Analysis of Risk Factors and Diagnosis for Anxiety Disorder in Older People with the Aid of Artificial Intelligence: Observational Study

Analysis of Risk Factors and Diagnosis for Anxiety Disorder in Older People with the Aid of Artificial Intelligence: 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[6], Mimi Zhang[6], James Ng[6], Geraldine Horigan[3], Paul Carlin[7], Kevin McCarroll[8], Conal Cunningham[8], Helene McNulty[3], Anne M Molloy[5]

[1]*School of Computing, Engineering and Intelligent Systems*
Ulster University
Derry~Londonderry, United Kingdom
(j.wang; mm.black; d.rankin1) @ulster.ac.uk

[2]*School of Computing*
Ulster University
Belfast, United Kingdom
jg.wallace@ulster.ac.uk

[3]*School of Biomedical Sciences*
Nutrition Innovation Centre for Food and Health
Ulster University
Coleraine, United Kingdom
(c.hughes; l.hoey; h.mcnulty; gb.horigan) @ulster.ac.uk

[4]*School of Geography and Environmental Sciences*
Ulster University
Coleraine, United Kingdom
a.moore@ulster.ac.uk.

[5]*School of Medicine,*
[6]*School of Computer Science & Statistics*
Trinity College Dublin
Dublin, Ireland
(amolloy; tobinjo; mimi.zhang; NGJA) @tcd.ie

[7]*School of Health, Wellbeing and Social Care*
The Open University
Belfast, United Kingdom
paul.carlin@open.ac.uk

[8]*Mercers Institute for Research on Ageing*
St James's Hospital
Dublin, Ireland
(kev_mccarroll@hotmail.com; CCunningham@stiames.ie)

***Abstract*— Anxiety disorders are the most common mental health problems particularly in older people who suffer from loneliness and social isolation, chronic health conditions, financial insecurity and other factors that can lead to anxiety disorders. The high prevalence and health risks of anxiety disorders, and the requirement for effective mental healthcare, integrated with recent advances in artificial intelligence, has resulted in an increase in the exploration of how machine learning can aid the diagnosis and prediction of mental health problems. Data from the Trinity-Ulster-Department of Agriculture (TUDA) study will be utilized to identify risk factors for anxiety in community dwelling older adults using machine learning techniques. The TUDA study includes detailed information on sociodemographic, clinical, biochemical, nutritional, and lifestyle factors in 5186 older people recruited from the Republic of Ireland and Northern Ireland. These characteristics could foster the prediction of anxiety disorders using supervised machine learning methods. Biomarker risk factor analysis was conducted to facilitate feature engineering. In this observational study, several classical machine learning models have been trained to predict anxiety disorders. Principal component analysis was used for further feature reduction, comparing the accuracy results of various features, and determining the impact of features on the predictions of each method. The models' performance was assessed on a held-out test set and achieved an accuracy of 85.4% (sensitivity: 67.0%, specificity: 90.3%) and 83.4% (sensitivity: 81.5%, specificity: 83.9%) for two best performing methods i.e., random forest and support vector machine respectively, using the standard Synthetic Minority Oversampling Technique. Risk factors such as female sex, loneliness, separated/divorced conditions, lifestyle-related, socio-economic low status, chronic diseases and family related diseases are identified. These results will aid in the early detection of anxiety disorder in future studies.**

Keywords—anxiety disorder, risk factor analysis, machine learning, diagnosis, ageing

I. Introduction

Anxiety signifies that a person is in anticipation of a future concern; it is a natural human response to the stress we feel about things that we think could happen in future. Mild levels of anxiety can be beneficial in some conditions, to alert us and help us pay attention to dangers. Anxiety disorders (AD) however involve persistent, excessive anxiety or fear, that profoundly affects normal daily functioning. The five main types of anxiety disorders include 1) generalised anxiety disorder (GAD), 2) panic disorder, 3) obsessive-compulsive disorder.4) social anxiety disorder, and 5) separation anxiety disorder. The cause of anxiety disorders seems to be multifactorial, involving genetic traits, and traumatic events as triggers [1-2].

Anxiety disorders are one of the most common mental health conditions. It is estimated that 4.05% of the global population suffer from an anxiety disorder [3], which affected 298 million or 3824.9 cases per 100,000 people prior to the COVID-19 pandemic [1]. The burden related to anxiety disorders has risen to 4802.4 cases per 100,000 in 2020 [1]. Following the pandemic, around 977 more cases have been added per 100,000 people. From 1990 to 2019, the number of affected people increased by over 55%, and it's still growing [3]. More than 8 million people in the UK are suffering from an anxiety disorder at any one time [4]. In England 6 out of 100 people are diagnosed with generalized anxiety disorder on any given week [5]. In Northern Ireland one in five adults have had a mental health issue, a single biggest cause of poor health and disability [6].

The AIM4HEALTH project gratefully acknowledges the support of the HEA, DFHERIS and the Shared Island Fund

There is variation in the results reported in several studies for the prevalence of anxiety disorders, ranging from 1.2% to 14.2% in older adults aged 55 years and over [7-10]. As the population is aging around the world, it is estimated that the population of people over 60 years of age will reach more than 2 billion by 2050 [11]. Anxiety disorder often causes reduced quality of life, distress, disability and even a risk of death in older adults, and has been associated with cardiovascular risks, cognitive decline, and other chronic diseases. The mechanisms of anxiety in older adults are mainly related to age-related neuropathology, as well as apparent loss such as due to retirement and economic hardship, bereavement and isolation in later life. The prevalence of anxiety-related disorders across the population is challenging for mental health service providers who struggle to deliver in-person therapy sessions in a timely manner to individuals in need [3].

