Prepared by Group 5

A Comparative Study of Machine Learning Approaches for Early Cardiac Disease Prediction

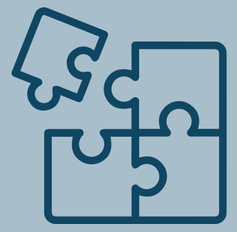
2 October, 2025

F21DL-Data Mining and Machine Learning

HERIOT WATT UNIVERSITY

About Dataset

- The dataset includes 70,000 patient records with variables such as age, gender, blood pressure, cholesterol, glucose levels, lifestyle habits, and diagnosis — ideal for exploring clinical and behavioral correlations.

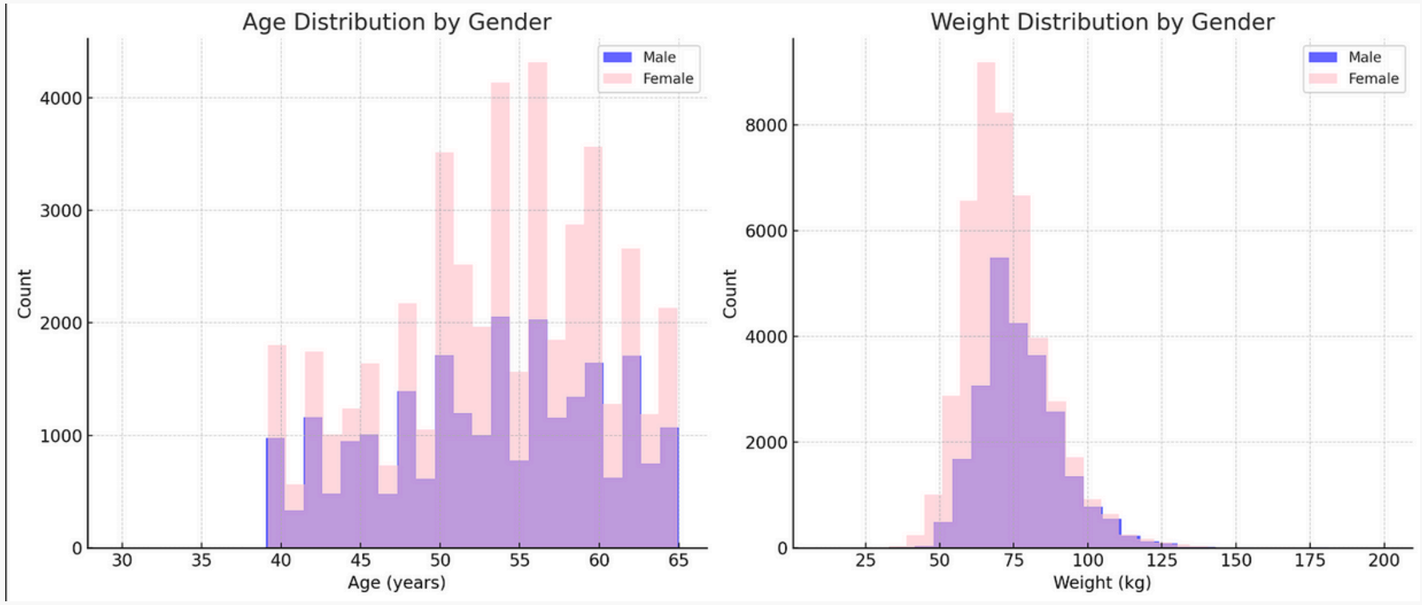


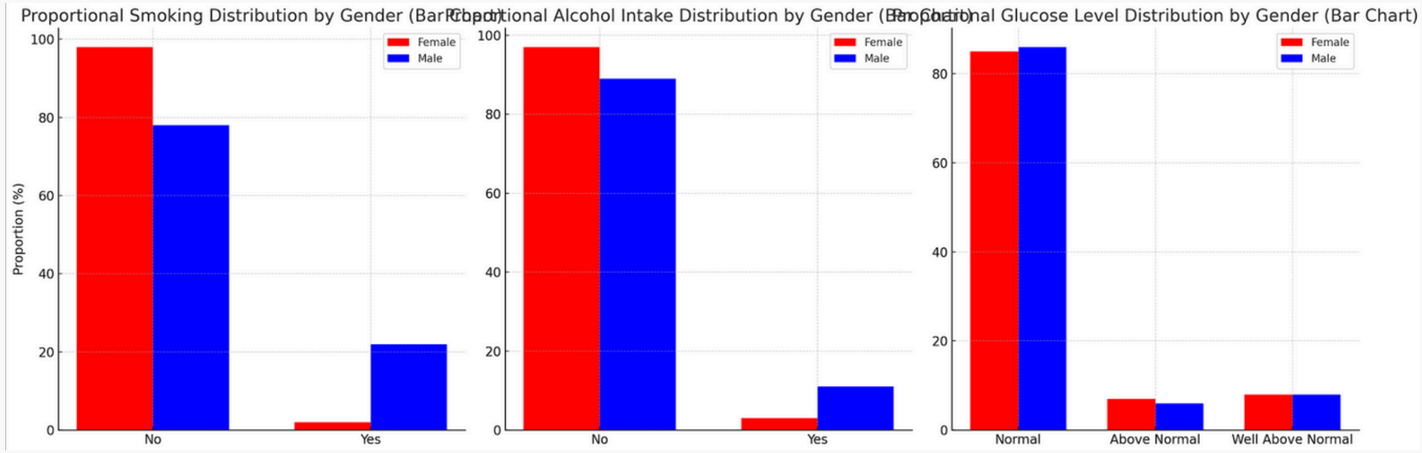
- The target variable (cardio) is binary, making it suitable for classification algorithms like logistic regression, decision trees, and ensemble methods.
- Includes both categorical and numerical features, allowing for diverse preprocessing techniques

- Aligns with public good and healthcare impact—ideal for projects with real-world significance.
