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

Manuscript No.: **IJAR-55140**

Title: HYBRID BOGA OPTIMIZATION OF XGBOOST FOR CALIBRATED AND ROBUST HYPERTENSION PREDICTION: A MULTI-COHORT VALIDATION STUDY,

Recommendation:
Accept after minor revision .....

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

Reviewer Name: Dr.K.ARUMUGANAINAR **Date:** 10.12.2025

Detailed Reviewer's Report

DETAILED REVIEW REPORT

1. Summary of the Manuscript

The manuscript presents a hybrid machine-learning optimization framework (BO–GA–XGBoost) aimed at improving discrimination, calibration, and cross-cohort robustness for hypertension prediction. The method integrates **Bayesian Optimization**, **NSGA-II multi-objective genetic algorithm**, and **isotonic regression calibration**. The approach is validated on three independent cohorts—NHANES, Framingham Heart Study, and Kaggle Clinical EMR—totaling 117,376 samples. The model demonstrates strong AUC (0.962), low calibration error (ECE = 0.021), and excellent cross-population stability (ΔAUC = 0.014). SHAP-based interpretability shows 94% alignment with ESC/ESH guidelines.

---

2. Major Comments

2.1. Strength of Methodological Framework (Strong Contribution)

The proposed BO–GA hybrid pipeline is well-justified and addresses a critical gap in hypertension prediction research. Most prior models only optimize discrimination, whereas this study incorporates calibration and robustness using multi-objective optimization. This is a significant methodological advancement.

2.2. Clarity and Organization

The manuscript is clearly written and logically organized, with detailed explanations of datasets, preprocessing strategies, optimization pipelines, and evaluation. However:

- The *Materials and Methods* section is long and slightly overloaded. Consider condensing repetitive explanations.
- Figure captions could be more concise, especially in calibration and SHAP visualizations.

2.3. External Validation

The tri-cohort validation is a major strength. The authors strictly freeze hyperparameters before external testing, preventing data leakage and ensuring rigorous validation.

However, the authors should:

- Provide a short table of **population demographics** across cohorts.
- Add a discussion on **potential demographic biases** (e.g., Framingham's heavy Caucasian composition).

2.4. Calibration Metrics and Clinical Utility

The inclusion of Brier Score, ECE, and calibration diagrams is commendable. But to improve clarity:

- Include **calibration plots for all three cohorts**, not just NHANES.
- Provide clinical interpretation of what ECE ≤ 0.03 means in practice.

2.5. Statistical Significance and Robustness

The manuscript uses DeLong tests, bootstrap CIs, and TOST equivalence testing—excellent practice for a clinical AI study.

However:

- The bootstrapping method (1,000 replications) should mention whether stratification was applied consistently across datasets.
- Sensitivity and specificity thresholds (≥ 0.80 and ≥ 0.70) are clinically reasonable but should be justified using ACC/AHA guidelines.

2.6. Interpretability (SHAP Analysis)

The SHAP analysis is strong and aligns well with clinical literature. However:

- The waterfall plot should highlight the **baseline risk score** more clearly.
- The authors may add a **subgroup-level SHAP interaction analysis** to validate fairness claims.

2.7. Limitations

The manuscript sufficiently mentions computational cost but should expand on:

- Possible overfitting risk in NSGA-II iterations.
- Variability in missing data patterns between cohorts.
- Limited representation of minority ethnic groups.

---

3. Minor Comments

1. **Formatting issues**
    - Some tables extend across multiple pages; compress or reposition.
    - Figure numbering consistency must be checked (some figures referenced, but not included in the PDF).
2. **Grammar & Language**
    - Minor language corrections needed (e.g., remove redundant words, check passive voice consistency).
    - Ensure consistent usage of terms: "hyperparameters," "hyper-parameters," etc.
3. **References**
    - Ensure all references follow the journal's formatting style.
    - Some entries include URLs; verify if this is allowed.
4. **Acronym Definitions**
    - Define ECE, NSGA-II, GP, and SBX at first use.
5. **Ethical Statement**
    - Add a sentence confirming the use of publicly available datasets with no identifiable personal information.

---

4. Novelty & Impact

Novelty

✔ First known model to combine:

- Bayesian Optimization
- NSGA-II Multi-Objective Optimization
- Isotonic regression
- Freeze-parameter tri-cohort validation
- Clinical constraint enforcement (sensitivity ≥0.80, specificity ≥0.70, ECE ≤0.05)

This is a meaningful step toward clinically deployable hypertension prediction.

Impact

✔ High impact for clinical AI, medical decision support systems, and real-world screening tools.

---

5. Overall Strengths

- Highly rigorous methodology
- Excellent multi-cohort validation
- Outstanding calibration
- Strong interpretability using SHAP
- Clear clinical relevance
- Model demonstrates high external robustness

---

6. Overall Weaknesses

- Manuscript verbosity; needs condensation
- Minor missing analyses (demographic table, subgroup calibration)
- Limited discussion on model deployment and real-world utility
- Computational requirements not explored in depth

---

7. Recommendation

Recommendation: ***MINOR REVISION***

The manuscript is strong and suitable for publication after addressing the minor issues listed. It has substantial scientific value and methodological contribution.

---

8. Overall Score (0–10)

Score: 8.8 / 10

Interpretation:
Minor Revision – High-quality work requiring only polishing and minor improvements.