid', maxiter=100000, probability=True KNeighborsClassifier: leafsize=1, njobs=-1, nneighbors=58, weights='distance', algorithm='auto', p=2
- RandomForestClassifier: criterion='entropy', maxdepth = 5, maxfeatures=2, nestimators=40, bootstrap=True
- DecisionTreeClassifier: classweight='balanced', criterion='gini', maxdepth=40, maxfeatures=8, minsamplesleaf=1, minsamplessplit=5, minweightfractionleaf=0.1, splitter='best'
- XGBClassifier: maxdepth=4, learningrate=0.05, minchildweight=20, gamma=0.7, subsample=0.6, colsamplebytree=0.7, nestimators=19, regalpha=0, reglambda=0.6
- CatBoostClassifier: depth=4, earlystoppingrounds = 1000, iterations=800, l2leafreg=4, learningrate=0.01, lossfunction='MultiClass', tasktype='GPU'
- CascadeForestClassifier: nestimators=4, criterion = 'entropy', maxlayers=10, ntrees=4, maxdepth=5

Abbreviations

ADCS-GI	Anxiety-Depression Comorbidity Syndrome in Gastroenterology Inpatients (ADCS-GI)
EEG	Electroencephalography
HCs	Healthy controls
KNN	K-nearest neighbors
SF-36	The 36-Item short form health survey
PSQI	The Pittsburgh sleep quality index questionnaire
SAS	Self-Rated Anxiety Scale
SDS	Self-Rated Depression Scale
AUC	Area under curve of ROC
SVM	Support vector machine
ML	Machine learning

Acknowledgements

We would like to thank the support of Xiangya Hospital of Central South University.

Authors' contributions

M.T. and J.Z. contributed equally to this work. M.T., J.Z., and J.X. contributed to the study concept and design. M.T., J.Z., Y.T., Q.Z., and L.F. contributed to data collection and analysis. M.T., J.Z., and Y.T. contributed to programming and writing. U.S., Y.Y., and L.X. provided important advice and assistance for manuscript drafting. Z.X., L.F., and J.X. supervised the study. All authors read and approved the final manuscript and were responsible for submitting it for publication.

Data availability

The data that support the findings of this study are available from the corresponding author, [CA], upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Xiangya Hospital of Central South University. All participants provided written informed consent. All procedures were in accordance with the ethical standards of the responsible committee on human experimentation.

Competing interests
The authors declare no competing interests.

Author details
[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. [2]Department of Gastroenterology and Hepatology, Shenzhen University General Hospital, Shenzhen University, Shenzhen 518055, China. [3]Department of Gastroenterology, Xiangya Hospital, Central South University, Changsha 410008, China. [4]Department of Gastroenterology, Futian District Second People's Hospital, Shenzhen 518049, China. [5]School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China. [6]University of Chinese Academy of Sciences, Beijing 101400, China.

Received: 8 March 2024 Accepted: 27 February 2025
Published online: 18 March 2025

