Scope

Provisional Title

Cardiovascular Risk Stratification through Multi-Objective Optimization: A Decision Support Model for Healthcare Resource Allocation

Introduction

Cardiovascular diseases remain one of the leading causes of morbidity and mortality worldwide, placing a significant burden on healthcare systems—especially in low- and middle-income countries. Early screening and accurate cardiovascular risk stratification are essential strategies for preventing critical events such as acute myocardial infarctions and cardiac arrests. However, the reality of public healthcare systems, such as Brazil's Unified Health System (SUS), presents considerable challenges in the efficient and equitable allocation of resources, including diagnostic tests, specialist consultations, and preventive hospital admissions.

Traditional screening models, often based on fixed rules or standardized protocols, frequently fail to capture the clinical complexity of individual patients or to adapt dynamically to operational and budgetary constraints. Moreover, such approaches are vulnerable to biases, subjectivity, and human error, which can compromise the effectiveness of patient prioritization and, consequently, health outcomes.

In this context, the application of data science and mathematical modeling techniques emerges as a promising alternative to support clinical and administrative decision-making. Integrating machine learning models, capable of predicting individual cardiovascular event risks based on clinical variables, with multi-objective optimization models enables the development of intelligent systems that strategically prioritize patient care while simultaneously considering estimated risk, associated costs, and the healthcare system's capacity constraints.

This project proposes the development of a hybrid computational model that combines cardiovascular risk prediction with Integer Linear Programming and Multi-Objective Optimization techniques. The goal is to provide a decision-support tool for the rational allocation of medical resources. The central research question guiding this proposal is: how can optimization models be used to maximize prioritized care for high-risk cardiovascular patients while adhering to the budgetary and operational constraints of the healthcare system?

By simultaneously addressing the challenges of accurate risk stratification and structural limitations within healthcare services, this project aims to offer an innovative and practical solution tailored to public healthcare settings, promoting greater effectiveness, equity, and rationality in health system management.

Research Question

How can multi-objective optimization techniques be used to maximize prioritized care for high-risk cardiovascular patients while adhering to the budgetary and capacity constraints of the healthcare system?

Objectives

General Objective

To develop a computational model based on the combination of machine learning and multi-objective optimization techniques, aiming to classify patients according to their cardiovascular risk level and to guide the efficient and equitable allocation of medical resources, while respecting the budgetary and operational constraints of the healthcare system.

Specific Objectives

- Apply supervised machine learning algorithms to estimate the individual probability of cardiovascular events based on clinical variables available in a public database;
- Establish criteria for categorizing patients into risk levels (low, medium, and high) based on percentiles of the predicted probability;
- Formulate a mathematical optimization model using Integer Linear Programming and multi-objective approaches to prioritize patient care according to estimated risk and resource limitations;
- Incorporate realistic constraints into the model, such as budget ceilings, operational capacities, and population equity criteria (e.g., minimum inclusion of women and elderly patients);
- Simulate different allocation scenarios and compare the outcomes with traditional screening methods based on fixed rules or clinical scoring systems;
- Evaluate the performance of the proposed model in terms of effectiveness in prioritizing high-risk patients, efficiency in resource utilization, and inclusion of vulnerable groups.

Justification

The rising incidence of cardiovascular diseases presents an ongoing challenge for healthcare systems worldwide, particularly within public systems operating under limited resources, such as SUS. Proper cardiovascular risk stratification and the efficient prioritization of care are crucial for reducing severe events and the economic burden associated with hospital admissions, specialized tests, and high-complexity treatments.

However, traditional screening strategies, based on rigid protocols or subjective clinical assessments, often fail to capture the individual complexity of patients or to dynamically adapt to budgetary, infrastructural, and workforce constraints. Furthermore, these approaches may inadvertently reinforce inequalities in access to care, especially among vulnerable populations such as the elderly, women, and low-income individuals.

In this context, integrating machine learning techniques with mathematical optimization models offers a relevant and innovative opportunity for developing decision-support tools in both clinical and administrative domains. By accurately predicting individual risk and formulating allocation strategies that account for multiple objectives and operational constraints, it becomes possible to enhance both the effectiveness and equity of healthcare resource management.

This project is justified by its direct applicability to improving decision-making processes in cardiovascular care, its methodological innovation in combining risk prediction with multi-objective optimization, and its potential contribution to evidence-based public health policies. Additionally, the use of public data and replicable techniques strengthens the scientific foundation of the proposal and supports its future expansion to other clinical conditions and institutional settings.

Methodological Design

Type of Research

This project is characterized as an applied research study with a quantitative approach and a descriptive-exploratory design. The study will be conducted through computational modeling, using publicly available secondary data, without any direct human intervention, therefore, submission to ethics committees is not required.

Methodological Steps

The project will be executed in the following sequential stages:

1. **Data acquisition and preparation**
   The dataset titled *Cardiac Arrest Dataset*, available on the Kaggle platform, will be used. This database compiles clinical and demographic variables associated with the risk of cardiac events. The data will undergo cleaning procedures, categorical variable encoding, handling of missing values, normalization, and exploratory analysis.

2. **Predictive modeling of cardiovascular risk**
   Supervised machine learning algorithms, such as Logistic Regression, Random Forest, and XGBoost, will be applied to estimate the individual probability of cardiovascular events. Model performance will be assessed using metrics such as AUC-ROC, F1-score, accuracy, and confusion matrix. The predicted probability will serve as an input variable for the optimization model.

