

Feature Name	Feature specification	Categories	No Heart Disease	Heart Disease	Total
age	Age	24-39	23	45	68
		40-54	155	283	438
		55-69	318	181	499
		70-85	3	17	20
		Total	499	526	1025
sex	Sex	Male	413	300	713
		Female	86	226	312
		Total	499	526	1025
cp	Chest Pain	1 = Typical angina	375	122	497
		2 = Atypical angina	33	134	167
		3 = Non-anginal pain	65	219	284
		4 = Asymptomatic	26	51	77
		Total	499	526	1025
Trestbps	Resting blood pressure	90-119	138	191	329
		120-149	287	294	581
		150-179	67	41	108
		180-210	7	0	7
		Total	499	526	1025
chol	Cholesterol	120-230	189	247	436
		231-341	299	264	563
		342-453	11	12	23
		454-564	0	3	3
		Total	499	526	1025
fbs	Fasting blood sugar	<120mg/dl	417	455	872
		>120mg/dl	82	71	153
		Total	499	526	1025
Restecg	Electrocardiographic results when testing	0=Normal	283	214	497
		1 = ST-T wave abnormality	204	309	513
		2 = LV hypertrophy	12	3	15
		Total	499	526	1025
Thalach	Maximum heart rate	70-104	40	6	46
		105-139	197	77	274
		140-174	248	357	605
		175-209	14	86	100
		Total	499	526	1025
Exang	Angina that is caused by anxiety	No	225	455	680
		Yes	274	71	345
		Total	499	526	1025
Oldpeak	ST depression due to exercise compared to rest	0-1.54	265	461	726
		1.55-3.00	167	59	226
		3.01-4.64	60	6	66
		4.65-6.2	7	0	7
		Total	499	526	1025
Slope	slope of the peak exercise ST segment	1 = Up sloping	46	28	74
		2 = Flat	324	158	482
		3 = Down sloping	129	340	469
		Total	499	526	1025
ca	Number of major blood vessels	0	163	415	578
		1	160	66	226
		2	113	21	134
		3	60	9	69
		4	3	15	18
		Total	499	526	1025
Thal	Thalassemia	0	4	3	7
		Normal	43	21	64
		Fixed defect	132	412	544
		Reversible	320	90	410
		Total	499	526	1025

Table 1. Description of the dataset.

Feature	Mann Whitney U test p-value	Chi Square test p-value
age	<0.001	<0.001
sex	<0.001	<0.001
cp	<0.001	<0.001
trestbps	<0.001	<0.001
chol	<0.001	<0.001
fbs	0.188	0.219
restecg	<0.001	<0.001
thalach	<0.001	<0.001
exang	<0.001	<0.001
oldpeak	<0.001	<0.001
slope	<0.001	<0.001
ca	<0.001	<0.001
thal	<0.001	<0.001

Table 2. Mann Whitney U test and Chi squared test outcomes.

Features	Value of r	Degree of Correlation
cp	0.435	+ Moderate
thalach	0.423	+ Moderate
slope	0.346	+ Moderate
restecg	0.134	No Correlation
fbs	-0.041	No Correlation
chol	-0.1	No Correlation
trestbps	-0.139	No Correlation
age	-0.229	No Correlation
sex	-0.28	No Correlation
thal	-0.338	- Moderate
ca	-0.382	- Moderate
exang	-0.438	- Moderate
oldpeak	-0.438	- Moderate

Table 3. Pearson correlation outcomes.

Features	Value of r	Degree of Correlation
cp	0.465	+ Moderate
thalach	0.43	+ Moderate
slope	0.369	+ Moderate
restecg	0.147	No Correlation
fbs	-0.041	No Correlation
trestbps	-0.115	No Correlation
chol	-0.133	No Correlation
age	-0.24	No Correlation
sex	-0.28	No Correlation
thal	-0.399	- Moderate
oldpeak	-0.438	- Moderate
exang	-0.438	- Moderate
ca	-0.453	- Moderate

Table 4. Spearman correlation outcomes.

	Permutation RF	Permutation CART	Permutation KNN	Decision Tree	Random Forest	XGBoost	Permutation XGBoost
age	0.019	0.087	0.022	0.081	0.086	0.039	0.059
sex	0.012	0.025	0	0.035	0.029	0.056	0.035
cp	0.061	0.184	0	0.273	0.132	0.249	0.099
trestbps	0	0.043	0.051	0.06	0.068	0.03	0.006
chol	0.005	0.077	0.091	0.069	0.08	0.031	0.027
fbs	-0.003	0	0	0	0.01	0.005	-0.003
restecg	-0.003	0	0	0	0.018	0.025	0.018
thalach	0.017	0.066	0.162	0.069	0.125	0.036	0.034
exang	0.009	0.027	0	0.016	0.042	0.069	0.019
oldpeak	0.024	0.107	0.002	0.087	0.123	0.062	0.049
slope	0.004	0.023	0	0.034	0.051	0.063	0.006
ca	0.068	0.074	0.001	0.123	0.123	0.131	0.099
thal	0.033	0.123	0	0.154	0.113	0.203	0.06

Table 5. Importance of features of the dataset with ML models.

