Using Machine Learning to Predict Fatality in Patients with Myocardial Infarction

Literature Survey

The increasing burden of myocardial infarction-related fatalities underscores the critical need for robust predictive models. Literature exist in this area. However, our project seeks to leverage advanced techniques to improve the performance of existing models. A brief review of 5 of the existing literature are presented in subsequent paragraphs.

Paper 1

The study applied machine learning (ML) to predict fatal complications within 72 hours of myocardial infarction (MI). A dataset of 1,699 patients (mean age 61.86 ± 11.26 years, 62.65% male) with 111 clinical, demographic, and ECG-based variables was analyzed. Recursive Feature Elimination (RFE) identified 50 key features. Models tested included logistic regression, support vector machine, random forest, and XGBoost. The XGBoost model outperformed others with an accuracy of 91.47%, sensitivity of 94.35%, F1-score of 95.14%, and AUC-ROC of 78.65%. Significant predictors were cardiogenic shock and complete left bundle branch block (LBBB). The unbalanced dataset (84.04% survivors, 15.94% deaths) was sourced from Krasnoyarsk Inter-District Clinical Hospital, Russia (1992–1995). Limitations include potential biases due to dated data and imbalance.

Paper 2

This study developed ML models to predict in-hospital mortality among Acute Myocardial Infarction (AMI) patients. Data comprised 1,749 patients (mortality rate: 12.5%) from a Portuguese hospital (2013–2015). Key features included demographic, administrative, and clinical variables, with cardiogenic shock, age, and elevated urea levels emerging as top predictors. Preprocessing involved SMOTE for class imbalance, imputation of missing values, and feature normalization. Six ML models were tested, with k-Nearest Neighbors (kNN) achieving the best performance (recall: 90%, AUC: 89%). The dataset's single-center origin limiting generalizability, and modest size.

Paper 3

This research focused on predicting in-hospital mortality in ST-segment elevation myocardial infarction (STEMI) patients with type 2 diabetes mellitus (T2DM). Data from 438 patients (2016–2020) were split into training (70%) and validation (30%) cohorts. The unbalanced dataset (9.5% mortality) featured clinical, demographic, and procedural variables, with GRACE scores and cardiac arrest status as key inputs. The models used and evaluated included logistic regression, Random Forest, XGBoost, CatBoost, and others. CatBoost achieved the highest accuracy (93%) and AUC-ROC (92%), surpassing the GRACE risk score. However, the single-center dataset, retrospective design, and absence of external validation limit the study's applicability.

Paper 4

This study assessed tree-based ML models (Random Forest, AdaBoost, and XGBoost) for predicting 14-year mortality in acute myocardial infarction (AMI) patients using clinical data and novel biomarkers (bPEP and bET). Data were drawn from 139 patients in Taiwan (2003–2004). Random Forest yielded the best AUC-ROC (83%), and bPEP/bET improved model performance by ~3%. AdaBoost achieved the highest sensitivity (90%) but poor specificity (67%). Small sample size and sex imbalance present significant limitations, suggesting the need for larger, more diverse datasets and exploration of advanced models.

Paper 5

This multi-center study predicted in-hospital mortality for first acute myocardial infarction (AMI) using data from 5,836 patients across two Chinese hospitals. Seventy clinical and laboratory indicators were analyzed, with eight key predictors (e.g., D-Dimer, BNP, cardiogenic shock) identified using feature selection. Seven ML models were tested, with the Bagging model delivering the best results: AUC-ROC of 93.2% (test set) and 89.3% (external validation), along with sensitivity/specificity of 88%/86%. The study's strengths include robust feature selection and external validation. However, regional bias and differences in laboratory standards between centers limit broader applicability.

Conclusion

The five studies demonstrate the potential of ML in predicting MI-related outcomes, employing diverse datasets and methodologies. While the models achieved promising results, common limitations include unbalanced datasets, single-center data sources, and dated information, highlighting the need for larger, multi-center studies to enhance generalizability and accuracy. Our project seeks to utilize modern ML algorithms with relevant modifications to improve on performance of the state-of-the-art work in the prediction of fatality in patients diagnosed with myocardial infarction.

References:

Paper 1: Ghafari R, Sorayaie Azar A, Ghafari A, Moradabadi Aghdam F, Valizadeh M, Khalili N, et al. Prediction of the Fatal Acute Complications of Myocardial Infarction via Machine Learning Algorithms. J Teh Univ Heart Ctr 2023;18(4):278-28

Paper 2: Mariana O, Joana S, Fausto JP, Roberto H and Teresa M. *Machine learning prediction of mortality in Acute Myocardial Infarction*. BMC Medical Informatics and Decision Making (2023) 23:70 https://doi.org/10.1186/s12911-023-02168-6

Paper 3: Panke C, Bine W, Li Z, Shuai M, Yanping W, Yunyue Z, Xin Z, Zhixun B and Bei S. *Machine learning for predicting intrahospital mortality in ST-elevation myocardial infarction patients with type 2 diabetes mellitus.* BMC Cardiovascular Disorders (2023) 23:585 https://doi.org/10.1186/s12872-023-03626-9

Paper 4: Bijan R, Boshra K, Hamid AM, Mohamad F. *Machine learning predicts long-term mortality after acute myocardial infarction using systolic time intervals and routinely collected clinical data*. Intelligent Medicine 4 (2024) 170–176

Paper 5: Xiaoli Z, Bojian X, Yijun C, Hanqian Z, Jinxi H. *Machine learning in the prediction of in-hospital mortality in patients with first acute myocardial infarction*. Clinica Chimica Acta 554 (2024) 117776