3. **Risk categorization**
   The predicted probabilities will be classified into three risk levels, based on the percentile distribution of the sample. This categorization will be used for comparative purposes and for analyzing equity in resource allocation.

4. **Optimization model formulation**
   An Integer Linear Programming (ILP) or Multi-Objective Optimization model will be

constructed with the goal of maximizing the coverage of patients at higher cardiovascular risk, while simultaneously respecting budgetary constraints, capacity limits, and equity criteria (such as minimum inclusion thresholds for women or elderly patients). The objective function may incorporate multiple criteria, including estimated risk, cost of care, and indicators of social vulnerability.

5. **Scenario simulation and sensitivity analysis**
   The model will be tested across different simulated scenarios, varying parameters such as available budget, service capacity, the presence or absence of equity constraints, and different risk classifiers. A sensitivity analysis will be conducted to assess the robustness of the decisions under varying operational conditions.

6. **Comparative evaluation**
   The results of the proposed model will be compared to traditional screening approaches based on fixed rules (e.g., prioritization by cholesterol levels or age), using metrics such as: number of high-risk patients served, efficiency in resource utilization, and inclusion of vulnerable groups.

Study Scope and Delimitations

This study will be conducted exclusively using secondary data, publicly available through the Cardiac Arrest Dataset hosted on the Kaggle platform. The research will not involve the collection of primary data, nor the execution of clinical or operational interventions in hospitals or healthcare units. Therefore, there will be no direct interaction with patients or healthcare professionals, characterizing the study as entirely simulated and computational.

The scope of the statistical and mathematical modeling will be limited to cardiovascular risk stratification and simulated allocation of medical resources, based on clinical and demographic variables present in the aforementioned dataset. While the proposed model holds potential for real-world application, its practical validation within public or private healthcare systems is not included in the scope of this project.

Additionally, risk categorization will be based on statistical criteria derived from the sample itself, and the medical care costs (such as exams or hospital admissions) will be simulated using approximate parameters obtained from the literature or secondary public data sources.

Equity considerations will be addressed through simulation scenarios incorporating various allocation constraints (such as gender, age, or other available variables), without the intention of fully encompassing the complex social and structural factors contributing to healthcare access disparities.

Finally, the results obtained will be exploratory and indicative in nature, and should not be interpreted as direct clinical recommendations without proper institutional and ethical validation. The central focus of this study lies in the development, analysis, and evaluation of a computational model that integrates machine learning and optimization techniques, aimed at enhancing healthcare management through resource rationalization and strategic prioritization of care.

Timeline

The project will be carried out in sequential stages to ensure methodological organization and feasible implementation within the proposed timeframe. The schedule below outlines the main activities by month, allowing for the development, analysis, and final drafting of the study based on realistic deadlines.

Project Stages	Week 1	Week 2	Week 3	Week 4	Week 5
Literature review and theoretical framework	✔️	✔️			
Data collection, cleaning, and exploratory analysis	✔️	✔️			
Training and evaluation of risk classification models		✔️	✔️		
Patient risk level categorization		✔️			
Optimization model formulation and implementation			✔️	✔️	
Scenario simulation and sensitivity analysis				✔️	
Comparative evaluation with traditional approaches				✔️	
Scientific article writing and submission preparation					✔️
Final manuscript revision and adjustments					✔️

References

[1] Weng, Stephen F et al. "Can machine-learning improve cardiovascular risk prediction using routine clinical data?." PloS one vol. 12,4 e0174944. 4 Apr. 2017, doi:10.1371/journal.pone.0174944

[2] Peng, M., Hou, F., Cheng, Z. et al. Prediction of cardiovascular disease risk based on major contributing features. Sci Rep 13, 4778 (2023). https://doi.org/10.1038/s41598-023-31870-8

[3] Kuo, R.J., Song, P.F., Nguyen, T.P.Q. et al. An application of multi-objective simulation optimization to medical resource allocation for the emergency department in Taiwan. Ann Oper Res 326, 199–221 (2023). https://doi.org/10.1007/s10479-023-05374-7

[4] Wan, F., Fondrevelle, J., Wang, T. et al. Two-stage multi-objective optimization for ICU bed allocation under multiple sources of uncertainty. Sci Rep 13, 18925 (2023). https://doi.org/10.1038/s41598-023-45777-x

[5] Bastian, N.D., Ekin, T., Kang, H. et al. Stochastic multi-objective auto-optimization for resource allocation decision-making in fixed-input health systems. Health Care Manag Sci 20, 246–264 (2017). https://doi.org/10.1007/s10729-015-9350-2

[6] Villamil, H.C.; Espitia, H.E.; Bejarano, L.A. Multiobjective Optimization of Fuzzy System for Cardiovascular Risk Classification. Computation 2023, 11, 147. https://doi.org/10.3390/computation11070147

[7] Wang Y, Aivalioti E, Stamatelopoulos K, et al. Machine learning in cardiovascular risk assessment: Towards a precision medicine approach. Eur J Clin Invest. 2025;55 Suppl 1:e70017. doi:10.1111/eci.70017

[8] Tanantong, T., Pannakkong, W. & Chemkomnerd, N. Resource management framework using simulation modeling and multi-objective optimization: a case study of a front-end department of a public hospital in Thailand. BMC Med Inform Decis Mak 22, 10 (2022). https://doi.org/10.1186/s12911-022-01750-8

[9] Mihan A, Pandey A, Van Spall HG. Mitigating the risk of artificial intelligence bias in cardiovascular care. Lancet Digit Health. 2024;6(10):e749-e754. doi:10.1016/S2589-7500(24)00155-9