

Machine Learning-Based Depression Risk Prediction Models for Older Adults Analyzing From the Perspective of the Health Ecology Model: A Scoping Review

Jie Yang^{1,2}, Xinyu Hao², Xiao Sang², Enshe Jiang^{1,2}, Xiaoguang Zhang¹

¹Department of Neurosurgery, The First Affiliated Hospital of Henan University, Henan University, Kaifeng, People's Republic of China; ²Institute of Nursing and Health, School of Nursing and Health, Henan University, Kaifeng, People's Republic of China

Correspondence: Enshe Jiang, Xiaoguang Zhang, Department of Neurosurgery, The First Affiliated Hospital of Henan University, Henan University, Kaifeng, People's Republic of China, Email esjiang@gmail.com; 13523781771@163.com

Objective: This review systematically examined the current research on machine learning-based depression risk prediction models for domestic and international older adults, aiming to provide references for methodological development and application in this field.

Design: A scoping review approach was used, following the Participants, Concept, and Context (PCC) framework and incorporating the Health Ecology Model (HEM) as the analytical theoretical framework.

Methods: Based on the PCC principles, the study scope was defined, and a systematic search was conducted in Web of Science, PubMed, Cochrane Library, Embase, CINAHL, CNKI, VIP Database, Wanfang Database, and the Chinese Biomedical Literature Database for literature related to depression in older adults, covering publications up to October 12, 2025. Two researchers independently screened the literature, extracted data, and summarized and analyzed the papers' content.

Results: After screening and full-text review, 15 studies on machine learning-based depression risk prediction models for older adults were included. A total of 90 risk prediction models were covered, with area under the curve (AUC) values ranging from 0.73 to 0.943. Most models were only internally validated, with only 20 models undergoing external validation. Predictors of depression in older adults were mainly at the individual intrinsic traits level (e.g. age, gender, number of chronic diseases, cognitive scores, pain) and the behavioral characteristics level (e.g. sleep duration, physical activity, smoking, drinking) of the HEM. At the same time, interpersonal network, community, environmental, and policy-level factors were less involved. These models are applicable in various settings, including community health service centers, outpatient clinics, long-term care institutions, and home health management.

Conclusion: Research on depression risk prediction models for older adults in China is still in its early stages. Existing models demonstrate good predictive performance, with a manageable risk of bias, and can provide more reliable decision support for healthcare professionals.

Keywords: depression, older adults, machine learning, depression risk, health ecology model

Introduction

The proportion of the global elderly population (aged 60 and above) continues to rise, reaching 2.1 billion by 2050, with the population aged over 80 expected to triple to 426 million over the same period.¹ As China's population ages rapidly, the risk of depression among older adults is increasingly prominent. The overall detection rate of depression among older adults in China is 20.6%, and the rate of seeking medical treatment is relatively low.² This indicates that a considerable proportion of elderly patients with depression experience delays in diagnosis and treatment, posing a significant challenge to their healthy aging process.

Depression in the elderly is a pathological condition that requires professional intervention, widely affecting emotions, cognition, behaviour and social functioning, and is not merely a matter of low mood or declining willpower.³



Compared with adult depression, its clinical manifestations are unique, with core features summarized as: significant somatic symptoms, accompanying cognitive impairments, and emotional vulnerability.⁴ The dual challenges of health decline and lack of social support in aging not only easily damage the self-esteem of older adults, but if not timely intervened, may also trigger depression.⁵ Studies show that depression is a common mental disorder among the elderly, which not only shortens patients' life expectancy, impairs their quality of life, but also increases family caregiving burdens and raises health economic expenditures for society as a whole.⁶ Currently, the clinical management of depression in the elderly focuses on strengthening the combined use of pharmacotherapy and comprehensive care models to improve patient outcomes.⁷

However, in clinical practice, early identification of depression in older adults (especially those with cognitive impairment) is significantly difficult. Traditional assessment methods mainly rely on self-report questionnaires and clinical interviews, which not only have limited sensitivity and specificity, potentially leading to missed diagnoses,⁸ but also highly depend on the cooperation of the assessed individual and the subjective experience of the clinician, making the assessment process complex and difficult to achieve rapid screening. Machine learning models can robustly predict the onset of depressive symptoms in healthy elderly individuals over the next two years and their trajectory over the following seven years, and the information relied upon by these predictive models is easily accessible.⁹

In recent years, several systematic reviews and scoping reviews on machine learning predictions of late-life depression have been published. Some studies have systematically reviewed risk prediction models for late-life depression, focusing on model performance and risk of bias assessment, but without deeply exploring the ecological dimensions of predictive factors;¹⁰ other studies have provided an overall overview of depression prediction machine learning techniques across age groups, but without particular attention to the older population or the introduction of theoretical analysis frameworks.¹¹ In addition, systematic reviews of depression prediction models based on electronic health records have largely focused on clinical populations and diagnostic outcomes, with less attention to community-dwelling older adults or risk prediction for screening purposes.¹² Therefore, this study adopts a scoping review approach, using the Health Ecology Model (HEM) as an analytical framework, to systematically review the application of machine learning in predicting depression risk in domestic and international older adults, summarize relevant algorithms and predictive factors, and provide evidence-based support for clinical decision-making.

Methods

Definition of Research Objectives QUADAS-2

Based on the PCC principle of participants, concept, and context,¹³ the study participants are older adults; the research concept is a depression risk prediction model constructed using machine learning methods; and the research context is the field of depression care for older adults. Based on a literature review, the final research objectives were determined: ① In elderly depression care, what are the commonly used machine learning algorithms? What predictive factors are usually included? ② How is the performance of these machine learning models? Have internal or external validations been conducted? ③ What are the application scenarios of machine learning in elderly depression care? This study strictly follows the PRISMA-ScR guidelines for reporting.¹⁴ Although the research protocol was not preregistered, the standardised application of these guidelines systematically presents key aspects such as the search strategy, screening process, data extraction and results synthesis, significantly enhancing the methodological transparency and reproducibility of the study.

