Heart Disease Diagnosis (UCI) — Decision Tree Report

ZIP file: /content/heart+disease (1).zip

Dataset inside ZIP: processed.cleveland.data

Target: diagnosis = num $\in$ {0,1,2,3,4}

Train/Test split: 75/25, stratified, random_state=42

Model:

- DecisionTreeClassifier(max_depth=4, min_samples_leaf=5, class_weight=balanced)

Dataset Overview

Samples: 303

Features: 13

Class distribution:
- Class 0: 164 (54.13%)
- Class 1: 55 (18.15%)
- Class 2: 36 (11.88%)
- Class 3: 35 (11.55%)
- Class 4: 13 (4.29%)
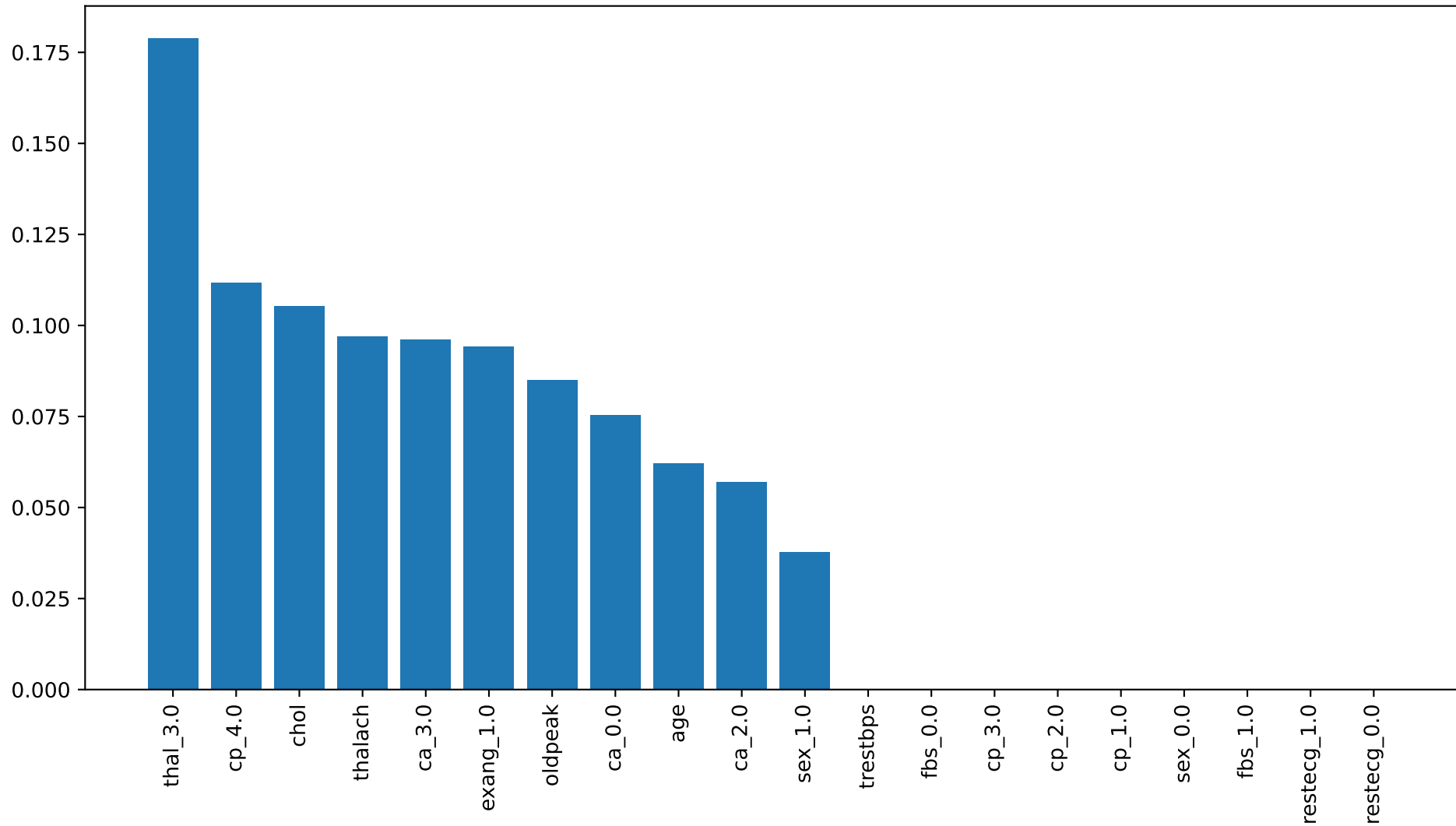
Evaluation (Test Set)

Accuracy: 0.4605

Balanced accuracy: 0.3316

Macro F1: 0.3073

Classification report:

	precision	recall	f1-score	support
0	0.897	0.634	0.743	41
1	0.250	0.357	0.294	14
2	0.182	0.222	0.200	9
3	0.500	0.111	0.182	9
4	0.071	0.333	0.118	3
accuracy			0.461	76
macro avg	0.380	0.332	0.307	76
weighted avg	0.613	0.461	0.505	76

Top Feature Importances

Decision Tree (Top Levels)