ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

ORIGINAL RESEARCH ARTICLE

A bagging ensemble machine learning method for imbalanced data to predict anxiety disorders and analyze risk factors in older people: An observational study

Jinling Wang[1]*, Michaela Black[1], Debbie Rankin[1], Jonathan Wallace[2], Catherine F. Hughes[3], Leane Hoey[3], Adrian Moore[4], Joshua Tobin[5], Mimi Zhang[5], James Ng[5], Geraldine Horigan[3], Paul Carlin[6], Kevin McCarroll[7], Conal Cunningham[7], Helene McNulty[3], and Anne M. Molloy[8]

[1]School of Computing, Engineering and Intelligent Systems, Ulster University, Derry-Londonderry, United Kingdom

[2]School of Computing, Ulster University, Jordanstown, United Kingdom

[3]School of Biomedical Sciences, Nutrition Innovation Centre for Food and Health, Ulster University, Coleraine, United Kingdom

[4]School of Geography and Environmental Sciences, Ulster University, Coleraine, United Kingdom

[5]School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

[6]School of Health, Wellbeing and Social Care, The Open University, Belfast, United Kingdom

[7]Mercers Institute for Research on Ageing, St James's Hospital, Dublin, Ireland

[8]School of Medicine, Trinity College Dublin, Dublin, Ireland

***Corresponding author:**
Jinling Wang
(j.wang@ulster.ac.uk)

Citation: Wang J, Black M, Rankin D, *et al*. A bagging ensemble machine learning method for imbalanced data to predict anxiety disorders and analyze risk factors in older people: An observational study. *Artif Intell Health*. 2026;3(1):116-137. doi: 10.36922/AIH025070009

Received: February 12, 2025

1st revised: June 27, 2025

2nd revised: July 7, 2025

Accepted: July 14, 2025

Published online: September 8, 2025

Abstract

Anxiety disorders (ADs) rank among the most prevalent mental health problems, especially in older people. The high risk and prevalence of ADs underscore the need for effective mental health care. Artificial intelligence has gained popularity in the diagnosis and prediction of medical conditions and diseases, including mental health problems. In this study, we developed an adapted bagging ensemble machine learning system that can be used for the diagnosis and prediction of ADs and can address the challenges posed by extremely imbalanced data from the Trinity-Ulster-Department of Agriculture study. Statistical techniques were used to identify the risk factors for ADs. Feature selection and feature engineering were conducted based on the analysis of biomarker risk factors. Five machine learning methods have been used in the developed system to build weak learner submodels, yielding promising prediction results. Some risk factors were identified. These findings will benefit the early prediction of ADs in our future studies.

Keywords: Anxiety disorder; Bagging ensemble machine learning; Risk factor analysis; Diagnosis; Imbalanced data; Aging

1. Introduction

Anxiety disorders (ADs) are one of the most common mental health issues and are characterized by anticipation of future concerns. In the course of our daily lives, instances of anxiety are not uncommon. Mild levels of anxiety can serve to alert us and sharpen our focus on potential dangers in certain situations. ADs encompass a group of conditions, such as excessive anxiety, worry, or fear, persistent worrying on most days for over 6 months without a logical cause, and difficulties in managing these feelings, which profoundly affect normal daily functioning. Symptoms of ADs include headaches, dizziness, muscle tension and aches, bowel problems, sweating, rapid heartbeat, and shortness of breath. This paper focuses on the diagnosis and prediction of ADs, which include five main types: (i) social AD, (ii) obsessive-compulsive disorder, (iii) panic disorder, (iv) generalized AD (GAD), and (v) separation AD. The causes of ADs appear to be multifactorial, including genetic traits and triggers such as traumatic events.[1,2]

Several studies on the prevalence of ADs in the elderly have reported variable results in the incidence rate for people aged 55 years and over, which ranges from 1.2% to 14.2%.[3-6] As the global population ages, the number of people over 60 years of age is expected to exceed 2 billion by 2050.[7] ADs often lead to distress and disability, reducing the overall quality of life. They can even pose a mortality risk in older adults. They are closely linked to cognitive decline, cardiovascular hazards, and other chronic illnesses. The mechanisms underlying anxiety in the elderly are primarily linked to age-related neuropathology and substantial life transitions, including economic hardship, retirement, isolation, and bereavement, that typically occur later in life. The high prevalence of anxiety-related disorders in the population poses a challenge for mental health service providers, who try to provide face-to-face therapy sessions to those who need them in a timely manner.[8] This underscores the importance of developing effective mental health-care strategies.

Machine learning (ML) and artificial intelligence (AI) technologies are becoming increasingly popular in disease diagnosis and are particularly important in the field of mental health, as there is a global shortage of qualified professionals who can handle these issues. The high cost of mental health services and the social stigma associated with these problems often deter people from seeking assistance. If left untreated, ADs can cause functional, mental, physical, cognitive, and social impairments. This, in turn, may result in decreased quality of life, delayed recovery from illness, and increased utilization of medical services.[9] As the availability of more complex health data increases, ML and AI methods are becoming increasingly valuable for analyzing risk factors to facilitate individualized treatment based on a patient's medical condition. Hence, the identification of relevant risk factors and the prediction of the prevalence of ADs among the elderly population will allow health-care providers to develop targeted strategies to reduce the incidence of ADs.

The Trinity-Ulster-Department of Agriculture (TUDA) study (ClinicalTrials.gov identifier: NCT02664584) dataset was used in our analysis. The main contributions of this study include: (i) Exploring and identifying the potential risk factors that contribute to the diagnosis of ADs; (ii) developing a bagging ensemble system for imbalanced data to help in the diagnosis and prediction of ADs; (iii) employing a threshold-moving strategy for prediction making; (iv) identifying appropriate base submodels by comparing the performance of several ML methods employed as weak learners in the system. The specific gap in identifying potential risk factors was addressed. The developed system may serve as a predictor of heightened vulnerability to ADs.

2. Related works

ML and AI play an important role in enhancing insights into health care, including mental health care, to support clinical decision-making. With the increasing availability of large amounts of complex data collected from patients in the health-care sector, and the ongoing advancements in computing power, ML can be used to identify illnesses at earlier or prodromal stages. Precision medicine, which involves personalized care and treatments tailored by health-care professionals based on an individual's unique characteristics, utilizes data to uncover knowledge and patterns, enhancing the effectiveness of early interventions.[10] Various factors, including environment, location, population, and medical knowledge, can impact the accuracy of data. Therefore, it is necessary to conduct appropriate preprocessing of the data to facilitate successful decision-making. In health care, ML can be applied in numerous ways. ML can aid health-care providers in predicting disease risks among patients, forecasting the likelihood of hospital re-admission of critically ill patients, and anticipating potential disease outbreaks.[5] Health-care professionals can use ML to help patients in their daily activities, aiming to enhance decision-making processes and minimize errors. Over time, this not only reduces costs but also improves workflow and contributes to the overall well-being of individuals. Many approaches have been developed in the medical and health field. Several review papers within this domain have explored the application of ML and AI in mental health across different domains and highlighted common gaps, trends, and challenges.[11-16]

Ancillon et al.[16] conducted a review focusing on the detection and prediction of ADs using various bio-signals and ML methods. The study provided an overview of the advantages and disadvantages of current research efforts, intending to offer guidance for future developments in the diagnosis of ADs. Notably, random forest (RF) and support vector machines (SVM) were two of the most popular ML methods, demonstrating promising performance after being combined with feature selection. Neural networks also achieved good performance and were widely used. The review emphasized the importance of features and highlighted the benefits of integrating multimodal approaches into the context of detecting and predicting ADs.

In their 2018 survey, Khan et al.[17] analyzed the mental state of social media users and made a depression prediction. They observed that certain symptoms associated with mental illness could be detectable on Twitter, Facebook, and web forums. They suggested the use of automated methods to identify signs of inactivity and other mental health conditions.

Agarwal et al.[18] created a new system designed for the early detection of mental health disorders using social media data, aiming to prevent them from escalating. The system tracked communication patterns on social networks to facilitate the timely identification of mental health issues. The analysis includes preprocessing steps such as stemming and stop word removal, feature extraction, and classification. Ensemble classifiers integrating principles from various models, including classifiers from the Bidirectional Gated Recurrent Unit, Improved Convolution Neural Network (ICNN), and Deep Maxout, were employed. A categorization was performed using the extracted characteristics, resulting in promising performance.

Nemesure et al.[19] presented an ensemble approach by combining ML and deep learning to predict psychiatric diseases. The study used electronic health records, including demographic and biometric data from 4184 undergraduate students. The model demonstrated predictive performance on a held-out test set with a sensitivity of 0.66 and a specificity of 0.7. The six most important features identified for predicting GAD were up-to-date vaccinations, control examination needed, the use of other recreational drugs, pre-hypertension or hypertension, systolic blood pressure, and marijuana use. The feature “control examination needed” refers to whether the student needed a follow-up for any reason.

Shen et al.[20] proposed a bagging algorithm, termed BPNN-Bagging, that integrates a backpropagation neural network for diagnosing GAD. Neuro-inflammatory biomarkers, specifically cytokines and S100 calcium-binding protein B (S100B), were combined in this approach. The activation of astrocytes and microglia, which are types of glial cells supporting the function of neurons and maintaining homeostasis in the central nervous system, is induced by the production of GAD-related cytokines, while neuronal growth and plasticity can be regulated by using S100B. ML techniques were employed to rank and classify cytokines and S100B, achieving a 94.47% diagnostic accuracy for GAD.

Byeon[21] proposed a stacking ensemble approach designed to identify high-risk older adults for ADs. Base models included RF, SVM, Adaboost, and Light Gradient Boosting (GB) methods, whereas XGBoost was used for the meta-model. He explored different combinations of base models and the meta-model to build stacking models. The results showed that after appropriate selection of the base model, the predictive performance of the stacking ensemble models achieved 87.4% prediction accuracy, 85.1% precision, and 87.4% recall. The highest risk predictors were identified, such as subjective family relations, subjective loneliness, the Self-Esteem Scale, family relationship and dissolution instability, instability in family support and caregiving, subjective frequency of communication with family, and the individual and their family’s experience of being a victim of a crime over the past year. This underscores a need for a tool capable of identifying older adults at high risk of developing ADs and managing them effectively.

Henry and Isa[22] proposed an implementation of ensemble methods using the open-sourced mental illness dataset to predict whether IT employees need treatment for mental health. Binary particle swarm optimization (BPSO) was used for feature selection. The performance results of Decision Tree Bagging (DT Bag), Naïve Bayes Bagging (NB Bag), and Logistic Regression Bagging (LR Bag) were presented. NB Bag obtained the highest accuracy performance at 87.86%. Naïve Bayes with BPSO feature selection had 88.44% accuracy. Based on these results, the ensemble methods, such as NB Bag, did not consistently outperform base NB with BPSO in terms of prediction accuracy.

The literature suggests that single models such as RF and SVM, combined with feature selection, can lead to effective diagnosis and prediction of ADs. In ML, ensemble techniques aim to enhance predictive results of models by combining predictions from multiple models, rather than using a single model. This approach reduces variance within a noisy dataset and addresses bias to improve the accuracy of models, while handling bias-variance trade-offs. Ensemble methods can combine models in two ways: A homogeneous or a heterogeneous ensemble model. A homogeneous ensemble model uses a single-base ML model across all submodels. A heterogeneous ensemble model uses multiple different base ML models for each submodel. The benefits of ensemble learning include increased reliability and stability in predictions. Boosting, bagging, and stacking are the three most popular ensemble models.[23]

In both stacking and bagging (known as bootstrap aggregation), multiple weak learners are trained in parallel. Bagging involves a simple voting mechanism to sum the output of each weak learner to compute the final prediction, typically with each weak learner being of the same type. In contrast, stacking uses a meta-learner trained on the predictions of previous weak learners to output the final prediction, and its weak learners are usually of different types. Stacking ensemble models tends to perform better when the individual models are stacked appropriately, and the designed stacking model, which combines different base models and the meta-model, can achieve the best predictive performance. In both bagging and stacking methods, the input data is randomly sampled with replacement from the original dataset, allowing some instances to be used repeatedly during the training stage.[24] However, boosting learns multiple weak learners sequentially, where each subsequent model assigns more weight to the data points misclassified by the previous weak learners. The weak learners can focus on specific data points and jointly reduce prediction bias.