Literature Search Strategy

Systematic searches were conducted in Web of Science, PubMed, Cochrane Library, Embase, CINAHL, China National Knowledge Infrastructure (CNKI), VIP Database, Wanfang Data, and Chinese Biomedical Literature Database (CBM). Searches combined subject headings, free-text terms, and Boolean, with the time range from database inception to October 12, 2025. Taking PubMed as an example, the search strategy was as follows: #1:machine learning, #2: machine learning OR supervised learning OR unsupervised learning OR intensive learning OR deep learning OR decision tree OR support vector machine OR Bayesian classifier OR ensemble learning OR dimensionality reduction OR metric learning

OR neural networks OR random forest, #3:#1 OR #2, #4:aged, #5:aged OR elderly OR older people OR older adult, #6:#4 OR #5, #7:Depressive Disorder, #8:Depressive Disorder OR Mental Disorders OR mental health* OR psycholog* OR depressi* OR negative emotions OR stress, #9:#7 OR #8, #10:#3 AND #6 AND #9. To ensure comprehensive coverage, manual searches using the reference-tracing method were also conducted.

Inclusion and Exclusion Criteria

Inclusion Criteria: ① Study Population: Older adults, age \geq 60 years; ② Study Content: Design of machine learning methods; ③ Study Context: Use of machine learning methods in clinical nursing, community nursing, and nursing in elderly care institutions to predict depression risk in older adults and address nursing issues for elderly patients with depression. Exclusion Criteria: ① Full text not accessible or published in languages other than Chinese or English; ② Document types such as reviews, commentaries, letters, patents; ③ Lack of independent research design and empirical data, or no clear data sources, or use of virtual data; ④ Duplicate publications.

Literature Screening and Data Extraction

The retrieved literature was imported into EndNote 20 software, and duplicate references were removed. In this study, two researchers independently completed literature screening and data extraction. During the literature screening stage, the two researchers conducted the screening using established inclusion and exclusion criteria. If disagreements arose during screening or subsequent data extraction, a third researcher was consulted and made the final decision.

Methodological Quality Assessment

This study followed the strategy of previous research and used the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies 2) tool to assess the risk of bias and applicability of the included studies.¹⁰ QUADAS-2 comprises four domains: patient selection, index test, reference standard, and flow and timing. Each domain evaluates the risk of bias, and the first three domains also consider applicability concerns. Two reviewers conducted the assessments independently, and any disagreements were resolved by a third reviewer. The rules for the overall rating of bias risk in each domain are: ① if all signaling questions in the domain are answered “yes,” the risk is rated as “low”; ② if any signaling question in the domain is answered “no,” the risk is rated as “high”; ③ if there are no “no” answers but at least one “unclear” answer, the risk is rated as “unclear.” The assessment of applicability in each aspect is based on a single question, and its direct result represents the final rating for that aspect, classified as “high,” “low,” or “unclear.” Two researchers independently evaluated the risk of bias and applicability of the included models. A third researcher resolved any disputes arising during the assessment.

Results

Literature Screening Results and Basic Characteristics of Included Studies

The initial search yielded 7,271 articles. After removing duplicates, conducting preliminary abstract screening, and performing full-text reviews in three rounds, 15 studies were ultimately deemed eligible for inclusion (Figure 1). The publication years of these 15 studies ranged from 2017 to 2025, with 3 published in Chinese and 12 in English. Geographically, most studies were conducted in China (66.7%), followed by the United States (20%), South Korea (13%), and India (6%). Study types included cross-sectional (n=7), longitudinal (n=7), and observational (n=1). The basic characteristics of the included studies are shown in Table 1.

Model Bias Risk Assessment Results

A comprehensive evaluation of the prediction model developed in this study was conducted based on the QUADAS-2 criteria. Although QUADAS-2 was originally developed for diagnostic accuracy studies, it has been widely used to assess clinical prediction models, including those based on machine learning.¹⁵ The assessment covers four key areas: patient selection, index test, reference standard, and flow and timing. For each area, the risk of bias is rated as “low,” “high” or “unclear” according to the following criteria: ① if all signalling questions are answered “yes,” the risk is rated