A number of psychotherapeutic methods have been used in the treatment of anxiety disorder. The annual worldwide cost associated with treating such disorders is estimated at $2.4 trillion as of 2010, and is estimated to reach $16.3 trillion by 2030 [12], exceeding that of other diseases such as cancer, cardiovascular disease, diabetes and chronic respiratory disease. In the UK, mental health related disorders accounted for 7% of all diseases and cost £117.9 billion as of 2019, a 21-fold increase in the economic cost since 2009 [13-14]. In Northern Ireland mental health issues cost the economy £3.4 billion each year [6]. Early intervention and diagnostics could potentially lead to reduction of the growing personal, social, and economic cost of common mental health disorders [15-16], and the high prevalence of mental illness requires effective mental healthcare strategies for older age groups. The pandemic was a catalyst for the adoption of technologies to support mental health and services [1].

Machine learning (ML) and artificial intelligence (AI) technologies have become increasingly popular for disease diagnosis, and are especially important in mental health, where there is a worldwide shortage of qualified professionals capable of dealing with these problems; the cost for these services is high and people suffering from these problems often refuse to take advantage of these services due to social stigma. If left untreated, people with anxiety disorders may experience a range of adverse consequences including the onset of physical, mental, functional, cognitive, and social impairments, as well as delayed recovery from illness, decreased quality of life, and increased utilisation of health care services [12]. As more complex health data is available, ML and AI methods can be used for risk factors analysis and diagnosis of mental health and provides a platform for doctors to conduct personalized treatment according to the patient's medical situation. Therefore, identifying relevant risk factors and predicting anxiety disorder prevalence among the ageing population will allow healthcare providers to develop strategies aimed at reducing anxiety.

In this paper, the potential risk factors contributing to an anxiety disorder diagnosis will be explored using statistical methods, ML and AI with the Trinity-Ulster and Department of Agriculture (TUDA) study (ClinicalTrials.gov identifier: NCT02664584) dataset. Identified risk factors may assist in the detection, diagnosis, and prediction of anxiety disorders. Several analytical models on the data were developed and compared to identify potential risk factors that may serve as predictors of poor anxiety disorders.

II. Related work and Risk factor analysis

Machine learning (ML) and artificial intelligence (AI) have become increasingly popular in supporting clinical decision making in mental healthcare, attributed to the availability of large amounts of complex data and improved computational capabilities [4]. Using ML, these illnesses can be identified at an earlier or prodromal stage when interventions may be more effective, and personalized treatments based on an individual's unique characteristics can be used. Although in the medical and health fields lots of algorithms have been developed and presented, review papers have been presented for application of AI in mental health and common trends, gaps and challenges have been highlighted [17-21].

In [21], the success of different ML methods for detecting and predicting AD with various bio-signals were reviewed. Perspectives on the advantages and disadvantages of current developments are provided to guide future advancements in anxiety detection. Support Vector Machines (SVM) and Random Forest (RF) were the most commonly used ML methods and have achieved good results when combined with feature selection. Neural networks also provide good results and are extensively used. The review also indicates the importance of features and benefits of multi-modality. A study by Nemesure et al. [22] used a novel ensemble ML pipeline that includes deep learning to diagnose and predict psychiatric illness with Electronic Health Records (EHRs) containing biometric and demographic data from 4184 undergraduate students. A moderate predictive performance on a held-out test set is achieved (sensitivity: 0.66, specificity: 0.7), and the top six most important features i.e., vaccinations being up to date, marijuana use, control examination needed, hypertension or prehypertension, systolic blood pressure and the use of other recreational drugs for predicting generalised anxiety disorder are identified.

In [23], an integrated back propagation neural network based on a bagging algorithm (BPNN-Bagging) was used for diagnosing GAD by combining S100 calcium-binding protein B (S100B) and Cytokines as Neuro-Inflammatory Biomarkers. S100B can regulate neuronal growth and plasticity, and astrocytes and microglia are activated through the production of GAD-related cytokines. ML techniques were used for feature ordering of cytokines and S100B, and the classification. An accuracy of 94.47% in diagnosing GAD has been achieved.

Byeon [24] presented a study that used SVM and ensemble learning to identify group at a high risk of anxiety disorders in old age. SVM, RF, Light Gradient Boosting Method, and Adaboost are used for the base model, a single predictive model, the predictive performance in terms of accuracy of the single model were compared, while XGBoost is used for the meta model. The stacking models that combining different base models and the meta model were explored. The results showed that the predictive performance of stacking ensemble models have achieved the best accuracy of 87.4%, with 85.1% precision and 87.4% recall after proper selecting base model. The predictors with the highest risk i.e., subjective loneliness, subjective family relations, Self Esteem Scale (SES), family relationship and dissolution instability, subjective frequency of communication with family,

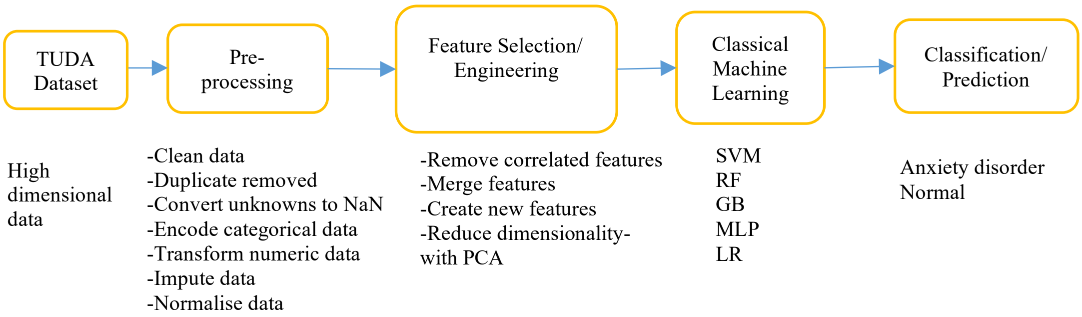

Fig. 1. Pre-processing and analysis of high dimensional TUDA dataset.

