Response to the reviewers and the chief editor

Thank you so much for your comments and giving us the opportunity to revise our paper!

Please do not hastate to give us another opportunity to revise our paper if needed

All the changes marked in red color and automatic track changes

Response to **Reviewer #1:**

I had the opportunity to review the article titled "Discriminating Type 1 and Type 2 Diabetes from Single-Lead ECG Using Machine Learning and Multinomial Regression." I congratulate the authors on their meticulous work. Overall, the article is well-written, but I have some suggestions for revision. These suggestions are listed below.

• Abstract The running title does not reflect the title of the article. Changing the running title might be a good idea. Keywords like ‘’Prevalence, Cardiovascular disease, and Metabolic syndrome” can be deleted. Too many keywords have been typed.

Thank you so much for your comments: by your guidance we made the following changes; We have modified the running title to fit the topic of the paper and removed the unnecessary keyword.

• Introduction The Introduction section is well-organized. The literature support, the purpose of the article, and any gaps in the literature are well explained. I have no additional suggestions or criticisms for this section.

Thank you so much for your comments.

Materials and Methods The biggest shortcoming of this section is that the diagnostic criteria for DM are not specified. Furthermore, the classification of patients diagnosed with DM is not specified.

Thank you so much for your comments: by your guidance we made the following changes; we added a detailed information on the threshold for the diagnosis of DM and elaborated on the group patients who

The results share the prevalence of coronary artery disease. However, the methodology does not specify how CAD is diagnosed. A section titled "Definition" should be created, and these definitions should be added.

Thank you so much for your comments: by your guidance we made the following changes; we added the criteria for the diagnosis of the coronary artery disease. The diagnosis of coronary artery disease based on the anamnesis and medication intake for it is management.

The study describes the use of artificial intelligence to diagnose and even type DM. However, this section should explain the AI's algorithm for making this distinction.

Thank you so much for your comments: by your guidance we made the following changes; It's a multiclass log regression with elastic net regularization. To assess associations with a quantitative outcome (glycemia), a linear regression model with combined L1 and L2 (Elastic Net) regularization was employed. This approach significantly mitigated the risk of multicollinearity and facilitated the selection of factors contributing most substantially to the outcome's variance. Quantitative data were normalized prior to analysis. The model's performance was evaluated on the validation set using the coefficient of determination (R^2), mean absolute error (MAE), maximum error, and the correlation coefficient between predicted and true values.

Furthermore, the potential pathophysiological explanation should also be stated. Adding the Clinical trial number URL to the methodology may make it easier for the reader.

Thank you so much for your comments: by your guidance we made the following changes; we added to the discussion section potential pathophysiological explanation on how diabetes effects the heart tissue and can show on the single lead ECG as well as a graphical representation. We highlighted the clinical trial number and added the link to the paper.

I think that the statistical methods used in the post-hoc analyses of ANOVA and Kruskal-Wallis tests should be specified.

Thank you so much for your comments: by your guidance we made the following changes; we are pointing to you that it is described in the Statistical Data Analysis section.

For related (paired) samples, comparative analysis of normally distributed quantitative variables was conducted using a paired Welch's t-test (for two groups) or repeated-measures ANOVA (for more than two groups), again followed by post-hoc pairwise comparisons. For related, non-normally distributed quantitative variables, the Wilcoxon signed-rank test (for two groups) or Fisher's test (for more than two groups) was used.

To analyze dynamic changes across visits, Cochran's Q test was applied. Where necessary, pairwise comparisons for these dynamic changes were performed using McNemar's test.

The Holm-Bonferroni method was used to correct for multiple comparisons. A significance level (α) of 0.05 was adopted for all analyses, corresponding to a 5% probability of committing a Type I error.

• Results Table 1 needs significant revisions. There are numerous abbreviations, none of which have explanations. Only p-values are listed, but no information is provided on the nature of the differences between groups. I believe Table 1 needs to be rewritten. To be honest, I didn't understand anything from Table 1.