Although state-of-the-art ML and AI techniques have been used in several studies for mental health problems, more efforts need to be made in this field. AI and ML can be promising solutions for precision medicine tailored to the needs of individual patients.

An analysis of both univariate and multivariate risk factors, conducted on the TUDA dataset, is described in the following sections. The results were then used for feature selection and feature engineering. The structure of the proposed approach is illustrated. The predictive performance, including specificity, sensitivity, accuracy, and Matthew's correlation coefficient (MCC),[25] was compared among adopted weak learners such as RF, SVM, multilayer perceptron (MLP), GB, and Linear regression (LR), and also compared to the results of the base approaches embedded oversampling technique. Efforts were made to develop an adapted bagging ensemble ML method for the prediction of ADs with the extremely imbalanced Hospital Anxiety and Depression Scale (HADs) variable of ADs diagnosis.

3. Methods

3.1. TUDA dataset

The TUDA cohort consists of detailed sociodemographic, lifestyle, biochemical, clinical, health, and nutritional data on 5186 older people aged between 60 and 102 years who were born on the island of Ireland (Figure 1 for details). Other relevant published works[7,26,27] also provide details regarding this dataset. Conducted between 2008 and 2012, the study recruited participants from general practice clinics or hospital outpatient departments in the Republic of Ireland or Northern Ireland. Standardized protocols were used for sampling, data assessment, data recording, and centralized laboratory analysis across participants. The inclusion of participants without diagnosed dementia and the collection of non-fasting blood samples allowed for the measurement of a broad spectrum of parameters, including hematological profiles, routine biochemistry, and biomarkers of micronutrient status. In addition, from a 90-min interview involving administering a comprehensive health and lifestyle questionnaire, medical and demographic details, as well as information on medication and vitamin supplement use, were collected. Blood pressure, bone health (dual-energy X-ray absorptiometry scans), physiological function tests, and cognitive function tests were also conducted.

The initial dataset contained 701 variables. During preprocessing, variables were grouped into categories including lifestyle, body measures, sociodemographics, diseases, medications, cognitive function, biochemistry, clinical, and nutritional data (Figure 1) based on domain knowledge to facilitate feature selection, feature engineering, and future analysis. Some characteristics of TUDA cohort study participants are summarized in Table 1. The preprocessing and feature selection performed on the original dataset to obtain a subset of data are described in the next subsection.

3.2. Preprocessing of the TUDA dataset

Initially, exploratory data analysis was performed, and the dataset was preprocessed following the pipeline shown in Figure 2.

First, in the initial cleansing and exploration phase, variable values were manually checked, identified, and corrected for issues such as spelling mistakes, incorrect units, coding inconsistencies, and invalid values. Duplicate and less relevant variables, as advised by domain experts, were identified and removed. Manual data processing can introduce noise, such as errors and inconsistencies. To mitigate this, a data dictionary was used to maintain consistent definitions of variables and formats. Acceptable

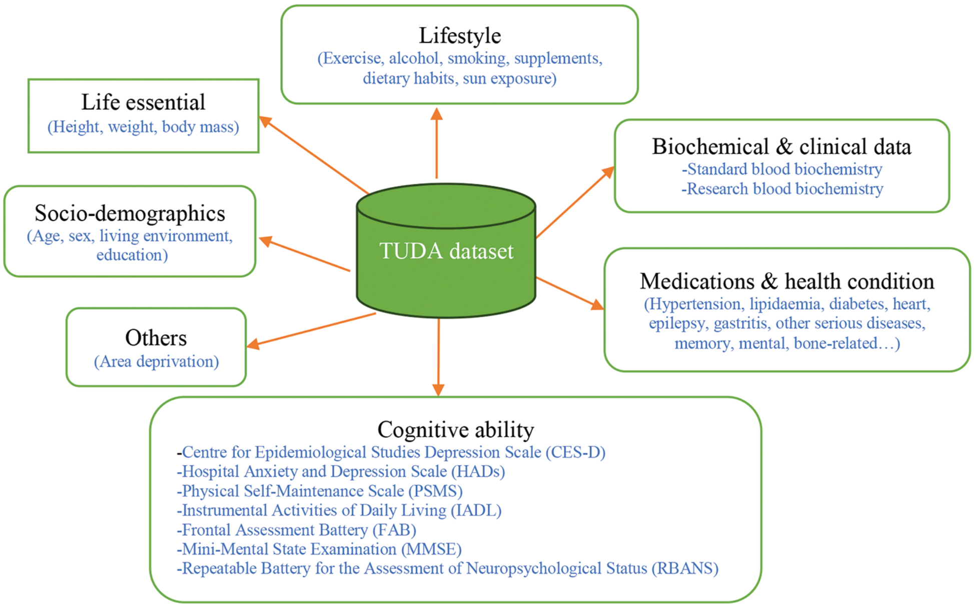

Figure 1. The features of the Trinity-Ulster-department of agriculture dataset were grouped into categories based on domain knowledge

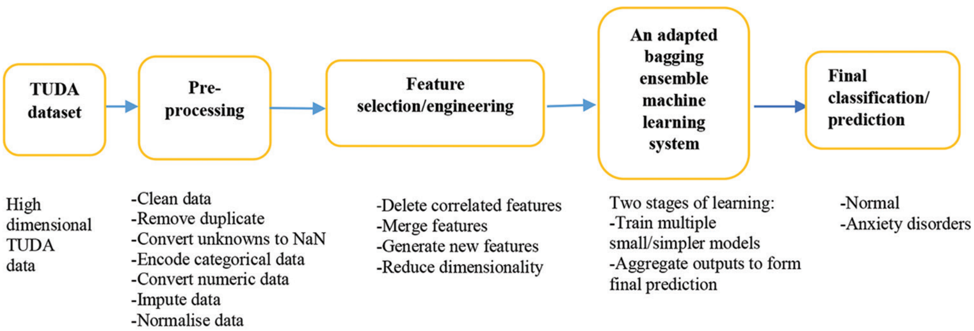

Figure 2. Pre-processing and analysis of the Trinity-Ulster-Department of Agriculture dataset with 701 features
Abbreviation: NaN: Undefined value.

values and formats for each variable were defined to prevent data entry errors. In addition, domain experts were invited to help group the data and identify and fix inconsistencies and errors in the dataset.

Second, when dealing with missing values, the aim is to retain as many valuable predictor variables as possible without introducing bias by needing to fill in more missing values. Analyses were performed with the cutoff threshold set at different percentages. By balancing, 10% was set as the cutoff threshold. Where the number of missing values for a specific variable was less than the threshold, the variable was retained; otherwise, it was deleted. The dataset was then split into the data subsets: Real numerical continuous variables and ordinal or nominal categorical variables.

Third, real numeric variables underwent transformation using two methods: The square root and the log transformation. Continuous numeric variables were divided into two groups based on their minimum and maximum values. Variables with a minimum value of zero were square-root processed, whereas those with a minimum value not equal to zero underwent a log transformation.

Next, the nominal subset was filtered for unique values to eliminate duplicates and retain maximal information from variables with high cardinality. Text information was recoded as numeric values according to the requirement defined in the dictionary, and unknown values were converted to an undefined value (NaN). Categorical variables were encoded into numeric values using the one-hot encoding method. As a result of these operations, the cleaned dataset was entirely numerically represented, with some missing values remaining, and ready for analysis to identify characteristics that could affect the fitting of an ML model in subsequent stages.

Table 1. General characteristics of the 5186 participants on the Trinity-Ulster and Department of Agriculture study

Characteristics	Males (*n*=1699)	Females (*n*=3487)
Age (years), mean (SD) (95% CI)	73.5 (8.1) (73.1–73.8)	74.3 (8.3) (74.0–74.6)
Education (years)[a], mean (SD) (95% CI)	16.0 (3.2) (15.9–16.2)	16.0 (2.9) (15.9–16.1)
Health and lifestyle		
BMI (kg/m^2), mean (SD) (95% CI)	28.5 (4.5) (28.3–28.7)	27.6 (5.8) (27.4–27.8)
Waist-to-hip ratio, mean (SD) (95% CI)	0.97 (0.07) (0.967–0.974)	0.88 (0.07) (0.877–0.882)
Timed up and go (seconds), mean (SD) (95% CI)	13.8 (9.7) (13.3–14.2)	14.2 (8.9) (13.9–14.5)
Living alone, *n* (%)	379 (22.3)	1372 (39.3)
Current smoker, *n* (%)	1185 (69.7)	1539 (44.1)
Current alcohol consumers, *n* (%)	1099 (64.7)	1876 (53.8)
Past alcohol consumers, *n* (%)	338 (19.9)	586 (16.8)
Socioeconomically most deprived, *n* (%)	438 (25.8)	886 (25.4)
Neuropsychiatric assessment		
Depression (CES-D score), mean (SD) (95% CI)	5.34 (6.8) (5.02–5.66)	6.41 (7.8) (6.16–6.67)
Anxiety (HADs score), mean (SD) (95% CI)	2.75 (3.5) (2.59–2.92)	3.39 (3.7) (3.26–3.51)
FAB score, mean (SD) (95% CI)	15.2 (2.7) (15.1–15.3)	15.2 (2.9) (15.1–15.3)
PSMS score, mean (SD) (95% CI)	23.1 (1.8) (23.1–23.2)	22.9 (2.0) (22.8–22.9)
IADL score, mean (SD) (95% CI)	24.3 (4.6) (24.1–24.5)	24.0 (4.1) (23.9–24.2)
RBANS score, mean (SD) (95% CI)	84.7 (15.9) (84.0–85.5)	85.8 (17.5) (85.2–86.4)
Clinical measures		
White cell count (109/L), mean (SD) (95% CI)	7.15 (3.2) (7.00–7.31)	7.26 (15.9) (6.74–7.79)
Hemoglobin (g/DL), mean (SD) (95% CI)	14.0 (1.6) (13.9–14.0)	12.9 (1.3) (12.8–12.9)
Mean corpuscular volume (FL)[b], mean (SD) (95% CI)	90.9 (5.6) (90.7–91.2)	91.0 (5.5) (90.8–91.2)
Platelet count (10^9/L), mean (SD) (95% CI)	233 (65.9) (230.1–236.4)	269 (71.7) (266.8–271.5)
Urea (mmol/L), mean (SD) (95% CI)	7.32 (3.1) (7.17–7.46)	6.79 (2.6) (6.70–6.87)
Creatinine (μmol/L), mean (SD) (95% CI)	98.6 (31.7) (97.0–100.1)	78.9 (24.7) (78.1–79.8)
Albumin (g/L), mean (SD) (95% CI)	41.9 (3.8) (41.7–42.1)	41.9 (3.5) (41.7–42.0)
Gamma GT (U/L), mean (SD) (95% CI)	43.5 (51.6) (41.1–46.0)	34.0 (37.4) (32.8–35.3)
Sodium (mmol/L), mean (SD) (95% CI)	139.3 (4.8) (139.1–139.5)	139.3 (20.6) (138.7–140.0)
Potassium (mmol/L), mean (SD) (95% CI)	4.29 (0.5) (4.27–4.31)	4.11 (0.4) (4.09–4.12)
Calcium (mmol/L), mean (SD) (95% CI)	2.30 (0.13) (2.29–2.30)	2.33 (0.14) (2.326–2.335)
Phosphate (mmol/L), mean (SD) (95% CI)	0.96 (0.2) (0.95–0.97)	1.04 (0.2) (1.04–1.05)
Alkaline phosphatase (U/L), mean (SD) (95% CI)	83.2 (38.3) (81.4–85.0)	82.0 (34.2) (80.8–83.1)
Low-density lipoprotein (mmol/L), mean (SD) (95% CI)	2.19 (0.8) (2.15–2.23)	2.56 (0.9) (2.53–2.60)
High-density lipoprotein (mmol/L), mean (SD) (95% CI)	1.25 (0.4) (1.24–1.27)	1.59 (0.5) (1.57–1.60)
Triglycerides (mmol/L), mean (SD) (95% CI)	1.70 (0.9) (1.66–1.75)	1.53 (0.8) (1.50–1.56)
C-reactive protein (mg/L), mean (SD) (95% CI)	7.38 (19.2) (6.47–8.29)	6.71 (16.1) (6.17–7.24)
Glycated hemoglobin (%), mean (SD) (95% CI)	5.97 (0.9) (5.92–6.01)	5.83 (0.7) (5.81–5.86)
Parathyroid hormone (pg/mL), mean (SD) (95% CI)	45.9 (29.6) (44.5–47.3)	47.4 (33.7) (46.3–48.5)
Glomerular filtration rate (mL/min), mean (SD) (95% CI)	74.6 (25.9) (73.4–75.8)	63.7 (23.0) (62.9–64.5)
Nutritional biomarkers		
Red blood cell folate (nmol/L), mean (SD) (95% CI)	1055 (622) (1025–1085)	1121 (623) (1101–1142)
Serum Vitamin B12 (pmol/L), mean (SD) (95% CI)	268 (173) (260–276)	300 (221) (293–307)

(Cont'd...)