instability in family support and caregiving, and your and your family's experience of being a victim of a crime over the past year, were identified. This indicates the need of a tool that can identify older adults at high risk of anxiety disorders and manage them effectively. From the literature, we can see that the single model such as SVM and RF, combined with feature selection, can lead to good diagnosis and prediction of anxiety disorders, a stacking ensemble model can work better when the single models are properly stacked, and the best predictive performance was identified by exploring the designed stacking model via combining different base models and the meta model.

Although several studies that apply state of art ML and AI techniques to mental health data have been described above, more work needs to be done in mental health. AI can be a promising solution for precision medicine that is capable of the needs of the individual patient. ML can be applied to address the challenges and identify inherent features of anxiety disorders.

In the following sections, analysis of multivariate risk factors is conducted on the TUDA dataset. The results are then used for feature selection; the predictive performance in terms of accuracy with several popular models such as SVM, RF, Gradient Boosting (GB), Multilayer Perception (MLP) and Linear Regression (LR) are compared with and without oversampling. Efforts are made to develop ML methods for the prediction of AD.

III. METHODS

In this section, the TUDA dataset is described and preprocessed, the key patient's predictors are identified to facilitate feature selection and feature engineering using statistical and machine learning methods.

A. *TUDA dataset*

The TUDA cohort comprises detailed nutrition and health data, along with related lifestyle, sociodemographic, clinical and biochemical details on a total of 5186 community-dwelling older people aged 60 to 102 years, making this cohort one of the most comprehensively characterised cohorts of its kind for aging research internationally. More details regarding this dataset are contained in relevant published works [11 25-26]. The original dataset includes 701 variables, which have been segmented into groups based on domain knowledge for facilitating future analysis. General characteristics of some features of study participants can refer to Table I in [11]. Fig. 1 shows the pre-processing and analysis steps undertaken for the high dimensional TUDA dataset. In this study we focus on diagnosis of anxiety disorders.

B. *Preprocessing of the TUDA dataset*

During the initial cleansing and exploration, invalid values, spelling mistakes, inaccurate values and coding inconsistencies were identified and corrected; duplicate variables were identified and removed; to deal with missing values, a cut-off threshold of 10% was set where if the number of missing values for a specific column is less than the threshold, the variable was retained, otherwise, it was deleted. Then the text information was recoded as numeric values, so the cleaned dataset contains only numerical variables with some missing values. Unknown values were converted to NaN; categorical data are encoded into numeric values using the one-hot encoding method.

The dataset was split into real numeric continuous and ordinal or nominal categorical variables. Two methods were used for the transformation of real numeric variables, i.e., square root and log transformation. The continuous numeric variables were divided into two groups based on the minimum and maximum values for each variable. Those variables whose minimum value equals zero were placed in one group and the values of these variables were square root processed; other numeric variables whose minimum values do not equal zero, were log processed.

Finally, the two data subsets are concatenated, the missing values of remaining variables are imputed using the K-Nearest Neighbors (KNN) algorithm, and then the dataset was normalised to the range of [0 1] using z-score normalisation.

C. *Feature selection/Feature engineering*

A correlation analysis is necessary to avoid potential unstable estimates when dealing with multicollinear and correlated predictors, as well as redundant predictors that do not contribute additional information during the development of classification models [11]. Correlations among variables and the diagnosis of anxiety disorders were explored for variable selection. An unimputed and unnormalised dataset was used initially for risk factor analysis to investigate the association between an anxiety diagnosis and other nominal variables such as diseases, medications, lifestyle factors. Based on the results, some variables are merged, and new variables are created. Variables that were highly correlated and therefore do not add meaningful information for prediction were removed. All 5186 records, 84 variables (83 predictors, 1 outcome) remained for initial analysis. The outcome indicates anxiety diagnosis of each participant, which was self-reported by the participant. In the following experiments, statistical biomarker risk factors analysis is conducted.

The Spearman nonparametric correlation coefficient was used to calculate correlations between numerical variables. Nominal and ordinal variables are the common types of categorical variables. To explore the association between these nominal variables, Cramer's V is used. Cramer's V is in the range of [0 1]. 0 indicates no association between the two variables whereas 1 indicates a strong association between the two variables. It can be calculated as shown in (1). Cramer's V was used to identify considerable variation and strong associations between the nominal variables, and the results were then used to identify redundant variables which can be removed from the dataset. For example, the variable 'lipid_meds' that is related to lipidaemia medication intake is removed as it has a strong association with variable 'Hyperlipidaemiadiagnosis' (Cramer's V=0.7168), which is related to diagnosis of hyperlipidaemia.

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

Where X^2 is the Chi-square statistic, n represents 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 anxiety disorder.

A nonparametric Kruskal-Wallis test was used to determine whether data from two groups differed from each other. A post-hoc test such as the Dunn-Bonferroni-Test or Wilcoxon test can be used to determine which of the groups differed if needed. The calculated p-value is compared to the significance level usually set at 0.05. If the p-value is bigger, the null hypothesis is retained, the Kruskal-Wallis test indicates that there is no significant difference between categories of independent variables with respect to the dependent variable. Otherwise it is rejected. After these statistical analyses were performed, 84 variables were remained.