- Encourages responsible AI use in sensitive domains like medicine.

Dataset Analysis

Attribute	Type	Description
age	Numeric	Age in **days**
gender	Categorical	1: female, 2: male
height	Numeric	Height in cm
weight	Numeric	Weight in kg
ap_hi	Numeric	Systolic blood pressure (upper value)
ap_lo	Numeric	Diastolic blood pressure (lower value)
cholesterol	Categorical	1: Normal, 2: Above normal, 3: Well above normal
gluc	Categorical	1: Normal, 2: Above normal, 3: Well above normal
smoke	Binary	0: non-smoker, 1: smoker
alco	Binary	0: doesn't drink, 1: drinks alcohol
active	Binary	0: not active, 1: physically active
cardio	Binary	Target: 0: no cardiovascular disease, 1: has disease





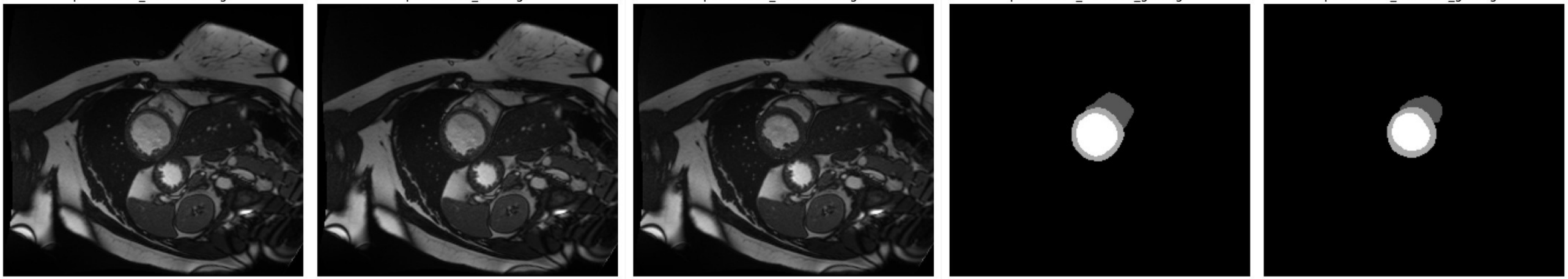
ACDC Cardiac MRI Dataset (MICCAI)

The ACDC dataset contains 150 cardiac MRI exams, evenly distributed across five classes—Normal, Myocardial Infarction, Dilated Cardiomyopathy, Hypertrophic Cardiomyopathy, and Abnormal Right Ventricle.