Table 1. ***(Continued)***

Characteristics	Males (*n*=1699)	Females (*n*=3487)
Plasma Vitamin B6 (nmol/L), mean (SD) (95% CI)	70.7 (50.5) (68.3–73.1)	80.5 (72.9) (78.1–82.9)
Riboflavin (EGRac)[c], mean (SD) (95% CI)	1.35 (0.2) (1.30–1.40)	1.34 (0.2) (1.30–1.40)
Total plasma homocysteine (μmol/L), mean (SD) (95% CI)	15.5 (6.0) (15.2–15.8)	14.5 (5.7) (14.3–14.7)
Total Vitamin D (nmol/L), mean (SD) (95% CI)	53.0 (27.9) (51.6–54.3)	62.1 (32.4) (61.0–63.2)

Notes: [a]Education refers to the age of stopping formal education. [b]FL: Femtolitre. [c]EGRac: Erythrocyte glutathione reductase activation coefficient, with a higher EGRac value indicating poorer riboflavin status.
Abbreviations: BMI: Body mass index; CES-D: Center for epidemiological studies depression scale; FAB: Frontal assessment battery; HADs: Hospital anxiety and depression scale; IADL: Instrumental activities of daily living; PSMS: Physical self-maintenance scale; RBANS: Repeatable battery for the assessment of neuropsychological status.

Finally, the two data subsets were concatenated, the K-Nearest Neighbors algorithm was used to fill missing values, and the dataset was normalized to the range between 0 and 1 using z-score normalization.

The primary objectives of this study were the diagnosis and prediction of ADs. In an existing work,[28] a self-reported anxiety variable was utilized as the outcome variable where participants reported that they had, at some point in their lifetime, either been diagnosed with anxiety by a doctor or not. The variable encompassed 4064 participants (78.36%) who reported that they did not have an anxiety diagnosis, and 1122 participants (21.64%) who reported that they did have an anxiety diagnosis. This resulted in an approximate 78:22 split in the class of this self-reported outcome variable. This formed the basis for using binary classification models to fit the data. A notable challenge was the inherent imbalance in the outcome variable, as only 22% of the patients had reported an anxiety diagnosis. To address this, synthetic records of the minority class in the training dataset were generated using the standard synthetic minority over-sampling technique (SMOTE) method[29,30] and its extension, the adaptive synthetic sampling (ADASYN) algorithm.[31] The test set was unchanged to preserve the representativeness of the original population. This approach ensures fair comparisons with other methods, as well as reliable predictions on the unseen test set.

In this study, we used the score from the HADs assessment as the outcome variable. HADs are a standard tool designed to measure levels of anxiety and depression in individuals. Scores range from 0 to 21, where a score of 11 or greater indicates ADs, whereas a score <11 is considered normal. Out of 5186 participants, 4918 (94.83%) were diagnosed as not having anxiety, 260 participants (5.01%) had an anxiety diagnosis, and 8 (0.16%) had missing values. The extreme imbalance in the class, with an approximate 95:5 split, hinders the direct construction of models using conventional ML methods. Using oversampling techniques could introduce bias due to the large proportion of records that would need to be oversampled. To address this challenge, we developed a novel diagnosis and prediction system specifically designed to handle imbalanced data.

3.3. Feature selection and engineering

Correlation analysis is required when dealing with correlated and multicollinear predictor variables to avoid potentially unstable estimates and redundant predictor variables that do not contribute additional information for developing models.[7] To select predictor variables, we explored correlations between the diagnosis of ADs and predictor variables. In the beginning, risk factor analysis was conducted using an unimputed and unnormalized dataset to investigate associations between an anxiety diagnosis and nominal predictor variables, such as medications, lifestyle factors, and diseases. Depending on the results, some predictor variables were merged, and new predictor variables were created. For example, a new predictor variable (LimitAct) was created by merging "Limithouseholdactivities" and "Limitoutsideactivities" that have a very strong association (Cramer C = 0.6088). The predictor variable "Limithouseholdactivities" has two values: 1 (Yes) and 0 (No), by answering the question "Does the participant limit any household activities because they are afraid they might fall?" The predictor variable "Limitoutsideactivities" has two values: 1 (Yes) and 0 (No), by answering the question "Does the participant limit outside activities because they are frightened, they might fall?" The created predictor variable took the value logical 1 if at least one of the values of these two predictor variables was 1; logical 0 was taken if the values for both are 0. Highly correlated predictor variables that did not add meaningful information for prediction were removed.

The Spearman non-parametric correlation coefficient was used to calculate correlations between numerical predictor variables. Table 2 shows the relationships among the cognitive function numerical predictor variables. Nominal and ordinal variables are common types of categorical variables. Cramer's V, a statistic in the range of

Table 2. Associations between cognitive function numerical variables

CED-S	1.00					
HADs	0.43	1.00				
IADL	−0.20	−0.03	1.00			
FAB	−0.16	−0.05	0.44	1.00		
PSMS	−0.17	−0.07	0.59	0.30	1.00	
RBANS	−0.21	−0.07	0.47	0.66	0.26	1.00
	CED-S	HADs	IADL	FAB	PSMS	RBANS

Notes: The colors indicate the associations between cognitive function numerical predictive variables. Different colors indicate different levels of intensity.
Abbreviations: CES-D: Center for epidemiological studies depression scale; FAB: Frontal assessment battery; HADs: Hospital anxiety and depression scale; IADL: Instrumental activities of daily living; PSMS: Physical self-maintenance scale; RBANS: Repeatable battery for the assessment of neuropsychological status.

[0, 1], was used to explore the association between these nominal variables. A value of 0 indicates no association between the two variables, whereas a value of 1 indicates a strong association. Cramer's V was calculated as shown in Equation I. Cramer's V was used to identify considerable variation and strong associations between the nominal variables. The results were then used to identify redundant variables, which were subsequently removed from the dataset. For example, the predictor variable "lipid_meds," representing lipidemia medication intake, was removed as it has a strong association with the predictor variable "Hyperlipidemiadiagnosis" (Cramer's V = 0.7168), which represents the diagnosis of hyperlipidemia.

$$\text{Cramer's V} = \sqrt{\frac{\chi^2}{n^*\min(c-1, r-1)}} \quad (1)$$

where χ^2 is the Chi-square statistic, n represents the total sample size, r represents the number of rows, and c represents the number of columns. Chi-square tests were used to compare groups between participants with and without a diagnosis of AD.

A non-parametric Kruskal–Wallis test was used to determine whether the data from the two groups differed from each other. The calculated *p*-value was compared to the significance level, usually set at 0.05. If the $p>0.05$, the null hypothesis can be retained, signifying that the Kruskal–Wallis test does not detect a significant difference between categories of independent variables with respect to the dependent variable. Otherwise, the null hypothesis can be rejected.

Upon completion of the preprocessing steps, a total of 5186 records with 84 variables (83 predictors and 1 outcome) were retained for the following analysis. The outcome denotes the anxiety diagnosis of each participant, determined using HADs. In the following experiments, an analysis of risk factors was conducted.

3.4. Risk factor analysis

While the exact reasons for mental health problems remain unclear, attempts have been made to identify potential risk factors.[32,33] Our previous research using self-reported ADs diagnoses[28] identified several potential risk factors for ADs in older adults. These include female sex, loneliness, separated/divorced status, low socioeconomic status, lifestyle-related factors such as smoking and alcohol intake, as well as medical conditions such as cardiovascular disease, lipidemia, diabetes, hypertension, some chronic inflammatory diseases, and family history of diseases such as stroke and presenile dementia. In this study, we used the HADs ADs diagnosis to analyse risk factors. Identifying and intervening in modifiable risk factors can contribute to the mitigation or prevention of ADs.[9]

Table 3 shows the results of the univariate analysis for participant characteristics related to ADs diagnosis based on the total HADs score. Since only the anxiety questions are included in the HADs score in the TUDA study, we used HADs to assess anxiety.

Notably, "Gender" emerges as a key feature, with 5.9% of females experiencing ADs, compared to 3.7% of males. These differences may be attributed to hormonal fluctuations and variations in brain chemistry throughout a woman's life, potentially linked to ADs. Marital status also appears influential, as participants in the separated/divorced and widow/widower categories showed a slight susceptibility to ADs. Interestingly, living with others seems to mitigate the prevalence of ADs in older people, potentially due to the emotional support from others in shared living arrangements. Participants residing in low socioeconomic areas demonstrated a higher likelihood of experiencing ADs, aligning with findings from previous studies.[26] Lifestyle factors such as smoking were identified as risk factors for ADs, consistent with existing research.[34] Our analysis revealed that approximately 5.0% of participants had ADs, and the presence of a family history of stroke, heart disease, or presenile dementia increased the risk of ADs (Table 3).

Figures 3 and 4 illustrate the percentages of frequency distribution of predictors related to the effect of food supplements, Vitamin B, and Vitamin D in terms of ADs diagnosis. Vitamin D helps our body better absorb calcium, a key building block for our bones and an essential nutrient for overall health. It has been reported that nearly one billion people worldwide have low levels

Table 3. Univariate analysis of general characteristics of the Trinity-Ulster Department of Agriculture study participants

Variables	HADs (anxiety diagnosis)	Effect size estimates	Effect size estimates
	Yes (*n*=260) n (%)	Cramer's V	Odds ratio
Gender		0.044	1.591
Male	62 (3.7)		
Female	198 (5.9)		
Marital status		0.017	
Single	25 (3.8)		1.377
Married/common law	134 (5.0)		1.031
Separated/divorced	13 (5.2)		0.958
Widow/widower	88 (5.6)		0.843
Area deprivation		0.041	1.737
Normal	159 (4.3)		
SESlow	95 (7.2)		
Accommodation status		0.022	
Alone	82 (4.7)		1.114
Spouse/partner	135 (5.0)		0.995
Other	8 (3.4)		0.691
Children	35 (6.8)		1.511
Smoking		0.014	1.204
No	112 (4.6)		
Yes	148 (5.4)		
Drinking alcohol		0.013	
Never	65 (5.1)		0.986
Past	54 (5.8)		0.820
Current	141 (4.7)		1.145
Vitamin D		0.024	1.356
>50 nmol/L	125 (4.4)		
≤50 nmol/L	135 (5.9)		
Fortified food and supplements		0.0057	0.948
No/low	106 (5.2)		
Medium/high	154 (4.9)		
Diagnosis of			
Hypertension			
Yes	184 (5.0)	0.016	0.930
No	76 (5.3)		
Hyperlipidemia			
Yes	150 (5.5)	0.016	1.200
No	106 (4.6)		
Diabetes			
Yes	30 (4.5)	0.011	0.884
No	230 (5.1)		

(Cont'd...)

