Name:Fahad Iqbal SID:220708016

ECS784P:Heart Disease Data Analysis

Question 1

This project explores cardiovascular health, aiming to identify causal relationships between clinical indicators and heart disease using structured learning methods.

The dataset from Kaggle includes 1,025 patient records with 14 variables. (Gangal, 2020) There was a preprocessing step where Continuous features like age and cholesterol are discretized, and no missing values remain. This dataset is -suited for structure learning because of its categorical variables and sufficient sample size. And because all variables are categorical, meeting Bayesys requirements. Finally, it reflects a real medical problem where uncovering dependencies (e.g., cholesterol's effect on heart disease) is valuable.

	age	sex	chest_pain_type	resting_blood_pressure	cholestoral	fasting_blood_sugar	rest_ecg	max_heart_rate	exercise_induced_angina	oldpeak	slope	vessels_colored_by_flourosopy	thalassemia	target
0	51to60	Male	Typical angina	normal	borderline	Lower than 120 mg/ml	ST-T wave abnormality	vhigh	No	low	Downsloping	Two	Reversable Defect	0
1	51to60	Male	Typical angina	normal	borderline	Greater than 120 mg/ml	Normal	high	Yes	moderate	Upsloping	Zero	Reversable Defect	0
2	61plus	Male	Typical angina	elevated	normal	Lower than 120 mg/ml	ST-T wave abnormality	moderate	Yes	moderate	Upsloping	Zero	Reversable Defect	0
3	61plus	Male	Typical angina	elevated	borderline	Lower than 120 mg/ml	ST-T wave abnormality	vhigh	No	low	Downsloping	One	Reversable Defect	0
4	61plus	Female	Typical angina	normal	vhigh	Greater than 120 mg/ml	ST-T wave abnormality	moderate	No	mild	Flat	Three	Fixed Defect	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1020	51to60	Male	Atypical angina	normal	borderline	Lower than 120 mg/ml	ST-T wave abnormality	vhigh	Yes	low	Downsloping	Zero	Fixed Defect	1
1021	51to60	Male	Typical angina	normal	high	Lower than 120 mg/ml	Normal	high	Yes	moderate	Flat	One	Reversable Defect	0
1022	41to50	Male	Typical angina	low	high	Lower than 120 mg/ml	Normal	moderate	Yes	low	Flat	One	Fixed Defect	0
1023	41to50	Female	Typical angina	low	high	Lower than 120 mg/ml	Normal	high	No	low	Downsloping	Zero	Fixed Defect	1
1024	51to60	Male	Typical angina	low	normal	Lower than 120 mg/ml	ST-T wave abnormality	moderate	No	mild	Flat	One	Reversable Defect	0