Feature	Total Counts	Order of Features
cp	80	1
ca	74	2
thal	71	3
oldpeak	67	4
age	62	5
thalach	59	6
chol	50	7
sex	38	8
trestbps	37	9
exang	37	10
slope	31	11
restecg	16	12
fbs	11	13

Table 6. Borda counts outcome.

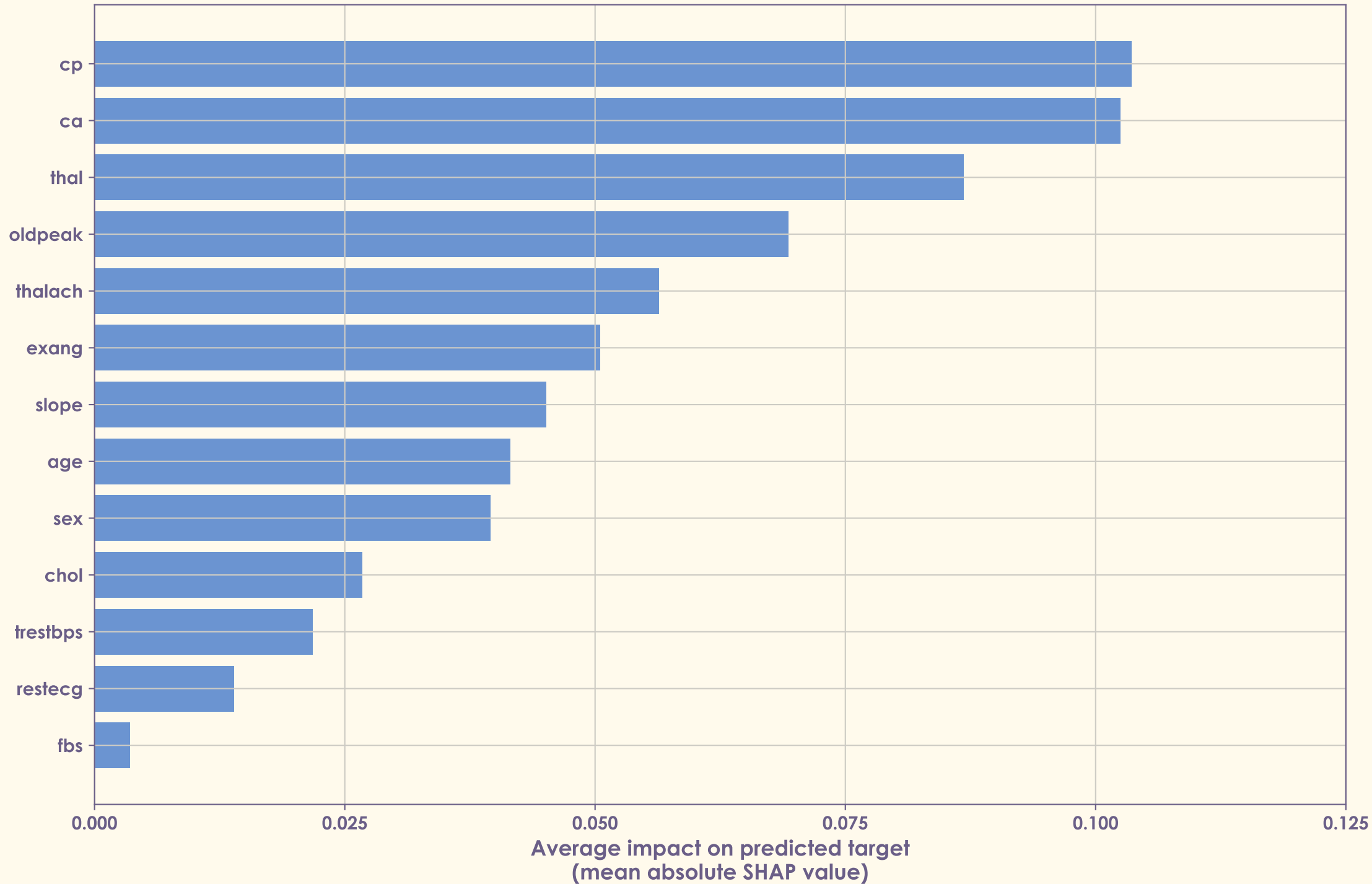
Feature	LightBoost	CatBoost	Random Forest	XGBoost	Gradient Boost
age	0.74	0.74	0.74	0.74	0.74
sex	0.727	0.744	0.744	0.744	0.737
cp	0.844	0.857	0.857	0.857	0.847
trestbps	0.903	0.886	0.906	0.886	0.906
chol	0.916	0.906	0.964	0.942	0.906
fbs	0.935	0.932	0.961	0.955	0.916
restecg	0.974	0.909	0.981	0.981	0.948
thalach	0.955	0.925	0.99	0.971	0.942
exang	0.977	0.919	0.981	0.971	0.955
oldpeak	0.968	0.935	0.981	0.99	0.964
slope	0.955	0.942	0.981	0.981	0.951
ca	0.971	0.945	0.99	0.981	0.964
thal	0.977	0.945	0.981	0.981	0.955

Table 7. Importance of features of the dataset.

