Machine Learning Prediction of 3-Year Survival in Patients Undergoing Percutaneous Coronary Intervention for Coronary Artery Disease

MSR 18

Lejia Hu[1], Xuan Zhang[1] , Hal Yapici[1], Kunal J. Lodaya[1], Sibyl H. Munson[1], Fabian D'Souza[1], Weiqi Jiao[1], Hayden W. Hyatt[1], Rahul Rajkumar[1], Nicholas Bettencourt[1]
[1] *Boston Strategic Partners, Inc.*

BACKGROUND

- Percutaneous coronary intervention (PCI) is a widely used procedure for treating coronary artery disease (CAD), yet long-term survival rate varies
- Identifying key risk factors is essential for improving patient outcomes
- Machine learning (ML) offers a powerful approach to uncover predictors of outcomes and regional disparities, guiding targeted interventions to enhance post-PCI survival

OBJECTIVE

- We developed an ML-based model to predict 3-year survival in PCI patients using real-world data
- By identifying key clinical predictors and examining regional patient distribution, we sought to understand patient risk profiles and ensure applicability across diverse populations

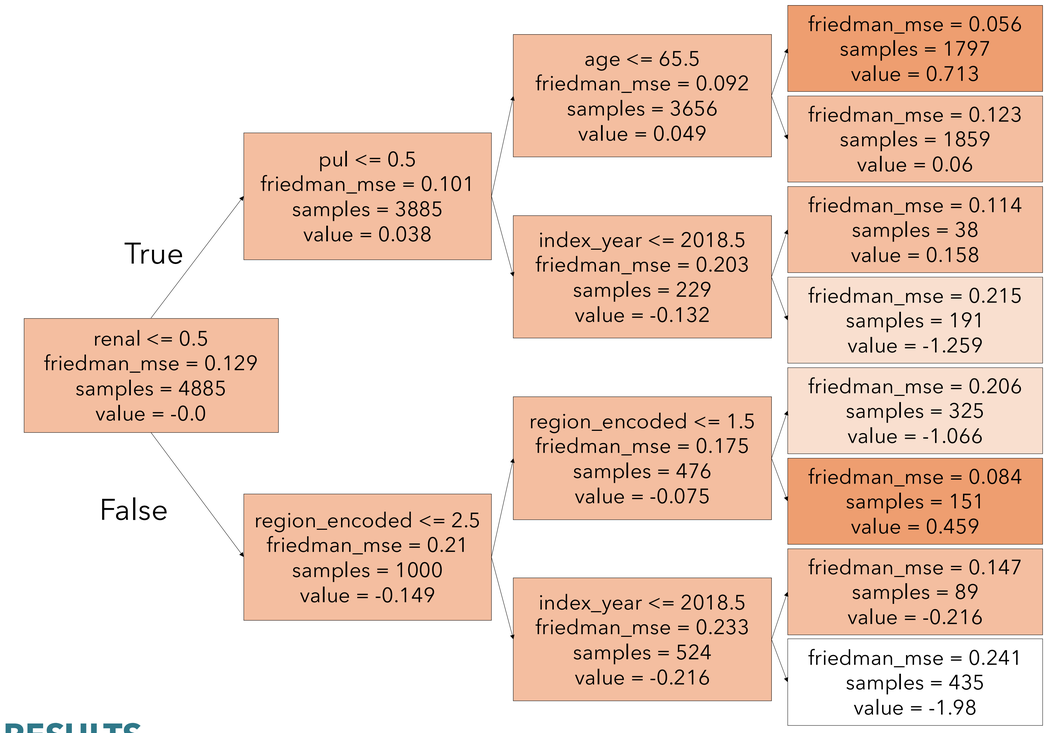
STUDY DATASET

We used the National COVID Cohort Collaborative (N3C) database, a nationwide clinical database, aggregating electronic health records, to enable comprehensive analysis of PCI outcomes and regional disparities[1].