1025 rows × 14 columns

Figure 1 :trainindata.csv after preprocessing

Name:Fahad Iqbal SID:220708016

Question 2

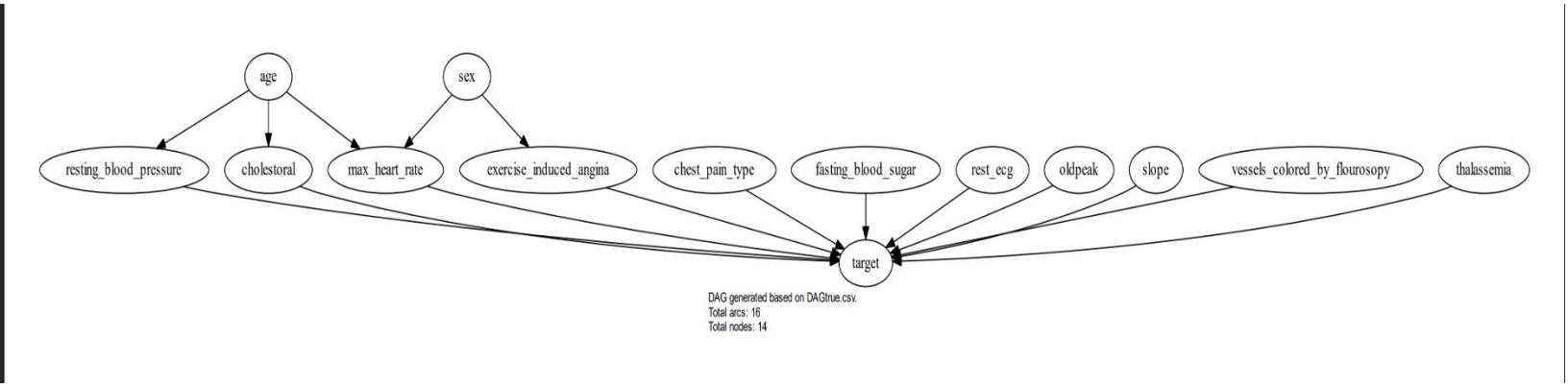

Figure 2: DAGtrue **(zoom in for better view)**

The structure of the DAGtrue graph is informed by both general cardiovascular knowledge and research evidence. As shown in the figure, assumptions such as age influencing resting blood pressure, cholesterol, and maximum heart rate are based on established domain knowledge and clinical reasoning. Similarly, modeling cholesterol as a cause of heart disease (target) reflects a well-known and straightforward causal relationship. Research also contributed to the knowledge graph: for example, exercise-induced angina is modeled as influencing heart disease, consistent with findings that describe it as “a relatively uncommon clinical scenario characterized by the occurrence of chest symptoms on exertion due to coronary vasospasm” (Tamura et al., 2019). Additionally, the relationship between thalassemia and heart disease is supported by literature stating, “Heart disease is the leading cause of mortality and one of the main causes of morbidity in β-thalassemia” (Aessopos, Kati & Farmakis, 2005).

Name:Fahad Iqbal SID:220708016

Question 3

Algorithm	CPDAG BSF	CPDAG SHD	CPDAG F1	Log-Likelihood (LL) score	BIC score	# free parameters	Structure learning elapsed time
HC	0.21	20.5	0.314	-17392.326	-18132.43	148.0	0
TABU	0.197	21.5	0.306	-17370.408	-18130.515	152.0	0
SaiyanH	0.179	21.0	0.286	17340.292	-18195.413	171.0	0
MAHC	0.192	20.0	0.303	-17468.778	18148.874	136.0	0
GES	0.21	20.5	0.314	17392.326	-18132.43	148.0	0

I compared CPDAG scores (BSF, SHD, F1) from my structure learning experiments to results from the Bayesys manual's Asia and Sports networks (8–9 nodes, 8–15 arcs). My real-world dataset, with 14 variables and 16 arcs, is of comparable scale but somewhat more complex. Consistently, my scores were lower than those reported in the manual: my best F1 score was 0.314 (HC and GES), versus 0.615–0.875 in the benchmarks; SHD values were around 20.5, much higher than the manuals 3.0–9.0 range; and my BSF scores did not exceed 0.210, while manuals often surpassed 0.600.

These results are expected, as my data(sourced from a real medical domain) contains more noise, possible latent confounders, and non-linear relationships than synthetic benchmark datasets. Preprocessing steps like discretizing continuous variables may have further introduced bias or information loss, contributing to reduced performance. The relatively high number of arcs (16) also increased model complexity, expanding the search space and making the algorithms more susceptible to false positives and incomplete discoveries—factors that likely raised SHD and lowered precision . However, certain performance trends remained consistent: HC and GES still outperformed SaiyanH-

on small-scale problems, supporting the idea that, while real-world data lowers absolute scores, the common strengths and weaknesses of structure learning algorithms largely persist.

