

Supplementary Material

S1. Mathematical definitions of quantitative biomarkers

All biomarkers were derived from the automated voxel-wise segmentation masks. Let M_{LV} , M_{RV} , and M_{MYO} represent the binary masks for the left ventricular cavity, right ventricular cavity, and LV myocardium, respectively. Let V_{voxel} denote the physical volume of a single voxel (product of pixel spacing and slice thickness). Features were extracted at two cardiac phases: End-Diastole (ED) and End-Systole (ES).

S1.1 Ventricular volumes (LVEDV, LVESV, RVEDV, RVESV)

Short-axis volumes were computed by summing slice-wise segmented areas multiplied by the through-plane spacing (slice thickness + inter-slice gap). Let Δz denote through-plane spacing (mm). For slice $s = 1 \dots S$, let $A_{LV,ED}(s)$ and $A_{LV,ES}(s)$ denote LV cavity area at end-diastole and end-systole. Similarly, $A_{RV,ED}(s)$ and $A_{RV,ES}(s)$ denote RV cavity area. Volumes were computed as:

$$LVEDV = \sum_{s=1..S} A_{LV,ED}(s) \times \Delta z \quad LVESV = \sum_{s=1..S} A_{LV,ES}(s) \times \Delta z \quad RVEDV = \sum_{s=1..S} A_{RV,ED}(s) \times \Delta z \quad RVESV = \sum_{s=1..S} A_{RV,ES}(s) \times \Delta z$$

S1.2 Stroke volume and ejection fraction (LVSF, RVSF, LVEF, RVEF)

Stroke volume represents the volume of blood pumped per beat, and ejection fraction represents the volumetric fraction relative to end-diastolic volume:

$$LVSF = LVEDV - LVESV \quad RVSF = RVEDV - RVESV \quad LVEF = (LVEDV - LVESV) / LVEDV \quad RVEF = (RVEDV - RVESV) / RVEDV$$

S1.3 Myocardial mass (MYO_mass_ED, MYO_mass_ES)

Myocardial mass was derived from the myocardial mask volume, assuming a standard tissue density (ρ) of 1.05 g/mL.

$$MYO_mass_ED = 1.05 \times \sum_{s=1..S} A_{MYO,ED}(s) \times \Delta z \quad MYO_mass_ES = 1.05 \times \sum_{s=1..S} A_{MYO,ES}(s) \times \Delta z$$

S1.4 Maximal wall thickness (Max_Wall_Thickness_mm)

A myocardial thickness map was computed at end-diastole from the LV myocardium mask using a Euclidean distance transform. Pixel distances were converted to millimeters using the in-plane voxel spacing. Maximal wall thickness was defined as the maximum value observed across the entire 3D myocardium mask at end-diastole:

$$Max_Wall_Thickness_mm = \max(\text{thickness_ED}(x)) \text{ over all myocardium voxels } x.$$

S1.5 Systolic thickening percentage (Systolic_Thickening_Pct)

To quantify global myocardial thickening, the mean myocardial thickness was computed at ED and ES from the corresponding thickness maps. Systolic thickening percentage was defined as the relative change in mean thickness:

$$MeanThickness_ED = \text{mean}(\text{thickness_ED}(x)) \quad MeanThickness_ES = \text{mean}(\text{thickness_ES}(x)) \quad Systolic_Thickening_Pct = 100 \times (MeanThickness_ES - MeanThickness_ED) / MeanThickness_ED$$

S1.6 Wall thickness variance (Wall_Variance)

To capture regional heterogeneity (e.g., asymmetric hypertrophy or focal scarring), the variance of the wall thickness distribution was computed at end-diastole. Let w_i denote the thickness of the i -th voxel in the myocardial mask and N be the total number of myocardial voxels:

$$Wall_Variance = (1/N) \times \sum_{i=1..N} (w_i - MeanThickness_ED)^2$$

S1.7 LV sphericity index (Sphericity_Index)

The sphericity index quantifies the remodeling of the LV from a prolate ellipsoid toward a sphere. It was calculated as the ratio of the observed LVEDV to the volume of a sphere with a diameter equal to the LV long-axis length (L). The long-axis length L was derived from the bounding box of the LV mask in the z -direction (corrected for slice spacing):

$$Sphericity_Index = LVEDV / [(\pi/6) \times L^3]$$

S1.8 Derived relational indices

Three composite markers were calculated to capture inter-chamber discordance and dominance phenotypes:

$$RV_LV_ED_ratio = RVEDV / LVEDV$$

$$MYO_LV_ED_ratio = MYO_mass_ED / LVEDV$$

$$EF_Difference = LVEF - RVEF$$

S2. Prompts strategies

S2.1 Prompt V1: intrinsic knowledge baseline

This baseline prompt evaluates the model's latent medical knowledge without external logical scaffolding. It forces the model to rely solely on its pre-training.