Table 3. *(Continued)*

Variables	HADs (anxiety diagnosis)	Effect size estimates	Effect size estimates
	Yes (*n*=260) n (%)	Cramer's V	Odds ratio
Other serious diseases			
Yes	62 (4.3)	0.019	0.816
No	198 (5.3)		
Self-memory concern			
Yes	113 (7.5)	0.048	1.875
No	138 (4.1)		
Family-memory concern			
Yes	67 (10.2)	0.061	2.461
No	184 (4.4)		
Family history			
Cancer			
Yes	70 (4.9)	0.0028	0.960
No	190 (5.1)		
Stroke			
Yes	27 (5.8)	0.0073	1.168
No	233 (5.0)		
Heart disease			
Yes	79 (5.2)	0.0033	1.048
No	181 (5.0)		
Presenile dementia			
Yes	7 (9.6)	0.017	2.014
No	253 (5.0)		
Senile dementia			
Yes	34 (4.9)	0.0023	0.958
No	225 (5.1)		

Abbreviations: HADs: Hospital anxiety and depression scale; SESlow: Low standard socioeconomic status.

of Vitamin D, and approximately 20% of the population in the UK suffers from a Vitamin D deficiency.[35] Numerous studies have highlighted the link between Vitamin D deficiency and increased anxiety.[36,37] Figure 3A shows the percentages of participants who have an anxiety diagnosis and their Vitamin D levels. 5.85% of participants with a Vitamin D deficiency experienced AD, compared to 4.38% with normal vitamin D levels. Participants with low Vitamin D levels appeared to have a higher chance of experiencing ADs. Research suggests that Vitamin D, influencing brain receptors and mental health, may play a role in cell growth promotion. In theory, a Vitamin D deficiency could potentially limit this behaviour, impeding overall brain function.[38,39] Vitamin D supplements could serve as preventive measures or potential treatments for

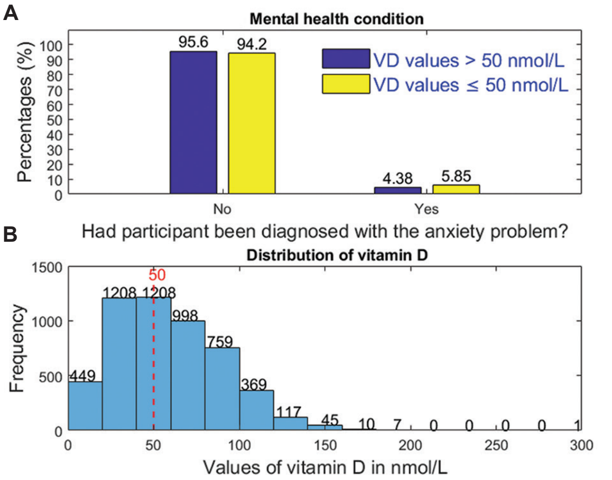

Figure 3. Bar plot showing percentages of participants with anxiety diagnosis and their Vitamin D levels (A) and distribution of Vitamin D values (B)

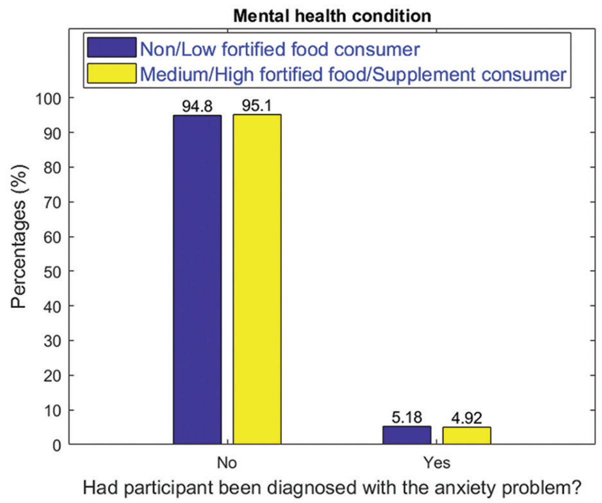

Figure 4. Bar plot showing percentages for non/low fortified food consumers and medium/high fortified food/supplement consumers classified based on their ADs diagnosis

anxiety and depression due to their strong associations. Another study found that individuals with anxiety tend to have lower levels of calcidiol, the most active form of Vitamin D. Calcidiol helps the body use more calcium from foods or supplements and regulates the production of parathyroid hormones in the body.[40]

Figure 4 shows the percentage distribution for predicting ADs diagnosis among consumers of non- or low-fortified foods and consumers of medium- or high-fortified foods or supplements. It should be noted that this feature was engineered by combining variables related to fortified food and supplements in the original dataset. Consumers of non- or low-fortified food included participants who did not consume fortified food and supplements or consumed 1–4 servings of fortified foods per week. Consumers of medium- or high-fortified foods or supplements included participants who consumed 5–7 servings of fortified foods per week (medium consumers), or more than 8 servings of fortified foods per week (high consumers), or users of supplements. Participants with non/low fortified food intake appeared to have a higher chance of experiencing ADs compared to those with medium/high fortified food/supplements intake. This suggests that fortified foods may offer mental health benefits.

The participant was asked if their family had any concerns with regard to their memory (family-memory concern), and if they themselves had any concerns about their memory (self-memory concern). Figure 5 shows percentages of anxiety diagnosis for participants with (a) self-memory concern and (b) family-memory concern. From the results, we can see that participants who had a concern about their memory, and/or whose family had

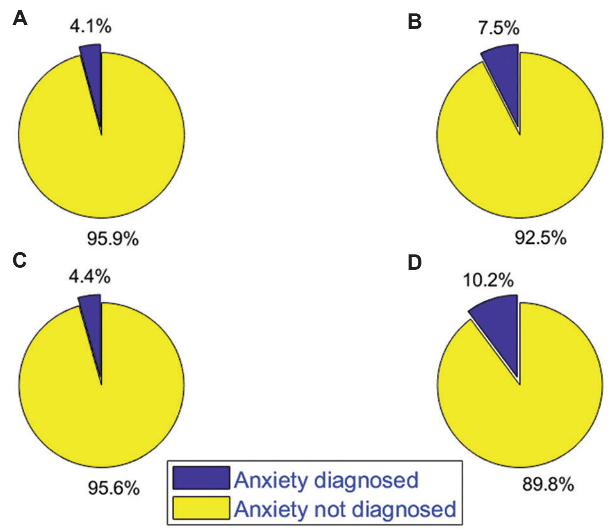

Figure 5. Pie chart illustrating percentages of anxiety diagnosis for participants with (A) non-self-memory concern, (B) self-memory concern, (C) non-family-memory concern and (D) family-memory concern

concerns about their memory, were more prone to ADs than those with no memory concerns reported.

Figures 6 and 7 present the results of multivariate analysis for lifestyle characteristics in terms of smoking status and alcohol intake of participants, and in relation to ADs diagnosis. Figure 6 shows that, regardless of gender, smoking poses a higher risk of ADs in older people. The diagnosis of ADs increases from 8.92% for non-smoking females to 11.2% for females who smoke. The difference is less pronounced for males at 2.25% for non-smokers

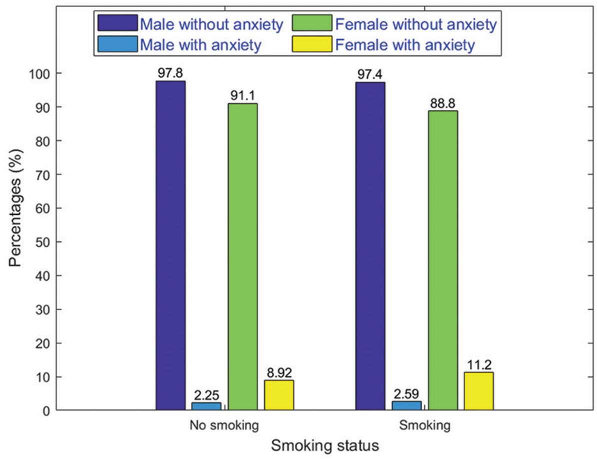

Figure 6. Bar plot showing percentages of participants who smoke versus those who do not, for males and females, respectively, in terms of their anxiety disorder diagnosis

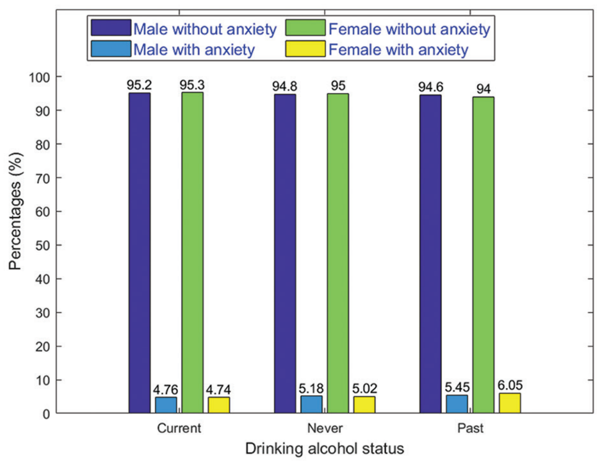

Figure 7. Bar plot showing percentages of participants based on their alcohol consumption status, for males and females, respectively, and in terms of anxiety disorder diagnosis. "Current:" They currently consume alcohol; "Past:" They consumed alcohol in the past but not currently; "Never:" They never consumed alcohol.

and 2.59% for smokers. Smoking appears to have a more pronounced impact on women.

Figure 7 shows how participants' alcohol intake status affects the diagnosis of ADs for males and females, respectively. The participants with a history of past alcohol consumption are more prone to ADs, especially among females.

The analysis highlights several factors associated with a higher risk of ADs, including being female, experiencing separation or divorce, being widowed, having self- or family-reported memory concerns, having a family history of chronic diseases (such as stroke or presenile dementia), smoking, and having a history of alcohol consumption. Furthermore, participants from areas with higher socioeconomic deprivation were found to have a higher prevalence of ADs, with a prevalence of 7.2% compared to 4.3% in areas with normal socioeconomic status.

Feature selection and engineering, which involves identifying and selecting the most relevant features in a high-dimensional dataset, is crucial due to the challenges posed by high dimensionality, noise reduction, and model interpretability requirements. This study used a variety of methods to reduce features: (i) Domain experts were involved in grouping the features and selecting the most relevant features; (ii) The percentage of missing values for a specific feature was compared to a threshold to determine whether to keep or remove the feature; (iii) Filter-based feature selection techniques such as statistical correlation analysis were used in the TUDA dataset to eliminate irrelevant or redundant features, and generate new features from existing similar features, for example, by removing a predictor that had a strong association with another predictor, retaining a predictor that had high correlation to the outcome feature, and by merging variables "Bone fracture" and "Hip fracture" into one predictor. These methods make feature selection and feature engineering an integral part of building effective models in this field.

4. Results

4.1. Preparation of training and test sets

The preprocessed dataset was randomly divided into two sets: A training set containing 70% of the data (3631 records) and a test set consisting of 30% of the data (1555 records). Figure 8 illustrates the process of preparing the training and test sets. To ensure consistency in model comparisons, each model followed the same procedure and was trained and tested on the same datasets. The response variable, anxiety diagnosis, relates to the diagnosis of ADs in participants using the HADs method. Among the 5186 participants, 4918 participants (94.83%) were not diagnosed with an AD, whereas 260 participants (5.01%) were diagnosed with an AD, resulting in an imbalanced class distribution of approximately 95:5. Due to this extreme imbalance, standard SMOTE oversampling could introduce bias, as a large proportion of records would require oversampling. To address this issue, we developed an adapted bagging ensemble diagnosis and prediction system, which involves reconstructing the training data.

The m value determines how many submodels are ensembled. We let *Nrn* represent all the records with "No anxiety" diagnosis, and *Nra* represent all the records with "Anxiety" diagnosis in the training set. *m* is calculated based on Equation II:

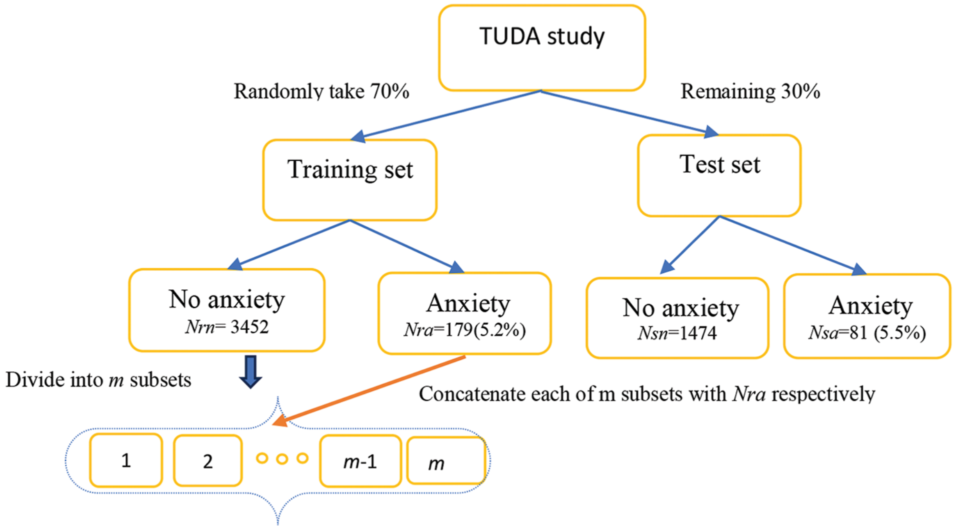

Figure 8. Preparation of the training and test sets, and construction of subsets in the training set

$$m = \mathrm{round}\left(\frac{\mathrm{Nrn}}{\mathrm{Nra}}\right) \quad \text{(II)}$$

Then, the records with "No anxiety" diagnosis in the training set were split into m chunks with no repetition. In this study, *m* equals 20.