Question 4

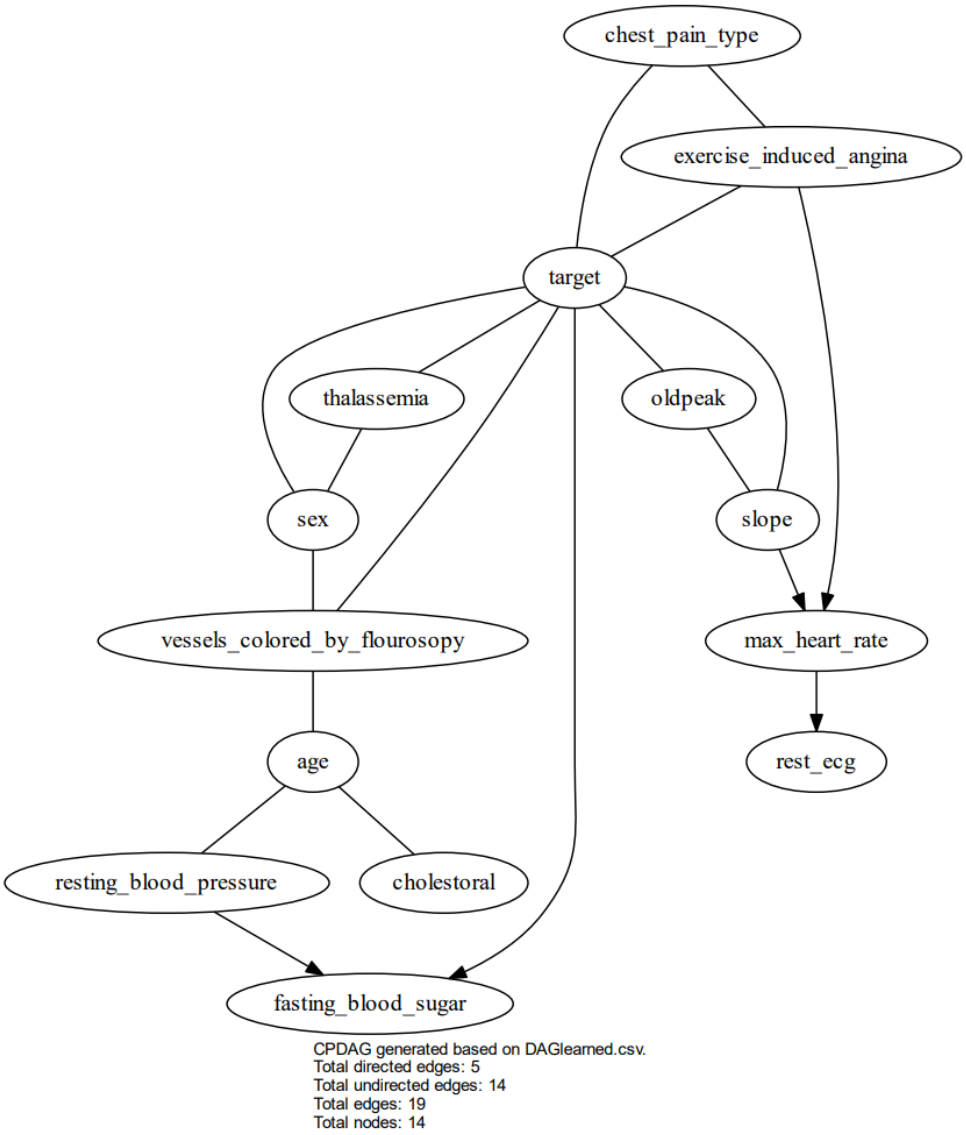

Figure 3:HC CPDAGlearned

Name:Fahad Iqbal SID:220708016

In the figure for CPDAG, all three of the causal clauses are evident. Firstly, an example of the Causal Chain is slope → max_heart_rate → rest_ecg. This suggests that the slope of the ST segment during exercise influences the maximum heart rate achieved, which in turn affects the resting ECG results. This reflects the expected result based on common knowledge.

In regard to Common Cause, age acts as a common cause, influencing both resting_blood_pressure and cholesterol. This implies that age is a factor in both blood pressure and common cholesterol which is also expected based on medical knowledge.

Finally with the Common Effect , A notable collider is target(heart diesease), which is influenced by both thalassemia and chest_pain_type. This supports the hypothesis that heart disease may arise due to the combined effects of blood disorder and chest pain symptoms.

Name:Fahad Iqbal SID:220708016

Question 5 – Score Comparisons

Rank	Your rankings			Rankings according to the Bayesys manual		
	BSF [single score]	SHD [single score]	F1 [single score]	BSF [average score]	SHD [average score]	F1 [average score]
1	HC[0.210]	MAHC (20.000)	HC[0.314]	SaiyanH [0.516]	MAHC [44.6]	SaiyanH [0.584]
2	GES[0.210]	SaiyanH (21.000)	GES [0.314]	TABU (0.515)	TABU[49.21]	TABU [0.569]
3	MAHC[0.192]	GES (20.500)	TABU [0.306]	HC (0.514]	HC [49.46]	HC [0.567]
4	TABU[0.197]	HC (20.500)	MAHC[0.303]	GES[0.505]	GES [50.56]	MAHC [0.562]
5	SaiyanH[0.179]	TABU (21.500)	SaiyanH [0.286]	MAHC (0.487)	SaiyanH [55.22]	GES [0.557]