D. Risk factor analysis

The risk factors for anxiety disorders in older people are a combination biological, socio-demographic, physical, economic and psychological. It is currently not clear what causes anxiety disorders, but it is likely a combination of factors including age, genetics, diet, lifestyle, psychological and environmental factors, and medical conditions such as cardiovascular disease, diabetes, hypertension, lipidaemia and some chronic inflammatory diseases [9-10]. A study suggests that interventions targeting modifiable risk factors for anxiety disorders may prevent or mitigate disease onset [12].

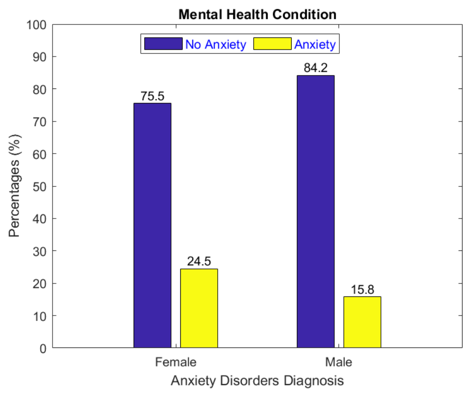

Fig. 2. Bar plot to show percentages for male and female in terms of anxiety disorders diagnosis.

Anxiety in the Figures in this study indicates anxiety disorders that are self-reported by the participants themselves. Figs. 2-8 illustrate the percentages of frequency distribution of the predictors in terms of anxiety diagnosis. Table I shows some characteristics of variables related to Figs. 2-7. Fig. 2 indicates that 'Gender' is a notable feature because 24.5% of women suffer with anxiety disorders compared to just 15.8% of men. It could be because of differences in brain chemistry and hormone fluctuations. Reproductive events across a woman's life are associated with hormonal changes, which have been linked to anxiety.

TABLE I. GENERAL CHARACTERISTICS OF THE TUDA STUDY PARTICIPANTS-- UNIVARIATE ANALYSIS..

Variables	Anxiety Disorders	Depression Diagnosis
	Yes (n=1122)	***Yes (n=1247)***
	Number (%)	Number (%)
Gender		
Male	269 (15.8)	332 (19.5)
Female	853 (24.5)	915 (26.2)
Marital status		
Single	127 (19.4)	149 (22.8)
Married/Common law	573 (21.2)	600 (22.1)
Separated/Divorced	78 (31.3)	92 (36.9)
Widow/Widower	343 (21.8)	405 (25.7)
Area deprivation		
Normal	755 (20.3)	822 (22.1)

[1]SESlow	340 (25.7)	383 (28.9)
Accommodation status		
Alone	407 (23.2)	472 (27.0)
Spouse/Partner	565 (21.2)	591 (22.0)
Children	109 (17.6)	134 (26.1)
Other	41 (21.0)	50 (21.5)
Smoking		
No	495 (20.1)	518 (21.1)
Yes	627 (23.0)	729 (26.8)
Drinking alcohol		
Never	245 (19.1)	266 (20.7)
Past	243 (26.3)	281 (30.4)
Current	634 (21.3)	699 (23.5)
Diagnosis (yes) of		
Hypertension	837 (22.6)	910 (24.5)
Hyperlipidaemia	666 (24.2)	744 (27.0)
Diabetes	141 (21.4)	176 (26.7)
Other serious diseases	319 (22.3)	337 (23.6)
Self-memory concern	407 (26.9)	483 (31.9)
Family-memory concern	166 (25.1)	216 (32.7)
Family history (yes) of		
Cancer	336 (23.6)	361 (25.3)
Stroke	111 (23.7)	127 (27.1)
Heart disease	362 (23.9)	412 (27.2)
Preseniledementia	24 (32.4)	30 (40.5)
Seniledementia	172 (24.5)	173 (24.6)

[1]SESlow: Standard Socio-Economic Status

Females with a marital status of single and separated/divorced seem to be more prone to anxiety problems than males (see Fig. 3). While the effects of family factors such as anxiety about the dissolution of family relations were significantly higher. It might be related to family support or caregiving.

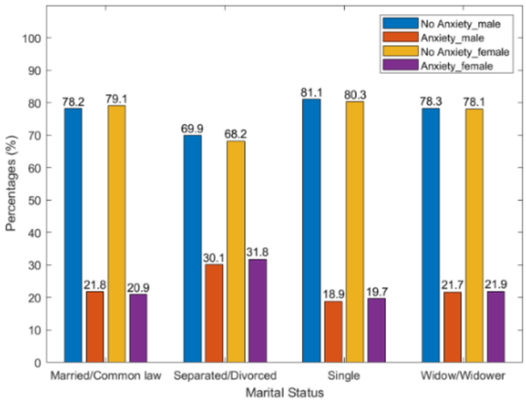

Fig. 3. Anxiety disorders diagnoses broken down by sex and marital status.

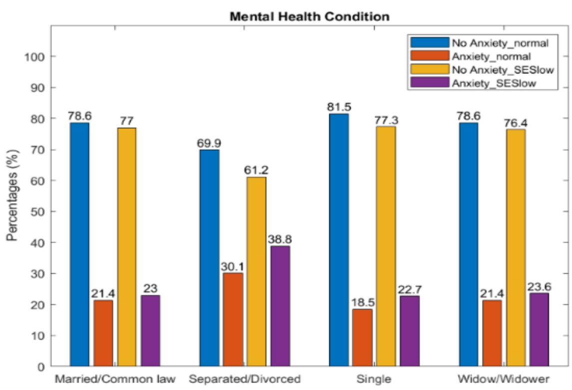

Fig. 4. Bar plot to show percentages of marital status in normal and social-economic deprivation area in terms of anxiety disorders diagnosis.