4.2. Structure of the system

Various techniques can be employed to address the challenge of imbalanced datasets. Two common methods are under-sampling and over-sampling, which aim to create a balanced dataset from an imbalanced one. Under-sampling is used when the volume of data is sufficient; it involves reducing the size of the over-represented class to balance the dataset. Conversely, over-sampling is used when the volume of data is insufficient, and it involves increasing the size of the under-represented class to balance the dataset, without removing abundant samples. Over-sampling techniques may generate new synthetic samples for the rare class. In previous work,[28] approximately 22% of participants self-reported as having ADs, and the class imbalance was addressed using two standard oversampling techniques: SMOTE[29,30] and ADASYN.[31] These methods synthetically generated additional records of the minority categories in the training dataset. Importantly, the test set was unchanged to preserve representativeness of the original population, ensure a fair comparison with other methods, and provide reliable predictions on the unseen test set. In this paper, we developed a procedure to overcome the challenge imposed by imbalanced data.

The bagging method involves generating *m* bootstrapped samples to construct models in parallel. Then, the individual predictions from the *m* models are aggregated through voting or averaging to obtain the final prediction from the ensemble of models. The main purpose of the bagging method is to minimise diversity and mitigate the risk of overfitting across the various models created. In this process, each bootstrapped sample is randomly generated from the given data records with replacement, allowing some individual records to be chosen more than once. In this paper, we use every record in the training set without repetition. We adapted the traditional bagging method and allocated the records with "No anxiety" to *m* groups; the number of records in each group roughly equals *Nra*. As 20 groups were allocated, the number of records in the first 19 groups is 173, and the number of records in the last group is 165. The flowchart illustrating this adaptation is presented in Figure 9.

First, we trained multiple small/simpler models instead of training one complex/large model for our training dataset. Normal records in the training set were divided into *m* small trunks, with the size of each trunk ensured to be the same or a size similar to that of the "Anxiety" records in the training set. Each divided subset, containing records with only "No anxiety" diagnosis, was then concatenated with records labeled with "Anxiety" diagnosis (*Nra*) in the training set. During the training phase, the generated *m* subsets were used to build *m* submodels using the same weaker learner, respectively. Second, after *m* submodels were trained, the prediction (1 = "Anxiety" diagnosis; 0 = "No anxiety" diagnosis) for each participant in the test set was obtained. During the testing phase, the held-out test set, which contains 1555 participants, was sent through to each of the generated *m* submodels, and a prediction was generated by each submodel for a participant. These predictions of *m* submodels for each participant in the test set were summed (*Sum*), which is in the range of [0, *m*]. Then, a threshold-moving strategy by the trial-and-error method could be employed for prediction making. After a threshold (*Th*) was set, the final prediction could be obtained. If the *Sum* exceeds *Th*, the participant can

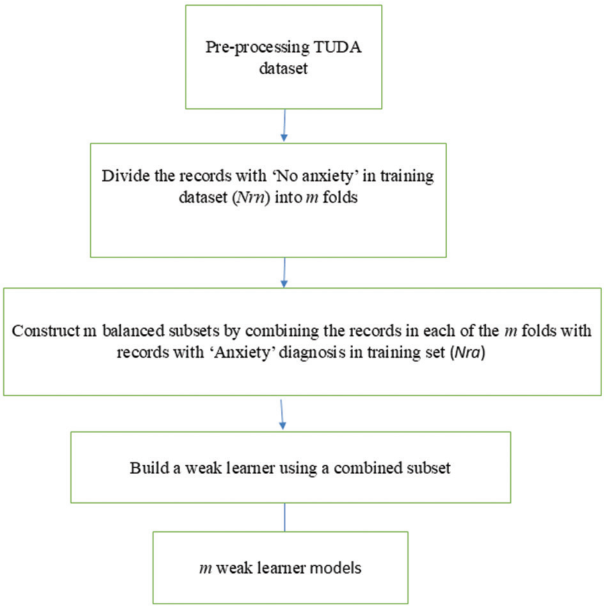

Figure 9. Flowchart for reconstructing *m* subsets and building *m* weak learners in parallel

be said to have an "Anxiety" diagnosis; otherwise, the participant is classified as having "No anxiety" diagnosis. The algorithm logic is based on concatenating predictions of each submodel and decision rule, ensuring the opinions of at least 80% submodels are respected. For example, if *Th* is taken as 16, the participant is diagnosed as having AD if the predictions of at least 80% of the submodels were 1. This is represented in Figure 10, showing the flowchart of the proposed bagging ensemble system for the analysis and prediction of ADs in the TUDA dataset using one specific weak learner across all submodels to predict the output. Figure 11 shows the prediction of the test set.

4.3. Evaluation

Evaluation metrics such as specificity (true negative rate, TNR), sensitivity (true positive rate, TPR), accuracy, F1 score, the area under the receiver operating characteristic curve (AUC-ROC), and MCC can be used to evaluate the performance of an approach. The AUC-ROC represents the relationship between TPR and false positive rate. F1 score, AUC, and accuracy are three of the most widely adopted metrics in binary classification tasks. However, for imbalanced datasets, these statistical measures can produce misleadingly over-optimistic information. In such cases, MCC can more accurately reflect general diagnosis and prediction problems.[25]

MCC is a correlation coefficient between predicted and observed binary classifications. It is a more reliable

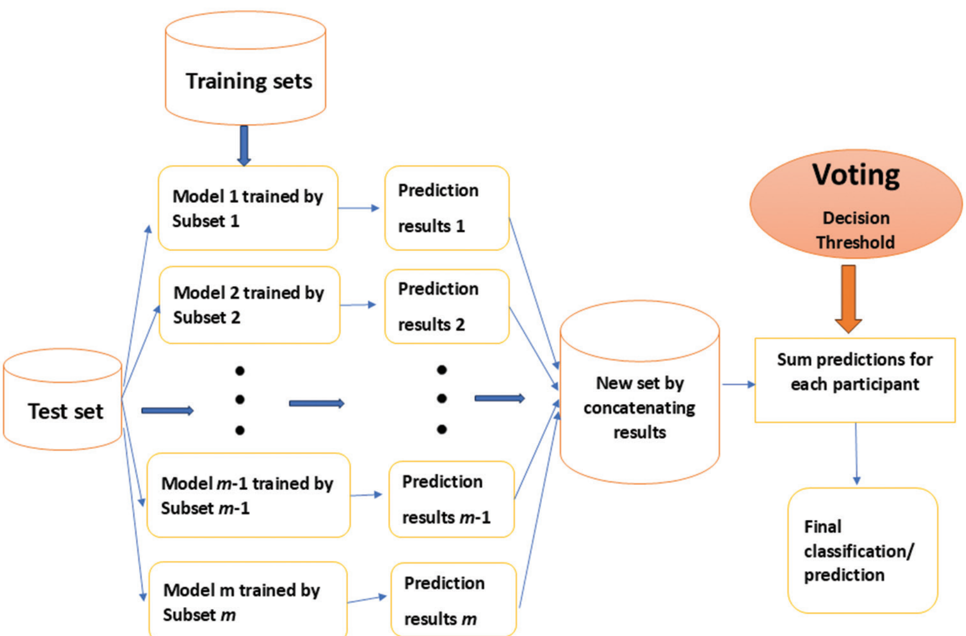

Figure 10. A pipeline for analysis and prediction of a high-dimensional Trinity-Ulster-Department of Agriculture dataset using an ensemble machine learning method (*m* = 20), *m* submodels with changing data records for the normal class in the training set

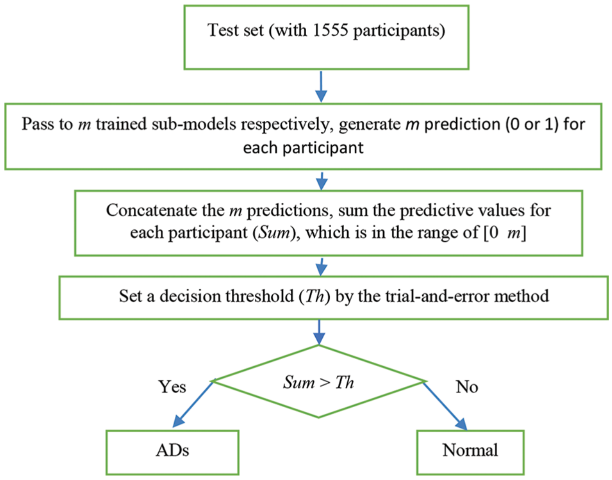

Figure 11. Flowchart of prediction on the test set
Abbreviation: ADs: Anxiety disorders.

statistical measure. MCC only provides a high value if good results are achieved in predictions across all four confusion matrix categories, *i.e.*, true positives (TP), true negatives (TN), false negatives (FN), and false positives (FP). It is proportional to the size of positive and negative elements in the dataset. A perfect classification is indicated by an MCC value of 1, whereas values close to 0 represent predictions made at random, and −1 represents an opposite prediction where all positive samples were predicted as negative and *vice versa*. The MCC is often normalized to the range of [0, 1], referred to as the normalized coefficient (normMCC), to match the value range and meaning of the other statistical rates. In evaluating binary classifications, MCC can produce a more informative and truthful score compared to accuracy and F1 score. In previous work,[41] the mathematical properties and use cases of MCC were explained and presented, establishing its preference over accuracy and F1 score as the standard metric for evaluating binary classification tasks. There has even been a suggestion that MCC should replace the AUC-ROC as the standard metric for assessing binary classification.[42] In this paper, we used specificity, sensitivity, accuracy, and normMCC as evaluation metrics. The formulas of MCC and normMCC are listed in Equations III and IV.

$$C = \frac{\mathrm{TP{*}TN - FP{*}FN}}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}} \quad \text{(III)}$$

$$\mathrm{normMCC} = \frac{\mathrm{MCC}+1}{2} \quad \text{(IV)}$$

In Equation IV, 0 represents the worst and minimum value, whereas 1 represents the best and maximum value.

4.4. Classical ML methods

Common ML techniques such as SVM, RF, GB, MLP, and LR were used in this work as weak learners to build submodels. Embedding these ML techniques into the pipeline of the proposed system can help with decision-making. Doctors can use it to recognize patterns of ADs, and distinguish an AD patient from a healthy patient. The accuracy of the results can be optimized, and adequate treatment can be provided. We briefly explore these ML methods below.

RF takes the mean of the results of a number of distinct decision trees, which work collectively as a group. A majority voting method is used to make the model's final prediction. Voting or averaging mechanisms may deal with the problem of overfitting. It performs well with missing values.

SVM is a supervised ML algorithm, which is good at finding the optimal decision boundary to best separate the hyperplane via linear separation. SVM transforms the input space into a high-dimensional feature space so that the non-linear problem can be solved. SVM is more accurate than other ML methods and is less likely to suffer from overfitting issues, suitable for modeling complex non-linear decision domains. However, SVMs are not suitable for larger datasets and are more sensitive to missing data.

LR predicts a dependent data variable by analyzing the relationship between one or more existing independent variables and helps model a binary dependent variable. LR is easier to interpret, implement, and very quick to train, with no assumptions needed to be made about distributions of classes in the entire feature space. Since linear boundaries are constructed, it is necessary to assume that there is a linear relationship between the independent variable and the dependent variables.

