

Stacking-Driven Explainable Machine Learning Method: Uncovering Anxiety Risk in Elderly Women under Home-based Care from CLHLS

Liam Anderson¹ Sophia Martinez² Ethan Reynolds³

¹Department of Cognitive Sciences, Faculty of Behavioral Sciences, Alderton University, Bristol, UK

²Department of Global Affairs, Faculty of Political Studies, Eastbridge College, Manchester, UK

³Department of Computer Systems, Faculty of Information Technology, Kingswell Institute of Technology, London, UK

⁴Department of Learning and Development, Faculty of Pedagogy, Northfield University, Sheffield, UK

* Correspondence:

Dr. Olivia Bennett

Department of Learning Sciences

Faculty of Education, Bristol University

BS8 1TH, Bristol, UK

Email: olivia.bennett@bristol.ac.uk

Phone: +44 (117) 555 4321

Fax: +44 (117) 555 8765

Abstract

Background

With the global population aging rapidly, elderly women receiving home-based care are increasingly vulnerable to anxiety disorders. Accurate predictive models are essential to improve mental health outcomes in this at-risk group. This study aims to develop an interpretable machine learning model to predict the likelihood of anxiety symptoms among elderly women under home care in China.

Methods

Data from 2,313 women aged 65 and older, collected during the fifth wave (2008) of the China Longitudinal Health Longevity Survey (CLHLS), were used for model development and internal validation. Forty variables were selected based on the biopsychosocial framework. Key predictors were identified using LASSO regression, followed by application of three data balancing techniques to address class imbalance. A Stacking ensemble model combining six machine learning algorithms was built to predict anxiety disorder risk. Model interpretability and feature importance were assessed using Shapley Additive exPlanations (SHAP).

Results

Anxiety disorder prevalence was substantial among elderly women receiving home-based care. After balancing the data, the stacking ensemble model outperformed individual algorithms, achieving an AUROC of 0.833 and an AUPRC of 0.818. SHAP analysis revealed that significant predictors spanned multiple domains including health status (e.g., self-rated health, chronic pain), lifestyle factors (e.g., sleep duration, physical activity), and social determinants (e.g., social support, financial stability).

Conclusion

The interpretable stacking ensemble model demonstrated superior predictive accuracy (AUC=0.818) relative to single-model approaches by integrating multidimensional risk factors from biological, psychological, and social domains. Key

predictors identified through SHAP—such as severity of body pain, sleep duration, self-rated health, and environmental factors like air quality satisfaction and hearing impairment—offer valuable insights for targeted interventions. This model can assist community healthcare providers in stratifying elderly women under home care into risk categories, enabling early psychological assessments and tailored support. The fusion of machine learning with the biopsychosocial model represents a promising approach for enhancing mental health management in geriatric populations through comprehensive intervention strategies.

Limitations

This study has several limitations: (1) the cross-sectional nature of the 2008 CLHLS data limits longitudinal assessment of risk factor trajectories; (2) reliance on self-reported GAD-7 may lead to recall bias and underestimation of subclinical anxiety; (3) LASSO-based feature selection may omit clinically relevant factors with weak statistical signals, such as financial strain and caregiving burden. Future work should incorporate longitudinal biomarkers and broader socioeconomic stressors to improve the model's clinical relevance and applicability.

Keywords: anxiety risk, elderly women, home-based care, China, machine learning, prediction model

Background

Late-life anxiety disorder poses a significant public health concern, especially among elderly women receiving home-based care. This population experiences a 1.5 to 2 times higher risk of anxiety following menopause, largely due to estrogen decline and social isolation triggered by factors such as spousal loss or the migration of younger family members to urban areas (Wei et al., 2019). With projections estimating that 22% of the global population will be aged 60 or older by 2050 (Wang et al., 2022), the growing number of elderly individuals reliant on home care highlights an urgent demand for accurate risk prediction tools.

Existing traditional models fall short in addressing the complex interplay of physiological vulnerabilities—such as neurotransmitter imbalances from hormonal changes—psychosocial stressors like chronic loneliness affecting nearly 30% of this group (Cui et al., 2024), and the substantial caregiving burdens reflected by a 40% anxiety rate among long-term family caregivers (Diaz-Gallo et al., 2021). These limitations hinder timely interventions, which are crucial to preventing the cascading effects of anxiety on functional decline, family dynamics, and escalating healthcare costs. Consequently, there is a pressing need to develop an interpretable and precise prediction model capable of estimating anxiety risk among elderly women receiving home-based care.

Previous research has attempted to create anxiety risk prediction models for elderly women (Tan et al., 2023; Li et al., 2025), but these studies exhibit several notable shortcomings. Many models rely on a limited set of risk factors, often focusing narrowly on physical health or social support, while neglecting other vital contributors such as psychological resilience and lifestyle habits (Kim et al., 2021). Such a limited perspective undermines prediction accuracy, given the multifactorial nature of anxiety. Additionally, the complex interactions between physiological, psychological, and social factors are frequently overlooked, resulting in incomplete understanding of anxiety risk mechanisms (Guo et al., 2021). Furthermore, few studies have specifically addressed the unique cultural and social characteristics of elderly women under home-based care in China, including how traditional values influence mental health outcomes (Wang et al., 2023). These gaps underscore the necessity for a more comprehensive and culturally sensitive investigation.

Machine learning has emerged as a powerful tool for disease risk prediction, offering unparalleled capabilities to uncover complex patterns in large datasets and transform healthcare decision-making (Ehtemam et al., 2024). Notably, machine learning excels at modeling non-linear relationships among variables, which traditional methods often fail to capture (Vu et al., 2025). However, despite its promise, current machine learning applications for anxiety prediction exhibit two key methodological challenges.

First, many rely on single-algorithm frameworks (e.g., logistic regression or random forest alone), limiting computational accuracy by failing to leverage the complementary strengths of multiple learners (Qiu et al., 2024). Second, insufficient consideration is given to the challenges posed by cross-sectional survey data, such as missing key variables and data imbalance—common issues in epidemiological datasets like the China Health and Retirement Longitudinal Study (CLHLS) (Wang et al., 2024). Few studies comprehensively analyze how different data balancing techniques (e.g., SMOTE vs. ADASYN) interact with feature selection processes and affect model interpretability (Zheng et al., 2024). This lack of algorithmic diversity and transparency in data preprocessing significantly limits the clinical utility of predictive models, particularly when handling complex geriatric data encompassing biological, psychological, and social determinants.