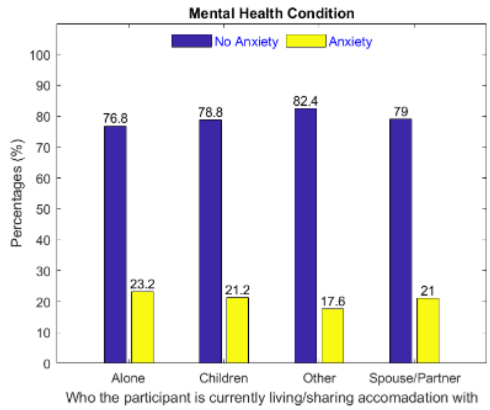

Fig. 5. Bar plot to show percentages for living/sharing accommodation status in terms of anxiety disorders diagnosis.

Fig. 4 clearly shows that no matter the marital status, participants who live in areas of high deprivation are at a higher risk of anxiety disorder. The results from this study are consistent with [25]. McCann et al. [25] presented a study on the TUDA dataset, which reported that socioeconomic differences were significantly related to anxiety in old age. From Fig.5, we can understand that relationships have an impact on the psychological health of older adults. Although it is not possible to assess the communication frequency, living with other people can provide emotional support for older adults, and the results of this study imply that the emotional support obtained from living with others can alleviate anxiety disorders in older people.

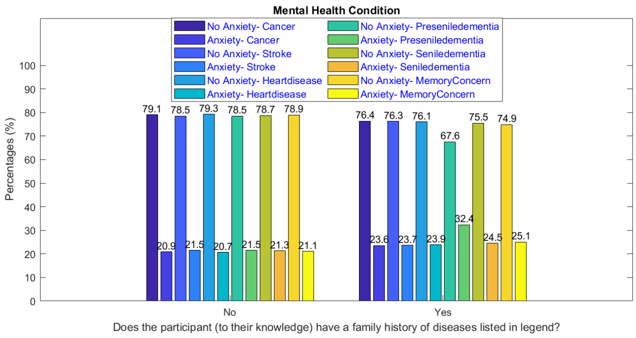

Fig. 6. Bar plot to show percentages of anxiety diagnosis for participants with family history related diseases.

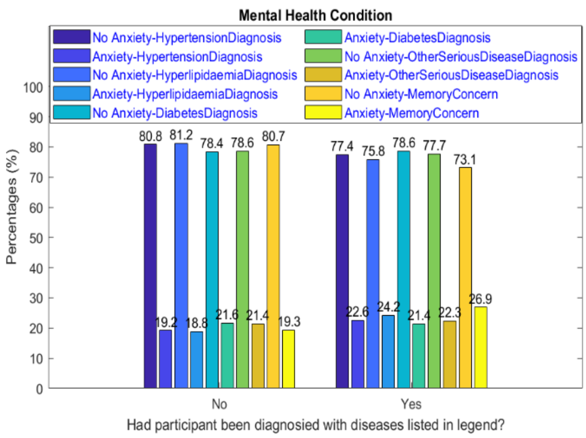

Fig. 7. Bar plot to show percentages of participants with anxiety diagnosis on different chronic diseases.

Multiple factors that are associated with anxiety disorder in older people were identified in [21]. Our analysis revealed that nearly 21.6% of participants had anxiety disorders, and the presence of family history of cancer, stroke, heart disease, presenile dementia and senile dementia diseases in family increased the risk of anxiety disorder (see Fig. 6). Older people often have chronic diseases, from Fig. 7 we can see those chronic diseases such as hypertension, hyperlipidaemia, other serious diseases like cancer contributed the risk of anxiety disorders.

People who suffer from other mental health conditions, such as depression, often also suffer from anxiety disorders. In the TUDA cohort, 1122 participants (21.64%) are diagnosed as anxiety disorders, 1247 (24.0%) participants are diagnosed as depression. 14.3% of all 5186 participants have both anxiety disorder and depression, 59.7% of the 1247 participants with depression diagnosed also have anxiety diagnosed.

Based on the aforementioned analysis, a higher risk of anxiety disorder is evident among participants who are living alone, being female, having experienced separation or divorce, having family history of chronic diseases, smoking, consuming alcohol, and having chronic diseases. The risk was higher in participants from areas with higher socioeconomic deprivation with a prevalence of 25.7% versus 20.3% in normal area.

IV. RESULTS

In the previous section risk factors analysis was conducted, based on the results features were selected and engineered. In this section supervised classification tasks on the pre-processed TUDA dataset that have all 5186 records, 83 predictors and one outcome variable were carried out as below to build models that can be used for diagnosis of anxiety disorder for the participants.

A. Preparation of training set and test set

The dataset was randomly split into two sets: a training set consisting of 70% of the data (3631 records), a test set consisting of 30% (1555 records) using the fully normalised and imputed dataset. Now the dataset is prepared to have models fitted. To ensure consistency in the comparisons across models, every model was fed the same training set and test set. The anxiety diagnosis, the outcome variable, was a variable that is related to diagnosis of anxiety disorders of the participants. There are 5186 participants, 4064 participants (78.36%) are diagnosed as non-anxiety disorders, and 1122 participants (21.64%) are diagnosed as anxiety disorders. It is a roughly 78:22 split in the class of the outcome variable.

