1) First of all, cross-sectional data cannot be used to develop prediction model this is because the predictors and outcomes to be predicted are measured at the same time-point but in prediction studies, we use baseline factors to predict an outcome that occurs in a future time-point.

Reply: In this study, we divided the data into training set and prediction set according to 8:2, and adopted a variety of machine learning methods for modeling, which is a mature and effective prediction model building method. At present, many prediction model studies still use cross-sectional questionnaire data, which does have some limitations, so we will do further research. This limitation will be fully explained in the conclusion.

2) Second, the authors used terms "anxiety", "depression", and "dementia" to denote diagnoses of mental disorders but in the methodology, they only assessed anxiety symptoms, anxiety symptoms, and cognitive impairments. The authors are wrong in these psychiatric terms.

Reply: We have revised this issue in the manuscript to achieve context unity.

3) Third, the authors used the term "mental health" to denote "anxiety", "depression", and "dementia", however, in psychiatry and mental health, dementia belongs to cognitive function, not a typical mental health outcome.

Reply: First, current mental health standards and specific disorders are not clearly defined, covering a wide range of issues, from the familiar anxiety and depression, to sleep disorders, mental disorders, as well as cognitive impairment and loneliness. In 2018, the Aging Health Service and Standardization Branch of the Chinese Geriatric Medical Research Society defined the mental health standards of the elderly in China. Mental health refers to the internal psychological harmony and consistency of the individual, and the stable mental state that is well adapted to the external. It includes five dimensions: cognitive efficacy, emotional experience, self-knowledge, interpersonal communication and adaptive ability. Among them, the elderly can maintain the basic daily cognitive function as the first criterion, is an important link to ensure the quality of life. Second, cognitive disorders and depression in the elderly are particularly easy to be confused, difficult to differentiate diagnosis, and often exist in the form of comorbidity. Finally, severe cognitive disorders will appear mental symptoms, known as dementia mental behavioral disorder, specific manifestations of hallucinations, delusions, depression, anxiety, agitation, personality change, early screening and intervention can reduce and avoid such problems. To sum up, we regard cognitive screening as an important item in mental health screening.

4)   Fourth, there have been time-efficient and relatively accurate self-report measures to screen for anxiety, depression, and cognitive impairments in the elderly population, such as the GAD7, PHQ9, and AD8; these scales can also be easily administered by primary care providers. So it is not challenging for these early identification of anxiety, depression, and cognitive impairments by using these simple measures. The rationale and clinical needs of this study, as described in the introduction of this study, are not convincing. Due to these methodology and professional problems, the paper cannot be corrected via extensive revisions.

Reply: We do have some convenient screening tools, such as GAD7, PHQ9, and AD8, but they are not widely used in Chinese community work. First, even with such a simple screening tool, the total screening time is 8 to 10 minutes, which is not easy for busy medical staff in developing countries to carry out. Second, these questionnaires still need to be completed by more professional doctors, but at present, the number of psychologists in China is extremely rare, unable to complete extensive screening work. Thirdly, in the actual survey, we found that the elderly have comprehension obstacles, or are too shy to talk about some of the problems, avoid questions or cannot give real answers, which is easy to miss diagnosis. Therefore, the prediction model we developed has only 8 questions, among which there are no overly sensitive questions. It is simple and easy to operate, and is widely applicable to the elderly of all ages.

Reviewer B

Data source is very important for machine learning research. In this study, the author uses many classic machine learning methods to establish prediction model, which has great use value in the field of elderly health. The article is good as a whole, with strong innovation and practicality. I only have a few questions:

1. Please supplement the inclusion and exclusion criteria of study subjects;
2. Please further elaborate on how to collect the questionnaire and complete the standardized quality control.
3. It is recommended that the significance of the important features of model screening be described in more detail in the discussion section.

Reply:
Thank you very much for your valuable advice, we have modified our text as advised
1. We have added Inclusion criteria and exclusion criteria, see Page 6, line 193-199)
2. We have added Investigation methods and standardization strategies for project implementation, see Page 6-7, line 199-208)
3. We have added the significance of the important features of model screening, see Page 15-16, line 505-510)

Reviewer C

The psychological problems of the elderly are becoming increasingly prominent, and it is necessary to have a model that can be used to predict psychological diseases. In this study, a

new model based on machine learning is proposed, which can be used to predict psychological problems such as anxiety and depression of the elderly. The sample size is sufficient, the manuscript is well written, and the statistics are appropriate. I have no new questions.

Reply: *Thank you very much for your recognition.*

Reviewer D

1. Authors should also state that the study conformed to the provisions of the Declaration of Helsink.

Reply: *Added.*

2. Please check all abbreviations in the main text, such as below. All abbreviated terms should be full when they first appear.

> 243   algorithms with the best ensemble learning performance, Random Forest, XGBoost
> 244   and LightGBM, were used to model the above data respectively, and then some
>
> 429   ROC curve is also called "receiver operating characteristic curve". The horizontal
> 430   coordinate of the curve is false positive rate (FPR) and the vertical coordinate is true

Reply: *We have revised the relevant content in the original article.*

3. Please check whether the full name of "PDPs" is correct.

> 450   the model's performance was studied. Partial correlation graphs (PDPs) (33,34) can
> 451   be used to analyze the importance of individual features and marginalize other

Reply: *We have revised the full name of "PDPs" in the original article.*

4. Please add reference.

> 248   Overview of Random Forest: Random Forest model (RF) was proposed by Leo
> 249   Breiman in 2001. Based on Bagging integrated learning method, RF can be used for

Reply: *We added the citation of reference in the original article.*

5. Please check if any more references need to be added in the below sentence since you mentioned "Studies", but only one reference was cited.

> 505   In addition to achieving high accuracy, our method was highly applicable Studies
> 506   have shown that age is related to the probability of mental diseases in older people
> 507   (35). To further prove that our model could be used for older people of different ages,

Reply: *We have revised the relevant content in the original article.*

6. Please check if the citations of references need to be added in the below 2 sentences since you mentioned the previous "studies".

132   interventional, and can be more easily changed. Numerous studies have shown that
133   old age is not a linear, gradual stage of decline but a malleable stage of development.
134   Hierarchical prevention and related intervention training can restore, maintain, and

253   prediction result through voting. A large number of theoretical and empirical studies
254   have proved that random forest has high prediction accuracy, good tolerance to
255   outliers and noise, and is not easy to overfit. In the training stage, random forest uses

Reply: *We have revised the relevant content in the original article.*

7. Figure 4:
1) Figure 4 not clear enough. Please send it in higher resolution.
2) Please check whether "auROC" should be "AUC". Please revise.

Reply: *Since the picture was compressed when inserted into the word document, it was not clear, so we further clarified the picture and submitted the original picture as an attachment. We also corrected "auROC" marked in the figure to "AUC".*