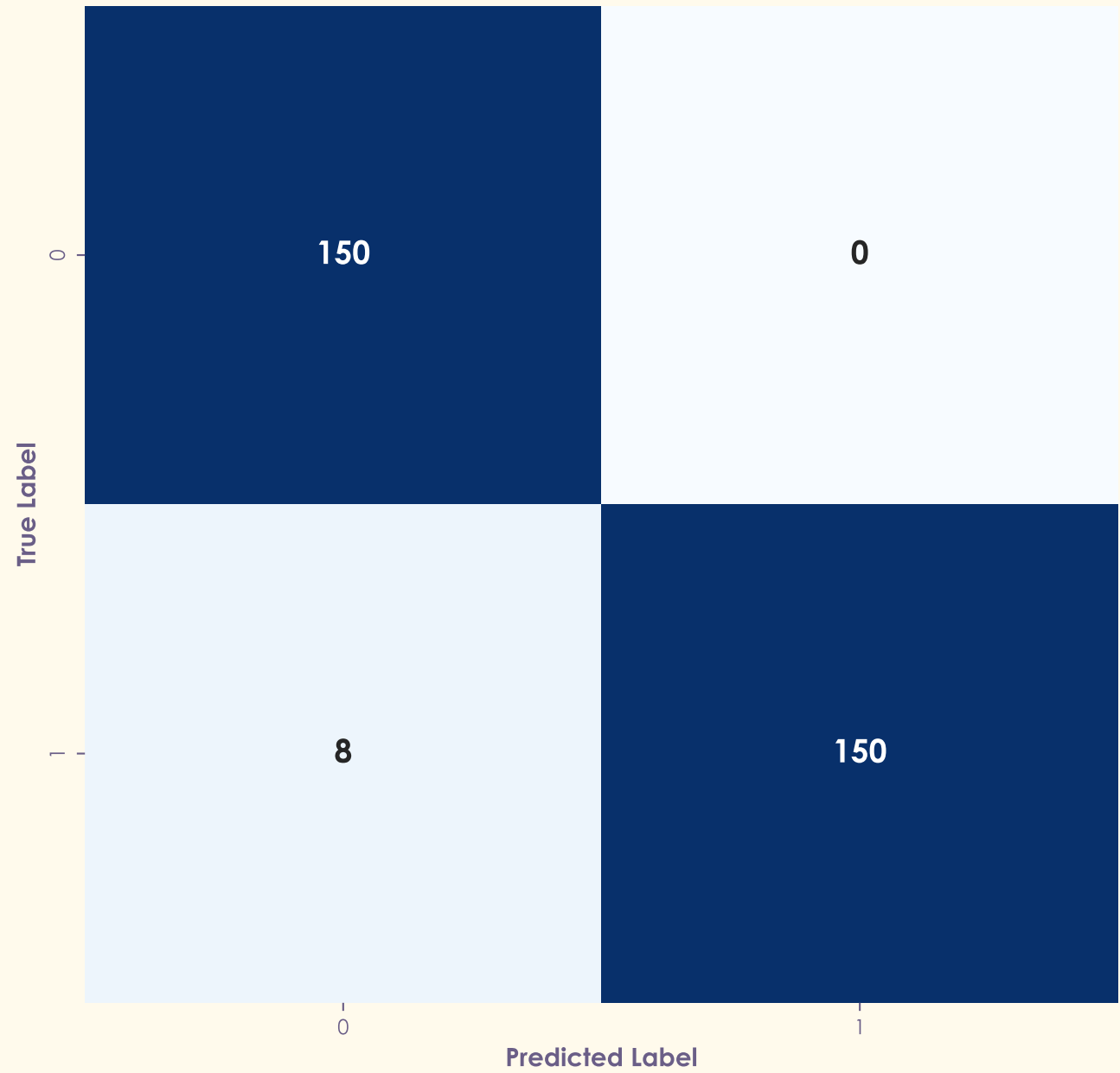
	precision	recall	f1-score	support
No Heart Disease	0.95	1	0.97	150
Heart Disease	1	0.95	0.97	158
weighted avg	0.98	0.97	0.97	308

Table 8. Classification Report.

SHAP value-based Feature Importance



Confusion Matrix



Classifier Accuracy vs Number of Features

