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 ANNÂ–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 ...✓.....
Do not accept (*Reasons below*) .....

Rating	Excel.	Good	Fair	Poor
Originality		✓		
Techn. Quality		✓		
Clarity		✓		
Significance		✓		

Reviewer's ID: JPR- **098**

Detailed Reviewer's Report

Decision: Accept after Major Revision

Major Revisions

1. **Strengthen the methodological description of the proposed Hybrid ANN–XGBoost model.**
   The manuscript should provide a clearer explanation of how the ANN generates latent features, how these features are combined with the original clinical variables, and how the resulting representation is passed to XGBoost. The complete workflow should be sufficiently detailed to allow independent reproduction of the study.
2. **Clarify the data preprocessing and class-imbalance handling.**
   Since class imbalance is a central motivation of the study, the authors should clearly explain the imbalance ratio, the method used to address it, and whether resampling, class weighting, or another strategy was applied. It should also be demonstrated that preprocessing and balancing were performed without information leakage between training and testing data.
3. **Strengthen the validation strategy and statistical comparison.**
   The two-stage evaluation using a 100,000-record public dataset and an independent 2,182-record clinical dataset is a strength. However, the authors should report confidence intervals, variability across repeated

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

runs or folds, and appropriate statistical tests to determine whether the improvements of the proposed model over the baseline algorithms are statistically meaningful.

4. **Provide a stronger justification for the selected evaluation metrics.**
   The study emphasizes recall and F2-score for model selection, which is reasonable for minimizing false negatives in screening. However, the authors should explain this choice more explicitly and report sensitivity, specificity, precision, F1-score, ROC-AUC, PR-AUC, and calibration-related measures where appropriate.
5. **Expand the external clinical validation analysis.**
   The independent clinical dataset from Cotonou is an important contribution, but the manuscript should provide more information about the patient population, inclusion/exclusion criteria, disease prevalence, data-collection procedure, missing values, and differences between the development and external-validation datasets. These factors are essential for assessing distribution shift and clinical generalizability.
6. **Improve the clinical interpretation and avoid overstating practical applicability.**
   The manuscript describes the proposed framework as a clinically relevant decision-support solution. However, high predictive performance alone does not establish clinical utility. The authors should discuss threshold selection, false-negative and false-positive consequences, calibration, prospective validation, and the role of clinicians before recommending deployment in healthcare settings.

Minor Revisions

7. **Improve the presentation of hyperparameters and reproducibility information.**
   The manuscript provides useful ANN and XGBoost parameters, including the number of hidden layers, dropout, learning rate, batch size, number of trees, tree depth, and regularization. The authors should also report the random seed, software/library versions, hardware environment, and complete tuning procedure to improve reproducibility.
8. **Clarify the feature-importance analysis.**
   Permutation Feature Importance identifies **HbA1c and blood glucose level** as the two most influential predictors. The authors should explain

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

how the importance values were calculated, provide uncertainty or variability where possible, and discuss whether the results remain consistent across the public and external clinical datasets.

9. **Strengthen the discussion of limitations.**
   The manuscript should explicitly discuss limitations related to dataset representativeness, potential selection bias, missing clinical variables, differences between datasets, and the absence of prospective clinical testing. The limitations should be connected directly to the claims of generalizability and clinical deployment.
10. **Improve language, formatting, and technical presentation.**
    The manuscript requires careful proofreading for grammatical and typographical errors. For example, expressions such as **"provises"** and **"ressource-constrained"** should be corrected. Tables, figures, abbreviations, mathematical notation, and references should also be checked for consistency throughout the manuscript.