The algorithm rankings derived from my results diverge from those presented in the Bayesys manual, particularly concerning SaiyanH, which ranked highest in the manual for BSF and F1 but performed worst for these two metrics in my analysis. Conversely, HC and GES, which ranked around 3rd and 4th in the manual, achieved the best scores in my dataset, exhibiting the highest F1 and BSF and slightly lower SHD. MAHC's performance, however, showed some alignment with the manual, maintaining the best SHD score. This outcome was anticipated, as the manual's rankings are averaged across diverse datasets of different sizes and nodes, whereas my analysis focused on a real-world heart disease dataset with 1000 sample size and 14 Nodes.

Algorithms such as SaiyanH, optimized for more intricate and complex graphs, may have been overfitted when applied to this simpler structure, resulting in diminished performance. In contrast, HC and GES, which rely on simpler assumptions, such as greedy hill climbing, were better suited to the characteristics of my dataset. These findings support that structure learning performance is largely data dependent.

Name:Fahad Iqbal SID:220708016

Question 6

The observed structure learning runtimes (0 seconds) contrast the Bayesys manual's average in Table 3.1. This differemce is expected due to my dataset's simplicity: 14 variables versus larger benchmark datasets like Pathfinder (109 nodes, 195 edges) and ForMed (88 nodes, 138 edges), which skew the average. Fewer nodes significantly reduce the search space, enabling rapid pruning and HC and GES use faster greedy approaches on small graphs. Compared to larger datasets my learning times are significantly shorter. However , when compared to datasets of similar sizes and nodes to mine like Sports and Asia, the runtime is 0 seconds.

Question 7

Algorithm	Task 3 Results			Algorithm	Task 4 Results		
	BIC Score	Log-Likelihood	Free Parameters		BIC Score	Log-Likelihood	Free Parameters
My knowledge-based graph	-4627300.772	-18251.82	921680	HC	-18132.430	- 17392.326	148
				TABU	-18130.515	-17370.408	152
				SaiyanH	-18195.413	-17340.292	171
				MAHC	-18148.874	-17468.778	136
				GES	-18132.430	-17392.326	148

Name:Fahad Iqbal SID:220708016

When comparing the BIC and Log-Likelihood scores of my knowledge-based graph to those of the five structure learning algorithms, a clear difference is seen. My knowledge-based model achieves a much lower BIC score (−4,627,300.772) and a competitive Log-Likelihood (−18,251.82), which at first suggests a better fit to the data. The LLH for my model outperforms even HC and GES(-17392.326). However, this performance is largely due to the model's extremely high complexity, as evidenced by its 921,680 free parameters, thousands of times larger than the 136–200 parameters used in the Structured Learning models. This is because I fully specified many directed relationships from domain knowledge and research.

In contrast to my graph, the algorithmically learned graphs are far more computationally efficient and still achieve reasonable data fit as seen in the attached rankings table. These findings display the trade-off: while my expert models may capture domain knowledge and fit the training data closely, they risk overfitting and inefficiency. The learned models, though simpler often generalize better and are more practical for real-world use.

Name:Fahad Iqbal SID:220708016

Question 8

Knowledge Approach	CPDAG Scores			LL	BIC	Free Parameters	Number Of Edges	Runtime(seconds)
	BSF	SHD	F1					
Without Knowledge	0.210	20.5	0.314	-17392.326	-18132.430	148	19	0
Directed Edge	0.335	18.5	0.417	-17373.433	-18153.542	156	20	0
Forbidden Edge	0.224	19.5	0.324	-17425.486	-18145.587	144	18	0

I selected two approaches: Directed and Forbidden . These methods were recorded above along with the default HC results.

Directed:

ID	Variable 1	Variable 2
1	age	cholestoral
2	sex	exercise_induced_angina
3	chest_pair	target

For the Directed approach, I chose constraints based on common assumptions such as age → cholesterol and chest_pain_type → target (heart disease). The constraint sex → exercise_induced_angina is supported by research indicating that sex influences the frequency of exercise-induced angina. (Reynolds et al., 2020).

