FAIKR/3
Bayesian Heart Failure Prediction

Mauro Dore, Leonardo Monti
Master's Degree in Artificial Intelligence, University of Bologna
{ mauro.dore, leonardo.monti3 }@studio.unibo.it

May 11, 2025

Abstract

Cardiovascular diseases (CVDs) are the number 1 cause of death globally. This mini-project aims to detect the presence of coronary heart disease and investigate the clinical conditions of the patients in order to assess the risk of developing the disease.

Introduction

Domain

Cardiovascular diseases (CVDs) are the leading global cause of death, with most fatalities due to heart attacks and strokes (Mariachiara Di Cesare 2024). Early diagnosis and identification are crucial for preventing the onset of heart failure.

We utilize the "Heart Failure Prediction" dataset, which comprises 918 unique patient records (fedesoriano 2021). The dataset includes 11 relevant features, such as demographic information (e.g. age, sex), clinical measurements (e.g., resting blood pressure, serum cholesterol, maximum heart rate), and electrocardiographic (ECG) findings (e.g., resting ECG results, ST slope).

Aim

The purpose of this project is to develop and evaluate Bayesian Network (BN) models and compare the effectiveness of four distinct BN modeling strategies at assessing the risk of suffering from CVDs and investigating the conditions that lead to it, observing the computational efficiency and runtime of the models. The model structures are: custom-built from general knowledge of the medical domain, search-based, and constraint-based.

Method

This project uses the pgmpy library (Ankan and Textor 2024) for heart disease prediction. The dataset preprocessing involves label encoding, discretization based on medical guidelines, and imputation of hidden missing values with meaningful statistics, such as the median of the available data. The data was split 80/20 into training and testing sets.

Four Bayesian Network structures were compared: one hand-crafted by the authors built upon knowledge of the clinical domain, one learned via TreeSearch, and two via HillClimbSearch using K2 score (the latter incorporating constraints). Model parameters (CPTs) were learned from the training data using Maximum Likelihood Estimation, and inference was performed using the Variable Elimination algorithm on the test set. Various queries were applied on each model to explore the differences.

Results

In terms of accuracy, the two models learned using the **"Hill Climb"** algorithm achieve the best performance based on their scores. The Tree Search model shows moderate performance, while the custom-built model performs the worst. Regarding run-time, the **"Tree Search"** model is the most efficient, followed by the Hill Climb model with constraints. The least efficient are the full Hill Climb model and the custom model, which has the longest execution times.

Model

Three out of the four Bayesian Networks (BNs) are learned from the data using a scoring function, specifically the **K2Score**, which has to be optimized during the learning process. We also tested the **Bayesian Dirichlet equivalent uniform (BDeuScore)** score.

- **Custom model**, a BN with all the features provided by the dataset. The causal links between the features have been designed by looking at the correlation plot characterizing our data and the authors' cardiology knowledge.
- **Tree model**, learned automatically from the dataset using a tree search algorithm without imposing additional structural constraints and with all the features provided by the dataset.
- **HillClimb model**, learned automatically from the dataset using hill-climb search algorithm, using all available features and without imposing structural constraints.
- **Constraint-Based** model, where the structure learning process was guided by enforcing specific constraints obtained from the study of the previous models and pruning the nodes that had the smallest correlation with Heart Disease.

The nodes in the Bayesian Network can be grouped into the following categories:

- **Root nodes** like Age and Sex, which are not influenced by anything else.

