1. Introduction

This report provides a streamlined medical validation of the 14 attributes in the UCI Heart Disease dataset for use in machine learning and Bayesian Network construction. The analysis confirms that the dataset represents a valid clinical cohort of patients referred for angiography at the Cleveland Clinic in the 1980s.[1] All attributes correspond to standard physiological markers used to predict coronary artery disease (CAD).

2. Attribute Analysis

1. age

- **Definition:** Patient age in years.
- **Analysis:** The range (29–77) is plausible. Age is a primary risk factor; the risk of CAD increases significantly with age.
- **BN Connection:** Age physiologically affects thalach (Max Heart Rate) and trestbps (Blood Pressure).

2. sex

- **Definition:** 1 = Male, 0 = Female.
- **Analysis:** Standard binary coding. Males historically have a higher baseline risk of CAD than pre-menopausal females.[3]
- **BN Connection:** Sex often influences chol (Cholesterol) levels and thalach limits.

3. cp (Chest Pain Type)

- **Definition:** 1 = Typical Angina, 2 = Atypical Angina, 3 = Non-anginal Pain, 4 = Asymptomatic.

- **Analysis:** Based on the **Diamond-Forrester classification**. "Typical Angina" (Value 1) is the strongest clinical predictor of CAD. "Asymptomatic" (Value 4) accounts for silent ischemia.[4]

4. trestbps (Resting Blood Pressure)

- **Definition:** Systolic blood pressure in mm Hg (on admission).
- **Analysis:** Values (94–200 mm Hg) are plausible. Values >130/140 mm Hg indicate hypertension, which increases heart muscle strain.[5]
- **BN Connection:** Chronic high BP leads to resected abnormalities (LVH).

5. chol (Serum Cholesterol)

- **Definition:** Total cholesterol in mg/dl.
- **Analysis:** Range (126–564 mg/dl) is valid. Values >200 mg/dl are elevated; values >240 mg/dl are high. The outlier (564 mg/dl) likely indicates genetic hypercholesterolemia.[7]

6. fbs (Fasting Blood Sugar)

- **Definition:** 1 if >120 mg/dl; 0 otherwise.
- **Analysis:** A binary marker for diabetes or pre-diabetes. High blood sugar damages blood vessels.[8]
- **BN Connection:** Diabetes (fbs=1) is a major risk factor for CAD (num).

7. restecg (Resting ECG)

- **Definition:** 0 = Normal, 1 = ST-T abnormality, 2 = Left Ventricular Hypertrophy (LVH).
- **Analysis:**
  - **Value 1:** Indicates ischemia or electrolyte imbalance (threshold > 0.05 mV).[9]
  - **Value 2:** Indicates "thickening" of the heart muscle (LVH), usually caused by long-term high blood pressure (trestbps).[10]

8. thalach (Max Heart Rate)

- **Definition:** Maximum heart rate achieved during stress test (bpm).
- **Analysis:** Range (71–202 bpm) is biologically sound. Lower values in this context may indicate "Chronotropic Incompetence" (inability to raise heart rate) or beta-blocker usage.[12]
- **BN Connection:** Directly inversely correlated with age.

9. exang (Exercise Induced Angina)

- **Definition:** 1 = Yes, 0 = No.
- **Analysis:** Determines if chest pain was reproduced *during* physical stress. This is a highly specific indicator of coronary blockages.[13]

10. oldpeak (ST Depression)

- **Definition:** ST depression induced by exercise relative to rest.
- **Units: Millimeters (mm)**, not millivolts.
- **Analysis:** Range (0–6.2) is valid. This is the most critical electrical sign of ischemia (oxygen starvation) in the heart muscle. Higher values = more severe disease.[14]

11. slope (Slope of Peak Exercise ST Segment)

- **Definition:** 1 = Upsloping, 2 = Flat, 3 = Downsloping.
- **Analysis:** Describes the *shape* of the oldpeak depression.
  - **Upsloping (1):** Least severe.
  - **Downsloping (3):** Most severe/dangerous indicator of ischemia.[16]

12. ca (Fluoroscopy Vessels)

- **Definition:** Number of major vessels (0–3) colored by fluoroscopy.
- **Analysis:** Specifically counts **calcified** vessels. The 3 major coronary arteries are LAD, LCx, and RCA. A value of 3 implies widespread plaque/calcium buildup.[17]

13. thal (Thallium Stress Test)

- **Definition:** 3 = Normal, 6 = Fixed Defect, 7 = Reversible Defect.
- **Analysis:** Nuclear imaging results.
  - **Fixed (6):** Scar tissue from a *past* heart attack.
  - **Reversible (7):** Active ischemia (blood flow is blocked during exercise but returns at rest).[18]

14. num (Target Variable)

- **Definition:** Diagnosis of heart disease (angiographic).
- **Analysis:** Defined as **>50% diameter narrowing** of a major vessel.
  - 0 = No disease.
  - 1–4 = Disease present (severity increases with number).[19]

3. Conclusion

The dataset is medically plausible. The units (mm Hg for BP, mm for ST depression, mg/dl for cholesterol) are correct, and the categorical encodings align with 1980s clinical standards. For a Bayesian Network, strong dependencies should be expected between **Age -> Max Heart Rate**, **Blood Pressure -> LVH (Rest ECG)**, and **Oldpeak/Slope -> Disease Presence**.