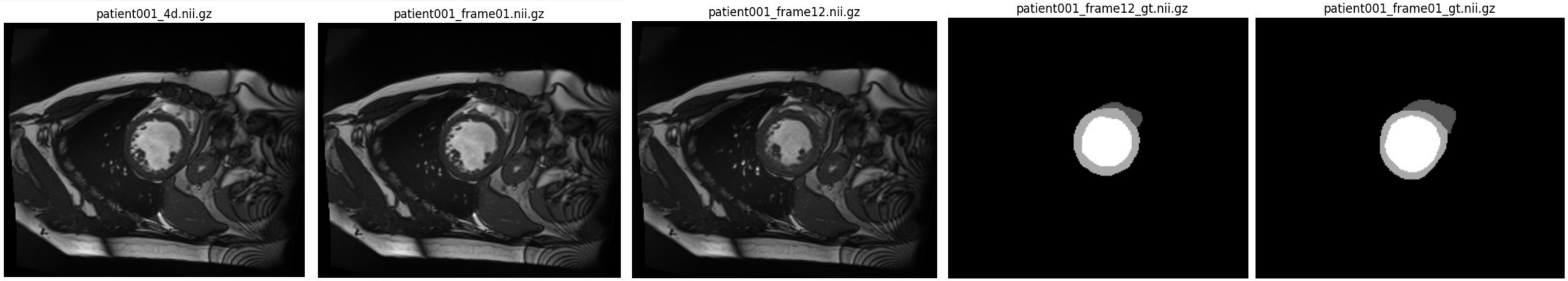
It includes cine MRI images in short-axis view, along with ground-truth segmentations of the left ventricle (LV), right ventricle (RV), and myocardium, as well as patient metadata such as height, weight, and diastolic/systolic phases.

This dataset is a widely used benchmark from the MICCAI ACDC Challenge 2017, supporting both segmentation and disease classification tasks, with balanced and clinically meaningful classes.

Testing -> patient101



Training -> patient001



Testing File Name	Description
patient101_4d.nii.gz	Full cine sequence of heart (3D + time).
patient101_frame01.nii.gz	First frame (likely End-Diastole).
patient101_frame14.nii.gz	Later frame (likely End-Systole).
patient101_frame01_gt.nii.gz	Segmentation at ED
patient101_frame14_gt.nii.gz	Segmentation at ES