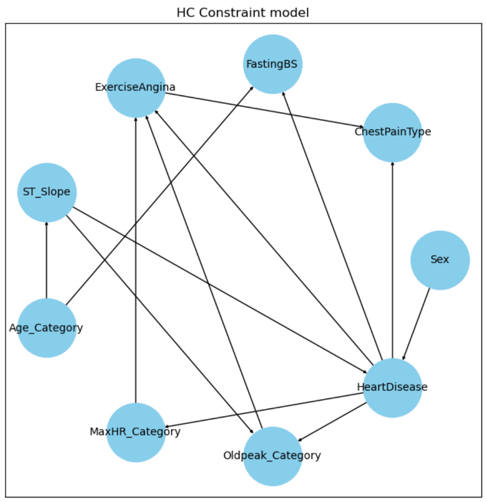

Figure 1: Bayesian network, model obtained with search and constraints

- **Clinical measurements** like RestingBP for resting blood pressure, and FastingBS that is blood sugar.
- **Stress test results** like MaxHR that is the maximum heart rate, STslope which express the difference in height between two points on the ECG signal and Oldpeak for the difference between ST slope at rest and during activity.

Analysis

Experimental setup

The first experiment was performed using the four models and a query to classify the risk of heart disease based on parameters not depending on other clinical complication. The causes are Sex and Age, and we expect the models to agree that older people are more prone to CVD.

Building on the results of the first query we investigate how the under-stress features influence the risk of HeartDesease on male elders. Knowing that these feature are predictors of ischemia and coronary artery disease. (Robert S. Finkelhor 1986) (ecg 2023). We expect a clear mapping between disease and the symptoms.

Third, we analyzed young people who are expected to have a lower probability of developing cardiovascular disease, in order to investigate the underlying causes of heart disease in this subgroup. This focus was motivated by the need to understand risk factors in cases where symptoms are mild or absent.

Lastly, we checked the accuracy of the model at straight diagnosis of heart disease from the symptoms.

Results

From the first query it appears that men are in general much more at risk of heart disease than women. All the models agree on this with a small difference of confidence. Although it is known that younger women have a lower risk profile for CVDs such a clear difference between males and females is unexpected.

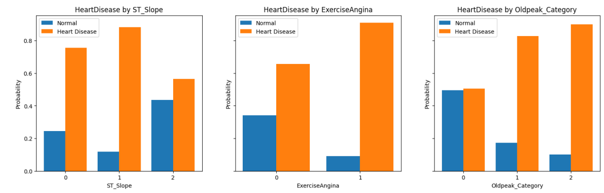

Figure 2: Results of the second query: how under stress symptoms are related to the presence of a CVD

From the second query we confirm the link between symptoms under stress and CVDs. Results show Flat and Downsloping ST segment patterns are associated with Heart Disease. Presence of Angina during exercise is a strong indicator of potential underlying heart disease, specifically coronary artery disease. High values of Oldpeak suggest ischemia. The results of the query are shown in Figure 2.

For the third query all the models agree that among young people not experiencing angina under stress is not a valid reason for excluding CVDs. In fact, the asymptomatic group is the one at highest risk.

In terms of accuracy, at the raw diagnosis, all the models produced satisfactory performances above 84%. The handmade custom model has the worst accuracy, this is expected since humans can easily overestimate or underestimate the importance of a causal link among the many possible ones in a complex domain. The HillClimb model with constraint scored above 87%.

Conclusion

In this project we show an application of BN for the diagnosis and explanation of CVDs starting from clinical features. Among the models the best performance is obtained by the Hill Climb model with constraints. This model is also the most efficient. Seeing the results and the nature of the dataset, we still advise further testing on newer data to check if the designed models work as stated and guidance from experts of the medical domain in order to provide better constraints to the Hill Climb model since we see its potential improvement.

Links to external resources

The dataset used in this project is available at https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction

References

Ankan, A., and Textor, J. 2024. pgmpy: A python toolkit for bayesian networks. *Journal of Machine Learning Research* 25(265):1–8.

2023. Evaluation of exercise stress test: Ecg, symptoms, blood pressure, heart rate, performance. Accessed: 2025-05-01.

fedesoriano. 2021. Heart failure prediction dataset. Accessed: 2025-05-01.

Mariachiara Di Cesare, e. a. 2024. The heart of the world. *Global Heart Journal*.

Robert S. Finkelhor, e. a. 1986. The st segment/heart rate slope as a predictor of coronary artery disease: Comparison with quantitative thallium imaging and conventional st segment criteria. *American Heart Journal* 112(2):296–304.