In this study, we focus on elderly women receiving home-based care in China, utilizing data from the 2008 CLHLS cohort. We employ LASSO regression for feature selection grounded in the Bio-Psycho-Social model and apply three data balancing techniques to construct a stacking ensemble model comprising five machine learning algorithms. Finally, SHAP analysis is used to interpret feature importance, aiming to develop a more accurate and interpretable anxiety risk prediction model. This model intends to facilitate early intervention strategies and improve mental health outcomes in this vulnerable population.

Method

2.1 Data Sources

The China Longitudinal Health Longevity Survey (CLHLS) is designed to collect high-quality microdata representing Chinese households and individuals, specifically focusing on elderly people aged 65 and above. This survey is conducted regularly across 23 provinces and cities in China, covering various dimensions such as health status, economic background, and lifestyle. As one of the largest longitudinal surveys of elderly populations worldwide, CLHLS data plays a crucial role in researching healthy aging, longevity mechanisms, and informing related policy development. The baseline and follow-up data were obtained with approval from the Ethics Review Committee of Peking University (IRB00001052-13074), with all participants providing informed consent. This study uses data from the fifth wave of CLHLS conducted in 2008 (<http://CLHLS.pku.edu.cn/en>). Initially, 19,717 participants were enrolled, but after applying exclusion criteria — (1) younger than 65 years, (2) males or females not receiving home-based care, (3) missing data on the Generalized Anxiety Disorder 7-item Scale (GAD-7), and (4) more than 20% missing or invalid individual variable data — the final sample included 2,313 eligible subjects, as illustrated in Figure 1.

2.2 Predictor Variables

Guided by the Bio-Psycho-Social model (Delshad et al., 2022) and supported by clinical and scientific findings on anxiety (Tan et al., 2023; An et al., 2024), along with previous research (Han et al., 2023; Ai et al., 2024), a comprehensive set of predictor variables related to psychological and social factors was selected. These variables are categorized into three main groups: biological function, psychological condition, and social function.

Biological Function: This category includes variables such as self-rated health status, total number of chronic diseases, level of body pain, history of inpatient care, eyesight and hearing problems, blood pressure or blood sugar issues, tooth loss, sleep duration, physical activity, history of falls and hip fractures, as well as disabilities measured by BADL (Basic Activities of Daily Living) and IADL (Instrumental Activities of Daily Living). BADL assesses fundamental self-care activities like eating, dressing, bathing, toileting, and indoor mobility. IADL evaluates more complex daily tasks such as using the phone, shopping, cooking, and managing finances. Disability in these areas is determined using the Katz weighted index, where any difficulty in performing these tasks is classified as BADL or IADL disability.

Psychological Condition: Variables in this group assess cognitive and mental health through measures including the Mini Mental State Examination (MMSE), Telephone Interview for Cognitive Status (TICS), Word Recall (WR) for memory evaluation, Retrieval Fluency (RF) to assess language and cognitive fluency, and the Community Screening Instrument for Dementia (CSI-D).

Social Function: This category encompasses demographic and lifestyle variables such as gender, age, ethnicity, religious beliefs, type of residence, marital status, education level, residential district, building type, number of steps to the household entrance, receipt of government pension, alcohol and cigarette consumption, participation in social activities, internet entertainment usage, involvement in agricultural work, fringe benefits, and pension type..

2.3 Role of Culture and Education in the Success of Premarital Screening Programs:

‘Consanguinity,’ which refers to relationships by blood or shared ancestry, increases the likelihood of inheriting recessive diseases. The risk grows with the closeness of the relationship, particularly in marriages between first cousins, which is common in certain cultures. In contrast, consanguineous marriages are rare in Western countries, where unions between first cousins are prohibited by the Orthodox Church and Roman Catholic Church and may be considered incestuous in the United States. Personal factors such as socioeconomic status can significantly affect the outcome of premarital screening programs. Educating couples and the entire screening team (including laboratory technicians, nurses, physicians, counselors, outreach workers, and social workers) is vital for the success of such programs. Schmidt emphasized that thorough educational planning before collecting the first blood sample can help avoid failures in the program. It is essential that the concept of ‘carrier status’ is explained to the public long before marriage, and that public education efforts involve collaboration between the government, religious and community leaders, health professionals, and school parent organizations.

People who are exposed to information about premarital screening generally show positive attitudes toward premarital counseling and screening for consanguineous marriages. This positive shift is linked to social changes, decreasing illiteracy, growing economic pressures, the rise of nuclear families, and delayed childbearing. Those with negative attitudes towards such screenings tend to be unmarried men. Eshra and colleagues proposed that educational campaigns about the benefits of premarital testing should focus on unmarried men, as they play a significant role in decisions regarding consanguineous marriages. In some communities, religious beliefs pose barriers to the effectiveness of screening programs. Despite the mandatory nature of premarital screening in southern Iran for over ten years, high-risk couples still choose to marry and have children affected by beta-thalassaemia. This is often due to cultural and religious traditions, as consanguineous marriages are allowed in Islam, contributing to the persistence of thalassaemia in certain areas. Many people also believe that their fate is determined by divine will, and thus accept the risks associated with having a child with genetic disorders. A recent report in *The Jordan Times* highlighted that some Jordanians view their union outcomes as destiny, with one interviewee commenting that despite having ten disabled children, they believe they will be rewarded in heaven. However, Islamic teachings also encourage healthy marriage practices and counseling to prevent genetic diseases.

AlKhaldi et al. assessed the attitudes of health science students in Saudi Arabia towards premarital screening and counseling. While most students held positive views, about 25% refused testing and counseling based on their interpretation of Islamic teachings. Awatif’s study of female students at King Saud University revealed that 86% of them had a positive opinion on premarital testing. El-Hazmi’s community-based study found that 94% of participants considered premarital counseling and testing crucial for preventing genetic diseases, with 87% believing such testing should be mandatory. Despite high levels of awareness, studies like Al Hamdan et al.’s on the first two years of Saudi screening programs revealed that approximately 90% of high-risk couples still proceeded with marriage, even after being informed of their genetic risks. This may be due to the study’s focus on university students. Similar findings have been reported in various Islamic countries, where religious beliefs often undermine the success of premarital screening programs, despite the influence of education. Reports by Karimi et al., Monaghan, and AlKhaldi et al. in different Islamic countries suggest that religious factors can act as significant obstacles to screening success, regardless of educational levels. A study by Angastiniotis and Hadjiminis in 1981 revealed that religious support played a central role in the success of screening programs in Cyprus and Greece.