Name:Fahad Iqbal SID:220708016

Forbidden:

ID	Variable 1	Variable 2
1	target	age
2	target	sex
3	target	cholestoral

For the Forbidden approach, I restricted three arcs that were biologically implausible or unrealistic for this dataset. The constraints target → age and target → sex are justified because heart disease cannot determine a person's age or sex(these are attributes present before disease onset). Target → cholesterol was also forbidden, as cholesterol is measured prior to diagnosis and serves as a predictor for heart disease.

Applying the Directed knowledge approach led to clear improvements across all major metrics: the F1 score increased from 0.314 to 0.417, SHD decreased, and the BSF score rose from 0.210 to 0.335. This demonstrates that enforcing key causal edges from my Task 3 knowledge graph helped the learner recover a structure more closely aligned with the true DAG. A slight increase in free parameters and edges shows the greater complexity from the enforced directions.

The Forbidden knowledge approach produced more modest gains. By prohibiting implausible arcs, it slightly reduced false positives, resulting in minor improvements in F1 (0.324) and BSF (0.224), and a reduction in parameters (144 vs. 148), making the model more compact. However, the Log-Likelihood decreased slightly as constraining the search space can improve generalization but limits model fit.

Overall, these results were as anticipated. Directed constraints was the strongest approach by guiding the structure learning process, while Forbidden constraints helped avoid misleading dependencies. Both methods show that using knowledge is a superior approach.