MLP is a fully connected multilayer neural network. It has three layers, including one hidden layer. It is an integral part of a deep neural network. When the number of hidden layers is more than one, they are called deep neural networks. MLP can model complex non-linear relationships and handle various types of data. However, MLP is sensitive to feature scaling; several hyperparameters need to be tuned, such as the number of hidden neurons, layers, and iterations.

GB is a high-performance algorithm that is mainly used for ML sorting or classification. It is generally more accurate compared to other models, but it may require more resources and time compared to simpler ML methods. In the case of high learning rates and complex models, GB can be prone to the overfitting issue, often considered a black box model that is less interpretable.

Table 4. Predictive performance of the proposed system with various weak learners for anxiety disorder using 83 variables of the Trinity-Ulster-department of Agriculture dataset

Models	Metrics				Voting
	TPR (%)	TNR (%)	Accuracy (%)	normMCC	*Th*
RF	63.0	87.8	86.5	0.6585	>17
SVM	58.0	89.1	87.5	0.6543	>17
GB	65.4	90.5	89.2	0.6885	>17
MLP	60.5	89.8	88.2	0.6668	>17
LR	59.3	87.7	86.2	0.6473	>17

Abbreviations: GB: Gradient Boosting; LR: Linear regression; MLP: Multilayer Perceptron; normMCC: Normalized Matthew's correlation coefficient; RF: Random Forest; SVM: Support vector machine; *Th*: Threshold; TNR: True negative rate; TPR: True positive rate.

4.5. Performance of the system

In this paper, we used the HADs score as the outcome variable. The ratio of the two classes is approximately 95:5, indicating an extremely imbalanced dataset. The system proposed in section 4.2. is applied to address the class imbalance. To ensure consistency in the comparisons using various weak learners, every weak learner follows the same procedure and is given the same subsets of data for the training and test sets.

Table 4 lists the predictive performance of the proposed system with weak learners of RF, SVM, GB, MLP, and LR, respectively, with a decision threshold taken as 17. The linear kernel function was used in SVM. In MLP, 200 hidden neurons were used with the adaptive moment estimation (Adam) optimizer; for both MLP and LR, 0.001 was the initial learning rate, and the mini-batch size was taken as 25. The number of decision trees was 300 in RF. Other parameters were employed in the standard procedures. Figures 12-16 illustrate the predictive performance in terms of accuracy, TPR (sensitivity), TNR (specificity), and normMCC against various decision thresholds with various weak learners.

The major difference between oversampling techniques and SMOTE is that SMOTE produces synthetic samples by interpolating samples among existing minority samples. However, oversampling techniques replicate existing minority samples to make the dataset balanced. Due to the extremely imbalanced data, SMOTE can cause problems when generating a large number of minority samples. Therefore, we opted to use an oversampling technique by replicating the existing minority samples in the training set 20 times. This approach resulted in a roughly balanced training set for model training.

Table 5 compares the performance in terms of sensitivity, specificity and normMCC between the oversampling method with repetition and the proposed system when the decision threshold is taken as 17, and RF, SVM, GB, MLP, and LR ML methods are used as weak learners, respectively, for ADs diagnosis. For fair comparison, each result used the same 83 predictor variables of the TUDA dataset.

From Table 5 , we can see that the proposed ensemble system achieves better performance than the base approach using the embedded oversampling technique. The proposed homogeneous ensemble system that used a single-base ML model across all submodels indeed increased reliability and stability in predictions compared

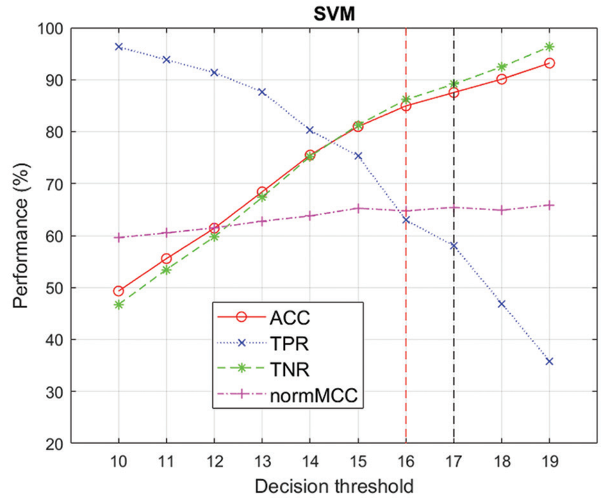

Figure 12. Predictive performance in terms of accuracy (ACC), true positive rate (TPR), true negative rate (TNR), and normalized Matthew's correlation coefficient (normMCC) against decision threshold with support vector machine (SVM) weak learner

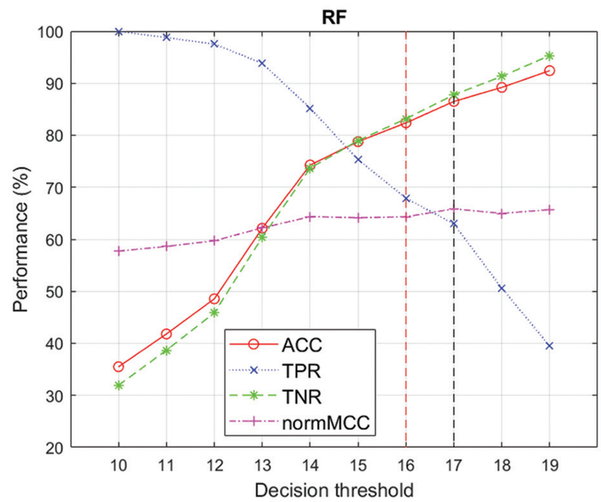

Figure 13. Predictive performance in terms of accuracy (ACC), true positive rate (TPR), true negative rate (TNR), and normalized Matthew's correlation coefficient (normMCC) with random forest (RF) weak learner against decision threshold

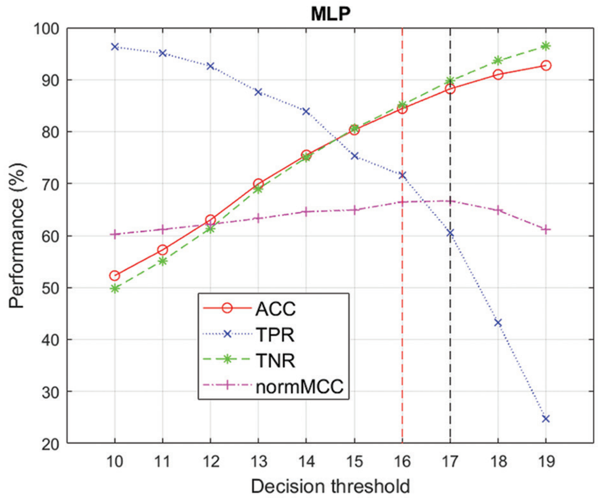

Figure 14. Predictive performance in terms of accuracy (ACC), true positive rate (TPR), true negative rate (TNR), and normalized Matthew's correlation coefficient (normMCC) with a multilayer perceptron (MLP) weak learner against a decision threshold.

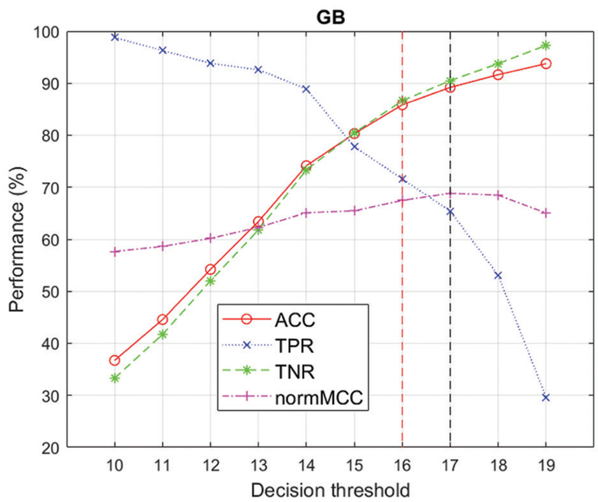

Figure 15. Predictive performance in terms of accuracy (ACC), true positive rate (TPR), true negative rate (TNR), and normalized Matthew's correlation coefficient (normMCC) with gradient boosting (GB) weak learner against decision threshold.

to using the oversampling technique. The numMCC value of the proposed system with GB as a weak learner is 0.6885, which is the highest among these five approaches.

In Figures 17-19, we use an interpretability tool, such as the Shapley function in Matlab, to interpret the prediction of a testing instance using a trained RF, GB, and SVM submodel, respectively. Shapley values are calculated to show how much each predictor variable contributes to that prediction.

Figure 20 shows the 30 most important predictor variables when a trained RF submodel was used to predict

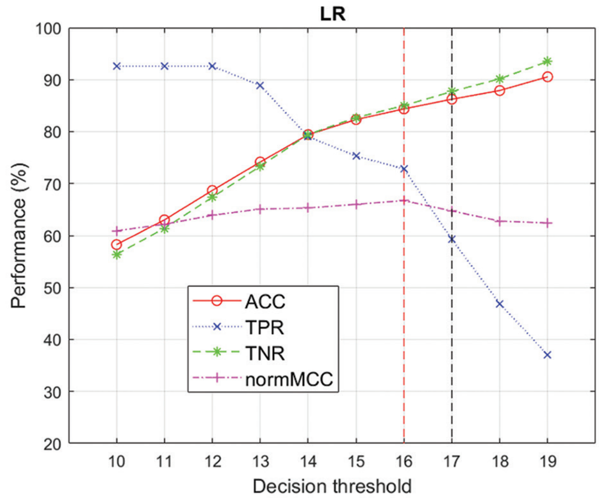

Figure 16. Predictive performance in terms of accuracy (ACC), true positive rate (TPR), true negative rate (TNR), and Matthew's correlation coefficient (normMCC) with linear regression (LR) weak learner against decision threshold

the test set. Note that variables relevant to anxiety diagnosis, depression diagnosis, total score of depression scale, and age were among the top 10 most important variables.

5. Discussion

In this paper, the imbalanced HADS variable was the outcome variable in data from the TUDA study. The results show that the proposed adapted bagging ensemble system, using a variety of sociodemographic, lifestyle, clinical, and biochemical factors, may effectively predict ADs in older adults with a high degree of accuracy. An accuracy of 89.2% (sensitivity: 65.4%, specificity: 90.5%) was achieved with the GB weak learner method when the decision threshold was set to more than 17. In this context, the sensitivity for the ADs class was 65.4% when the decision threshold was more than 17, meaning that 65.4% of true ADs diagnoses are correctly identified as ADs. Therefore, 34.6% (1–65.4%) of true AD diagnoses are unfortunately missed. For fairness, the oversampling method was investigated on the same dataset, and the results show that the proposed bagging ensemble system achieved significant improvements over the oversampling method. The threshold-moving strategy was adopted to add the predictions of multiple submodels for the final prediction, which reduced the sensitivity of each submodel to outliers or noise and avoided overfitting. Compared with using a single model, better generalization performance and higher accuracy can be achieved.

To enhance the sensitivity of ADs (the TPR) and reduce AD losses (minimize FNR), a threshold-moving strategy can be employed. By lowering the decision threshold to >16,

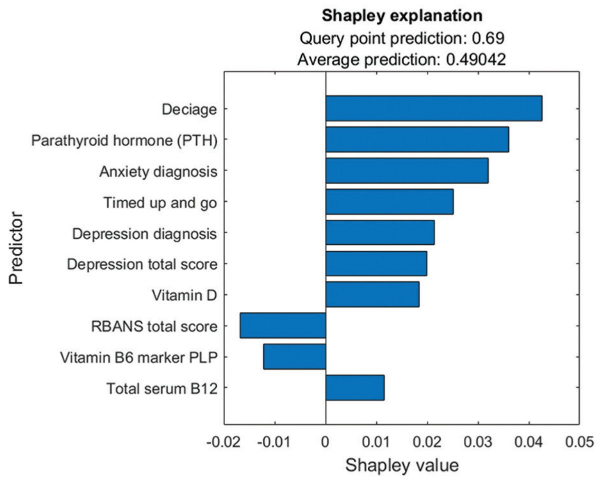

Figure 17. Shapley's explanation of the top 10 predictors predicted for a testing instance using a random forest (RF) submodel

Abbreviations: PLP: Pyridoxal phosphate; RBANS: Repeatable battery for the assessment of neuropsychological status.