2.4 Premarital Screening Programs in Asia and the Indian Subcontinent:

In Asia, voluntary premarital screening programs for inherited and sexually transmitted diseases began in countries like China, Taiwan, Malaysia, India, Indonesia, Maldives, Singapore, and Thailand. However, there is no available data regarding premarital screening programs in countries such as Pakistan, Nepal, Sri Lanka, Bhutan, and Bangladesh. In India, a few private clinics offer premarital screening packages.

In countries like Bangladesh, Pakistan, Nepal, Bhutan, Sri Lanka, and India, inherited diseases, including hemoglobinopathies, are prevalent. Although these governments are focused on preventing more common communicable and chronic diseases due to economic constraints, as their economies improve, there is growing recognition that inherited diseases should also receive attention. Hemoglobinopathies and beta-thalassaemia are more common in these regions, and it is crucial for governments to formulate strategies and action plans to reduce the incidence of these inherited diseases.

2.3 Outcome Definition

The primary outcome variable of this study is the presence of anxious symptoms. These symptoms were assessed using the Generalized Anxiety Disorder 7-item Scale (GAD-7), a brief yet reliable psychological tool specifically designed for measuring anxiety symptoms in Chinese elderly populations (Wang et al., 2022; Wang et al., 2023). The GAD-7 uses a 4-point scale ranging from 0 to 3 for each of the seven items. A score of 0 indicates that the symptom did not occur at all in the past week, while a score of 3 indicates the symptom was present nearly every day, severely impacting daily functioning and mental well-being. The overall score is the sum of all item scores, with higher totals indicating more severe anxiety symptoms. In general, a total score above 5 is considered clinically significant and suggests the need for further assessment and intervention. The scale demonstrated excellent reliability in this study, with a Cronbach's alpha of 0.919.

2.5 Data Preprocessing and Feature Engineering

To prepare the data, we first stratified the dataset based on anxious symptom outcomes and then randomly split it into a training set and a test set at a 70:30 ratio to minimize overfitting. For missing data, we applied the random forest imputation technique. This approach builds multiple decision trees using all available features to predict missing values from other observed data points. Compared to simpler methods like mean or mode imputation, random forest imputation better captures complex data patterns, thereby reducing imputation bias (Ouet et al., 2024).

Feature selection plays a vital role in improving model training by reducing dimensionality, lowering computational demands, accelerating training and prediction processes, and eliminating irrelevant or redundant features that could harm model generalization (Chen et al., 2024). To avoid data leakage, feature selection was conducted exclusively on the training set. We employed Least Absolute Shrinkage and Selection Operator with cross-validation (Lasso CV) to select key features, identifying variables with non-zero coefficients through 10-fold cross-validation (Tang et al., 2021).

Addressing class imbalance was another critical step, as imbalanced datasets can lead to poor performance on minority classes. The Synthetic Minority Over-sampling Technique (SMOTE) is a popular method to synthetically increase minority class samples, enhancing classification accuracy (Nguyen et al., 2023). Building on this, advanced techniques such as Borderline-SMOTE (AlMahadin et al., 2023) and SMOTE combined with Edited Nearest Neighbors and Tomek Links (SMOTE-ENN+Tomek) (Alsinglawi et al., 2022) have been developed. Borderline-SMOTE focuses on generating samples near the decision boundary of the minority class, avoiding excessive and unnecessary oversampling in safe regions, thereby improving boundary classification. Meanwhile, SMOTE-ENN+Tomek combines oversampling with noise removal (ENN) and elimination of redundant samples (Tomek Links) to optimize the sample distribution and further enhance classification performance on imbalanced data.

2.6 Model construction for risk prediction

To develop the optimal prediction model, we initially applied five individual machine learning algorithms along with an ensemble learning approach that combined these base models into a meta-model. Additionally, we explored the integration of these machine learning methods with three resampling techniques to assess and compare model performance. Specifically, six algorithms were employed to build the prediction models: Logistic Regression (LR), k-Nearest Neighbors (kNN), Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), and Neural Network (NN). The ensemble method, Stacking, leverages the prediction outputs of the five base models as inputs for a meta-model, enhancing overall performance through hierarchical learning.

This process involves two main stages:

Stage 1 (Base Model Training): The training dataset is split into five subsets using 10-fold cross-validation. Each base model is trained on nine folds and then generates Out-of-Fold predictions on the remaining fold to prevent data leakage.

Stage 2 (Meta-Model Training): The Out-of-Fold predictions from the base models serve as new features for training the meta-model, which produces the final prediction outcomes based on these inputs.

A brief overview of the algorithms used:

Logistic Regression (LR): A supervised algorithm that models classification probabilities using a logistic function based on linear regression.

k-Nearest Neighbors (kNN): A non-parametric instance-based classifier that predicts the class of new samples based on the majority class of their nearest k neighbors.

Random Forest (RF): An ensemble method that constructs multiple decision trees through random sampling and aggregates their votes or averages for prediction.

Support Vector Machine (SVM): An algorithm grounded in statistical learning theory that identifies the optimal hyperplane to separate classes for classification and regression tasks.

Neural Network (NN): A computational model inspired by biological neural networks, capable of modeling complex nonlinear relationships.

2.6 Model Evaluation and Interpretation

Model performance was assessed on the test dataset using standard classification metrics including Area Under the Receiver Operating Characteristic Curve (AUC), accuracy, recall, and F1-score. While the Stacking method effectively integrates the strengths and diversity of multiple models to boost classification performance, its hierarchical learning structure reduces interpretability, rendering it somewhat of a black box.