Name:Fahad Iqbal SID:220708016

Question 9

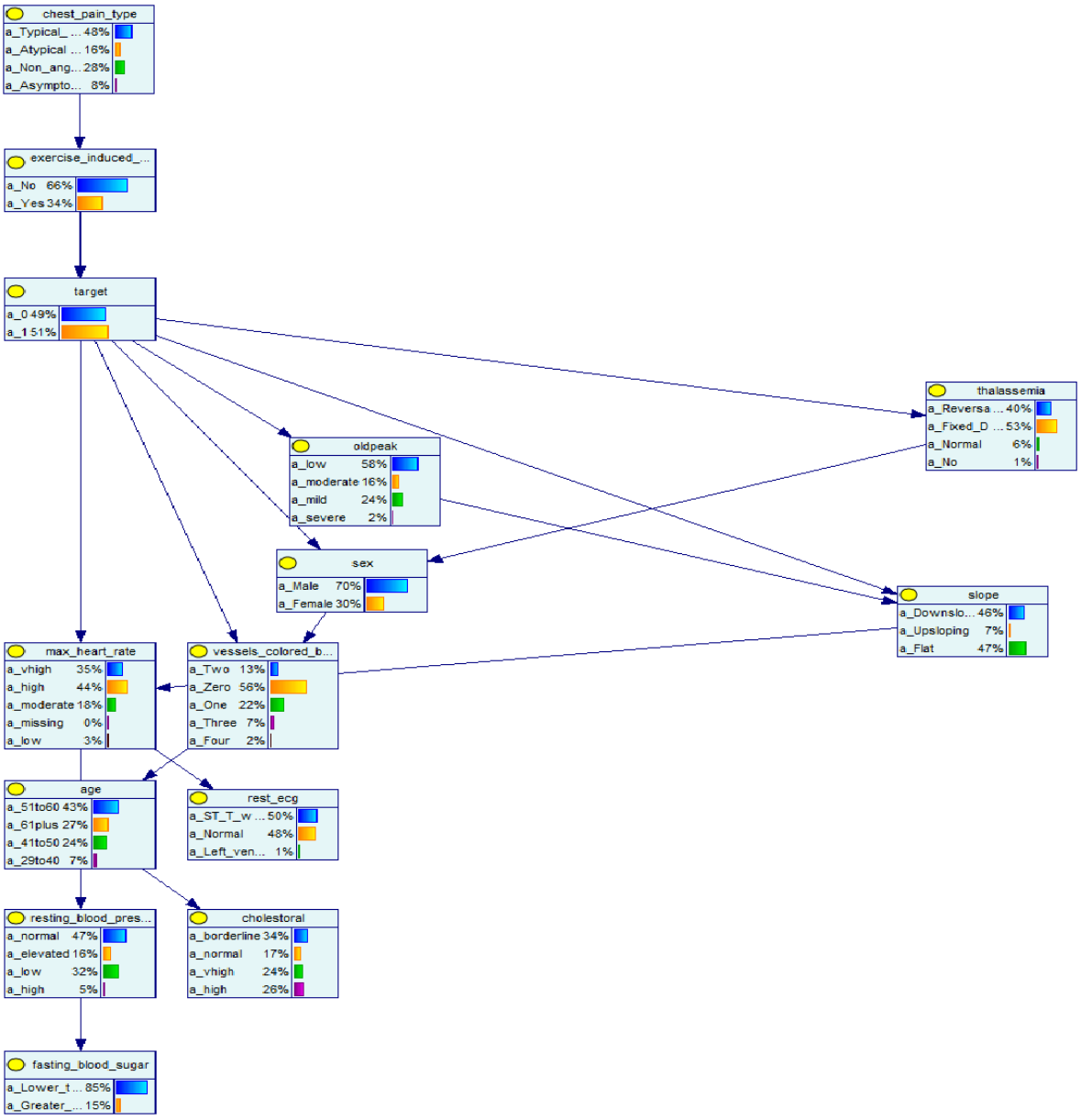

Figure 4:Network Visualisation

Following the instructions in the Bonust task, I selected a structure learned using HC algorithm.

Name:Fahad Iqbal SID:220708016

Accuracy:

Using 10-fold cross-validation, the Bayesian Network was validated on four nodes: chest_pain_type, exercise_induced_angina, oldpeak, and target. The overall accuracy across all nodes was 0.6956 (2852 correct out of 4100). Individually, the model performed best on the target node with an accuracy of 0.8283, followed by exercise_induced_angina (0.7649), oldpeak (0.6293), and chest_pain_type (0.56). This proves that the model is highly effective at predicting heart disease presence (target) and moderately reliable on other heart health indicators.

Confusion Matrices:

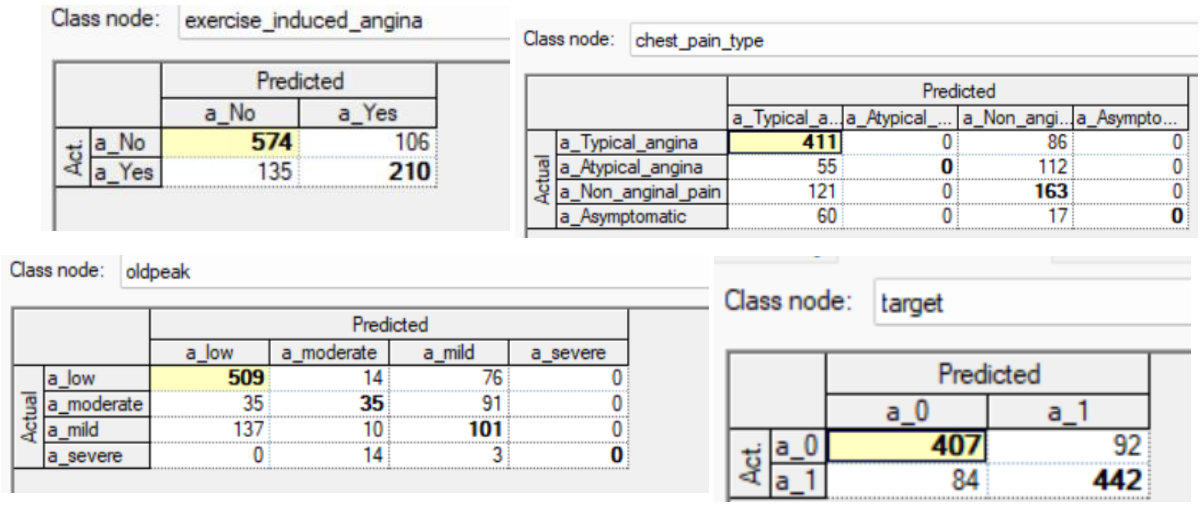

Figure 5: Confusion Matrices

The confusion matrices reveal more insights. For target, both classes (a_0 and a_1) are well predicted (407/499 and 442/526 correct, respectively). For exercise_induced_angina, the "No" class is predicted with 84% accuracy (574/680), while the "Yes" class reaches 61% (210/345). In contrast, chest_pain_type shows significant misclassification in a_Atypical_angina and a_Asymptomatic, both of which had 0 correct predictions. Similarly, oldpeak shows strong performance on a_low (509/599 correct) but poor accuracy for moderate, mild, and severe levels due to overlapping characteristics in the data.