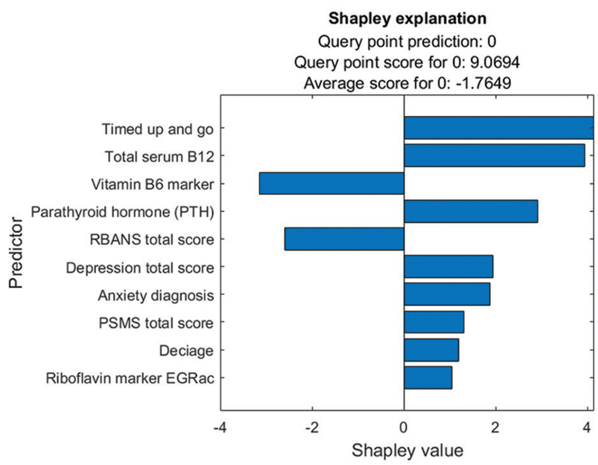

Figure 18. Shapley's explanation of the top 10 predictors predicted for a testing instance using a gradient boosting (GB) submodel

Abbreviations: EGRac: Erythrocyte glutathione reductase activation coefficient; PSMS: Physical self-maintenance scale; RBANS: Repeatable battery for the assessment of neuropsychological status.

we can identify instances as ADs when the summarized predictive value surpasses the decision threshold. Table 3 presents the results for the decision threshold values taken as more than 17. The results indicate that the proposed system achieved an accuracy of 86.5% (sensitivity: 63.0% and specificity: 87.8%) with the decision threshold taken at >17, using RF as the weak learner. The results indicate that models incorporating a combination of features, including nutrition, health, clinical, biochemical, and lifestyle factors, should be encouraged. For GB as a weak learner, we have reduced the false negative rate from 34.6%

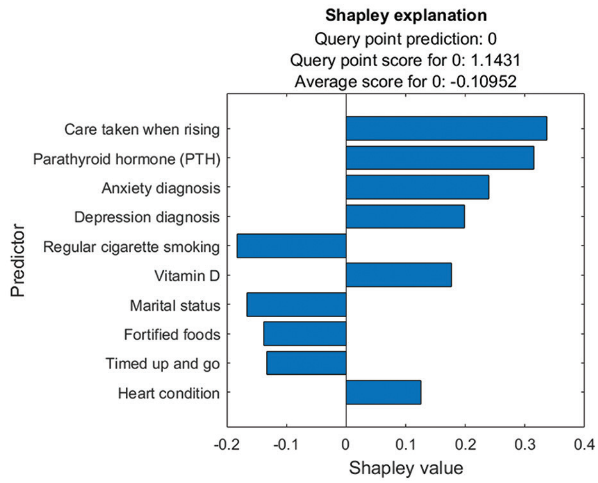

Figure 19. Shapley's explanation of the top 10 predictors predicted for a testing instance using a support vector machine submodel

down to 28.4% whereas the decision threshold changed from more than 17 to more than 16. The trade-off of misclassifying approximately 3.9% of the normal class may be deemed worthwhile to correctly classify more than 6.2% of the ADs.

As a trade-off, the number of FP will inevitably increase as we adjust to decrease the decision threshold that we apply to the model's prediction. To elucidate this trade-off and assist in threshold selection, Figures 12-16 illustrate the predictive performance in terms of the TPR and TNR against various decision thresholds. This approach is a variant of the ROC curve, with a focus on stressing the decision threshold.

Several potential risk factors for a diagnosis of ADs were identified. Understanding risk factors for ADs in people with specific chronic diseases can aid health-care professionals in immediately identifying at-risk patients, allowing improved screening activities for psychological assessment and the introduction of personalized treatments within the care settings for specific illnesses. In the long term, it will be crucial to assess the impact of real-time feedback and identify specific triggers that lead to inappropriate and high levels of anxiety.

In this study, potential risk factors were identified. Older people who have a marital status of separated/divorced or widow/widower seem to be more prone to anxiety problems. Other risk factors for ADs were identified, including areas of higher deprivation, being female, smoking, and alcohol drinking in the past. Vitamin D and fortified food/Vitamin B supplements intakes may be beneficial to ADs. Furthermore, consistent with evidence from other large cohort studies, the variables including lack of formal education,[7] functional and cognitive

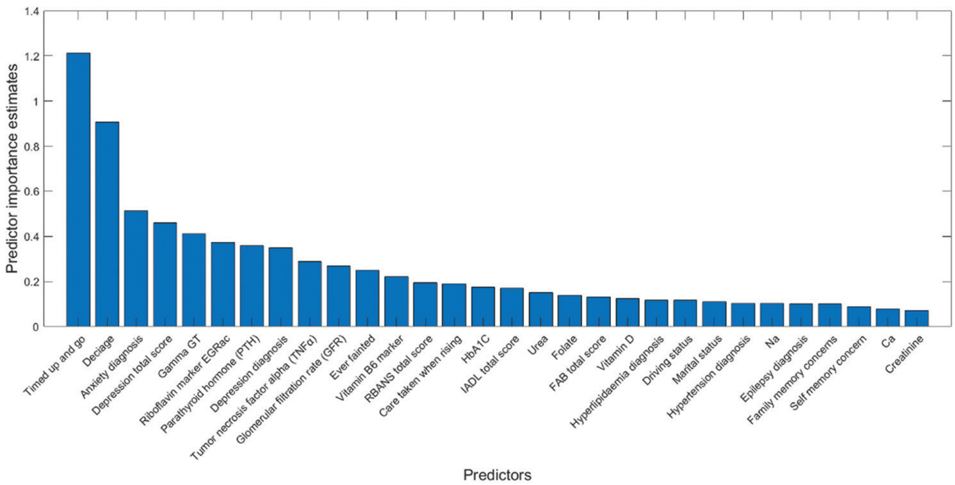

Figure 20. The 30 most important predictor variables using a random forest (RF) submodel

Abbreviations: EGRac: Erythrocyte glutathione reductase activation coefficient; FAB: Frontal assessment battery; Gamma GT: Gamma-glutamyl transferase; IADL: Instrumental activities of daily living; RBANS: Repeatable battery for the assessment of neuropsychological status.

Table 5. Performance comparison between the proposed system and oversampling method

Models	Metrics						pa
	Oversampling			Proposed			
	TPR (%)	TNR (%)	normMCC	TPR (%)	TNR (%)	normMCC	
RF	4.94	99.9	0.5861	63.0	87.8	0.6585	<2.5e-16****
SVM	79.0	76.2	0.6387	58.0	89.1	0.6543	<2.2e-16****
GB	18.5	98.6	0.6285	65.4	90.5	0.6885	<2.2e-16****
MLP	28.4	96.5	0.6301	60.5	89.8	0.6668	<2.2e-16****
LR	81.5	75.8	0.6430	59.3	87.7	0.6473	<2.2e-16****

Note: **** *p*-value indicating that Kruskal–Wallis's test rejects the null hypothesis at the 0.01% significance level.

Abbreviations: GB: Gradient boosting; LR: Linear regression; MLP: Multilayer perceptron; normMCC: Normalized Matthew's correlation coefficient; RF: Random forest; SVM: Support vector machine; TNR: True negative rate; TPR: True positive rate.

impairment,[7] poor quality of life,[19] low status of folate and metabolically related B vitamins, namely Vitamin B6, Vitamin B12, and riboflavin deficiencies have been identified as risk factors for ADs among older people.[27] It has also been considered that fortified foods could play a role in optimizing B vitamin status and potentially reduce the risk of these mental health disorders.[27] An intriguing finding is the association between Vitamin D and fortified food/Vitamin B supplements and ADs, indicating their potential value as contributing factors.

This developed system trains multiple models from generated multiple subsets of the training data, which can reduce variance, mitigate overfitting, and improve the ability of generalization of the system. The training process is scalable because each submodel is trained independently, and the developed system has the ability of expansion to handle increasing amounts of data efficiently. However, compared to training a single model, training multiple submodels may increase computational overhead for large datasets; the aggregated prediction output can reduce interpretability, too, and the reasoning behind a specific prediction can be hard to understand. In addition, there is a risk of the diluted insights if a rare but important pattern is captured by a single submodel; the impact might be diluted by aggregating it with other submodels. The choice of base model affects the performance of the system. The limitations also include that the observational studies are subject to bias due to their inherent flaws; there were fewer positive cases in the training set, resulting in a decrease in the ability of the system to predict the occurrence of positive cases. The findings of this study, focusing on a specific cohort, are not generalizable to other racial groups. The developed system is able to learn from different subsets of data and reduce variance, making it suitable for various external applications, such as those in clinical settings. Identified risk factors, such as the link between certain lifestyle factors and ADs, may have clinical implications and help health-care professionals invest more effort in educating people about these factors. The system could help doctors identify and treat patients earlier and even improve treatment outcomes.

In Figures 17-19, we show Shapley's explanation of the top 10 predictors predicted for the same testing instance using RF, GB, and SVM submodels trained with the same subdataset, respectively. The interpretability of each trained RF, GB, and SVM submodel varied, with predictors such as anxiety diagnosis, depression diagnosis, total score of depression scale, age, and timed up and go, ranking among the top ten factors contributing most to the prediction.

ML algorithms such as RF, MLP, SVM, GB, and LR are widely used and typically achieve good performance.[18] The success of these algorithms depends on the quality and quantity of features used, as well as the characteristics of the available dataset. However, when dealing with extremely imbalanced datasets, using these methods directly for specificity in the disadvantaged class can lead to poor performance. In this study, we demonstrated how to adapt an ensemble bagging ML method to deal with imbalanced classes. Through this adaptation, satisfactory performance was achieved.

6. Conclusion

ML technologies have become increasingly popular for disease diagnosis in the field of mental health. ADs are one of the major health burdens facing older adults worldwide. Despite evidence[43-46] that the impact of ADs can be reduced through prevention and intervention, the prevalence remains high worldwide, highlighting that people with such disorders need intervention.[47-49] The widespread prevalence of anxiety-related disorders presents many challenges for mental health-care providers, who find it difficult to provide face-to-face treatments to those who need it in a timely manner. As more complex health data is becoming available, ML can deal with data that can be from multiple sources, predict the risk of developing ADs by identifying individual characteristics and risk factors, and perform mental health diagnoses. ML can help personalize treatment plans and ensure that individuals receive the most appropriate care. In this study, an adapted bagging ensemble approach was proposed to identify ADs, in which traditional ML methods act as weak learners. In future work, the goal is to identify ADs in its earlier or prodromal stage, when interventions may be more effective, and treatments can be personalized based on individuals' unique characteristics. This could pave the way for developing a "mental health status indicator" to monitor an individual's mental health, establish different alert levels, and efficiently address emerging issues.

In this study, key predictors were identified that could effectively predict ADs using the proposed learning system, achieving a satisfactory level of accuracy. Some variables were determined to be closely associated with an increased risk of ADs, such as gender, marital status, accommodation status, lifestyle-related factors, quality of life, fortified food/supplements intake level, and family history of certain diseases and chronic diseases. Efforts were made to assess risk factors of anxiety in older adults. More studies are needed to fully understand the characteristics of anxiety in this population.

Future works involve conducting a longitudinal study by studying the individuals in the TUDA dataset over the years to observe how they develop over time, to understand physical and cognitive developmental processes, and to predict future development of ADs. The long-term consequences of interventions can be investigated.

ADs and depression are commonly found to coexist. Participants with one condition were generally at higher risk for the other condition. Future studies should explore the relationship between ADs and depression, especially given that older adults often suffer from comorbidities. Maintaining mental health is crucial, not only to improve daily functioning, strengthen relationships, and enhance self-image, but also to address physical health issues related to mental health conditions. It could help reduce the prevalence of ADs by supporting older adults' participation in physical activities and reducing social isolation.[49] Encouraging proper intake of fortified food and supplements to prevent deficiencies of necessary vitamins is also essential. Given the burden on health-care resources, these efforts may promote inclusivity in policymaking, especially in identifying strategies in the public health field to promote reduced inequalities and improved mental health.

Acknowledgments

The authors would like to acknowledge the support of all who are involved in the AIM4HEALTH programme.