To address this, we employed the SHapley Additive exPlanations (SHAP) technique for model interpretability. SHAP quantifies the average contribution of each feature to the prediction outcome, visualized through feature importance bar charts and summary plots (Wang et al., 2021). Furthermore, to elucidate the prediction mechanism behind anxious symptoms in elderly women receiving home-based care, we developed a separate prediction model based on the best-performing algorithm. By varying the values of a target feature across a specific range while holding others constant, we generated corresponding predictions and plotted Individual Conditional Expectation Curves (ICEC) to illustrate feature effects.

2.7 Statistical Analysis

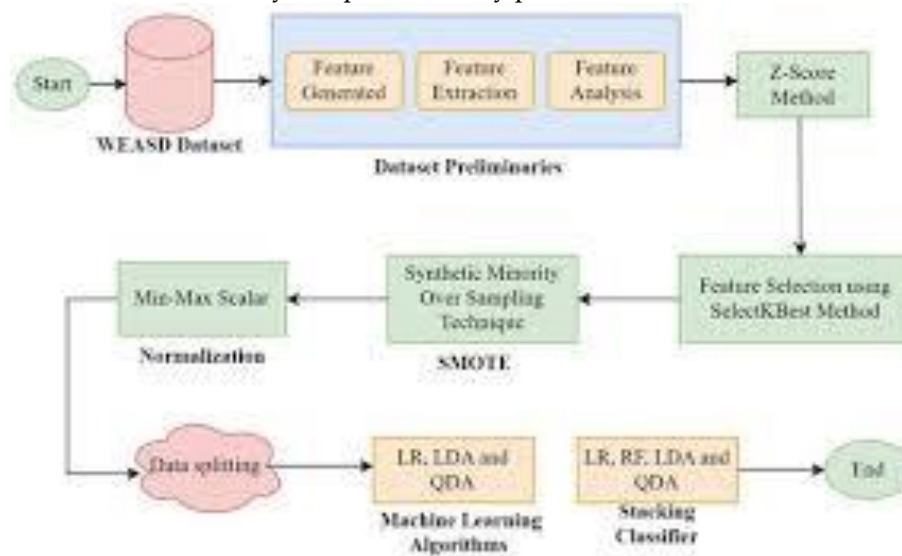
All study variables were categorized and presented as frequencies and percentages. Group comparisons for

categorical data utilized the chi-square test, while continuous data comparisons employed the t-test, chosen according to data distribution characteristics. These analyses aimed to explore the relationships between predictor variables and anxious symptoms. Statistical analyses were performed using R version 4.3.0, with a significance threshold set at $p < 0.05$. Table S2 in Appendix 1 presents the selected characteristics of the CLHLS participants.

Result

Geographical distribution of anxiety risk

Figure 2 illustrates the prevalence of anxiety risk across China, based on data from the fifth wave of the CLHLS conducted in 2008. Provinces such as Hainan, Shanxi, Chongqing, Heilongjiang, Jiangxi, and Anhui reported notably high anxiety risk rates, each exceeding 20% (28.1% [16/57], 27.3% [6/22], 22.6% [19/84], 20.6% [7/34], 20.4% [10/49], and 20.0% [5/25], respectively). Conversely, Shandong, Shanghai, Jilin, and Liaoning demonstrated relatively low prevalence, with rates under 10% (8.6% [31/359], 8.3% [2/24], 6.7% [1/15], and 5.6% [3/54], respectively). The other provinces showed moderate levels of anxiety risk, ranging between 10% and 20%. Comprehensive details on anxiety risk prevalence by province can be found in Table S3 of Appendix 1.



Baseline characteristics of anxiety risk in elderly women receiving home-based care

This study involved a total of 2,313 elderly women participants, of whom 962 lived in urban areas and 1,351 resided in rural or urban-rural fringe areas. Among these participants, 382 exhibited anxious symptoms, defined by a GAD-7 score greater than 5, resulting in an overall anxiety prevalence rate of 16.5%. Out of 40 variables examined, several were found to have a significant association with anxious symptoms. These included age level, self-rated health status, total number of chronic diseases, body pain level, eyesight and hearing problems, blood pressure or sugar issues, disabilities in basic and instrumental activities of daily living (BADL and IADL), sleep duration, history of falls or hip fractures, alcohol and cigarette consumption, cognitive function assessments (MMSE, TICS), various social and family satisfaction measures, religious beliefs, government pension status, marital status, education level, residential and building characteristics, and household economic factors.

Identification of Predictive Factors

Using the training dataset, key predictive factors were selected through Lasso regression with ten-fold cross-validation, which optimized the lambda parameter to 0.0134. This process identified 20 variables with non-zero coefficients as the most important predictors for anxious symptoms. These variables were: age level, self-rated health, chronic disease count, body pain level, eyesight and hearing problems, blood pressure or sugar issues, BADL and IADL disabilities, sleep duration, cognitive function scores (MMSE, TICS, WR, RF), satisfaction with children and marriage, religious beliefs, government pension, and marital status. To predict anxiety risk among elderly women receiving home-based care, a Stacking ensemble learning model was developed. This model combined six machine learning algorithms and was evaluated using three different

data balancing methods to address class imbalance. Parameter tuning was performed via a 10-fold grid search (details in Table S4, Appendix 1). The model's performance was assessed through ROC and calibration curves on the test dataset (Figures 3 and 4). The results revealed that the choice of data balancing method significantly impacted classifier performance. Notably, the Stacking-SET and GB-SET models demonstrated good calibration, particularly at higher predicted probability intervals (0.6–1.0), while other models like SVM-SET showed greater deviation, especially at lower prediction probabilities (0–0.4).

Among all models, the Stacking-SET method achieved the highest accuracy in predicting positive anxiety cases. The k-Nearest Neighbors model combined with the SMOTE-ENN+Tomek data balancing method (kNN-SET) performed best among single classifiers, achieving an AUC of 0.818. This outperformed the best Gradient Boosting classifiers using SMOTE (AUC = 0.762) and BorderlineSMOTE (AUC = 0.757). Gradient Boosting showed stable performance across balancing methods, with an average AUC of 0.776.