Thank you so much for your comments: by your guidance we made the following changes; table 1 shows the mathematically obtained parameters from the QardioQVARK device. These parameters obtained after performing some mathematical formulas using the single lead ECG by integration of machine learning. These parameters are more important for the mathematical calculations and it is important the machine learning model does not know about the parameter to avoid bias. All the list of the parameters ranges from 47-48, and for our study only 25 parameters showed importance. The whole version of these parameters published in previous work https://pmc.ncbi.nlm.nih.gov/articles/PMC9421020/ .

Figure 1 illustrates the differences between groups. I prefer using precise numbers in the table rather than this diagram. Exact numbers should be presented rather than large and small figures and estimates.

Thank you so much for your comments: by your guidance we made the following changes; for better presentation and easy understanding, our head biostatistics suggest the graphical presentation for the results. And we hope on your understanding and give a positive comment regarding this moment.

An examination of Table 2 reveals that the regression analysis results are presented. However, the table's format is not very clear. I strongly recommend revising the tables.

Thank you so much for your comments: by your guidance we made the following changes; we have revised the table and revised the whole paper for clarity issues.

Abbreviations should definitely be added with their full names below the tables. This article's format presents an image that will be overwhelming for the reader.

Thank you so much for your comments: by your guidance we made the following changes; we have added link to the abbreviation in the table and the figures, where the reader can find the whole mathematically obtained parameters from the single lead electrocardiography based on the machine learning integration.

There's another shortcoming in the presentation of the results. This section lists the relevant clinical pathology and ECG differences between the groups. However, no reason is given.

Thank you so much for your comments: by your guidance we made the following changes; we added to the discussion section a potential explanation on the relevant clinical pathology and ECG differences between the groups

The presentation of the results appears confusing. This may be due to the sheer number of results presented. However, I must point out that the article, in its current form, appears very confusing and doesn't provide a clear flow for the reader.

Thank you so much for your comments: by your guidance we made the following changes; we tried our best to improve the paper presentation and be more readable and easier to digest for the readers.

Placing the headings of Table 3 in the row on the left instead of the column above may make the article more understandable for the reader.

Thank you so much for your comments: by your guidance we made the following changes; we tried our best to enhance the table readability.

• Discussion The study should explain the relationship between DM and changes detectable on a single-lead ECG. This pathophysiological explanation needs to be detailed. I believe the existing explanations are inadequate. Reading the study in this way will raise many doubts in the reader. Overall, the article is very well designed, but I think the writing and presentation of the results are not very good.

Thank you so much for your comments: by your guidance we made the following changes; we added a detailed information and figure representing the potential pathophysiological pathway of diabetes mellitus in the single lead electrocardiography alterations.

Response to **Reviewer #2:**

This study investigates the use of single-lead ECG and machine-learning models to detect and classify diabetes mellitus (type 1, type 2, and non-diabetic states). The topic is novel and clinically relevant, yet several critical aspects of study design, data description, and presentation require clarification before the findings can be properly interpreted :

1. Diagnostic Standards for Diabetes Mellitus The manuscript does not specify how diabetes and its subtypes were diagnosed or confirmed. It is essential to state whether the diagnostic criteria followed ADA, WHO, or national guidelines and which tests (fasting plasma glucose, OGTT, HbA1c, medical record diagnosis) were used. Without these details, the reliability of the diagnostic labels used for model training and validation cannot be assessed.

Thank you so much for your comments: by your guidance we made the following changes; we added to the material and methods section the diagnostic threshold for DM.

2. Enrollment Strategy The inclusion process remains unclear. Please indicate whether all eligible patients were consecutively enrolled during the study period or if the sample was selected or filtered (e.g., only those with complete ECG and glucose data). This information is critical to evaluate representativeness and potential selection bias.

Thank you so much for your comments: by your guidance we made the following changes; patient who have a complete ECG recording with interpretable quality and glucose data are included according to the inclusion and exclusion criteria.