Funding

The TUDA study was supported by government funding from the Irish Department of Agriculture, Food and the Marine, and Health Research Board (under the Food Institutional Research Measure), as well as from the Northern Ireland Department for Employment and Learning (under its Strengthening the All-Island Research

Base Initiative). The AIM4HEALTH project gratefully acknowledges the support of the higher education authority, Department of Further and Higher Education, Research, Innovation and Science, and the Shared Island Fund, and the SFI grant 21/RC/10295_P2.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Jinling Wang, Michaela Black, Debbie Rankin
Formal analysis: Jinling Wang
Investigation: Catherine F. Hughes, Leane Hoey, Geraldine Horigan, Helene McNulty, Anne M. Molloy
Methodology: Jinling Wang, Michaela Black, Debbie Rankin, Catherine F. Hughes, Leane Hoey, Anne M. Molloy, Helene McNulty, Mimi Zhang
Writing–original draft: Jinling Wang, Debbie Rankin
Writing–review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The data will not be shared due to some concerns. For more data information, please refer to the relevant research works.[7,26-28]

References

1. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet*. 2021;398:1700-1712.

   doi: 10.1016/S0140-6736(21)02143-7

2. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: A systematic analysis for the global burden of disease study 2019. *Lancet Psychiatry*. 2022;9:137-150.

   doi: 10.1016/S2215-0366(21)00395-3

3. Lauderdale SA, Sheikh JL. Anxiety disorders in older adults. *Clin Geriatr Med*. 2003;19(4):721-741.

   doi: 10.1016/s0749-0690(03)00047-8

4. Sheikh JI. Investigations of anxiety in older adults: Recent advances and future directions. *J Geriatr Psychiatry Neurol*. 2005;18(2):59-60.

   doi: 10.1177/0891988705276253

5. Andreescu C, Varon D. New research on anxiety disorders in the elderly and an update on evidence-based treatments. *Curr Psychiatry Rep*. 2015;17(7):53.

   doi: 10.1007/s11920-015-0595-8

6. Ishikawa RZ, Vyas C, Okereke O. Anxiety disorders among older adults: Empirically supported treatments and special considerations. In: Bui E, Charney ME, Baker AW, editors. *Clinical Handbook of Anxiety Disorders: From Theory to Practice*. United States: Humana Press/Springer Nature; 2020. p. 175-189.

   doi: 10.1007/978-3-030-30687-8_9

7. Rankin D, Black M, Flanagan B, *et al*. Identifying key predictors of cognitive dysfunction in older people using supervised machine learning techniques: Observational study. *JMIR Med Inform*. 2020;8(9):e20995.

   doi: 10.2196/20995

8. Javaid SF, Hashim IJ, Hashim MJ, Stip E, Samad MA, Ahbabi AA. Epidemiology of anxiety disorders: Global burden and sociodemographic associations. *Middle East Curr Psychiatry*. 2023;30:44.

   doi: 10.1186/s43045-023-00315-3

9. Fusar-Poli P, Correll CU, Arango C, Berk M, Patel V, Ioannidis JP. Preventive psychiatry: A blueprint for improving the mental health of young people. *World Psychiatry*. 2021;20:200-21.

   doi: 10.1002/wps.20869

10. Jorm AF, Patten SB, Brugha TS, Mojtabai R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. *World Psychiatry*. 2017;16:90-99.

    doi: 10.1002/wps.20388

11. Jain PR, Quadri SMK. Emerging role of intelligent techniques for effective detection and prediction of mental disorders. In: Hemanth J, Bestak R, Chen JIZ, editors. *Intelligent Data Communication Technologies and Internet of Things. Lecture Notes on Data Engineering and Communications Technologies*. Vol. 57. Singapore: Springer; 2021.

    doi: 10.1007/978-981-15-9509-7_16

12. Meehan AJ, Lewis SJ, Fazel S, *et al*. Clinical prediction models in psychiatry: A systematic review of two decades of progress and challenges. *Mol Psychiatry*. 2022;27:2700-2708.

    doi: 10.1038/s41380-022-01528-4

13. Graham S, Depp C, Lee EE, *et al*. Artificial intelligence for mental health and mental illnesses: An overview. *Curr Psychiatry Rep*. 2019;21:116.

    doi: 10.1007/s11920-019-1094-0

14. Cearns M, Hahn T, Baune BT. Recommendations and future

directions for supervised machine learning in psychiatry. *Transl Psychiatry*. 2019;9:271.

doi: 10.1038/s41398-019-0607-2

15. Thieme A, Belgrave D, Doherty G. Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ml systems. *ACM Trans Comput Hum Interact*. 2020;27(5):1-53.

doi: 10.1145/3398069

16. Ancillon I, Elgendi M, Menon C. Machine learning for anxiety detection using biosignals: A review. *Diagnostics (Basel)*. 2022;12(8):1794.

doi: 10.3390/diagnostics12081794

17. Khan A, Husain MH, Khan A. Analysis of mental state of users using social media to predict depression: A survey. *Int J Adv Res Comput Sci*. 2018;9:100-106.

doi: 10.26483/ijarcs.v9i0.6146

18. Agarwal D, Singh V, Singh AK, Madan P. Stacked ensemble model for analyzing mental health disorder from social media data. *Multimed Tools Appl*. 2023;83:53923-53948.

doi: 10.1007/s11042-023-17395-2

19. Nemesure MD, Heinz MV, Huang R, Jacobson NC. Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. *Sci Rep*. 2021;11(1):1980.

doi: 10.1038/s41598-021-81368-4

20. Shen ZX, Cui LJ, Mou SQ, *et al*. Combining S100B and cytokines as neuro-inflammatory biomarkers for diagnosing generalized anxiety disorder: A proof-of-concept study based on machine learning. *Front Psychiatry*. 2022;13:881241.

doi: 10.3389/fpsyt.2022.881241

21. Byeon H. Exploring factors for predicting anxiety disorders of the elderly living alone in South Korea using interpretable machine learning: A population-based study. *Int J Environ Res Public Health*. 2021;18(14):7625.

doi: 10.3390/ijerph18147625

22. Henry M, Isa SM. Mental health treatment prediction for Tech Employee with the implementation of ensemble methods. *J Theor Appl Inf Technol*. 2022;100(8):2675-2685.

23. Rocca J. *Ensemble Methods: Bagging, Boosting and Stacking--Understanding the Key Concepts of Ensemble Learning*; 2019. Available from: https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205. [Last accessed on 2024 Jan 27].

24. Patel A. *Ensemble Learning- the Heart of Machine Learning*. Available from: https://medium.com/ml-research-lab/ensemble-learning-the-heart-of-machine-learning-b4f59a5f9777 [Last accessed on 2020 Jan 03].

25. Chicco D, Tötsch N, Jurman G. The matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min*. 2021;14(1):13.

doi: 10.1186/s13040-021-00244-z

26. McCann A, McNulty H, Rigby J, *et al*. Effect of area-level socioeconomic deprivation on risk of cognitive dysfunction in older adults. *J Am Geriatr Soc*. 2018;66(7):1269-1275.

doi: 10.1111/jgs.15258

27. Moore K, Hughes CF, Hoey L, *et al*. B-vitamins in relation to depression in older adults over 60 years of age: The trinity ulster department of agriculture (TUDA) cohort study. *J Am Med Dir Assoc*. 2019;20(5):551-557.e1.

doi: 10.1016/j.jamda.2018.11.031

28. Wang J, Black M, Rankin D, *et al*. Analysis of Risk Factors and Diagnosis for Anxiety Disorder in Older People with the Aid of Artificial Intelligence: Observational Study. In: *2023 the 31st Irish Conference on Artificial Intelligence and Cognitive Science,* Letterkenny, Ireland, IEEE. p. 1-8.

doi: 10.1109/aics60730.2023.10470782

29. Larsen BS. *Synthetic Minority Over-Sampling Technique (SMOTE)*. Available from: https://github.com/dkbsl/matlab_smote/releases/tag/1.0),github [Last accessed on 2023 May 31].

30. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321-357.

doi: 10.1613/jair.953

31. He H, Bai Y, Garcia EA, Li S. ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. In: *2008 IEEE International Joint Conference on Neural Networks*; 2008. p. 1322-1328.

doi: 10.1109/ijcnn.2008.4633969

32. Edwards R. *Causes and Risk Factors of Anxiety*; 2021. Available from: https://www.verywellhealth.com/anxiety-causes-and-risk-factors-5191778 [Last accessed on 2023 Dec 03].

33. Narmandakh A, Roest AM, De Jonge P, *et al*. Psychosocial and biological risk factors of anxiety disorders in adolescents: A TRAILS report. *Eur Child Adolesc Psychiatry*. 2021;30:1969-1982.

doi: 10.1007/s00787-020-01669-3

34. *UK Statistics on Vitamin and Mineral Deficiency*; 2023. Available from: https://vitall.co.uk/health-tests-blog/statistics-vitamin-mineral-deficiency-uk [Last accessed on 2023 Aug 23].

35. *Vitamin D: The Connection to Depression and Anxiety*. Available from: https://montarebehavioralhealth.com/vitamin-d-the-connection-to-depression-and-anxiety [Last accessed on 2023 Aug 23].

36. Chang S, Lee H. Vitamin D and health - the missing vitamin in humans. *Pediatr Neonatol*. 2019;60(3):237-244.

doi: 10.1016/j.pedneo.2019.04.007

37. Menon V, Kar SK, Suthar N, Nebhinani N. Vitamin D and depression: A critical appraisal of the evidence and future directions. *Indian J Psychol Med*. 2020;42(1):11-21.

    doi: 10.4103/ijpsym.ijpsym_160_19

38. Kowalówka M, Gówka AK, Karaniewiczada M, Kosewski G. Clinical significance of analysis of vitamin D status in various diseases. *Nutrients*. 2020;12(9):2788.

    doi: 10.3390/nu12092788

39. Chiang JJ, Park H, Almeida DM, *et al*. Psychosocial stress and C-reactive protein from mid-adolescence to young adulthood. *Health Psychol*. 2019;38(3):259-267.

    doi: 10.1037/hea0000701

40. Anjum I, Jaffery SS, Fayyaz M, Samoo Z, Anjum S. The role of vitamin D in brain health: A mini literature review. *Cureus*. 2018;10(7):e2960.

    doi: 10.7759/cureus.2960

41. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020;21(6):6.

    doi: 10.1186/s12864-019-6413-7

42. Chicco D, Jurman G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Min*. 2023;16(1):4.

    doi: 10.1186/s13040-023-00322-4

43. Adeniji OD, Adeyemi SO, Ajagbe SA. An improved bagging ensemble in predicting mental disorder using hybridized random forest - artificial neural network model. *Informatica*. 2022;46(4):543-550.

    doi: 10.31449/inf.v46i4.3916

44. Ogunseye EO, Adenusi CA, Nwanakwaugwu AC, Ajagbe SA, Akinola SO. Predictive analysis of mental health conditions using adaboost algorithm. *Paradigmplus*. 2022;3:11-26.

    doi: 10.55969/paradigmplus.v3n2a2

45. Alabi EO, Adeniji OD, Awoyelu TM, Fasae EO. Hybridization of machine learning techniques in predicting mental disorder. *Int J Hum Comput Stud*. 2021;3(6):22-30.

    doi: 10.31149/ijhcs.v3i6.2083

46. Obiedat R, Toubasi SA. A combined approach for prediction employee's productivity based on ensemble machine learning methods. *Int J Comput Inform*. 2022;46:49-58.

    doi: 10.31449/inf.v46i5.3839

47. Moreno C, Wykes T, Galderisi S, *et al*. How mental health care should change as a consequence of the COVID-19 pandemic. *Lancet Psychiatry*. 2020;7:813-824.

    doi: 10.1016/S2215-0366(20)30307-2

48. Campion J, Javed A, Sartorius N, Marmot M. Addressing the public mental health challenge of COVID-19. *Lancet Psychiatry*. 2020;7(8):657-659.

    doi: 10.1016/S2215-0366(20)30240-6

49. Baxter AJ, Scott KM, Vos T, Whiteford HA. Global prevalence of anxiety disorders: A systematic review and meta-regression. *Psychol Med.* 2013;43(5):897-910.

    doi: 10.1017/S003329171200147X