Additional metrics including accuracy, recall, F1-score, and Matthews correlation coefficient (MCC) are detailed in Table S3. The Precision-Recall and Lift curves are presented in Figure 5. Overall, all six base classifiers improved their performance when combined with the SMOTE-ENN+Tomek balancing technique. SVM exhibited the largest improvements across multiple metrics, indicating sensitivity to different data balancing strategies. Similarly, kNN showed the most significant increase in accuracy. Given the urban-rural disparities in elderly care contexts within China, rural and urban-rural fringe areas were grouped as large rural areas. Separate predictions were conducted for urban and large rural participants. The Stacking-SET model maintained strong predictive performance in both groups, with AUC values of 0.793 in urban areas and 0.786 in large rural areas.

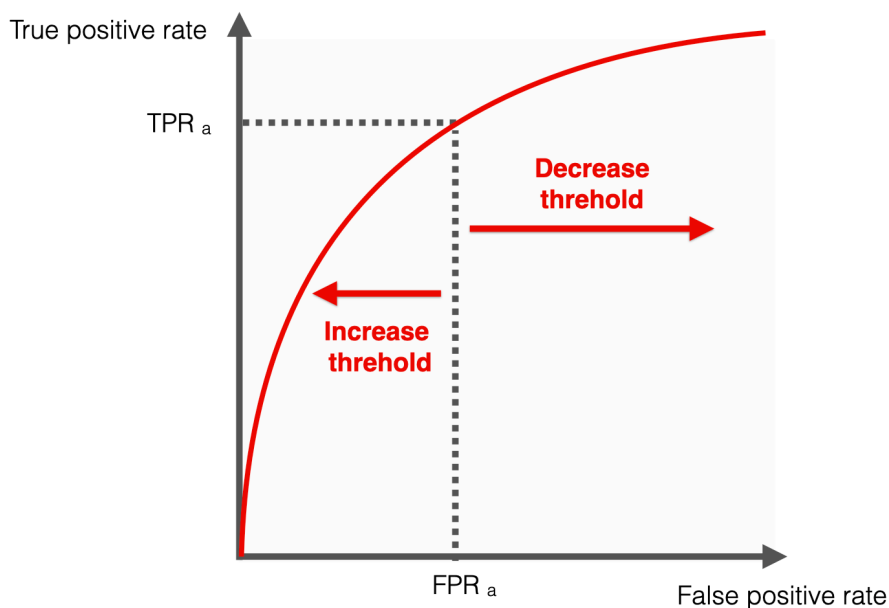


Figure 3. The ROC curve generated on the test dataset.



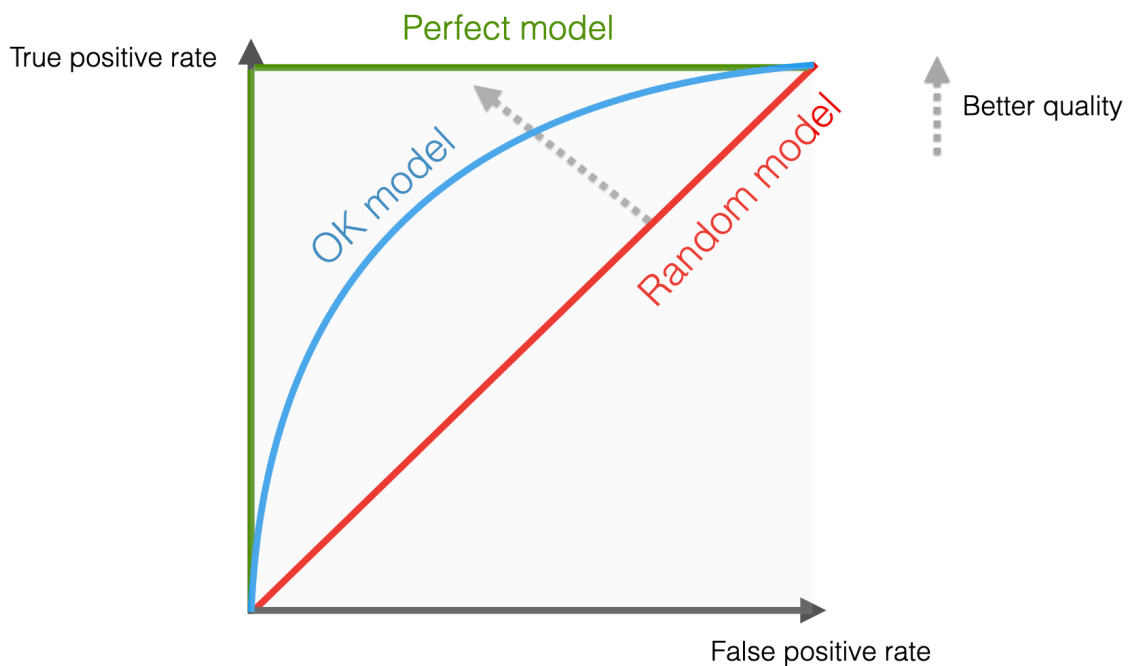


Figure 4. Calibration curve generated on the test dataset.

3.5 Model Interpretation

We analyzed and ranked the average SHAP values from the Stacking-SET model to identify the key predictive variables and understand their influence on the risk of anxious symptoms across different groups. Figure 6 displays the top 10 most important features driving the prediction for the entire target population. Among these, **body pain level** emerged as the most influential feature, followed by air quality satisfaction, sleep duration, self-rated health status, and hearing problems. Features with higher SHAP values have a stronger impact on the model's predictions.

Notably, body pain level consistently ranked highest. The SHAP summary plot shows that high values of body pain (indicated by red) correspond to a positive impact on the model's output. This means that individuals experiencing greater body pain are more likely to be predicted as at risk for anxiety. Conversely, the feature ranked lowest—**chronic disease total**—shows mostly low values (blue) clustering around zero or slightly negative impact, suggesting that a smaller number of chronic diseases weakly influences, or may even slightly decrease, the predicted probability of anxious symptoms. While lower-ranked features contribute less overall, they still help explain anxiety risk in specific cases.

To further validate how core features influence the model's predictions, Figure 7 presents Individual Conditional Expectation (ICE) plots for **body pain level**, the feature with the strongest impact. These plots illustrate how changes in body pain and self-rated health status affect the predicted probability of anxiety risk. The red curves—representing the anxious risk group (category 2.0)—demonstrate that higher pain levels and poorer self-rated health correspond to higher predicted anxiety risk. Compared to the non-anxious group (category 1.0), those at risk are more sensitive to these features.