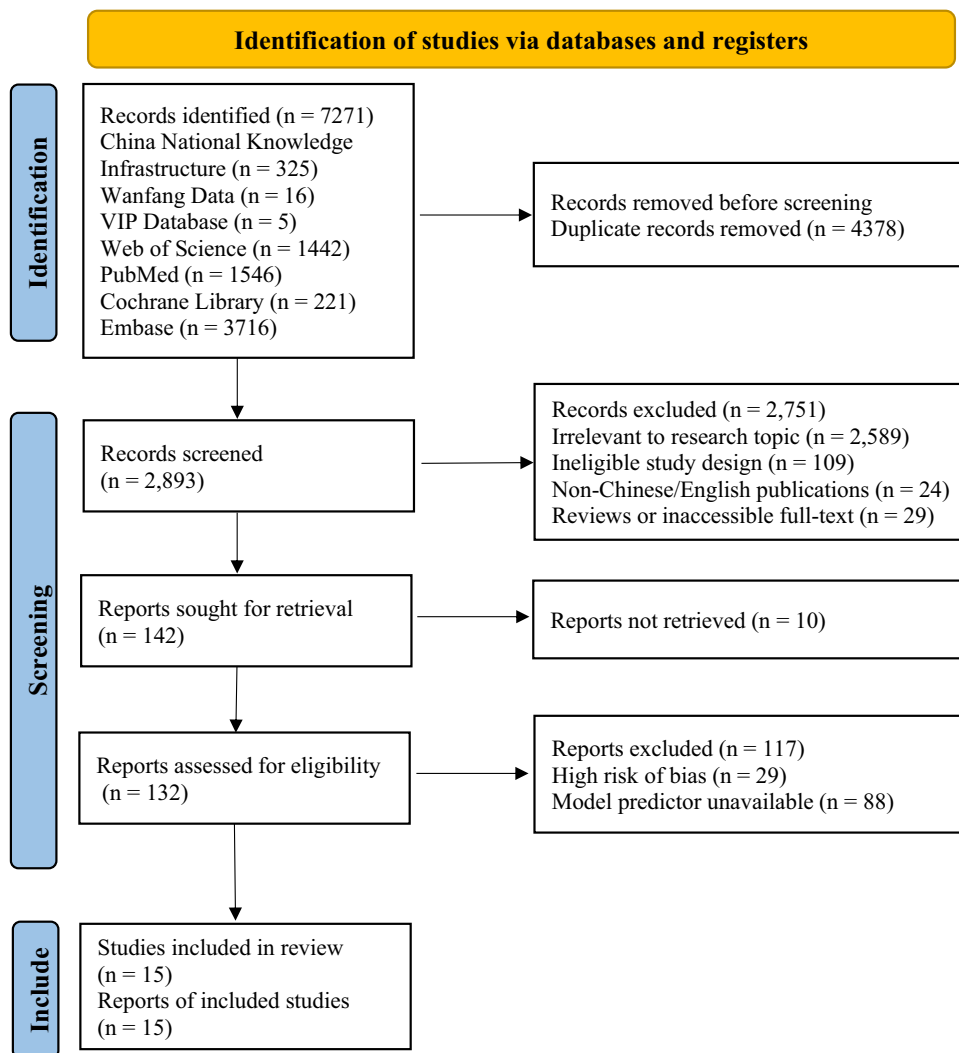


Figure 1 Flow diagram of the literature screening process. This flow diagram was adapted from the PRISMA-ScR flow diagram template. **Abbreviation:** PRISMA-ScR, Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews.

“low”; ② if any signalling question is answered “no”, the risk is rated “high”; ③ if there are no “no” answers but at least one “unclear” answer, the risk is rated “unclear”. Applicability is assessed for the first three areas. Two researchers independently evaluate each study, and any disagreements are resolved through consultation with a third researcher. Research shows¹⁶ the importance of adapting QUADAS-2 for machine learning-based predictive models, especially in terms of algorithm transparency and external validation. The results indicate that the risk of bias in critical areas such as case selection, flow, and timing is low, and the overall applicability is good. The evaluation results are shown in [Figure 2](#).

Types of Machine Learning Algorithms and Predictors Used in Depression Care for the Elderly

Among the included studies, 15 used supervised machine learning algorithms, of which 14 used random forests,^{8,9,17–28} Logistic regression appeared in 11 studies,^{8,9,18,19,21–24,26–28} support vector machines in 9 studies,^{8,18,19,23,24,26–29} decision trees^{18,19,21,22,25–27} and extreme gradient boosting^{8,18–20,22,25,26} in 7 studies each, K-Nearest Neighbors,^{9,22,25,27} light gradient boosting machine^{8,18,22,27} and multilayer perceptron^{8,18,24,29} in 4 studies each, gradient boosting^{9,20,22} and naive Bayes^{9,22,24} in 3 studies each, gradient boosted decision trees^{25,26} in 2 studies, and categorical variable boosting,¹⁸ bootstrap aggregation,¹⁸ histogram gradient boosting,¹⁸ neural

Table 1 Basic Characteristics of the Included Literature (n=15)

Study	Study Design	Country	Study Population	Sample Size	Machine Learning Algorithm(s)	Best model/ Discriminative Ability (AUC)	Predictive Factors	Validation Method	Model Presentation Form
¹⁷ Liu et al 2025	CS	China	OA-FD	4322	LR, RF, GBC, KNN, NB	LR/0.75	Age, sex, sleep duration, life satisfaction, cognitive score, residential area, self-rated health status, among 12 variables in total	Internal validation	Nomogram, SHAP plot
¹⁸ Xin et al 2022	CS	China	UR-DOA-China	1460	RFC	RFC/rural elderly with disabilities 0.79, urban elderly with disabilities 0.78	Age, sex, sleep duration, life satisfaction, education level, smoking, and alcohol consumption, among 30 variables in total	Internal validation and external validation	SHAP plot
¹⁹ Jin et al 2025	LS	China	DOA-China	8637	LR, HistGBM, MLP, XGBoost, Bagging, DT, LightGBM, RF, SVM, CatBoost	HistGBM//0.779	Age, sex, sleep duration, life satisfaction, education level, smoking, alcohol consumption, chronic diseases, BMI, and self-rated health, among 21 variables in total	External validation	SHAP plot
²⁶ Hong et al 2024	LS	China	DOA	2753	LR, RF, SVM, DT, XGBoost	XGBoost/ 0.76	Sleep duration, self-rated health, social interaction, cognition, and place of residence, among five key variables	Internal validation and external validation	SHAP plot
⁸ Li et al 2025	CS	USA	OA-CI	3472	LR, RF, SVM, LightGBM, MLP, XGBoost, KNN	XGBoost/ 0.868	Age, sex, sleep duration, marital status, education level, and body mass index (BMI), among others	Internal validation	SHAP plot
²⁰ Xu et al 2025	LS	China	C-OA-CD	1,921	RF, GBC, XGBoost	XGBoost /0.893	Sleep duration, life satisfaction, education level, gastrointestinal health, and decline in daily living, among others	Internal validation	SHAP plot
²¹ Kim et al 2023	CS	Korea	K-LI-MA /OA	2975	LR, RF, DT, NN	DT/Accuracy 97.3%	Age, sex, life satisfaction, education level, number of household members, and disability status, among others	Internal validation	SHAP plot
²² Park et al 2025	LS	Korea	OA	5153	RF, GBC, LightGBM, ET, LR, LDA, QDA, NB, ADA, DT, KNN, XGBoost,	RF /0.74	Age, sex, low-income status, pain, frequency of meeting friends, participation in social gatherings, residence in a large city, hearing discomfort, and visual discomfort, among others	Internal validation	–

(Continued)

Table I (Continued).