System Instruction:

You are an expert radiologist.

Apply your internal medical knowledge to the provided patient metrics.

Output EXACTLY one line:

LABEL=<one of DCM,HCM,MINF,NOR,RV>

No other text.

S2.2 Prompt V2: static constraint optimization (Simultaneous pattern matching)

This prompt embeds specific diagnostic criteria for all pathologies. Crucially, the agent is explicitly instructed that there is "no strict hierarchy" and it must find the "best fit" among the profiles simultaneously.

System Instruction:

You are an expert radiologist. Output EXACTLY one line: LABEL ≤ one of DCM,HCM,MINF,NOR,RV>. No other text.

Instructions:

Evaluate the patient against ALL the following diagnostic profiles simultaneously. There is no strict hierarchy; select the diagnosis that represents the best fit for the data provided.

DIAGNOSTIC PROFILES:

- **HCM:** Characterized by significant wall thickening with preserved function.
 - Criteria: Max_Wall_Thickness_mm > 15.5 AND LVEF > 45%
- **RV:** Characterized by right ventricular dominance or isolated systolic dysfunction.
 - Criteria: RV_LV_ED_ratio > 1.35 OR (LVEF > 45% AND Systolic_Thickening_Pct < 30.0).
- **DCM:** Characterized by significant heart failure, massive volume overload, and thinned walls.
 - Criteria: LVEF ≤ 45% AND LVESV > 130 AND Max_Wall_Thickness_mm ≤ 12.0.
- **MINF:** Characterized by heart failure but with preserved wall thickness (edema/scarring) or significant imbalance between LVEF and RVEF.
 - Criteria: LVEF ≤ 45% AND (Max_Wall_Thickness_mm > 12.0 OR EF_Difference ≤ -45%).
- **NOR:** Characterized by normal geometry and function.
 - Criteria: Patient does not fit the pathological profiles above (LVEF > 45%, Wall ≤ 15.5, Normal Volumes).

S2.3 Prompt V3: hierarchical protocol (Veto logic)

This protocol implements a strict sequential decision tree. It forces the calculation of derived indices and uses "stop-logic" (Veto) to rule out structural phenotypes before assessing functional failure.

System Instruction:

You are an expert radiologist. Apply the decision tree STRICTLY and output EXACTLY one line: LABEL ≤one of DCM,HCM,MINF,NOR,RV>. No other text.

DECISION TREE:

APPLY DIAGNOSTIC RULES (in order)

A. CHECK HYPERTROPHY (HCM):

- IF Max_Wall_Thickness_mm > 16.0 AND LVEF > 45% Diagnosis is **HCM**. (STOP).

B. CHECK RV ABNORMALITY (RV):

- IF RV_LV_ED_ratio > 1.4 Diagnosis is **RV**. (STOP).
- IF LVEF > 45% AND Systolic_Thickening_Pct < 35.0 Diagnosis is **RV**. (STOP).

C. CHECK FUNCTIONAL FAILURE (LVEF <= 0.45):

If the patient has not been diagnosed yet, they have heart failure. Distinguish DCM vs MINF using Volume and EF Balance:

- 1. CHECK VOLUME (Is the heart massive?):
 - IF LVESV > 135:
 - IF EF_Difference <= -45% Diagnosis is **MINF**.
 - ELSE Diagnosis is **DCM**.
- 2. CHECK WALL (Is the wall preserved?):
 - IF Max_Wall_Thickness_mm > 11.0 Diagnosis is **MINF**.
 - ELSE Diagnosis is **DCM**.

D. NORMAL (NOR):

- If none of the above apply Diagnosis is **NOR**.

S2.4 Narrative Generator Prompt

This prompt was used in the second stage of the agentic-style workflow to generate the explainable clinical report based on the determined label.

System Instruction:

You are an expert academic radiologist writing a clinical narrative.

You are given: Patient metrics and a final label that is already decided.

Task:

Write ONLY the clinical narrative in the requested VOICE. Do NOT change the label. Do NOT output the label. Do NOT use bullet points.

THE VOICE (How to Speak):

- Do NOT use bullet points or "Step 1, Step 2". Write a coherent paragraph.
- Use clinical adjectives: "robust," "sluggish," "preserved," "severely depressed," "disproportionate," "global remodeling."
- Argue *against* the alternatives. (e.g., "The absence of dilation argues against DCM...").