In contrast, the ICE plots for **daily labor time** and **chronic disease total** show considerable variability among individuals, indicating complex and heterogeneous effects. For instance, increased daily labor time raises anxiety risk for some individuals (red lines trending upward), potentially due to fatigue and stress accumulation. However, for others (blue lines trending downward), more labor time may reduce anxiety risk by providing a sense of achievement or social engagement. This variation suggests that the relationship between labor time and anxiety is influenced by multiple factors such as the type of labor (voluntary vs. forced), personal coping mechanisms, and working conditions, rather than a simple direct effect.

Additionally, when predicting anxiety risk in high-risk populations, the conditional expectations of the top 10 important features are consistently higher than those of less important features. This highlights that anxious risk results from the interplay of multiple dimensions rather than any single factor alone. These findings reinforce the validity of the SHAP analysis. Furthermore, the presence of non-zero conditional expectations across many features indicates their meaningful contributions to explaining anxiety risk at an individual level.

To better demonstrate how characteristic variables of varying importance contribute to the development of anxiety risk in individuals, we calculated LIME values (Local Interpretable Model-agnostic Explanations) for the top 10 features in each sample. This analysis helps reveal how each feature influences the model's output. Figure 8 shows the LIME results for the Body Pain Level feature. The visualization displays distinct blue and red regions, which likely correspond to positive and negative impacts on the prediction, respectively. The blue region consistently appears near the top of the vertical axis, while the red region remains near the bottom, indicating a stable and regular pattern. This suggests that, when ranked by their output contribution, the influence of physical pain level on predicting depression risk exhibits similar behavior across different samples. In most cases, physical pain has a comparable direction and magnitude of effect—either positively or negatively—on the model's prediction of anxiety risk.

Discussion

This study integrates the bio-psycho-social model framework with stacking ensemble learning to develop a predictive model for anxiety risk among elderly women receiving home-based care. The model reveals the nonlinear effects of multi-dimensional factors on anxiety symptoms.

The Stacking-SET ensemble model (AUC = 0.818) significantly outperformed individual models, such as SVM-SET (AUC = 0.777), demonstrating the advantage of ensemble learning in handling complex feature interactions. Incorporating data balancing techniques like SMOTE-ENN+Tomek further improved model performance by up to 20.9% (notably with kNN-SET), highlighting the critical role of high-quality data preprocessing in addressing class imbalance common in medical datasets (Maharana et al., 2022). SHAP interpretability analysis identified physiological factors—particularly body pain severity (highest SHAP value), sleep duration, and self-rated health status—as core predictors. Notably, environmental and sensory factors such as air quality satisfaction and hearing problems emerged as significant contributors for the first time, offering fresh insights beyond traditional biomedical perspectives (Hao et al., 2024; Park et al., 2024).

The SHAP findings align with prior research, confirming that physical health strongly drives anxiety risk (Rahmati et al., 2024). Importantly, environmental adaptability factors like air quality and hearing issues ranked among the top ten predictors, suggesting that these dimensions impact mental health outcomes for elderly women receiving home care. This expands the scope of biomedical models to include environmental considerations.

Practical implications of this study include providing clinicians, nurses, and community workers with actionable tools for early risk identification and intervention. For example, by integrating multi-dimensional data from electronic health records (e.g., chronic conditions, ADL functions, social support), healthcare providers can swiftly screen high-risk individuals. Elderly women reporting body pain scores ≥ 3 combined with sleep durations ≤ 5 hours should be prioritized for psychological assessment. Community workers can tailor intervention strategies informed by SHAP results. Environmentally sensitive groups reporting low air quality satisfaction could benefit from collaborations with environmental agencies to improve living conditions (Chen et al., 2024). Additionally, for those with poor social support (low satisfaction with children), strengthening family-community linkage mechanisms is recommended (Żyrek et al., 2024). The model's applicability across urban and rural settings offers a foundation for region-specific mental health resource allocation, addressing disparities in medical service distribution (Zhuang et al., 2025).

By leveraging a stacking ensemble framework, this study overcomes limitations of traditional logistic regression models (AUC = 0.746) in capturing nonlinear relationships. Key innovations include the observed heterogeneity in the effects of chronic disease count and labor time on anxiety risk, as demonstrated by diverging ICE curves. This supports Wang et al.'s (2024) "dynamic threshold effect" theory and suggests future research should explore personalized dynamic modeling. Additionally, the confirmed link between air quality and mental health reinforces the exposome theory in environmental epidemiology (Chen et al., 2024), opening new interdisciplinary research avenues. The hierarchical integration of base and meta-models balances prediction accuracy with generalizability, akin to Qi et al.'s (2025) ensemble approach in cardiovascular risk prediction, but with enhanced interpretability through combined SHAP and ICE analyses—offering a model example of "white-box" medical AI. However, this study has limitations. It relies on cross-sectional data from the 2008 CLHLS, limiting assessment of long-term predictive validity. Anxiety symptoms were self-reported via the GAD-7 scale, which may introduce recall bias or underreporting. Feature selection via Lasso regression

may have excluded variables with weak but clinically relevant associations, such as economic pressure. Future work should enhance clinical utility by incorporating multimodal data fusion (e.g., wearable device sleep monitoring), dynamic prediction models using longitudinal data (e.g., recurrent neural networks), and cross-cultural validation comparing East Asian, European, and American populations (Chen et al., 2024). Plans also include validating temporal stability with multi-wave follow-up data, integrating objective biomarkers (e.g., inflammatory markers) to boost accuracy, and refining interpretability to elucidate feature interactions for targeted interventions. Additionally, developing a streamlined assessment tool for this population and testing its applicability in other East Asian groups will promote translation of research findings into practice. In summary, this study presents an interpretable and robust tool for early detection of anxiety risk among elderly women receiving home care, highlighting the importance of multi-dimensional interventions spanning physiological, psychological, and social domains. These findings have significant implications for optimizing community mental health services.