Study	Study Design	Country	Study Population	Sample Size	Machine Learning Algorithm(s)	Best model/ Discriminative Ability (AUC)	Predictive Factors	Validation Method	Model Presentation Form
⁹ Lin et al 2023	LS	China	H-OA	2650	LR, RF, GBDT, SVM	GBDT/ 0.770	Sex, sleep duration, life satisfaction, marital status, geographic region, education level, per capita household income, household registration type, and medical insurance, among others	Internal validation	–
²³ Sau et al 2017	LS	India	OA	510	BN, NB, LR, MLP, SMO, KS, RS, J48, RF, RT	RF/0.943	Age, sex, marital status, employment status, recent bereavement, multi-site/widespread pain, diabetes, hypertension, visual impairment, and hearing impairment, among others	External validation	–
²⁹ Xu et al 2019	LS	USA	OA	20000	MLP,SVM,DBN,Single-task LSTM,Multitask LSTM (without auxiliary variables),Multitask LSTM (full model)	Multitask LSTM (full model) / 0.873	Age, sex, marital status, mobility index, smoking status, self-rated health status, history of cancer, history of diabetes, and employment status, among others	Internal validation	–
²⁷ Ding et al 2024	CS	China	OA	8258	XGBoost, RF, DT, KNN, GNB	XGBoost /0.804	Age, sex, nighttime sleep duration, life satisfaction, marital status, place of residence, education level, pension insurance, and medical insurance, among others	Internal validation	SHAP plot
²⁸ Su et al 2025	OS	China	OA	5954	LR, RF, DT, SVM, XGBoost	XGBoost-RFE-RF /0.732	Age, sex, nighttime sleep duration, life satisfaction, marital status, education level, household registration type, and support from children, among others	Internal validation	SHAP plot
²⁴ Song et al 2025	CS	China	OA	7880	LR, KNN, DT, SVM, RF, LightGBM	LightGBM/ 0.73	Age, sex, nighttime sleep duration, nap duration, marital status, education level, region, living alone status, and self-rated health, among others	Internal validation	SHAP plot
²⁵ Lin et al 2025	CS	China	C-OA	92	RF, SVM, GBDT, LR,	RF /0.911	Body sway, right arm swing, left step length, right step length, left step height, gait speed, step width, right stride length, and left step height	Internal validation	SHAP plot

Note: “-”: not reported.

Abbreviations: CS, cross-sectional; LS, longitudinal study; OS, observational study; USA, the United States of America; OA-FD, older adults with functional disability; UR-DOA (China), urban and rural - disabled older adults; DOA-China, disabled older adults in China; DOA, disabled older adults; OA-CI, older adults with cognitive impairment; C-OA-CD, Chinese older adults with cognitive decline; K-LI-MA/OA: middle-aged and older adults from low-income households in south Korea; OA, older adults; H-OA, home-dwelling older adults; C-OA, community-dwelling older adults; NN, Neural Network.

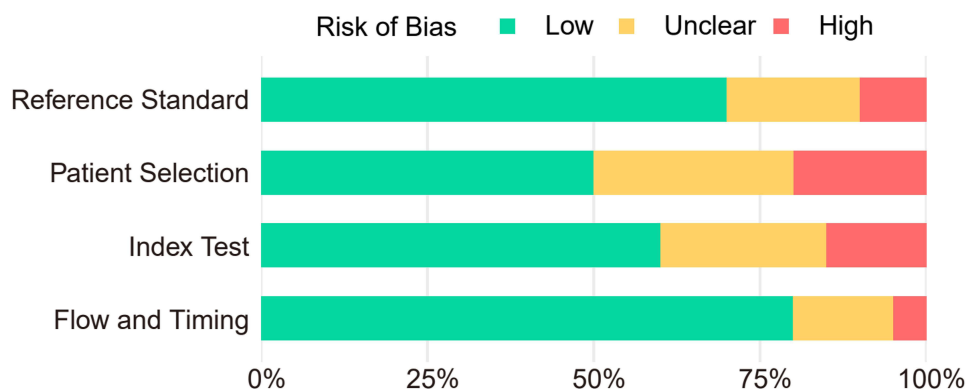


Figure 2 The quality assessment for the included literature (n=15). This figure was generated using R software (Version 4.5.2) based on study data. R is a programming language for statistical computing and data visualization.

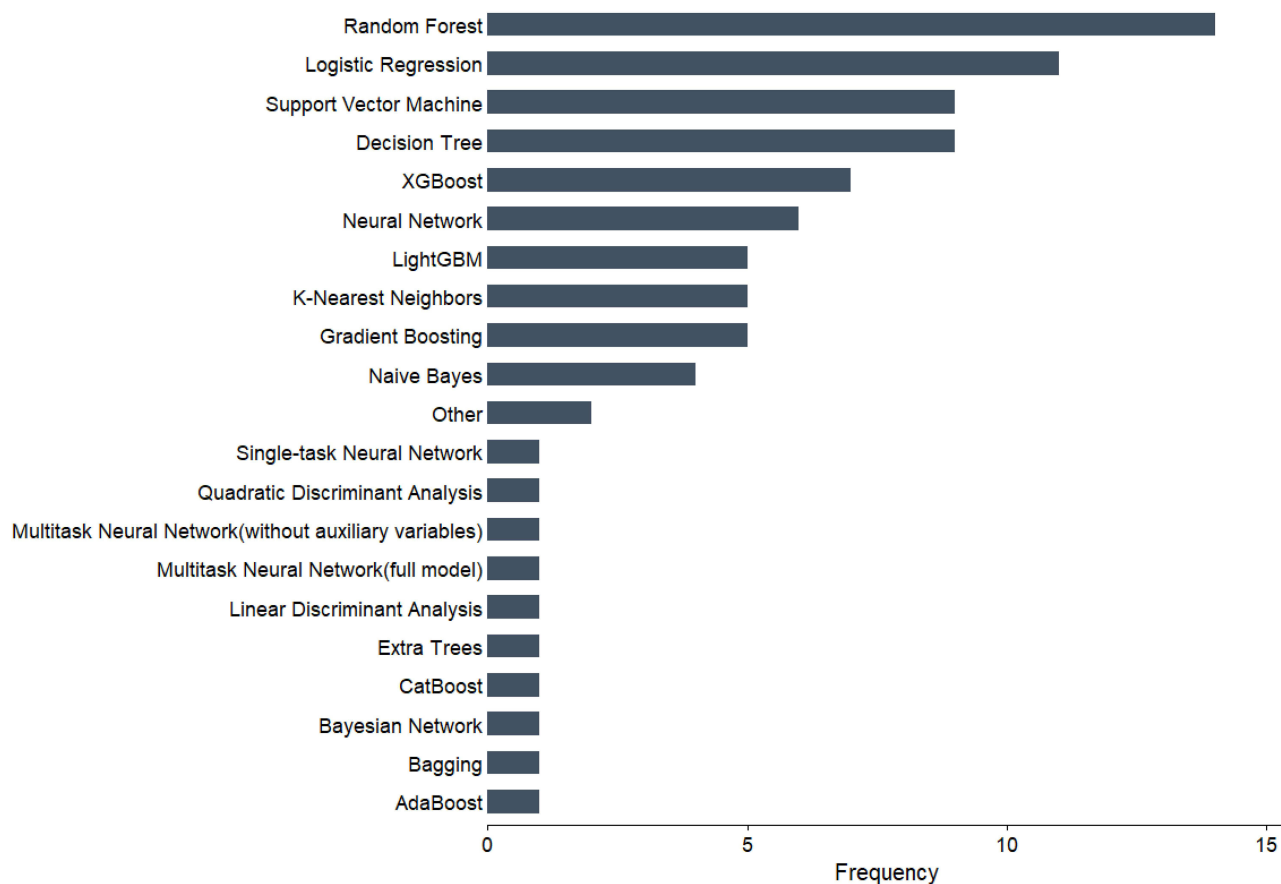


Figure 3 The use frequency of the included machine learning model algorithms. This figure was generated using R software (Version 4.5.2) based on study data.

