Enhancing Prediction of Systolic Heart Failure Outcomes through Genetic Algorithm-Optimized Laboratory Tests Bhumi Patel, MS, Lemba Priscille Ngana, MPH, Janusz Wojtusiak, PhD Health Informatics Program, George Mason University, Fairfax, VA

Introduction

Heart failure, a major cause of ICU admissions with high mortality rates, necessitates effective risk stratification to identify high-risk patients for optimal patient care [1]. Previous studies have used machine learning methodologies to predict mortality and hospital readmissions among heart failure patients [2]. Leveraging clinical and laboratory data for adverse outcome prediction in systolic heart failure patients, our study addresses the challenge of variable test readings across different time intervals. We aimed to optimize these variables to obtain the most informative readings at crucial times, thereby improving mortality predictions through the strategic use of laboratory test data. This involved identifying the best combination of test results to enhance model performance significantly.

Methods

Study population: We utilized data from the MIMIC-IV (Medical Information Mart for Intensive Care) version 2.2 database. Our study focused on patients admitted to the ICU with systolic heart failure between 2008 and 2019. The sample included adult patients aged 18 years and above with an ICU stay exceeding 24 hours and laboratory test results collected within the first 24 hours of admission. Our analysis involved 1,037 patients meeting these criteria.

Clinical measurements: The input variables comprised demographic information and clinical data. 48 baseline variables were collected, of which 38 were laboratory test results.

Optimization: A Genetic Algorithm (GA) was used to optimize feature selection and laboratory test timing for predicting systolic heart failure in-hospital mortality. The study utilized a random forest as the base model and the evaluation metric was Area Under Curve (AUC). Starting with a randomly generated population of test timing intervals, we underwent a 1000-generation evolutionary process aimed at enhancing model accuracy, measured by AUC. The GA improved model performance through fitness calculation, selection, crossover, and mutation.

Statistical analyses: Descriptive statistics were conducted to obtain the characteristics of patients. Baseline comparisons were conducted using the t-test, Mann-Whitney U test, or the Chi-square test. Additionally, Spearman's correlation coefficient was used to assess correlations between the variables. The results are presented as mean (SD) or number (percentage).

Results

The implementation of the genetic algorithm significantly improved the random forest model performance in predicting in-hospital mortality among patients diagnosed with systolic heart failure. The optimization process increased the AUC from an initial baseline of 0.70 to an optimized value of 0.78, with observed variations ranging from 0.71 to 0.87. Additionally, the results showed the optimal time at which each laboratory test reading results in high prediction. The findings highlight the effectiveness of incorporating genetic algorithm to enhance model performance. Lastly, by exploring and refining feature space, we identified key predictors amongst all the variables.

Conclusions

The findings demonstrate that utilizing genetic algorithms to optimize laboratory test results effectively enhances model performance in predicting in-hospital mortality among patients with heart failure. This approach highlights the importance of precise feature selection and laboratory test results timing to accurately predict patient outcomes. Integrating genetic algorithm optimization in model building has the potential to advance clinical practice by enhancing early risk stratification and improving patient prognostic outcomes. Continued research is currently underway to further enhance model performance.

References

- Centers for Disease Control and Prevention [Internet]. Heart Disease Facts. [Last Reviewed 2023 May 15] [cited 2024 Mar 17]. Available from: <u>https://www.cdc.gov/heartdisease/facts.htm#:~:text=Heart%20disease%20is%20the%20leading.groups%20in%20the%20United%20States&text=One%20person%20dies%20every%2033,United%20States%20from%20cardiovascular%20disease.</u>
- Mpanya D, Celik T, Klug E, Ntsinjana H. Predicting mortality and hospitalization in heart failure using machine learning: A systematic literature review. Int J Cardiol Heart Vasc. 2021 Apr 12;34:100773. doi: 10.1016/j.ijcha.2021.100773. PMID: 33912652; PMCID: PMC8065274.