

REVIEWER'S REPORT

Manuscript No.: IJAR-59491

Title: Using Artificial Intelligence for the Early Diagnosis of Vehicle Malfunctions: A Leakage-Aware Evaluation and Meta-Analysis.

Recommendation:

Accept as it is

Accept after minor revision.....

Accept after major revision yes...

Do not accept (*Reasons below*)

Rating	Excel.	Good	Fair	Poor
Originality			Y	
Techn. Quality			Y	
Clarity			Y	
Significance			y	

Reviewer's ID: JPR-Dr.Shaweta sachdeva

Detailed Reviewer's Report

1. The authors have employed a blocked validation strategy to address potential data leakage. However, further details regarding the implementation of the five-fold group split are required. The authors should clarify how the blocks were assigned to different folds and whether temporal dependence between adjacent blocks could still influence the evaluation results.
2. The study relies on a single laboratory dataset obtained from one C14NE spark-ignition engine, which limits the generalizability of the findings. The authors should discuss the applicability of the proposed approach to different engine types, vehicle platforms, and real-world operating conditions.
3. The reported AUC of 1.000 for binary fault detection is noteworthy. However, the authors should provide further explanation regarding the factors contributing to this near-perfect performance and clarify whether the result reflects generalizable fault characteristics or differences between the normal and fault acquisition runs.
4. The substantial improvement in classification accuracy following the merging of fault types F2 and F3 requires further justification. The authors should explain the rationale for combining these classes and discuss the implications of this approach for practical fault diagnosis, where the identification of specific fault types may be necessary.
5. The meta-analysis exhibits very high heterogeneity ($I^2 = 98.47\%$), while the within-study variance was estimated using an assumed effective sample size. The authors should justify this methodological choice more thoroughly and discuss its potential impact on the pooled accuracy estimate and confidence interval.
6. The meta-analysis combines studies with potentially different datasets, classification tasks, numbers of classes, and validation protocols. Furthermore, an F1-score was used in place of accuracy for one study. The authors should clarify the rationale for combining these performance measures and discuss whether the resulting pooled estimate can be meaningfully interpreted as a summary of diagnostic performance.
7. The marked reduction in classification accuracy under blocked validation is an important finding. However, the authors should provide a more detailed explanation of whether this reduction is

REVIEWER'S REPORT

primarily attributable to temporal dependence, changes in operating conditions, or other characteristics of the dataset.

8. The term "early diagnosis" requires careful consideration because the dataset contains steady-state fault conditions rather than incipient faults. The authors should ensure that the terminology used throughout the manuscript accurately reflects the scope of the study and avoids overstating the capability of the proposed models to detect faults before they become critical.
9. Additional methodological details are required to improve reproducibility. The authors should clarify the hyperparameter selection procedure, random seed settings, preprocessing steps, and model-selection strategy. Providing the implementation code or supplementary methodological information would further strengthen the study.
10. The authors report mean performance and standard deviations across five folds, but the statistical significance of the differences between classifiers remains unclear. Appropriate statistical comparisons or a stronger justification for the interpretation of these differences would improve the robustness of the conclusions.
11. The introduction provides relevant background information; however, the research gap and novelty of the study should be articulated more explicitly. The authors should clearly distinguish the contribution of leakage-aware validation from that of the meta-analysis.
12. The dataset description would benefit from a table summarizing the 14 variables, their units, measurement sources, and availability through standard OBD-II interfaces. This would improve the clarity and practical relevance of the feature selection.
13. The permutation importance analysis identifies CO concentration as the most influential variable. The authors should clarify whether this analysis was performed for the three-class or four-class classification task and discuss the potential influence of correlations among the input variables on the interpretation of feature importance.
14. The decision-window analysis examines windows containing 1–50 consecutive records. However, the sampling frequency is not specified sufficiently to translate these window lengths into actual diagnostic latency. The authors should clarify the temporal meaning of the selected window sizes and discuss their practical implications.
15. Overall accuracy may not adequately reflect the performance of the classifiers across individual fault categories. The authors should consider reporting class-wise precision, recall, and F1-scores under blocked validation to provide a more comprehensive assessment of fault-isolation performance.
16. Given the very high heterogeneity observed in the meta-analysis, the pooled accuracy of 96.50% should be interpreted cautiously. The authors should consider reporting a prediction interval, where appropriate, and discuss the limitations of generalizing the pooled estimate to new datasets and operating conditions.
17. The proposed two-stage diagnostic framework, involving fault detection followed by fault isolation, is potentially useful. However, the authors should elaborate on its practical implementation, particularly how uncertain predictions would be handled and when additional diagnostic measurements or workshop-based testing would be required.
18. The authors should clarify whether the selected six on-board signals are consistently available across different vehicle models and discuss how sensor calibration, measurement variability, and differences in operating conditions might affect the model's performance when deployed across vehicle platforms.
19. The limitations and future directions could be strengthened by outlining a more explicit external validation strategy involving independent vehicles, multiple engine types, and real-world operating conditions. Such validation would provide stronger evidence of the generalizability of the proposed approach.

International Journal of Advanced Research

Publisher's Name: Jana Publication and Research LLP

www.journalijar.com

REVIEWER'S REPORT

20. The manuscript requires careful English-language editing to improve grammatical accuracy, readability, and technical precision. The authors should also ensure consistency in terminology, abbreviations, figure captions, and reference formatting throughout the manuscript.
21. The authors should clarify the criteria used to identify and select the nine studies included in the meta-analysis. Providing additional information about the literature search strategy, eligibility criteria, and study selection process would improve transparency and reproducibility.
22. The manuscript addresses an important problem in AI-based vehicle fault diagnosis, and the comparison between random and blocked validation provides useful insights into the limitations of conventional evaluation protocols. However, the methodological concerns regarding validation, interpretation of the near-perfect fault-detection performance, and assumptions underlying the meta-analysis should be addressed before the manuscript can be considered for publication.