networks,²¹ extreme trees,²² linear discriminant analysis,²² quadratic discriminant analysis,²² adaptive boosting,²² Bayesian networks,²⁴ sequential minimal optimization,²⁴ K*²⁴ random subspace,²⁴ J48,²⁴ random trees,²⁴ dynamic Bayesian networks,²⁹ Gaussian naive Bayes²⁹ and multitask LSTM models²⁹ in 1 study each. The specific distribution of each algorithm is shown in Figure 3.

This study analyzed 90 predictive models, with the study population primarily older adults and sample sizes ranging from 92 to 20,000. The number of predictive factors included in the models gradually increased from 4 to 30. The

predictive factors in existing models are highly concentrated in the layer of personal intrinsic characteristics (such as age, sex, number of chronic diseases, cognitive scores, pain) and behavioral factors (such as sleep duration, physical activity, smoking, alcohol consumption), which appeared most frequently in the included studies.^{24,25} Some studies included variables related to interpersonal networks.²¹ Only a few studies indirectly included community and environmental layers, and variables at the policy and social levels were not reflected in any of the included studies. In addition, among the 9 models with an area under the curve ($AUC \geq 0.75$) for participant work characteristics, sleep duration was included as a predictor in all cases. Research has shown that abnormal sleep duration is a potential risk factor for depressive symptoms in the elderly population, and this association is generally observed across different genders.³⁰ The distribution of depressive risk factors in older adults, based on the health ecological model, is shown in [Table 2](#).

Prediction Model Validation and Performance

In this study, 15 studies were ultimately included. Regarding model validation, 12 studies conducted internal validation, while only 4 studies^{18–20,25} conducted external validation. The performance evaluation showed that 14 studies reported model discrimination metrics, and 11 also reported calibration results. Among them, 3 studies^{24,28,29} used the Brier score to assess calibration, with all scores ≤ 0.1 , indicating good overall predictive performance of the models. Research³¹ shows that, when comparing two risk prediction models, the Brier score,²⁷ which comprehensively captures model performance, can serve as a suitable single evaluation metric. Across the 15 studies analyzed, all models reported the area under the receiver operating characteristic curve, with values ranging from 0.738 to 0.943, indicating generally good predictive performance.

Application Scenarios of Machine Learning in Predicting Depression Risk in the Elderly

Among the 15 included studies, 14 models^{9,17–29} are suitable for community health service centers; 8 models^{8,9,21,24–28} are appropriate for outpatient settings, including geriatric, rehabilitation and psychiatric outpatient departments in tertiary, secondary and community hospitals; 6 models^{9,17–19,23,27} are suitable for long-term care institutions such as nursing homes and care homes; 6 models^{17,18,20–22,27} are ideal for home health management. Among them, 4 models^{20–22,27} are suitable for multiple environments.

Table 2 Distribution of Predictive Factors for Depression Risk in the Elderly Based on the Health Ecological Model (n = 15)

HEM Level	Typical Predictors	Studies Reporting Levels	Illustrative Studies
Intrapersonal Level	Age, sex, number of chronic conditions, cognitive score, and pain, among others	15	¹⁷ Liu et al, ¹⁸ Xin et al, ¹⁹ Jin et al, ²⁶ Hong et al, ⁸ Li et al, ²⁰ Xu et al, ²¹ Kim et al, ²² Park et al, ⁹ Lin et al, ²³ Sau et al, ²⁹ Xu et al, ²⁷ Ding et al, ²⁸ Su et al, ²⁴ Song et al, ²⁵ Lin et al
Behavioral Level	Sleep duration, physical activity, smoking, and alcohol consumption, among others	10	¹⁸ Xin et al, ¹⁹ Jin et al, ²⁶ Hong et al, ²⁰ Xu et al, ²¹ Kim et al, ²² Park et al, ⁹ Lin et al, ²⁹ Xu et al, ²⁴ Song et al, ²⁵ Lin et al
Interpersonal Network Level	Marital status, social support, and contact with children, among others	7	¹⁸ Xin et al, ¹⁹ Jin et al, ²⁶ Hong et al, ²⁹ Xu et al, ²⁷ Ding et al, ²⁸ Su et al
Community and Environmental Level	Place of residence (urban/rural), among others	3	¹⁸ Xin et al, ²⁷ Ding et al, ²⁴ Song et al
Policy and Societal Level	—	—	—

Note: “—”: not reported.

Abbreviation: HEM, Health Ecological Model.

Discussion

Older-Adult Depression Machine-Learning Models Primarily Employ Supervised Learning Algorithms

In the 15 studies included in the analysis, supervised machine learning algorithms were employed, indicating that this method has a solid foundation in the care of elderly patients with depression. Research shows that, when evaluating existing elderly depression prediction models, the vast majority rely on supervised learning algorithms.²⁷ Supervised machine learning algorithms can analyze large datasets and comprehensively evaluate models using metrics such as accuracy, sensitivity, and specificity, ultimately selecting the best predictive model.²¹ These primarily include decision trees, random forests, neural networks, and logistic regression; each algorithm has distinct advantages and disadvantages.³¹ All 15 studies used multiple algorithms to construct models and conduct comparisons, and the recommended best algorithms varied between studies. For example, Sau et al²³ and Xu et al²⁰ respectively recommended the use of random forests and extreme gradient boosting. This heterogeneity may be related to research objectives, model functions, predictive variables and dataset characteristics.

