Predictive Horizons: Unveiling Cardiovascular Insights with Conditional Inference Trees

Bharadwaz M , Hood D – Axtria

OBJECTIVE

This study investigates the utility of machine learning techniques for early heart disease prediction. By identifying key risk factors associated with cardiovascular diseases , we aim to enhance our understanding of population-level health disparities. Our findings will inform targeted intervention strategies, improved patient outcomes, and reduced burdens.

DATA SOURCE

- An open dataset on Kaggle, focusing on heart disease occurrence and encompassing factors like Age, Gender, Race, BMI, General, Physical and Mental health, various comorbidities, Sleep time etc., is a significant component of the Behavioral Risk Factor Surveillance System (BRFSS).
- BRFSS collects comprehensive health status data from 400,000+ US residents annually, the largest continuous health survey system globally.

METHODS

Sample Imbalance:

- Addressing sample imbalance is crucial due to the lower prevalence of heart disease within the dataset. Balancing the sample ensures unbiased analysis and reliable model performance.
- The Near Miss algorithm addresses sample imbalance by under-sampling units with higher prevalence indicator values. This enhances variability among retained units.

Focus on True Negative Rate:

- Unlike traditional models that emphasize risk factors, this study prioritizes the identification of protective factors. Focusing on the true negative rate informs preventive strategies and improves cardiovascular health.

Conditional Inference Tree (CIT):

- Utilizing the CIT model, this study optimizes performance through meticulous tuning of the CTree hyperparameter. Bootstrap methods determine the minimum criterion for CIT estimation, enhancing robustness. The Bonferroni correction ensures statistical validity, enabling the CIT to capture complex relationships and provide insightful predictions regarding heart disease incidence.

THE ADVANTAGES OF EMPLOYING A CONDITIONAL INFERENCE TREE

Statistical Foundation
CITs utilize conditional permutation tests for variable selection, establishing a robust statistical foundation. This enhances the reliability of findings and mitigates the risk of overfitting.

Rigorous Testing
Potential splits in CITs undergo rigorous statistical testing to determine their significance. This ensures that only meaningful splits are included in the tree.

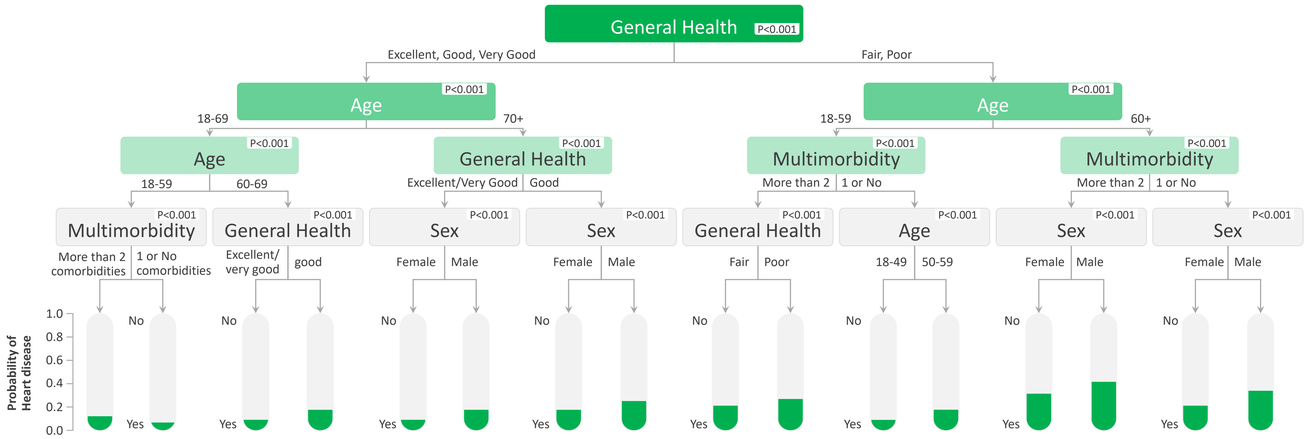
Non-Parametric Flexibility
CITs exhibit non-parametric characteristics, accommodating a wide range of data distributions. They handle both categorical and continuous variables without extensive preprocessing.

Sample Size Resilience
CITs remain effective even with limited sample sizes. This resilience is valuable in real-world scenarios.

Automated Variable Selection
CITs autonomously choose variables based on their statistical significance. This streamlines model construction.

Interpretability and Missing Data Handling
CITs offer interpretability and effectively handle missing data. Their practical utility is enhanced by these features.

RESULTS



- **General Health:** The CIT analysis identified general health as the most influential factor in predicting heart disease incidence; notable disparities were observed between individuals with excellent, good, or very good health and those with fair or poor health.
- **Age:** Age plays a critical role, particularly among individuals with excellent, good, or very good general health, revealing distinct risk profiles for those below 69 years and those aged 70+.
- **Morbidity:** Multimorbidity significantly impacts heart disease incidence, especially among individuals aged 70+, emphasizing the importance of managing comorbidities in preventive healthcare strategies.
- **Gender Disparities:** Males show higher risk profiles, particularly among those aged 70+ and with multimorbidity.
- **Limited Influence of Other Factors:** BMI, race, addiction, mental health, and sleep have minimal impact on heart disease incidence, indicating the robustness of the predictive model.
- **Mitigating Effects of General Health, Gender, and Age:** The CIT highlights the mitigating effects of general health, gender, and age on heart disease incidence, emphasizing the importance of targeted preventive measures.

CONCLUSIONS

Key Findings: The Conditional Inference Tree (CIT) analysis revealed general health as the most influential factor in predicting heart disease incidence, with distinct risk profiles observed across different age groups and multimorbidity.

Significant Factors: Factors such as age, multimorbidity, and physical activity emerged as important predictors of heart disease, emphasizing the multifaceted nature of cardiovascular risk.

Interplay of Factors: The CIT elucidated complex interactions between various risk factors, providing insights into the synergistic effects of age, health status, and lifestyle behaviors on heart disease incidence.

Utility of CIT: The CIT proved to be a valuable tool for identifying at-risk individuals and informing targeted intervention strategies tailored to specific risk factors.

Improving Patient Outcomes: By leveraging CIT insights, healthcare practitioners can develop personalized care plans aimed at mitigating cardiovascular risk factors and improving patient outcomes.

Importance of Early Intervention: The findings highlight the importance of early detection and intervention in reducing the burden of cardiovascular disease and promoting better heart health outcomes.

Future Directions: Further research utilizing CIT and similar methodologies can enhance our understanding of cardiovascular risk factors and inform more effective preventive strategies in clinical practice. The method can also be applied across different therapeutical areas as well.

DISCLOSURES
- Mrinmoy P Bharadwaz and David Hood are employees of Axtria.
- This study was funded by Axtria.

POSTER PRESENTED AT
ISPOR US, Atlanta, Georgia, USA; 5 May – 8 May 2024.