Peer Review File

Article information: https://dx.doi.org/10.21037/jmai-22-94

Reviewer Comments

Comment 1

The title needs to indicate the research design of this study, i.e., a narrative review.

Reply 1

We amended the title accordingly and changed it to "Machine learning models for automated interpretation of 12-lead electrocardiographic signals: A narrative review of techniques, challenges, achievements and clinical relevance."

Changes in text

We replaced the title in all relevant parts of the submission.

Comment 2

The abstract needs to briefly describe the literature search strategies and results of the search. The authors need to summarize the findings on the diagnostic accuracy parameters of the ML models, clinical diagnoses of cardiac diseases involved in the available studies, and datasets used. The conclusion needs comments on how to improve the generalizability of the available studies.

Reply 2

Many thanks for these helpful recommendations. We rephrased the technical term "predictive accuracy" into the more clinically relevant and comprehensible "diagnostic accuracy", while retaining the associated score, to add clarity. We also added all the missing information (regarding methods, the diagnostic labels, included datasets and proposed solutions to problems mentioned in the Conclusions), and also reformatted the abstract according to the journal's standards.

Changes in text

Lines 33-58 in Abstract (changes with tracked changes).

Comment 3

The introduction needs to briefly review on the history of ML, DL, and related methods in medical practice.

Reply 3

We added appropriate content in the Introduction, relevant to ML and DL history and related applications, generally in medical practice, so as to narrow down to ECG analysis in the following introductory paragraph.

Changes in text

Lines 84-94 ("Machine learning... ...the most basic ML...")

Comment 4

In the main text of the review, please describe the findings on the external validity and accuracy of the ML models and have comments on these findings. It is bad to let the authors to read the tables without any descriptions and analyses.

Reply 4

Many thank for these valid points. We expanded more on the accuracy and other metric scores of the different models on ECG diagnostic tasks, based on the table-included findings. We reformed the text accordingly.

Changes in text

Lines 219 ("PPV and NPV, respectively"), 220-221 ("namely the percentage of correct to overall attempted diagnoses"), 223-227 ("While the mean specificity... ...and 85.7%, respectively").

Comment 5

Please discuss the potential reasons for the limited accuracy and poor external validity of the ML models including factors affecting the accuracy.

Reply 5

Aligning with your comments, we emphasized the reasons for low diagnostic accuracy and generalizability in this field, and added relevant content to the Discussion.

Changes in text

Lines 293-304 ("One of the main... ...problem of low generalizability...")