In summary, apart from the frequency of algorithm usage, this review reveals the trade-off between model interpretability and predictive performance. Linear models such as logistic regression are highly interpretable, allowing clinicians to understand the direct contribution of each predictive factor (such as sleep duration, age, and the number of chronic diseases); whereas ensemble methods like random forests and extreme gradient boosting, although offering superior discriminative performance (with AUC reaching up to 0.943), often exhibit “black box” characteristics that obscure the relationship between inputs and outputs. This trade-off has direct implications for clinical practice: in community screening scenarios, highly interpretable models facilitate doctor-patient communication; in complex risk stratification studies, acceptable predictive accuracy may take precedence over interpretability. Future research should focus on developing interpretable artificial intelligence techniques suitable for predicting geriatric depression to bridge the gap between accuracy and interpretability. Furthermore, algorithm selection should be based on specific data characteristics and application scenarios, rather than simply applying common methods.

The Predictive Factors of the Depression Risk Prediction Model for the Elderly Exhibit Centralized Distribution

This study, through a systematic analysis using the HEM framework, found that the factors included in existing machine learning prediction models show a clear “centralized” distribution pattern. As shown in Table 2, the predictive factors are primarily concentrated on personal intrinsic traits (eg., age, chronic diseases) and behavioral characteristics (eg., sleep duration), which underpin risk identification in current models. Our findings corroborate previous research, with age, gender, number of chronic diseases, and cognitive scores being the most common predictive factors.³² This centralized distribution is not merely based on empirical observation, but is increasingly supported by neurobiological and genetic evidence. For example, consistently including physical activity as a predictor can be linked to its modulatory effects on neural circuits involved in stress responses.³³ Studies have shown that physical activity affects the prefrontal cortex-amygdala pathway, a core circuit in the pathophysiology of depression, providing a mechanistic explanation for its predictive value.³⁴ Similarly, the prominent role of interpersonal factors such as social support can be interpreted through the lens of gene–environment interactions. Research has found that variations in the OXTR gene, associated with social connectivity, moderate the impact of social environments on depressive symptoms. These findings suggest that the predictors identified in current models are not merely statistical associations but exist due to their direct effects on known aetiological pathways. Incorporating this mechanistic understanding into future model development can enhance the interpretability and clinical credibility of machine learning predictions. Research shows that the quality of sleep at night is a key predictor of the occurrence of depression.²⁷ The risk prediction model for depression in the elderly is a powerful tool that can assist healthcare professionals in more accurately identifying related risk factors, thereby optimizing care decisions and ultimately providing patients with a higher-quality medical service experience.³⁵

Predictive factors in risk models for depression in older adults are critical.³⁶ Still, existing models have not adequately integrated multidimensional risk factors, including environmental factors, dynamic social networks, and genetics, thereby

limiting the comprehensiveness of predictions. Before conducting depression prediction modelling, data mining techniques should be fully utilised to systematically explore patients' baseline information, related scales, laboratory indicators and medical history in order to identify key predictive factors. Research has found that, compared with demographic variables, psychological factors (such as satisfaction with leisure activities, social support, subjective health awareness and family support) have a more prominent impact on depression, providing an important basis for subsequent variable selection and model construction.²¹ However, from the perspective of ecosystem integrity, existing models exhibit "incomplete ecology". Firstly, the singularity and retrospective nature of data sources are the fundamental reasons. Among the studies included in the research institute, Three^{8,18,19} were retrospective cohort studies, and four^{20,22-24} were prospective cohort studies. In addition, five studies^{20,22,26,27,29} excluded missing data directly, which may lead to measurement errors and an increased risk of overfitting.^{35,36} It is recommended to handle missing data using imputation, filling, or retaining the missing value methods. Research shows that implicit inclusion criteria and variables based on patient self-reports can yield unrepresentative samples and misclassification, thereby overestimating the model's predictive performance.³⁷ Finally, due to limitations in model validation strategies, only four studies^{17-19,24} conducted external validation.

Research has shown that the degree of greening in the living environment is significantly associated with the risk of depression in older adults.³⁸ The lack of such validation makes it difficult for us to assess whether a model trained on "individual-centered" data remains robust when applied to new populations with different community characteristics, cultural backgrounds, and policy environments.

The Use of Machine Learning Models for the Care of Depression in Older Adults Is Extensive, but Their Construction Methods Require Further Optimization

Future research should guide the entire model-construction process by drawing on theoretical frameworks such as HEM. At present, depression prediction models for older adults in China are still in the early stages of development. Subsequent efforts should focus on methodological standards for model development and validation, on constructing prediction models based on large-sample, multicenter, prospective studies, thereby improving research design, enhancing the scientific validity and credibility of the prediction models, and ultimately providing a more reliable predictive tool for clinical practice.

The results of this study indicate that the predictive models from 15 included studies have been attempted in various settings, such as community health service centers, outpatient departments of medical institutions at all levels, long-term care facilities, and home health management, demonstrating the broad applicability of machine learning techniques in identifying depression risk in older adults. However, several key issues in the construction methods of current models remain and require urgent improvement. The samples in this study are relatively homogeneous, and the procedures for handling missing data are not sufficiently standardized, which may lead to selection bias and an increased risk of overfitting.³⁹ The final establishment of the model must undergo rigorous external validation and continuous iterative optimization.³⁶

In the absence of a logical explanation consistent with clinical mechanisms, the model's interpretability is insufficient. However, the findings in our study are consistent with previous research, with the core of model evaluation lying in its applicability and generalizability.⁴⁰ In model construction, external validation should be strengthened to assess the ecological validity of the model, thereby promoting the translation of a "full-ecological" model into clinical practice and providing a basis for personalized interventions. Research indicates that there is a significant association between the level of greenery in the living environment and the risk of depression among the elderly. Due to the lack of such validation, it is difficult for us to assess whether models trained on "individual-centred" data can maintain their effectiveness when applied to new populations with different community characteristics, cultural backgrounds and policy environments.⁴¹

Limitations

This scoping review has several limitations. First, although we conducted a systematic and comprehensive search of nine international and Chinese databases, only studies published in English or Chinese were included, and this language restriction may result in publication bias and the omission of relevant studies published in other languages. Second, due to the methodological design of a scoping review, this study did not perform a formal meta-analysis or quantitative synthesis of model performance metrics, limiting the ability to directly compare the predictive accuracy between different algorithms and populations. Third, most of the included studies were cross-sectional in design, which cannot infer causality and may overestimate the predictive ability of certain factors due to unclear temporal sequences. Fourth, there is considerable heterogeneity across studies in terms of outcome definitions, predictor selections and reporting standards, posing challenges to cross-study comparisons and synthesis. Fifth, although this study used the health ecology model as an analytical framework to map the distribution of predictors, the original studies were not designed based on this model, which may limit the depth of theoretical interpretation. Finally, although the QUADAS-2 tool was used to assess study quality, there is currently no dedicated risk of bias assessment tool specifically for machine learning prediction models, which may affect the accuracy of methodological quality evaluation. As this field develops, future review studies should consider using updated specialised assessment tools.

