Data Quality Report

Team ID: SWTID1720067156

July 10, 2024

Data Collection and Preprocessing Phase

Team ID: SWTID1720067156

Project Title: Lymphography Classification Tool

Maximum Marks: 2 Marks

Data Quality Report

The Data Quality Report will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Raw Data Sources Report

Source Name	Description	Location/URL	Format	Size
UCI	The dataset comprises patient details (age,	UCI	CSV	3.5 kB
	lymph nodes status), and lymphography	Lymphography		
	conditions (normal, metastases, malign	Dataset		
	lymph, fibrosis).			

Attribute Information

— NOTE: All attribute values in the database have been entered as numeric values corresponding to their index in the list of attribute values for that attribute domain as given below.

• class: normal find, metastases, malign lymph, fibrosis

• lymphatics: normal, arched, deformed, displaced

• block of affere: no, yes

• bl. of lymph. c: no, yes

• bl. of lymph. s: no, yes

• by pass: no, yes

• extravasates: no, yes

• regeneration of: no, yes

• early uptake in: no, yes

• lym.nodes dimin: 0-3

• lym.nodes enlar: 1-4

• changes in lym.: bean, oval, round

 \bullet defect in node: no, lacunar, lac. marginal, lac. central

• changes in node: no, lacunar, lac. margin, lac. central

• changes in stru: no, grainy, drop-like, coarse, diluted, reticular, stripped, faint

• special forms: no, chalices, vesicles

dislocation of: no, yesexclusion of no: no, yes

• no. of nodes in: 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, ξ =70

Missing Attribute Values

None

Class Distribution

Class	Number of Instances	
normal find	2	
metastases	81	
malign lymph	61	
fibrosis	4	