Name:Fahad Iqbal SID:220708016

ROC Curves:

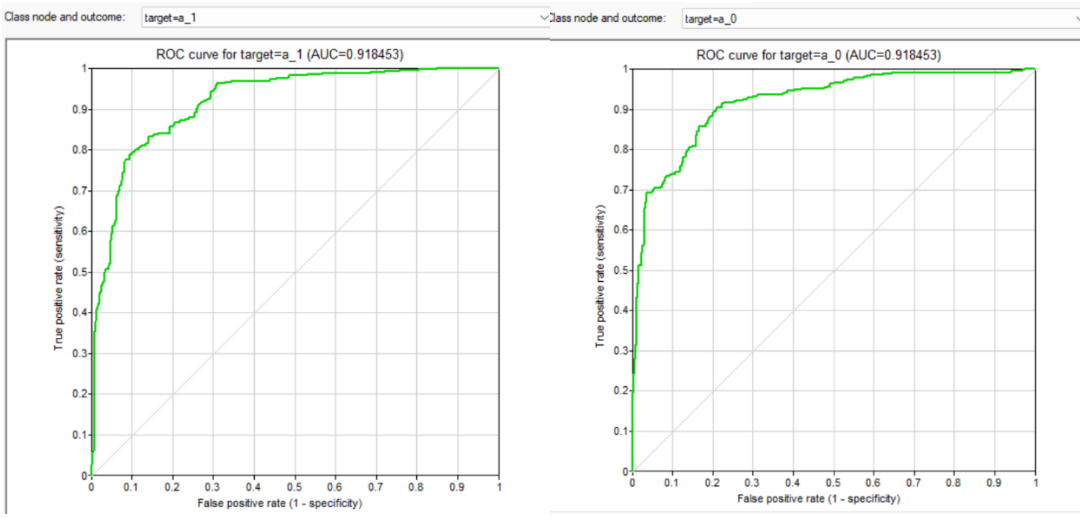

Figure 6: Target ROC

ROC curves for the target node demonstrate strong classification ability for both classes, each with an AUC of 0.918. This reflects excellent model performance with high sensitivity and specificity and shows the model reliably distinguishes between those with and without heart disease. The ROC curve's steep incline and proximity to the top-left corner again support the confusion matrix with minimal false positives and high true positive rates.

Bibliography:

Gangal, K. (2020) *Heart Disease Dataset UCI* Available at: https://www.kaggle.com/datasets/ketangangal/heart-disease-dataset-uci

Tamura, A., Nagao, K., Inada, T. and Tanaka, M., 2019. *Exercise-induced vasospastic angina with prominent ST elevation: a case report*. European Heart Journal - Case Reports, 3(1), ytz006. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6426112/

Aessopos, A., Kati, M. and Farmakis, D., 2005. **Heart disease in thalassemia intermedia: a review of the underlying pathophysiology**. *International Journal of Cardiology*, 105(2), pp.179–184. Available at https://doi.org/10.1016/j.ijcard.2004.12.010

Reynolds, H.R., Shaw, L.J., Min, J.K., Spertus, J.A., Chaitman, B.R., Berman, D.S., Picard, M.H., Kwong, R.Y., Bairey-Merz, C.N., Cyr, D.D., Lopes, R.D., Lopez-Sendon, J.L., Held, C., Szwed, H., Senior, R., Gosselin, G., Nair, R.G., Elghamaz, A., Bockeria, O., Chen, J., Chernyavskiy, A.M., Bhargava, B., Newman, J.D., Hinic, S.B., Jaroch, J., Hoye, A., Berger, J., Boden, W.E., O'Brien, S.M., Maron, D.J., Hochman, J.S. & ISCHEMIA Research Group (2020) 'Association of Sex With Severity of Coronary Artery Disease, Ischemia, and Symptoms Among Patients With Stable Ischemic Heart Disease: A Secondary Analysis of the ISCHEMIA Trial', JAMA Cardiology, 5(7), pp. 777–786. .Available at

https://pubmed.ncbi.nlm.nih.gov/32227128/