Response to Reviewer 1 Comments

Point 1: Thanks for the work to improve the article.

However, the article is still very weak in discussing relevant literature, explaining research design, particularly the results are not presented in the standard format of the sensor journal.

Response 1:

Authors have added more information relating relevant literature within introduction and discussion sections.

Related to research design, in Section 2.2.1, related to the data acquisition process, the first paragraph has been modified as follows:

All included patients performed the full battery of ocular movement tests selected by protocol [referencia de paper de sanos]. This protocol was design to characterise ocular movement of voluntary participants with no previous cognitive impairment in order to establish a control group. In this case, the biomedical procedure established the following steps:

1. Revision of the patient's condition.
    a. PHES evaluation and clinical diagnosis.
    b. Make-up and any kind of lenses removal is asked to the participants before ocular movement tests in order to guarantee the precision and accuracy of the eye tracking algorithm.
2. Ocular movement tests: Each test must be clearly explained to the patient and there is a demo version available if needed.
    a. Visually guided saccades tests (VGST).
    b. Antisaccades tests (AST).
    c. Memory guided saccades tests (MGST).
    d. Smooth pursuit tests (SPT).
    e. Fixation test (FIXT).

Each participant in the experiment followed this protocol based in the analysis and design of a group of specialists that tries to guarantee the attention of the participant. A simple test, VGST, is used to present the experiment. Then, most difficult tests are captured (AST and MGST) and, finally, the easiest ones are presented to end up the ocular movement experiment.

Results are re-structured to be aligned with the instructions for authors. New index is:

1. Classification results.
2. Ocular features description and significance.
3. Limitations: samples with missing data.

Also, the interpretation of these results is gathered in this section.

Point 2: This article is missing information on data preparation for ML (i.e., cross-validation), this is a major flaw.

Response 2: We have added more information on how ocular features are configured at the end of the feature extraction process just before the Section 2.2.3 Classification algorithm and validation:

To summarise, at the end of this process the number of samples remains constant, it is equal as the number of participants in the experiment. Each sample length is reduced from 160 ocular

features as established in Table 5 to the number of features that get over the threshold of the significant analysis. Table 7 shows the number of features per test that gets over each threshold.

Also, data preparation regarding cross validation is also detailed from:

The best classifiers were selected depending on accuracy, area under the ROC curve 298
(AUC) and the number of features used. After that, cross-validation was done using five 299
folds or subsets. For additional details of ROC, please check [47]. 300
Therefore, the training data set and test data set are evaluated and the probability of 301
belonging to one of both classes is calculated via predicted class scores. More information 302
can be checked in [48]. 303
This factor allows to know if a sample is in the borderline between sets or, even worse, 304
if it is classified in the wrong set. 305
In this paper, we used cross validation method where the data set is aleatory divided 306
into k subsets. While ki subsets with i = 1, 2...n − 1 are used for training, kn subset is used 307
for testing the algorithm. The error is computed as the mean value of the errors in each 308
iteration. 309
In order to compute the significance and statistical accuracy of the procedure, the 310
selected classifier algorithm was executed in an iterative loop which corresponds with the 311
two last stages shown in the flowchart of Figure 2. 312

To:

The best classifiers were selected depending on accuracy, area under the ROC curve (AUC) [47] and the number of features used. After that, cross-validation was done using five folds or subsets. One of these folds, which corresponds to 20% of the available samples, is used for testing, while the four remaining were used for training. The samples of each class were equally divided in the different folds.

In this study, we used cross validation method where the data set is aleatory divided into 5 subsets following a stratified k-fold division strategy. In order to compute the significance and statistical accuracy of the procedure, the selected classifier algorithm was executed 1000 times in an iterative loop which corresponds with the two last stages shown in the flowchart of Figure 2. All result metrics are computed as the mean value of the errors in each iteration, Figures 7 and 8 also shows the distribution over these iterations.

Point 3: Experimental setup should be mentioned clearly as well.

Response 3: The experimental setup of the study is clearly described in Section 2.1.1 with a very detailed description. However, we have identified some missing elements about protocol description. In Section 2.2.1, related to the data acquisition process, the first paragraph has been modified as follows:

All included patients performed the full battery of ocular movement tests selected by protocol [referencia de paper de sanos]. This protocol was design to characterise ocular movement of voluntary participants with no previous cognitive impairment in order to establish a control group. In this case, the biomedical procedure established the following steps:

3. Revision of the patient's condition.
    a. PHES evaluation and clinical diagnosis.
    b. Make-up and any kind of lenses removal is asked to the participants before ocular movement tests in order to guarantee the precision and accuracy of the eye tracking algorithm.
4. Ocular movement tests: Each test must be clearly explained to the patient and there is a demo version available if needed.
    a. Visually guided saccades tests (VGST).
    b. Antisaccades tests (AST).
    c. Memory guided saccades tests (MGST).
    d. Smooth pursuit tests (SPT).
    e. Fixation test (FIXT).

Each participant in the experiment followed this protocol based in the analysis and design of a group of specialists that tries to guarantee the attention of the participant. A simple test, VGST, is used to present the experiment. Then, most difficult tests are captured (AST and MGST) and, finally, the easiest ones are presented to end up the ocular movement experiment.

Point 4: Moreover, how this study allows to draw a scientific conclusion is not clear.

Response 4: We appreciate all the comments done by the Reviewer, all of them have been contributed to improve the article. Besides the full organization of the article, including the results section following all the reviewer recommendation, we also improve the discussion section with additional bibliography and comments. We think that the article is clearly full in scope with the topic of the special issue that look for new approaches for diagnosis of diseases based on innovative technologies and methods. Indeed, our proposal of MHE diagnosis by combining eye movements measures with machine learning tool is novel, feasible and reliable tool.

Point 5: There are lot of necessary and necessary plots in the paper without proper story which generate little interest among the audience to read the paper. Less important table and image should be moved to supplementary materials.

Response 5:

Figures 7 and 8 have been grouped in just one.

Figure 9 (ROC figure) have been deleted from the paper.

Figures 10 and 11 have been grouped in just one.

The story that figures and tables tell is:

Flowchart of the full experiment, visual representation of ocular movement tests, statistical results of the classification task in two figures (AUC and Accuracy) and visual representation of the differences between the classes (posterior probabilities and a scatter plot with two of the most significance ocular movement features) to support the discussion.

The different tables of the paper act as a summary of the different contents or results.

Point 6: Authors should re-structure the article to make it suitable for Sensor journal and highlight novelty of their work in respect to literature clearly in introduction and discussion section.

Response 6: We have identified some weak points on how the discussion is organised, so it is re-structured as asked. Here we propose this schedule to organise the different topics:

1. Summarization of the problem
2. Discussion of the main findings, classification results.
3. Missing data discussion
4. Limitations of the study

We also have improved the contents of introduction and discussion Section.

Point 7: Mention limitation of their work as well.

Response 7: Discussion Section is re-structured. The information related to the limitation of the work is between lines 405-432.