Conclusions

This scoping review systematically examined 15 studies on machine learning-based models for predicting depression risk in older adults and analysed them from the perspective of the Health Ecology Model (HEM). The main conclusions are as follows: first, the predictive factors are primarily concentrated at the level of individual intrinsic traits (such as age, chronic diseases, cognitive scores) and behavioural characteristics (such as sleep duration, physical activity), while factors at the level of interpersonal networks, community, environment and policy are significantly underestimated. Second, although these models demonstrated good discriminative performance (AUC range: 0.73–0.943), only 20 models underwent external validation, raising concerns about their generalisability across different populations and settings. Third, supervised learning algorithms, particularly random forests and logistic regression, were used most frequently, but the trade-off between model interpretability and predictive accuracy remains inadequately addressed.

These findings are significant. However, the effects of insufficient external validation have yet to be fully articulated from an ecological perspective. Given the highly context-dependent nature of community and policy factors, the lack of cross-ecological validation may substantially undermine the generalizability of models, which deserve the further investigation.

Future research should focus on three directions. First, model development should be guided by ecological frameworks such as HEM, systematically incorporating multi-level determinants, including macro-environmental factors. Second, external validation across different geographical, cultural, and policy contexts is crucial for establishing the model's reliability and ecological validity. Third, improving model interpretability through explainable artificial intelligence techniques, while enhancing predictive accuracy, will help facilitate the application of the model in clinical practice. Addressing these key issues will aid in developing models that are more ecologically sound, more generalisable, and better suited for predicting depression in older adults.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by the National Science Foundation of Henan Province (No. 252300423149), the Medical Science and Technology Program of Henan Province (SBGJ202502084), and the Open Project Research of the First Affiliated Hospital of Henan University (KFMS24004).

Disclosure

The authors declare that they have no conflicts of interests in this work.

References

1. WHO. Ageing and health. 2025. <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>. Accessed Mar 12, 2026.
2. Yue W, Qing C, Lurong L. Detection Rate of Depression and Its Influencing Factors in Chinese Elderly: a Meta-analysis. *Chinese General Practice*. 2023;26(34):4329–4335. doi:10.12114/j.issn.1007-9572.2023.0269
3. Paiva TC, Soares L, Faria AL. Depression in Elderly People. *Encyclopedia*. 2023;3(2):677–686.
4. Vyas CM, Donneyong M, Mischoulon D, et al. Association of Race and Ethnicity With Late-Life Depression Severity, Symptom Burden, and Care. *JAMA Network Open*. 2020;3(3):e201606–e201606. doi:10.1001/jamanetworkopen.2020.1606
5. He ZF, Tan WY, Ma H, et al. Prevalence and factors associated with depression and anxiety among older adults: a large-scale cross-sectional study in China. *J Affective Disorders*. 2024;346:135–143. doi:10.1016/j.jad.2023.11.022
6. Stecher C, Cloonan S, Domino ME. The economics of treatment for depression. *Ann Rev Public Health*. 2023;45:1.
7. Kasper S, Incalzi RA, Bozzali M, Morgado P, Siwek M. Managing depression in the elderly: real-world clinical considerations and perspectives. *Int Clin Psychopharmacol*. 2025;40(6):339–352. doi:10.1097/yic.0000000000000577
8. Li E, Ai F, Tian Q, Yang H, Tang P, Guo B. Develop and validate machine learning models to predict the risk of depressive symptoms in older adults with cognitive impairment. *BMC Psychiatry*. 2025;25(1):219.
9. Lin S, Wu Y, He L, Fang Y. Prediction of depressive symptoms onset and long-term trajectories in home-based older adults using machine learning techniques. *Aging Mental Health*. 2023;27(1):8–17. doi:10.1080/13607863.2022.2031868
10. Tan J, Ma C, Zhu C, et al. Prediction models for depression risk among older adults: systematic review and critical appraisal. *Ageing Res Rev*. 2023;83:101803. doi:10.1016/j.arr.2022.101803
11. Rahimapandi H, Maskat R, Musa R, Ardi N. Depression prediction using machine learning: a review. *IAES*. 2022;11:1108. doi:10.11591/ijai.v11.i3.pp1108-1118
12. Nickson D, Meyer C, Walasek L, Toro C. Prediction and diagnosis of depression using machine learning with electronic health records data: a systematic review. *BMC Med. Inf. Decis. Making*. 2023;23(1):271. doi:10.1186/s12911-023-02341-x
13. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. *Int J Evid Based Healthc*. 2015;13(3):141–146. doi:10.1097/xeb.0000000000000050
14. D Vinnakota. Mastering the Art of Scoping Reviews: a Comprehensive Guide for Public Health and Allied Health Students. *Asian J Public Health Nursing*. 2024;1(2):52–64. doi:10.62377/j544ed47
15. Chng C-L, Zheng K, Kwee AK, et al. Application of artificial intelligence in the assessment of thyroid eye disease (TED) - a scoping review Original Research. *Front Endocrinol*. 2023;14:1300196. doi:10.3389/fendo.2023.1300196
16. Jayakumar S, Sounderajah V, Normahani P, et al. Quality assessment standards in artificial intelligence diagnostic accuracy systematic reviews: a meta-research study. *Npj Digital Med*. 2022;5(1):11. doi:10.1038/s41746-021-00544-y
17. Liu D, Tian Y, Liu M, Yang S. Developing an interpretable machine learning model for screening depression in older adults with functional disability. *J Affective Disorders*. 2025;379:529–539. doi:10.1016/j.jad.2025.02.110
18. Xin Y, Ren X. Predicting depression among rural and urban disabled elderly in China using a random forest classifier. *BMC Psychiatry*. 2022;22(1):118.
19. Jin T, Halili A. Predicting the risk of depression in older adults with disability using machine learning: an analysis based on CHARLS data. *Front Artificial Intelligence*. 2025;8:1624171.
20. Xu J, Zhang W, Liu W. A machine learning-based approach to predict depression in Chinese older adults with subjective cognitive decline: a longitudinal study. *Sci Rep*. 2025;15(1):26985. doi:10.1038/s41598-025-10645-3
21. Kim K-M, Kim J-H, Rhee H-S, Youn B-Y. Development of a prediction model for the depression level of the elderly in low-income households: using decision trees, logistic regression, neural networks, and random forest. *Sci Rep*. 2023;13(1):11473. doi:10.1038/s41598-023-38742-1
22. Park JW, Ko CW, Lee DY, Kim JC. Prediction of late-onset depression in the elderly Korean population using machine learning algorithms. *Sci Rep*. 2025;15(1):1196. doi:10.1038/s41598-025-85157-1
23. Sau A, Bhakta I. Predicting anxiety and depression in elderly patients using machine learning technology. *Healthcare Technol Lett*. 2017;4(6):238–243. doi:10.1049/htl.2016.0096
24. Song YLQ, Chen L, Liu H, Liu Y. Machine learning algorithms to predict depression in older adults in China: a cross-sectional study. *Front Public Health*. 2025;12:1462387.
25. Lin S, Li S, Fang Y. Interpretable machine learning for depression recognition with spatiotemporal gait features among older adults: a cross-sectional study in Xiamen, China. *BMC Geriatr*. 2025;25(1):453. doi:10.1186/s12877-025-06101-6
26. Hong S, Jiang Y. Construction and evaluation of depression risk prediction model for disabled elderly people. *Mod Preventive Med*. 2024;51(15):2818–2823.
27. Ding F, Zhang Y, Tang R, Liu W, Yang X. Analysis of Influencing Factors of Depressive Symptoms in the Elderly in China Based on Explainable Machine Learning. *J Ningxia Med Univ*. 2024;46(12):1259–1264.
28. Su WY, Dong SH, Qing Y, Guifeng M. Prediction of depression symptoms in seniors and analysis of influencing factors based on explainable machine learning. *Chin J Epidemiol*. 2025;46(2):316–324.