Conclusions

This study developed a predictive model for assessing the risk of anxiety among elderly women receiving home-based care using Stacking ensemble learning. Utilizing cohort data from the 2008 China Longitudinal Healthy Longevity Survey (CLHLS), variables were selected based on the bio-psycho-social framework. Following feature selection through LASSO regression and application of multiple data balancing techniques, the model was constructed and interpreted with SHAP analysis. The findings indicate a notable prevalence of anxiety in this population. The Stacking-SET ensemble model demonstrated superior predictive performance compared to individual models. Moreover, incorporating data balancing methods significantly enhanced model accuracy. Key predictive factors identified included the intensity of bodily pain, sleep duration, and self-assessed health status. These insights offer valuable guidance for early screening, targeted interventions, and efficient allocation of healthcare resources.

Data Availability Statement The raw data underpinning the conclusions of this study will be provided by the authors upon reasonable request.

Ethics Approval and Consent to Participate

This research involving human participants received approval from the Biomedical Ethics Review Committee of Peking University (Beijing, China). All procedures adhered to relevant local laws and institutional policies. Written informed consent was obtained from all participants prior to their involvement in the study.

Funding

The authors declare that this study did not receive any specific financial support for its research, authorship, or publication.

Acknowledgements

We extend our gratitude to the Center for China Longitudinal Healthy Longevity Survey at Peking University for providing access to the database.

Conflict of Interest

The authors report no commercial or financial conflicts of interest that could have influenced this research.

Supplementary Material

Additional materials related to this article are available online.

References

1. Ai F., Li E., Ji Q., Zhang H., 2024. Construction of a machine learning-based risk prediction model for anxiety in middle-aged and elderly hypertensive people in China: a longitudinal study. *Front Psychiatry*. 15, 1398596. doi: 10.3389/fpsy.2024.1398596.
2. AlMahadin G, Lotfi A, Carthy MM, Breedon P., 2021. Task-Oriented Intelligent Solution to Measure Parkinson's Disease Tremor Severity. *J Healthc Eng*. 9624386. doi: 10.1155/2021/9624386.
3. (This preprint research paper has not been peer reviewed. Electronic copy available at: <https://ssrn.com/abstract=5201523>)
4. Alsinglawi B, Alshari O, Alorjani M, Mubin O, Alnajjar F, Novoa M, Darwish O., 2022. An explainable machine learning framework for lung cancer hospital length of stay prediction. *Sci Rep*. 12, 607. doi: 10.1038/s41598-021-04608-7.
5. An R, Gao Y, Huang X, Yang Y, Yang C, Wan Q., 2024. Predictors of progression from subjective cognitive decline to objective cognitive impairment: A systematic review and meta-analysis of longitudinal studies. *Int J Nurs Stud*. 149, 104629. doi: 10.1016/j.ijnurstu.2023.104629.
6. Chen Q, Zhang X, Wang Z, et al., 2024. Data-driven prediction of dimensionless quantities for semi-infinite target penetration by integrating machine-learning and feature selection methods. *Defence Technology*, 40, 105-124.
7. Chen T, Dai K, Wu H., 2024. Persistent organic pollutants exposure and risk of anxiety: A systematic review and meta-analysis. *Environ Res*. 263, 120160. doi: 10.1016/j.envres.2024.120160.
8. Chen YX, Zheng J, Zhang XF., 2024. Association analysis between organophosphorus flame retardants exposure and the risk of anxiety: Data from NHANES 2017-2008. *J Affect Disord*. 355, 385-391. doi: 10.1016/j.jad.2024.04.004.
9. Cui L., Li S., Wang S., Wu X. F., Liu Y., Yu W., Wang Y., Tang Y., Xia M. S., & Li, B. M., 2024. Major anxious disorder: hypothesis, mechanism, prevention and treatment. *Signal Transduction and Targeted Therapy*, 9, 30. <https://doi.org/10.1038/s41392-024-01738-y>.
10. Delshad V, Stueck M, Ebadi A, Bidzan M, Khankeh H., 2022. Bio-psycho-social health assessment in prehospital emergency technicians: A systematic review. *J Educ Health Promot*. 11, 41. doi: 10.4103/jehp.jehp_498_21.
11. Diaz-Gallo LM, Brynedal B, Westerlind H, Sandberg R, Ramsköld D., 2021. Understanding interactions between risk factors, and assessing the utility of the additive and multiplicative models through simulations. *PLoS One*. 16, e0250282. doi: 10.1371/journal.pone.0250282.
12. Ehtemam H, Sadeghi Esfahlani S, Sanaei A, Ghaemi MM, Hajesmaeel-Gohari S, Rahimisadegh R, Bahaadinbeigy K, Ghasemian F, Shirvani H., 2024. Role of machine learning algorithms in suicide risk prediction: a systematic review-meta analysis of clinical studies. *BMC Med Inform Decis Mak*. 24, 138. doi: 10.1186/s12911-024-02524-0.
13. Guo J, Wang J, Sun W, Liu X., 2022. The advances of post-stroke anxiety: 2021 update. *J Neurol*. 269, 1236-1249. doi: 10.1007/s00415-021-10597-4.
14. Han Y, Wang S., 2023. Disability risk prediction model based on machine learning among Chinese healthy older adults: results from the China Health and Retirement Longitudinal Study. *Front Public Health*. 11, 1271595. doi: 10.3389/fpubh.2023.1271595.
15. Hao Y, Xu L, Peng M, Yang Z, Wang W, Meng F., 2024. Synergistic air pollution exposure elevates anxiety risk: A cohort study. *Environ Sci Ecotechnol*. 23, 100515. doi: 10.1016/j.ese.2024.100515.
16. Kim JY, Lim MH., 2021. Psychological factors to predict chronic diarrhea and constipation in Korean high school students. *Medicine (Baltimore)*. 100, e26442. doi: 10.1097/MD.00000000000026442.
17. Li E, Ai F, Tian Q, Yang H, Tang P, Guo B., 2025. Develop and validate machine learning models to predict the risk of anxious symptoms in older adults with cognitive impairment. *BMC Psychiatry*. 25, 219. doi: 10.1186/s12888-025-06657-y.
18. Maharana, K., Mondal, S., & Nemade, B., 2022. A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3, 91–99. <https://doi.org/10.1016/j.gltp.2022.04.020>.
19. Nguyen T, Mengersen K, Sous D, Lique B., 2023. SMOTE-CD: SMOTE for compositional data. *PLoS One*. 18, e0287705. doi: 10.1371/journal.pone.0287705.
20. (This preprint research paper has not been peer reviewed. Electronic copy available at: <https://ssrn.com/abstract=5201523>)
21. Ou H, Yao Y, He Y., 2024. Missing Data Imputation Method Combining Random Forest and Generative Adversarial Imputation Network. *Sensors (Basel)*. 24, 1112. doi: 10.3390/s24041112.
22. Park W, Jang H, Ko J, Sohn J, Noh Y, Kim SY, Koh SB, Kim C, Cho J., 2024. Physical Activity-Induced Modification of the Association of Long-Term Air Pollution Exposure with the Risk of Anxiety in Older Adults. *Yonsei Med J*. 65, 227-233. doi: 10.3349/ymj.2023.0292.
23. Qiu Y, Ma X., 2024. Using machine learning models to identify the risk of anxiety in middle-aged and older adults with frequent and infrequent nicotine use: A cross-sectional study. *J Affect Disord*. 367, 554-561. doi: 10.1016/j.jad.2024.08.185.
24. Qi X, Wang S, Fang C, Jia J, Lin L, Yuan T., 2025. Machine learning and SHAP value interpretation for predicting