3. Non-Inclusion vs. Exclusion Criteria The current distinction between "Non-Inclusion" and "Exclusion" criteria is unnecessary and confusing, as both indicate reasons for not including participants. These should be merged into a single Exclusion Criteria section to avoid redundancy.

Thank you so much for your comments: by your guidance we made the following changes; we combined them to be, exclusion criteria.

4. Section Structure The paragraph beginning with "The patients underwent a comprehensive examination, including:" follows immediately after the exclusion list, which makes it appear to be part of the exclusion criteria. Adding a clear sub-heading such as "Clinical Assessment" or "Baseline Evaluation" would improve readability and logical flow.

Thank you so much for your comments: by your guidance we made the following changes; we added clinical subheading "clinical data collection.

5. Title Accuracy vs. Model Functionality The title "Discriminating Type 1 and Type 2 Diabetes from Single-Lead ECG Using Machine Learning" implies that the model distinguishes between T1DM and T2DM. In fact, the model primarily discriminates between diabetic vs. non-diabetic states, with limited

performance in differentiating T1DM from T2DM (AUC = 0.69). Consider revising the title and abstract to reflect the true scope and strength of the model.

Thank you so much for your comments: by your guidance we made the following changes; we revised the tittle and the abstract and conducted some modifications.

6. Timing of ECG Acquisition It is unclear whether ECG recordings were obtained before or after diabetes diagnosis. If ECGs were collected after confirmed diagnosis, the model reflects secondary electrophysiological changes caused by established disease rather than early screening potential. This limitation should be clarified and discussed.

Thank you so much for your comments: by your guidance we made the following changes; the ECG recording was performed after the diagnosis have been made. We mentioned that issue in the discussion section.

7. Potential Confounding Cardiac Conditions The study does not address whether patients with arrhythmias, ischemic heart disease, cardiomyopathies, or other ECG-altering disorders were excluded or adjusted for. These comorbidities can strongly influence ECG features and may confound model performance. Please clarify how such conditions were handled.

Thank you so much for your comments: by your guidance we made the following changes; according to the exclusion criteria, such pathologies are not excluded only individuals with significant QRS complex morphology changes (such as bundle branch block and ventricular extrasystole)

8. Structure of Discussion and Conclusion The last paragraph of the Discussion begins with "In conclusion…" and repeats the statements in the formal Conclusion section. This redundancy can be avoided by merging or re-phrasing the final paragraph so that the main conclusions are only presented once in the designated Conclusion section.

Thank you so much for your comments: by your guidance we made the following changes; we revised the final paragraph in the discussion section.

Response to reviewers and chief editor

Response to the reviewer comments

The authors have addressed several structural and presentation-related issues in the revised manuscript, including consolidation of inclusion and exclusion criteria, addition of methodological subheadings, clarification of the timing of ECG acquisition, and moderation of statements related to “early screening.” These changes have improved the overall readability and internal consistency of the manuscript. The research topic remains of interest and has potential clinical value, particularly in the context of combining wearable single-lead ECG technology with machine-learning approaches for diabetes identification.

However, from the perspective of methodological rigor and interpretability of the results, the current revision does not yet adequately resolve several key concerns. These issues directly affect the reliability of model training labels, the representativeness of the study population, and the independence of the association between ECG features and diabetes status. Without further clarification or additional analyses, the manuscript is not yet suitable for acceptance.

1. Diagnostic and classification criteria for diabetes mellitus remain insufficiently defined
Although the revised manuscript provides additional information on glucose measurement, the diagnosis and classification of diabetes mellitus still rely primarily on capillary glucose thresholds combined with medical history and treatment information. It remains unclear whether internationally accepted standards (e.g., ADA or WHO criteria) were followed, and no systematic description is provided regarding the use of HbA1c, OGTT, C-peptide, or autoantibody testing for subtype classification. This approach may introduce label noise and misclassification, thereby directly affecting the reliability of machine-learning model training and validation.
It is recommended that the Methods section clearly specify the diagnostic and classification guidelines used and the biochemical criteria applied. If standardized diagnostic or classification criteria were not consistently implemented or cannot be supplemented, the resulting classification bias should be explicitly acknowledged and discussed in the Limitations.