4. Reference List

- Detrano, R. et al. (1989) 'International application of a new probability algorithm for the diagnosis of coronary artery disease', *American Journal of Cardiology*, 64(5), pp. 304–310.[1]
- Diamond, G.A. and Forrester, J.S. (1979) 'Analysis of probability as an aid in the clinical diagnosis of coronary-artery disease', *New England Journal of Medicine*, 300(24), pp. 1350–1358.[4]
- Gibbons, R.J. et al. (2002) 'ACC/AHA 2002 guideline update for exercise testing', *Circulation*, 106(14), pp. 1883–1892.[12]
- Romhilt, D.W. and Estes, E.H. (1968) 'A point-score system for the ECG diagnosis of left ventricular hypertrophy', *American Heart Journal*, 75(6), pp. 752–758.[22]
- Whelton, P.K. et al. (2018) '2017 ACC/AHA... Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults', *Journal of the American College of Cardiology*, 71(19), pp. e127–e248.[6]

Works cited

1. UCI Heart Disease Data - Kaggle, accessed on November 16, 2025, https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data
2. I. Coronary Artery Disease feature selection - Kaggle, accessed on November 16, 2025, https://www.kaggle.com/code/lourenswalters/i-coronary-artery-disease-feature-selection
3. Performance of Traditional Pretest Probability Estimates in Stable Patients Undergoing Myocardial Perfusion Imaging, accessed on November 16, 2025, https://www.ahajournals.org/doi/pdf/10.1161/CIRCIMAGING.118.008473
4. European Society of Cardiology–Recommended Coronary Artery Disease Consortium Pretest Probability Scores More Accurately Predict Obstructive Coronary Disease and Cardiovascular Events Than the Diamond and Forrester Score | Circulation, accessed on November 16, 2025, https://www.ahajournals.org/doi/10.1161/circulationaha.116.023396
5. New BP guideline: 5 things physicians should know - American Medical Association, accessed on November 16, 2025, https://www.ama-assn.org/public-health/prevention-wellness/new-bp-guideline-5-things-physicians-should-know
6. 2017 Guideline for High Blood Pressure in Adults - American College of Cardiology, accessed on November 16, 2025, https://www.acc.org/Latest-in-Cardiology/ten-points-to-remember/2017/11/09/11/41/2017-Guideline-for-High-Blood-Pressure-in-Adults
7. Lipid Panel | Johns Hopkins Medicine, accessed on November 16, 2025, https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/lipid-pa

nel
8. Diagnosing Diabetes With Glucose Criteria: Worshipping a False God - PMC - NIH, accessed on November 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC3024380/
9. Clinical utility of aVR lead T-wave in electrocardiogram of patients with ST-elevation myocardial infarction - NIH, accessed on November 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC8555143/
10. The Use of Romhilt-Estes Criteria in the Presumptive Electrocardiographic Diagnosis of Left Ventricular Hypertrophy in Comparison to Voltage-Based Criteria - PMC - NIH, accessed on November 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC9470005/
11. Strain as Predictor of Outcomes: ECG LVH Criteria beyond Voltage - GE Healthcare, accessed on November 16, 2025, https://www.gehealthcare.com/insights/article/strain-as-predictor-of-outcomes-ecg-lvh-criteria-beyond-voltage
12. Exercise Standards for Testing and Training | Circulation - American Heart Association Journals, accessed on November 16, 2025, https://www.ahajournals.org/doi/10.1161/cir.0b013e31829b5b44
13. Angina - Wikipedia, accessed on November 16, 2025, https://en.wikipedia.org/wiki/Angina
14. ST depression - Wikipedia, accessed on November 16, 2025, https://en.wikipedia.org/wiki/ST_depression
15. ST Segment - StatPearls - NCBI Bookshelf - NIH, accessed on November 16, 2025, https://www.ncbi.nlm.nih.gov/books/NBK459364/
16. The Value of Upsloping ST Depression in Diagnosing Myocardial Ischemia - PMC - NIH, accessed on November 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC6932726/
17. Machine Learning Approach for Cardiovascular Risk and Coronary Artery Calcification Score - PMC - PubMed Central, accessed on November 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC9246606/
18. Stress-induced reversible and mild-to-moderate irreversible thallium defects: are they equally accurate for predicting recovery of regional left ventricular function after revascularization? - PubMed, accessed on November 16, 2025, https://pubmed.ncbi.nlm.nih.gov/9714106/
19. Angiographic Disease Progression and Residual Risk of Cardiovascular Events While on Optimal Medical Therapy | Circulation - American Heart Association Journals, accessed on November 16, 2025, https://www.ahajournals.org/doi/10.1161/circinterventions.110.960062
20. International application of a new probability algorithm for the diagnosis of coronary artery disease | CoLab, accessed on November 16, 2025, https://colab.ws/articles/10.1016/0002-9149%2889%2990524-9
21. Peer Review File Article information: https://dx.doi.org/10.21037/jmai-25-47 Reviewer A The potentially interesting and importan - AME Publishing Company, accessed on November 16, 2025, https://cdn.amegroups.cn/journals/jlpm/files/journals/33/articles/10222/public/102

22-PB2-4618-R1.pdf?filename=JMAI-25-47-Peer+Review+File.pdf&t=1761037452
22. The Romhilt-Estes electrocardiographic score predicts sudden cardiac arrest independent of left ventricular mass and ejection fraction - NIH, accessed on November 16, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC5495629/