METHODS

- Patients undergoing PCI for CAD (2017–2021) were analyzed using the N3C database
- A Gradient Boosting Decision Tree (GBDT) model was trained to predict 3-year survival based on demographics, comorbidities, and treatment-related features (Figure 1)
- GBDT, which constructs an ensemble of decision trees, was selected for its ability to handle complex interactions and variable importance ranking
- Model performance was evaluated using accuracy, precision, recall, and the confusion matrix

Figure 1. Gradient Boosting Decision Tree



RESULTS

- The final cohort included 7,505 patients, 6,243 (83.2%) of whom survived within 3 years of the index date
- Among the 22 variables analyzed, renal impairment, age, pulmonary hypertension, and index year were the strongest predictors of 3-year survival (Figure 2). One limitation to note is that the observed decrease in 3-year survival may be partially influenced by the onset of the COVID-19 pandemic (Figure 4)
- Survival rates varied significantly by region, with the South and Midwest experiencing the lowest rates, while the Northeast and West had higher rates. These disparities suggest potential regional differences in healthcare access, treatment quality, or patient risk profiles (Figure 3)
- The Gradient Boosting Classifier achieved an accuracy of 0.86, with a precision of 0.86 and a recall of 0.98 (Figure 5)

Figure 2. Feature Importance from Gradient Boosting Classifier

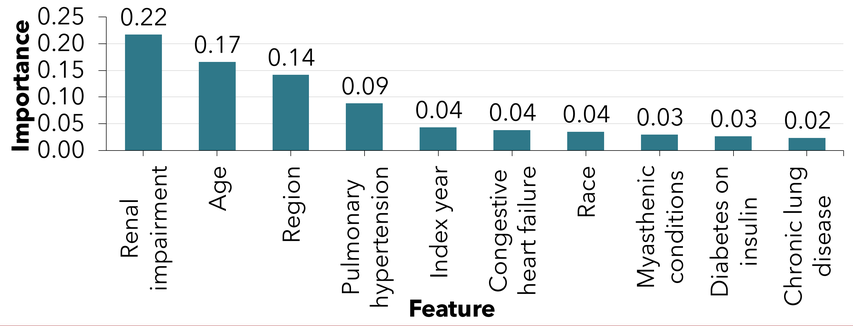


Figure 3. PCI Patients by Region and Mortality

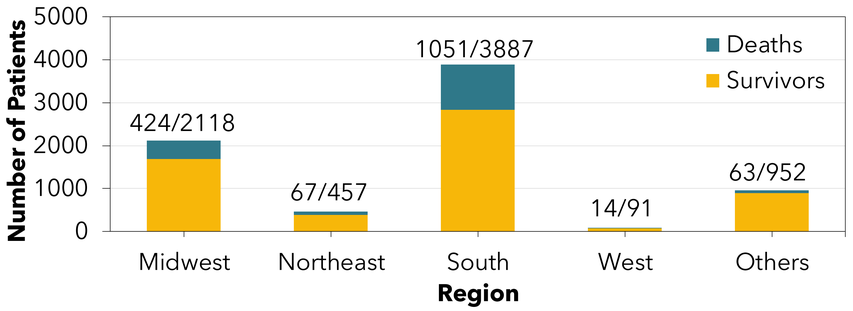


Figure 4. 3-Year Survival Rate

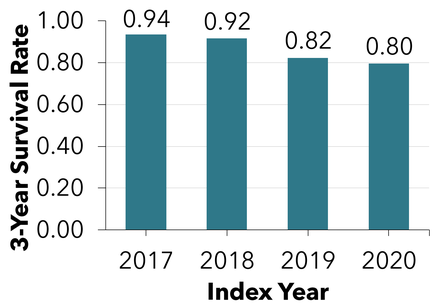
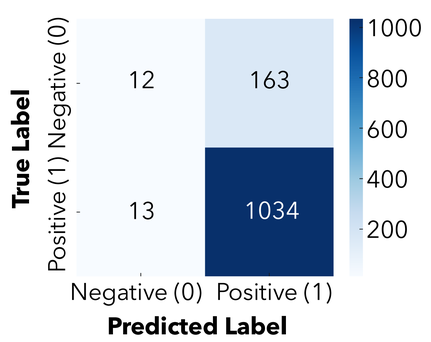


Figure 5. Confusion Matrix



CONCLUSIONS

- ML effectively identified key predictors of a 3-year survival rate in PCI patients, with renal impairment and age being dominant risk factors
- Significant regional variations in survival rates suggest underlying healthcare disparities that warrant further investigation
- These insights can inform risk stratification strategies and policy interventions to improve PCI outcomes and reduce geographic inequities in cardiovascular care

REFERENCES

1. National Center for Data to Health (CD2H). (n.d.). About. Retrieved March 26, 2025, from https://covid.cd2h.org/about/