Thank you so much for your comments. With your guidance we made the following changes: we added to the material and methods section a clarification on how the DM was diagnosed.

2. Enrollment strategy and potential selection bias remain unclear
The revised manuscript does not clearly state whether participants were enrolled consecutively during the study period or selected based on data completeness and signal quality. In addition, no information is provided regarding the numbers and reasons for exclusion at each stage of screening, inclusion, and analysis. The absence of this information prevents assessment of sample representativeness and the direction of potential selection bias.
It is recommended that the authors clarify in the Methods whether enrollment was consecutive and specify the clinical settings from which participants were recruited. A flow diagram detailing screening, inclusion, and exclusion (with corresponding numbers and reasons) should be provided. If feasible, a comparison of baseline characteristics between included and excluded participants (as supplementary material) would further help assess selection bias. If these additions are not possible, the limited representativeness of the study population should be explicitly stated in the Limitations.

Thank you so much for your comments. By your guidance we made the following changes; The study was conducted at University Clinical Hospital Number 2 between 02/05/2022-23/08/2025. Patients were recruited from both outpatient and inpatient settings, including those undergoing examination in the endocrinology department. Patients were consecutively recruited from University Clinical Hospital Number 2 during routine outpatient visits and inpatient admissions between 02/05/2022-23/08/2025. Recruitment was not limited to patients with optimal signal quality at enrollment; however, single exclusions were made based on predefined signal quality criteria as detailed below.

3. Insufficient control of ECG-related confounding factors
The current exclusion criteria address only pronounced QRS morphology abnormalities and signal quality issues, but do not systematically account for cardiac comorbidities (such as ischemic heart disease, cardiomyopathies, or arrhythmias) or medications known to affect ECG features, particularly the QT interval. As a result, part of the model's discriminative signal may reflect underlying cardiovascular comorbidity rather than diabetes itself.
It is recommended that the authors clarify whether major cardiac comorbidities and QT-modifying medications were recorded and either excluded or adjusted for in the analysis. Where possible, key conditions (e.g., coronary artery disease, heart failure, atrial fibrillation) should be included as covariates or explored through stratified or sensitivity analyses. If data limitations preclude such analyses, this confounding risk should be clearly acknowledged, and causal interpretations of diabetes-specific ECG signatures should be appropriately tempered.

Thank you so much for your comments. With your guidance we made the following changes; we included the whole list of the exclusion criteria.

4. Study positioning requires further refinement
Although the authors have clarified that ECG recordings were obtained after diabetes diagnosis, some statements may still be interpreted as suggesting an early screening application. Given the data source and current model performance, a more appropriate positioning would be as an adjunctive identification or risk-enrichment tool rather than an independent screening or classification method.

Thank you so much for your comments. By your guidance we made the following changes; added to the material and methods this paragraph "The main logic of the study was to determine the presence or absence of diabetes mellitus without adjustments for covariates—based solely on ECG data and transformations of averaged complexes. The rationale is straightforward: the algorithm should work quickly, simply, in a screening concept. Including numerous traditional cardiology covariates would have taken the algorithm out of the screening framework. Additionally, some covariates can be determined very imprecisely (many patients do not know if they had a myocardial infarction, when it occurred, its location, whether they have heart failure, etc.). This inevitably leads to significant bias due to retrospective analysis and degradation of model quality. Such an approach would no longer qualify as a screening method that could be used with an electrocardiograph."

It is recommended that wording in the Abstract and Discussion be further harmonized to emphasize the model's role in supporting identification or prompting confirmatory biochemical testing, while avoiding language that could be construed as replacing standard diagnostic approaches or enabling population-level early screening.

Thank you so much for your comments. With your guidance we made the following changes: we added to the discussion section a paragraph covering the stated concern and we modified the conclusion section of the abstract.