comorbidity of cardiovascular disease and cancer with dietary antioxidants. *Redox Biol.* 79, 103470. doi: 10.1016/j.redox.2024.103470.

25. Rahmati M, Lee S, Yon DK, Lee SW, Udeh R, McEvoy M, Oh H, Butler L, Keyes H, Barnett Y, Koyanagi A, Shin JI, Smith L., 2024. Physical activity and prevention of mental health complications: An umbrella review. *Neurosci Biobehav Rev.* 160, 105641. doi: 10.1016/j.neubiorev.2024.105641.
26. Tan J, Ma C, Zhu C, Wang Y, Zou X, Li H, Li J, He Y, Wu C., 2022. Prediction models for anxiety risk among older adults: systematic review and critical appraisal. *Ageing Res Rev.* 101803. doi: 10.1016/j.arr.2022.101803.
27. Tan J, Ma C, Zhu C, Wang Y, Zou X, Li H, Li J, He Y, Wu C., 2023. Prediction models for anxiety risk among older adults: systematic review and critical appraisal. *Ageing Res Rev.* 101803. doi: 10.1016/j.arr.2022.101803.
28. Tang G, Qi L, Sun Z, Liu J, Lv Z, Chen L, Huang B, Zhu S, Liu Y, Li Y., 2021. Evaluation and analysis of incidence and risk factors of lower extremity venous thrombosis after urologic surgeries: A prospective two-center cohort study using LASSO-logistic regression. *Int J Surg.* 105948. doi: 10.1016/j.ijsu.2021.105948.
29. Vu T, Dawadi R, Yamamoto M, Tay JT, Watanabe N, Kuriya Y, Oya A, Tran PNH, Araki M., 2025. Prediction of anxious disorder using machine learning approaches: findings from the NHANES. *BMC Med Inform Decis Mak.* 25, 83. doi: 10.1186/s12911-025-02903-1.
30. Wang H, Chen G, Sun D, Ma Y., 2024. The threshold effect of triglyceride glucose index on diabetic kidney disease risk in patients with type 2 diabetes: unveiling a non-linear association. *Front Endocrinol (Lausanne).* 15, 1411486. doi: 10.3389/fendo.2024.1411486.
31. Wang K, Tian J, Zheng C, Yang H, Ren J, Liu Y, Han Q, Zhang Y., 2021. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. *Comput Biol Med.* 137, 104813. doi: 10.1016/j.combiomed.2021.104813.
32. Wang L, Harris R, Simoni JM, Yue Q, Fu J, Zheng H, Ning Z, Xavier Hall CD, Burns PA, Wong FY., 2023. Health Service Utilization and Its Associations with Anxiety and Sexual Risk Behaviors Among Transgender Women in Shanghai, China. *Transgend Health.* 8, 516-525. doi: 10.1089/trgh.2021.0009.
33. Wang Q, Fan K, Li P., 2022. Effect of the Use of Home and Community Care Services on the Multidimensional Health of Older Adults. *International Journal of Environmental Research and Public Health*, 19, 15402. doi: 10.3390/ijerph192215402.
34. Wang W, Liu Y, Ji D, Xie K, Yang Y, Zhu X, Feng Z, Guo H, Wang B., 2024. The association between functional disability and anxious symptoms among older adults: Findings from the China Health and Retirement Longitudinal Study (CLHLS). *J Affect Disord.* 351, 518-526.
35. (This preprint research paper has not been peer reviewed. Electronic copy available at: <https://ssrn.com/abstract=5201523>) doi: 10.1016/j.jad.2024.01.256.
36. Wang X, Hu J, Wu D., 2022. Risk factors for frailty in older adults. *Medicine (Baltimore).* 101, e30169. doi: 10.1097/MD.00000000000030169.
37. Wang Y, Zhang Y, Ni B, Jiang Y, Ouyang Y., 2023. Development and validation of a anxiety risk prediction nomogram for US Adults with hypertension, based on NHANES 2007 - 2008. *PLoS One.* 18, e0284113. doi: 10.1371/journal.pone.0284113.
38. Wei J, Hou R, Zhang X, Xu H, Xie L, Chandrasekar EK, Ying M, Goodman M., 2019. The association of late-life anxiety with all-cause and cardiovascular mortality among community-dwelling older adults: systematic review and meta-analysis. *Br J Psychiatry.* 215, 449-455. doi: 10.1192/bjp.2019.74
39. Zheng Y, Zhang C, Liu Y., 2024. Risk prediction models of anxiety in older adults with chronic diseases. *J Affect Disord.* 359, 182-188. doi: 10.1016/j.jad.2024.05.078
40. Zhuang X, Chan CP, Yang X., 2025. A network comparison analysis of socio-ecological protective and risk factors of anxiety between Chinese urban and rural adolescents. *Soc Sci Med.* 365, 117628. doi: 10.1016/j.socscimed.2024.117628.
41. Żyrek J, Klimek M, Apanasewicz A, Ciochoń A, Danel DP, Marcinkowska UM, Mijas M, Ziomkiewicz A, Galbarczyk A., 2024. Social support during pregnancy and the risk of postpartum anxiety in Polish women: A prospective study. *Sci Rep.* 14, 6906. doi: 10.1038/s41598-024-57477-1.