29. Xu Z, Zhang Q, Li W, Li M, Yip PSF. Individualized prediction of depressive disorder in the elderly: a multitask deep learning approach. *Int J Med Inform.* 2019;132:103973. doi:10.1016/j.ijmedinf.2019.103973
30. Liao F, Wang W, Zhou B, et al. Longitudinal Cohort Study of the Relationship between Sleep Duration and Depressive Symptoms in Older People in China. *Sichuan Da Xue Xue Bao Yi Xue Ban.* 2022;53(1):109–113. doi:10.12182/20220160204
31. Salzberg SL. C4.5: programs for Machine Learning by J. Ross Quinlan. In: *Machine Learning*. Morgan Kaufmann Publishers, Inc;1994:16:235–240. doi:10.1007/BF00993309
32. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J.* 2021;14(1):49–58. doi:10.1093/ckj/sfaa188
33. Liu Y, Peng S, Wu X, et al. Neural, cognitive and psychopathological signatures of a prosocial or delinquent peer environment during early adolescence. *Developl Cognitive Neurosci.* 2025;73:101566. doi:10.1016/j.dcn.2025.101566
34. Liu H, Zhi J, Zhang C, et al. Association between Weight-Adjusted Waist Index and depressive symptoms: a nationally representative cross-sectional study from NHANES 2005 to 2018. *J Affect Disord.* 2024;350:49–57. doi:10.1016/j.jad.2024.01.104
35. Sisk R, Sperrin M, Peek N, van Smeden M, Martin GP. Imputation and missing indicators for handling missing data in the development and deployment of clinical prediction models: a simulation study. *Stat Methods Med Res.* 2023;32(8):1461–1477. doi:10.1177/09622802231165001
36. Saelmans A, Seinen T, Pera V, et al. Implementation and Updating of Clinical Prediction Models: a Systematic Review. *Mayo Clin Proc Digit Health.* 2025;3(3):100228. doi:10.1016/j.mcpdig.2025.100228
37. Ku WL, Min H. Evaluating Machine Learning Stability in Predicting Depression and Anxiety Amidst Subjective Response Errors. *Healthcare.* 2024;12(6). doi:10.3390/healthcare12060625
38. Wu Q, Li C, Zhang L, Zhou Y. The mitigation effects of residential green space and low air pollution on socioeconomic inequalities in depression. *Npj Mental Health Research.* 2025;4(1):33. doi:10.1038/s44184-025-00152-8
39. Gascon M, Triguero-Mas M, Martínez D, et al. Residential green spaces and mortality: a systematic review. *Environ Int.* 2016;86:60–67. doi:10.1016/j.envint.2015.10.013
40. Yuan L, Mengxi Z, Yuqian D, Haiye R, Jiaxin L, Liping Z. Machine learning for fall prediction and monitor of older adults: a systematic review. *Chin J Nurs Educ.* 2023;20(12):1513–1519.
41. Ayoub M, Liao Z, Li L, Wong KKL. HViT: hybrid vision inspired transformer for the assessment of carotid artery plaque by addressing the cross-modality domain adaptation problem in MRI. *Comput Med Imaging Graph.* 2023;109:102295. doi:10.1016/j.compmedimag.2023.102295

Neuropsychiatric Disease and Treatment

Publish your work in this journal

Neuropsychiatric Disease and Treatment is an international, peer-reviewed journal of clinical therapeutics and pharmacology focusing on concise rapid reporting of clinical or pre-clinical studies on a range of neuropsychiatric and neurological disorders. This journal is indexed on PubMed Central, the 'PsycINFO' database and CAS, and is the official journal of The International Neuropsychiatric Association (INA). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/neuropsychiatric-disease-and-treatment-journal>

Dovepress
Taylor & Francis Group