B. Modelling

Given that a relatively small portion of participants have anxiety disorder, a challenge arises when attempting to model this imbalanced dataset. Various techniques can be employed to overcome this challenge, and in this study we compare the performance of single model SVM, ensemble learning models such as RF, and Gradient boosting (GB), both without sampling, and with two standard oversampling techniques: Synthetic Minority Oversampling Technique (SMOTE) [27] and ADASYN [28], which were used to generate synthetic records for the minority class in the training dataset. Please note test set are kept the same, are not oversampled so that the representativeness of the original population is retained and fair comparison with other methods can be maintained, reliable predictions could be provided on the unseen test set. ADASYN, an extension of SMOTE, can generate more samples in the vicinity of the boundary between the two classes than in the interior of the minority class. Table II lists the predictive performance of models for RF, SVM, GB, MLP and LR.

Upon completion of pre-processing and feature selection we have 83 predictors in the dataset. Next, principal component analysis (PCA) is applied. PCA is a dimensionality reduction technique that combines input variables in a way such that they explain the maximum variance of the data and the least important components can be dropped. Note that PCA retains the most valuable information from all the variables, but the interpretability of features is lost.

TABLE II. PERFORMANCE OF VARIOUS MODELS FOR ANXIETY DISORDER PREDICTION IN TUDA DATASET USING 83 VARIABLES .

Models	**Metrics**			**Sampling**
	Sensitivity (%)	***Specificity (%)***	***Accuracy (%)***	***(yes/No)***
RF	47.8	94.7	85.0	No
	67.0	90.3	85.4	Yes/SMOTE
	66.7	90.1	85.2	Yes/ADASYN
SVM	61.4	91.1	84.6	No
	81.5	83.9	83.4	Yes/SMOTE
	83.0	79.6	80.3	Yes/ADASYN
GB	53.1	90.7	82.4	No
	54.9	91.1	83.5	Yes/SMOTE
	55.6	89.6	82.5	Yes/ADASYN
MLP	57.4	88.2	81.8	No
	55.9	88.2	81.5	Yes/SMOTE
	53.7	90.2	82.6	Yes/ADASYN
LR	42.0	94.9	83.9	No
	79.0	83.2	82.3	Yes/SMOTE
	81.2	79.1	79.5	Yes/ADASYN

Fig. 8 illustrates all the 83 principal components (PCs); it can be seen from the plot that 35 principal components explain 95.2% of the variability in our data. Table III shows the results of the three models that have best performance with 35 PCs after features reduction using the PCA. Fig.9 shows the 20 most important features for classification of anxiety disorders using RF classifier with standard Classification and Regression Tree (CART) method. It is not surprised that the predictors related to depression diagnosis, HADS-anxiety questionnaire, anxiety medications, antidepressant medications and depression scale are in top five of the most important predictors.

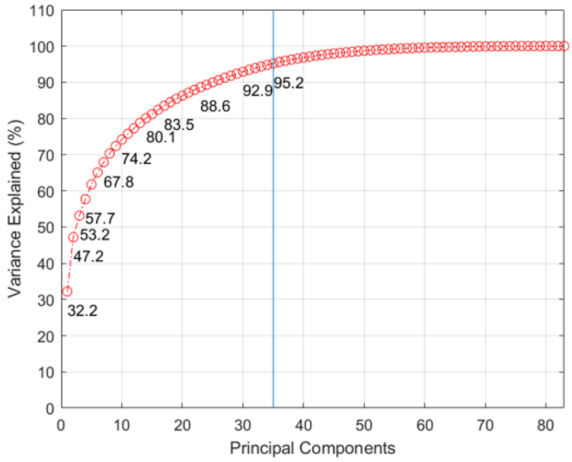

Fig. 8. Cumulated variance explained for principal components.

C. Discussion

In this preliminary analysis using data from the TUDA study, the results indicate that a variety of lifestyle, sociodemographic, biochemistry, clinical factors may aid to predict anxiety disorders in older people using machine learning techniques with a quite good level of accuracy (85.4% for RF model with SMOTE oversampling).

TABLE III. PERFORMANCE AFTER PCA FEATURES REDUCTION WITH 35 PCs—WITHOUT SAMPLING.

Models	**Evaluation Metrics**		
	Sensitivity (%)	***Specificity (%)***	***Accuracy (%)***
PCA+GB	53.7	90.7	83.0
PCA+RF	46.3	93.8	83.9
PCA+SVM	54.6	93.1	85.1

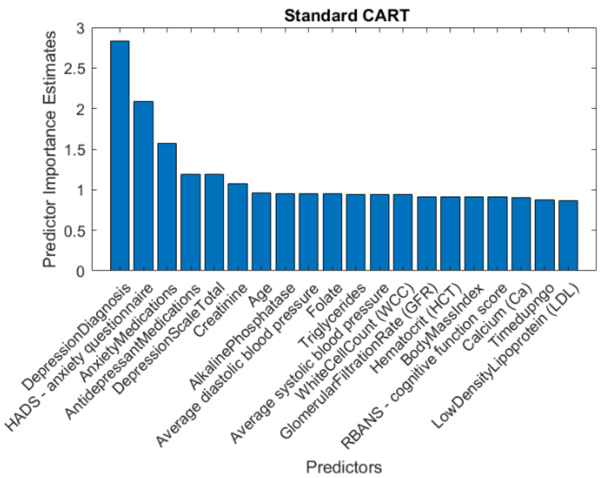

Fig. 9. The 20 most important features of RF models with standard CART methods.