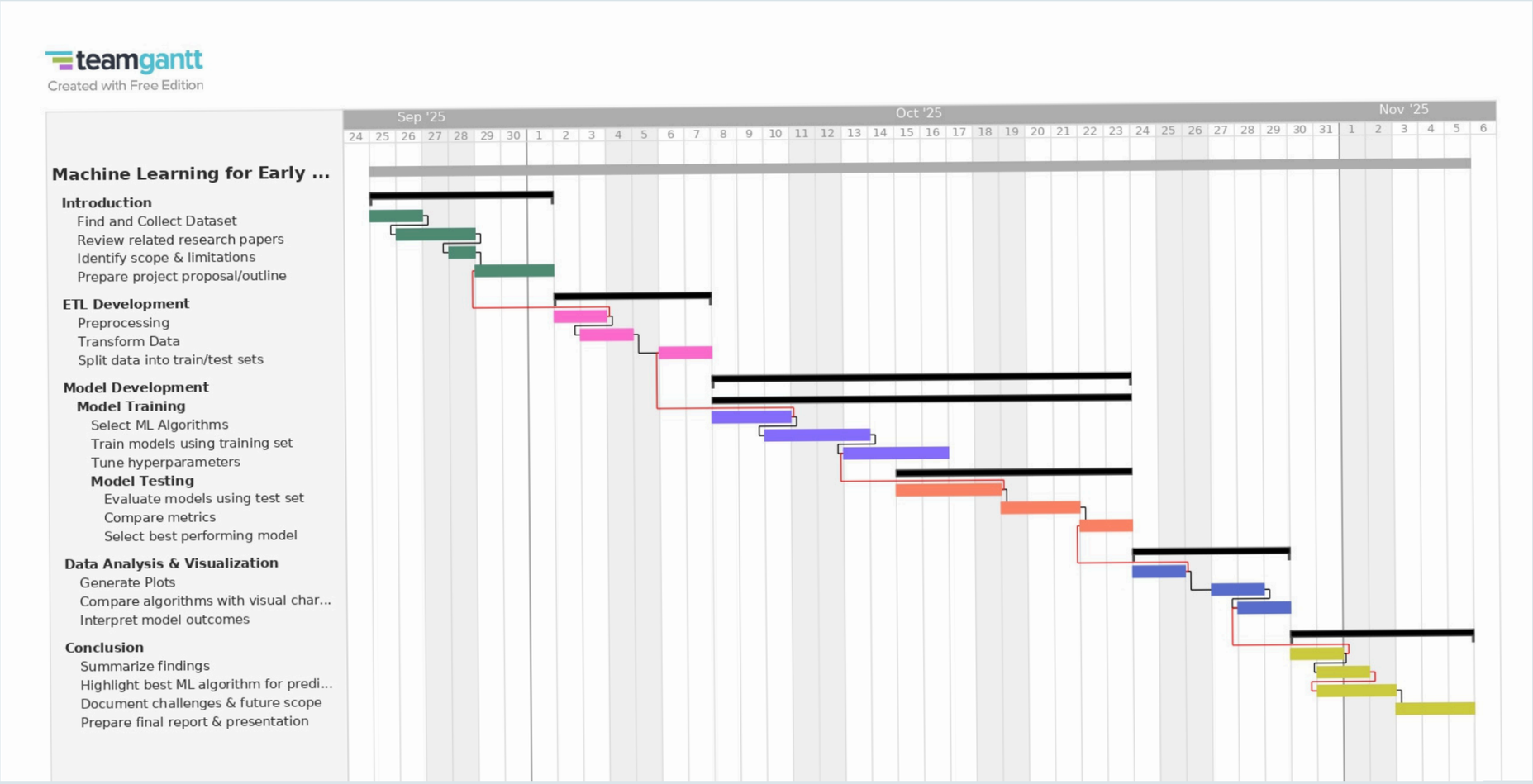
Training File Name	Description
patient001_4d.nii.gz	Full cine sequence of heart (3D + time).
patient001_frame01.nii.gz	First frame (likely End-Diastole).
patient001_frame12.nii.gz	Later frame (likely End-Systole).
patient001_frame12_gt.nii.gz	Segmentation at ED
patient001_frame01_gt.nii.gz	Segmentation at ES

Literature survey

Author	Dataset Format	Model	Accuracy
Gupta et al.	UCI Cleveland dataset	Random Forest (RF) + Convolutional Neural Network(CNN)	95%
Bernard et al.	ACDC (Cardiac MRI)	Convolutional Neural Network(CNN)	96 %
Al-Adhaileh et al.	Public datasets UCI	k-Nearest Neighbors	92%
Saikumar, K., & Rajesh, V.	Cardiovascular Disease datase	Region-based Convolutional Neural Network	95%

Project Deliverables

Gantt Chart



MoSCoW Analysis

References

- Gupta, I., Bajaj, A., Malhotra, M., Sharma, V., & Abraham, A. (2025). Heart Disease Prediction Using a Hybrid Feature Selection and Ensemble Learning Approach. IEEE Access.
- Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P. A., ... & Jodoin, P. M. (2018). Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved?. IEEE transactions on medical imaging, 37(11), 2514-2525.
- Al-Adhaileh, M. H., Ahmed Al-mashhadani, M. I., Alzahrani, E. M., & Aldhyani, T. H. (2025). Improving Heart Attack Prediction Accuracy Performance Using Machine Learning and Deep Learning Algorithms. Iraqi Journal for Computer Science and Mathematics, 6(2), 3.
- Saikumar, K., & Rajesh, V. (2024). A machine intelligence technique for predicting cardiovascular disease (CVD) using Radiology Dataset. International Journal of System Assurance Engineering and Management, 15(1), 135-151.
- https://www.projectcubicle.com/what-is-moscow-analysis-and-moscow-method/
- https://humanheart-project.creatis.insa-lyon.fr/database/#collection/637218c173e9f0047faa00fb/folder/637218e573e9f0047faa00fc
- https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset/data

Thank you