ISSN(O): 2320-5407 | ISSN(P): 3107-4928

ISSN NO. 2320-5407

International Journal of Advanced Research

Publisher's Name: Jana Publication and Research LLP

www.journalijar.com

REVIEWER'S REPORT

Manuscript No.: IJAR-58860

Title: A Hybrid ANN–XGBoost Framework for Early Diabetes Risk Prediction Under Class Imbalance: External Clinical Validation in an African Healthcare Setting

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 manuscript presents an interesting hybrid ANN–XGBoost framework, but the novelty should be articulated more clearly. The authors should explain more explicitly how the proposed architecture differs from existing ANN–XGBoost, stacking, or hybrid representation-learning approaches.
2. The source and characteristics of the 100,000-record Kaggle dataset should be described in greater detail, including data collection methodology, inclusion/exclusion criteria, duplicate handling, missing-value statistics, and class distribution.
3. The manuscript reports a substantial difference in diabetes prevalence between the development dataset (8.8%) and the external clinical dataset (88.9%). The authors should provide a more rigorous statistical analysis of this distribution shift, rather than primarily describing the difference descriptively.
4. The rationale for not applying class weighting or resampling techniques should be strengthened. Since class imbalance is a central theme of the study, comparison with methods such as class-weighted learning, SMOTE, or other imbalance-aware approaches would make the experimental evaluation more convincing.
5. The manuscript states that hyperparameters were selected through iterative manual tuning. This approach may introduce researcher bias and limits reproducibility. A systematic strategy such as grid search, random search, or Bayesian optimization should be considered or discussed.
6. The authors should provide more information about the train/validation/test split proportions and random seeds used in the experiments. These details are essential for reproducibility.
7. Since the proposed model uses ANN-generated latent features concatenated with the original variables, an ablation study with different latent feature dimensions and ANN architectures would help establish which components are actually responsible for the observed improvement. The manuscript itself identifies this as an area for future work, but stronger evidence would substantially improve the current study.

ISSN(O): 2320-5407 | ISSN(P): 3107-4928
International Journal of Advanced Research
Publisher's Name: Jana Publication and Research LLP
www.journalijar.com

REVIEWER'S REPORT

8. The improvement over standalone XGBoost is relatively modest: recall increases from 0.6792 to 0.7086 and F2-score from 0.7200 to 0.7402. The authors should conduct statistical significance testing or confidence-interval analysis to determine whether these improvements are statistically and practically meaningful.
9. The manuscript reports excellent external validation performance, including 95.37% accuracy, 99.30% precision, 95.46% recall, and 98.23% AUC. Such high performance warrants a more detailed investigation of potential dataset characteristics, leakage, feature availability, and differences in diagnostic procedures between the development and clinical datasets.
10. The clinical validation dataset is obtained from a single specialized endocrinology center. Therefore, the claims concerning generalizability to African healthcare settings should be moderated. The current evidence supports external validation at one institution rather than broad generalization across Africa.
11. The use of HbA1c and blood glucose as highly influential predictors raises an important issue regarding the distinction between early risk prediction and diagnosis. The authors should clarify whether the proposed model is intended for early screening before conventional diagnostic testing or as a complementary diagnostic-support model.
12. The comparison with ADA diagnostic thresholds is currently described largely as a qualitative comparison. A quantitative comparison using sensitivity, specificity, PPV, NPV, ROC-AUC, and potentially decision-curve analysis would provide stronger evidence for the claimed clinical utility.
13. The manuscript uses a fixed 0.5 classification threshold despite emphasizing recall and minimizing false negatives. Threshold optimization based on clinical objectives should be explored, and the authors should report how sensitivity and specificity change at clinically relevant thresholds.
14. Probability calibration should be assessed because the proposed system is presented as a clinical decision-support/risk prediction tool. Calibration curves, Brier score, or expected calibration error would help establish whether predicted probabilities are clinically reliable.
15. The feature-importance analysis is based on permutation importance calculated on the external test set. The authors should explain the statistical stability of the importance estimates and ideally provide confidence intervals or repeated permutation results.
16. The discussion occasionally uses strong terms such as "robust," "excellent generalization," and "reliable clinical decision-support tool." These claims should be moderated given the single-center validation, cross-sectional design, and absence of prospective clinical evaluation.
17. The manuscript would benefit from a clearer description of the ethical approval, informed consent/data-governance procedures, and anonymization process for the 2,182 clinical records collected at CNHU-HKM. This information is particularly important for a study involving patient data.
18. The authors should provide a complete demographic and clinical characteristics table comparing the development and external validation populations. This would make the distribution shift and differences between the two cohorts much easier to interpret.
19. The manuscript would benefit from reporting specificity, negative predictive value, false-positive rate, and false-negative rate, particularly because the clinical motivation emphasizes minimizing missed diabetic cases.
20. There are several language and grammatical issues that require careful proofreading. For example, phrases such as "provises," "ressource-constrained," and "a clinical validation" should be corrected. Terminology and capitalization of ANN–XGBoost should also be made consistent throughout the manuscript.

ISSN(O): 2320-5407 | ISSN(P): 3107-4928
International Journal of Advanced Research
Publisher's Name: Jana Publication and Research LLP
www.journalijar.com

REVIEWER'S REPORT

21. The manuscript should ensure consistent terminology for "diabetes prediction," "diabetes risk prediction," "early detection," and "screening," as these concepts have different clinical implications.
22. Figure captions and tables should be made sufficiently self-contained. In particular, the figures presenting class distributions, the hybrid architecture, confusion matrix, feature importance, and ROC curve should clearly specify the dataset, sample size, and evaluation setting.
23. The reference list should be carefully checked for formatting consistency, completeness, DOI accuracy, and correspondence between citations and references. The manuscript should also ensure that all important recent studies on hybrid deep-learning and boosting-based diabetes prediction are adequately covered.
24. The limitations section is appropriate, but the authors should additionally discuss potential selection bias, spectrum bias, label/diagnostic bias, and differences in data acquisition between the Kaggle and hospital datasets.
25. Overall, the manuscript has a promising methodological contribution and an important external-validation component, but additional methodological clarification, stronger statistical validation, quantitative clinical comparison, and more cautious interpretation of generalizability are required before publication.

Overall recommendation: Major Revision**.**