The predictive performance of five ML algorithms in terms of accuracy to predict anxiety disorders in older people were compared on the TUDA dataset. The results showed that RF with SMOTE sampling had the best predictive performance. The results indicate that models that incorporate a combination of features that include life essential, lifestyle, biochemistry, clinical should be encouraged. Some of the potential risk factors contributing to an anxiety disorder diagnosis are identified. Understanding the risk factors for anxiety disorder in persons with specific chronic diseases can aid healthcare professionals immediately identify patients who are at risk, thus screening activities can be made better for psychological assessment and personalized treatments can be introduced in disease-specific care settings. In the long term, the impact of real-time feedback may be considered and even specific triggers that contribute to inappropriate and high levels of anxiety are identified.

The various machine learning approaches investigated in this study are identifying risk factors for anxiety that are well established in the literature. Older people who are in single and separated/Divorced status seem to be more prone to anxiety problems. In this study, living alone and areas of higher deprivation were identified as risk factors for anxiety alongside accommodation status, smoking, drinking alcohol, being female. Furthermore, the following variables have been identified important determinant of anxiety disorders consistent with evidence from other large cohorts studies, these include poor quality of life [25], functional and cognitive impairment [11], lack of formal education [11], low status of folate and metabolically related B vitamins, i.e., vitamin B12, vitamin B6, and riboflavin deficiencies have been identified as risk factors for anxiety disorder among older people [26], the role of fortified foods as a means of optimising B-vitamin status and potentially reducing the risk of these mental health disorders are also considered in [26]. It is interesting to know loneliness was also identified as a risk factor in [24], which is same as in our study despite differences in data sets and variables available for analysis from our study. The finding is particularly interesting that separated/devoiced marital status is an important factor related to anxiety disorders.

Machine learning algorithms such as SVM, RF, GB, MLP and LR are widely used and achieve good performance [24], and their performance depends on the quality and quantity of features used. Without oversampling on the training dataset, comparable performance (accuracy) is achieved on the results using 35 PCs after PCA feature reduction and that of using 83 predictors using GB (83.0%-82.4%) and SVM (85.1%-84.6%) respectively. While RF achieved an accuracy of 83.9% on 35 PCs, and accuracy of 85.0% on 83 predictors, The results of the classification presented here demonstrate that RF and SVM methods indeed provide useful tools for anxiety disorder prediction. PCA is capable of reducing number of variables and also achieve comparable performance.

V. Conclusion

Anxiety disorder is one of the leading health burdens in older people globally. Despite evidence that prevention and intervention can reduce the impact of anxiety disorder, global prevalence remains high and so intervention is crucial for people with such disorders [29-30].

In this preliminary study of the TUDA dataset, some of the key predictors were assessed that can predict anxiety disorders using the derived classification models with a satisfied level of accuracy. Variables such as gender, marital status, lifestyle-related, accommodation status, quality of life, chronic diseases and family related diseases are closely related to increased risk of anxiety disorders.

The results in [24] showed that the stacking ensemble model that combines several models to reduce variance and improve predictions, could have a better performance than that of the single predictive ML models depending on how a base model and a meta model are combined. The results were consistent to previous studies [22-23], which showed that the stacking ensemble model had a lower root-mean-square error (RMSE) than the single machine learning model. In this study, we used the TUDA dataset that is different from what was used in [22-24], hence in our preliminary analysis, we applied single base models for our analysis. In our future study, the stacking ensemble model would be explored by combining various base models and meta models including deep learning, and unsupervised learning.

We made an effect to evaluate risk factors influencing the anxiety of the older people. More studies are required to understand the characteristics of anxiety in older people. These factors will also be further explored in our future work.

Participants who suffer from other mental diseases such as depression normally have a higher risk for anxiety disorder. Future work should explore the relationship between anxiety disorder and depression. As older people often suffer from comorbidities, the impact of anxiety disorders on acute and long-term complications will be evaluated. Maintaining mental health not only improves daily functioning and relationships, enhances self-image, but also helps people address some physical health issues that are related to mental health conditions such as heart diseases. Supporting older people to take part in activities that reduce social isolation could also be beneficial in reducing the prevalence of anxiety [31]. Given the burden on health care resources, this may contribute to inclusiveness in policy development, and particularly in identifying public health strategies that promote better mental health and reduce inequalities.

Acknowledgment)

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

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, 398, pp. 1700–1712, 2021.

[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, 9, pp. 137-150, 2022.

[3] S. F. Javaid, I. J. Hashim, M.J. Hashim et al., “Epidemiology of anxiety disorders: global burden and sociodemographic associations,” Middle East Curr Psychiatry, 30, 44, 2023.

[4] Mental Health UK, “What is anxiety disorder?”, Retrieved from: https://mentalhealth-uk.org/help-and-information/conditions/anxiety-disorders/what-is-anxiety/ [Accessed 03/10/2023]

[5] Mind, “Mental health facts and statistics,” Retrieved from: https://www.mind.org.uk/information-support/types-of-mental-health-problems/statistics-and-facts-about-mental-health/how-common-are-mental-health-problems [Accessed 03/10/2023]

[6] J. Betts & J. Thompson, “Mental Health in Northern Ireland: Overview, Strategies, Policies, Care Pathways, CAMHS and Barriers to Accessing Services”, Northern Ireland Assembly Research and Information Service Research Paper, 2017.

[7] S.A. Lauderdale & J. L. Sheikh, “Anxiety disorders in older adults”. Clin Geriatr Med, 19(4), pp. 721-41, Nov. 2003.

[8] J. I. Sheikh, “Investigations of anxiety in older adults: recent advances and future directions”, J Geriatr Psychiatry Neurol, 18(2), pp. 59-60, Jun. 2005.

[9] C. Andreescu & D. Varon, “New research on anxiety disorders in the elderly and an update on evidence-based treatments”, Curr Psychiatry Rep. 17(7), 53, Jul. 2015.

