Qeios

Peer Review

# Review of: "FGSCare: A Feature-driven Grid Search-based Machine Learning Framework for Coronary Heart Disease Prediction"

**Milan Toma**[1]

1. College of Osteopathic Medicine, New York Institute of Technology, United States

The authors claim to have used an 8:1:1 split ratio for training, validation, and testing. However, the methodology section does not provide sufficient details about how this split was implemented. If the split was not stratified (i.e., the dataset was divided into training, validation, and testing subsets in a way that didn't maintain the same proportion of classes (e.g., positive and negative cases) in each subset as in the overall dataset) or if SMOTE was applied before splitting, this could lead to data leakage, where synthetic samples influence both training and testing, resulting in over-optimistic performance metrics. The authors must clarify this in the methodology section. If SMOTE was applied before splitting, the results should be re-run with proper data splitting to avoid leakage.

The study does not include testing on an external dataset from a different institution or under different circumstances. This is a limitation, as the model's generalizability to unseen, real-world data cannot be assessed. The authors should discuss this limitation in the discussion section and emphasize that the results may not generalize well to other populations or datasets due to the lack of external validation.

The article does not include learning curves for accuracy and loss during training and validation. These curves are essential for assessing whether the model has converged and whether the model suffers from overfitting or underfitting. The authors should include these plots in the results section to allow readers to evaluate the model's training dynamics and generalizability.

The article does not discuss the public perspectives and ethical implications of using AI in healthcare. The authors should include a discussion on these topics, supported by recent and relevant citations, to address the broader implications of their work.



doi.org/10.32388/KH6WBW

Include more details on hyperparameter tuning, including the ranges explored during GridSearchCV.

Discuss the implications of the relatively low recall scores for some models, as missing CHD cases could have severe consequences.

Since this is a classification problem, confusion matrices should be included to evaluate the model's performance in terms of true positives, true negatives, false positives, and false negatives. The authors should include them for both validation and test sets.

## Declarations

**Potential competing interests:** No potential competing interests to declare.