References

1. Koloski NA, Jones M, Kalantar J, Weltman M, Zaguirre J, Talley N. The brain-gut pathway in functional gastrointestinal disorders is bidirectional: a 12-year prospective population-based study. Gut. 2012;61(9):1284–90.
2. Kawoos Y, Wani ZA, Kadla SA, Shah IA, Hussain A, Dar MM, et al. Psychiatric Co-morbidity in Patients With Irritable Bowel Syndrome at a Tertiary Care Center in Northern India. J Neurogastroenterol Motil. 2017;23:555–60.
3. Prieto DC, Mencia PM. The effects of gastrointestinal disturbances on the onset of depression and anxiety. PLoS ONE. 2022;17:e0262712.
4. Herrman H, Patel V, Kieling C, Berk M, Buchweitz C, Cuijpers P, et al. Time for united action on depression: a Lancet-World Psychiatric Association Commission. Lancet. 2022;399:957–1022.
5. Daros AR, Haefner SA, Asadi S, Kazi S, Rodak T, Quilty LC. A meta-analysis of emotional regulation outcomes in psychological interventions for youth with depression and anxiety. Nat Hum Behav. 2021;5(10):1443–57.
6. Keefer L, Palsson OS, Pandolfino JE. Best Practice Update: Incorporating Psychogastroenterology Into Management of Digestive Disorders. Gastroenterology. 2018;154:1249–57.
7. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960;23(1):56.
8. Zung WWK. A Self-Rating Depression Scale. Arch Gen Psychiatr. 1965;12(1):63–70.
9. Testa MA, Simonson DC. Assessment of quality-of-life outcomes. New Eng J Med. 1996;334(13):835–40. https://doi.org/10.1056/NEJM199603283341306.
10. Buysse DG, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. Psychiatr Res. 1989;28(2):193–213. https://doi.org/10.1016/0165-1781(89)90047-4.
11. Feng L, Li Z, Gu X, Jiang J, Liu X. Psychosomatic Disorders in Patients with Gastrointestinal Diseases: Single-Center Cross-Sectional Study of 1186 Inpatients. Gastroenterol Res Pract. 2021;2021:1–9.
12. Li X, La R, Wang Y, Niu J, Zeng S, Sun S, et al. EEG-based mild depression recognition using convolutional neural network. Med Biol Eng Comput. 2019;57:1341–52.
13. Mumtaz W, Qayyum A. A deep learning framework for automatic diagnosis of unipolar depression. Int J Med Inform. 2019;132:103983.
14. Xie W, Wang C, Lin Z, Luo X, Chen W, Xu M, et al. Multimodal fusion diagnosis of depression and anxiety based on CNN-LSTM model. Comput Med Imaging Graph. 2022;102:102128.
15. Saberi-Karimian M, Khorasanchi Z, Ghazizadeh H, Tayefi M, Saffar S, Ferns GA, et al. Potential value and impact of data mining and machine learning in clinical diagnostics. Crit Rev Clin Lab Sci. 2021;58(4):275–96.
16. Song X, Zhang Z, Zhang R, Wang M, Lin D, Li T, et al. Predictive markers of depression in hypertension. Medicine. 2018;97(32):e11768.
17. Song X, Zheng Q, Zhang R, Wang M, Deng W, Wang Q, et al. Potential biomarkers for predicting depression in diabetes mellitus. Front Psychiatry. 2021;12:731220.
18. Deberneh HM, Kim I. Prediction of type 2 diabetes based on machine learning algorithm. Int J Environ Res Public Health. 2021;18(6):3317.
19. Muhammad G, et al. Enhancing prognosis accuracy for ischemic cardiovascular disease using K nearest neighbor algorithm: A robust approach. IEEE Access. 2023;11:97879–95. https://doi.org/10.1109/ACCESS.2023.3312046.
20. Yu T, Pei W, Xu C, et al. Prediction of violence in male schizophrenia using sMRI, based on machine learning algorithms. BMC Psychiatr. 2022;22:676. https://doi.org/10.1186/s12888-022-04331-1.
21. Priya A, Garg S, Tigga NP. Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. Procedia Comput Sci. 2020;167:1258–67.
22. Leightley D, Williamson V, Darby J, Fear NT. Identifying probable post-traumatic stress disorder: applying supervised machine learning to data from a UK military cohort. J Ment Health. 2018;28:34–41.
23. Sau A, Bhakta I. Screening of anxiety and depression among the seafarers using machine learning technology. Inform Med Unlocked. 2019;16:100149.
24. Kotsiantis SB. Decision trees: a recent overview. Artif Intell Rev. 2011;39:261–83.
25. Ali N, Neagu D, Trundle P. Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets. Appl Sci. 2019;1(12):1559. https://doi.org/10.1007/s42452-019-1356-9.
26. Peterson LE. K-nearest neighbor Scholarpedia. 2009;4(2):1883.
27. Duda RO, Hart PE, et al. Pattern classification and scene analysis, vol. 3. New York: Wiley; 1973.
28. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20:273–97. https://doi.org/10.1007/BF00994018.
29. Breiman L. Random forests. Mach Learn. 2001;45:5–32.
30. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. New York: Association for Computing Machinery; 2016. https://doi.org/10.1145/2939672.2939785.
31. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. Proceedings of the 32nd international conference on neural information processing systems. Red Hook, NY, USA: Curran Associates Inc.; 2018. p. 6639–6649. https://doi.org/10.5555/3327757.3327770.
32. Zhou ZH, Feng J. Deep forest. Natl Sci Rev. 2019;6(1):74–86.
33. González-Sendino R, Serrano E, Bajo J. Mitigating bias in artificial intelligence: Fair data generation via causal models for transparent and explainable decision-making. Futur Gener Comput Syst. 2024;155:384–401.
34. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st international conference on neural information processing systems. Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 4768–4777. https://doi.org/10.5555/3295222.3295230.
35. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.

Publisher's Note