[10] R. Z. Ishikawa, C. Vyas & O. Okereke, “Anxiety Disorders Among Older Adults: Empirically Supported Treatments and Special Considerations”, In: Bui, E., Charney, M., Baker, A. (eds) Clinical Handbook of Anxiety Disorders. Current Clinical Psychiatry. Humana, Cham, 2020.

[11] D. Rankin, M. Black, B. Flanagan, C. F. Hughes, A. Moore, L. Hoey, J. Wallace, C. Gill, P. Carlin, A. M. Molloy, C. Cunningham, H. McNulty, “Identifying Key Predictors of Cognitive Dysfunction in Older People Using Supervised Machine Learning Techniques: Observational Study”, JMIR Med Inform, 16; 8(9):e20995, Sep. 2020.

[12] P. Fusar-Poli, C. U. Correll, C. Arango, M. Berk, V. Patel, J. P. Ioannidis, “Preventive psychiatry: A blueprint for improving the mental health of young people. World Psychiatry”, 20, 200–221, 2021.

[13] D. McDaid, A. L. Park, K. Wahlbeck, “The Economic Case for the Prevention of Mental Illness”, Annu. Rev. Publ. Health, 40, pp. 373–389, 2019.

[14] D. J. Schofield, R. N. Shrestha, R. Percival, M. E. Passey, E. J. Callander, S. J. Kelly, “The personal and national costs of mental health conditions: Impacts on income, taxes, government support payments due to lost labour force participation”, Bmc Psychiatry, 11, 72, 2011.

[15] T. Hahn, “Nierenberg, A. & Whitfield-Gabrieli, S. Predictive analytics in mental health: applications, guidelines, challenges and perspectives”, Mol Psychiatry 22, pp. 37–43, 2017.

[16] A. F. Jorm, S. B. Patten, T. S. Brugha, R. Mojtabai, “Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries”, World Psychiatry, 16, pp. 90–99, 2017.

[17] P. R. Jain, S. M. K. Quadri, “Emerging Role of Intelligent Techniques for Effective Detection and Prediction of Mental Disorders”, In: Hemanth, J., Bestak, R., Chen, J.IZ. (eds) Intelligent Data Communication Technologies and Internet of Things. Lecture Notes on Data Engineering and Communications Technologies, vol 57. Springer, Singapore, 2021.

[18] A. J. Meehan, S. J. Lewis, S. Fazel, et al., “Clinical prediction models in psychiatry: a systematic review of two decades of progress and challenges”, Mol Psychiatry, 27, pp. 2700-2708, 2022.

[19] S. Graham, C. Depp, E. E. Lee, et al., “Artificial Intelligence for Mental Health and Mental Illnesses: an Overview”, Curr Psychiatry Rep, 21, 116, 2019.

[20] M. Cearns, T. Hahn, & B. T. Baune, “Recommendations and future directions for supervised machine learning in psychiatry”, Transl Psychiatry, 9, 271, 2019.

[21] I. Ancillon, M. Elgendi, C. Menon, “Machine Learning for Anxiety Detection Using Biosignals: A Review”, Diagnostics, 12(8), 1794, 2022.

[22] M. D. Nemesure, M. V. Heinz, R. Huang, N. C. Jacobson, “Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence”, Sci Rep., 21;11(1):1980, Jan. 2021.

[23] Z X. Shen, L. J. Cui, S. Q. Mou, L. Ren, Y. G. Yuan, X. H. Shen, G. Gang, “Combining S100B and Cytokines as Neuro-Inflammatory Biomarkers for Diagnosing Generalized Anxiety Disorder: A Proof-of-Concept Study Based on Machine Learning”, Frontiers in Psychiatry, Vol. 13, 2022.

[24] H. Byeon, “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, 18, 7625, 2021.

[25] A. McCann, H. McNulty, J. Rigby, C. F. Hughes, L. Hoey, A. M. Molloy, C. J. Cunningham, M. C. Casey, F. Tracey, M. J. O'Kane, K. McCarroll, M. Ward, K. Moore, J. J. Strain, A. Moore, “Effect of Area-Level Socioeconomic Deprivation on Risk of Cognitive Dysfunction in Older Adults”, J Am Geriatr Soc., 66(7), pp. 1269-1275, Jul. 2018.

[26] K. Moore, C. F. Hughes, L. Hoey, M. Ward, C. Cunningham, A. M. Molloy, J. J. Strain, K. McCarroll, M. C. Casey, F. Tracey, E. Laird, M. O'Kane, H. McNulty, “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, 20(5), pp. 551-557, May 2019.

[27] N. V. Chawla, K. W. Bowyer, L. O. Hall & W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research”, 16, pp. 321-357, 2002.

[28] H. He, Y. Bai, E. A. Garcia & S. Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning”, In 2008 IEEE International Joint Conference on Neural Networks, pp. 1322-1328, 2008.

[29] C. Moreno, T. Wykes, S. Galderisi, M. Nordentoft, N. Crossley, N. Jones, M. Cannon, C. U. Correll, L. Byrne, S. et al., “How mental health care should change as a consequence of the COVID-19 pandemic”, Lancet Psychiatry, 7, pp. 813–824, 2020.

[30] J. Campion, A. Javed, N. Sartorius, M. Marmot, “Addressing the public mental health challenge of COVID-19”, Lancet Psychiatry, 7, pp. 657–659, 2020.

[31] A. J. Baxter, K. M. Scott, T. Vos, H. A. Whiteford, “Global prevalence of anxiety disorders: A systematic review and meta-regression”, Psychol. Med., 43, 897, 2013.