CirCor-2022: A multi-site heart sound dataset for murmur and outcome classification at the patient level.
Yaseen: A single-site heart sound dataset for diagnosing five valve disorders.

Shared Front-End **Feature Extraction** **Aggregation** **Classification Heads**

Augmentations (noise, time-stretch, SpecAugment)

Log-Mel Spectrogram (224×224) + z-score

DWT denoise (db4, level-5)

Fixed window length (12 s CirCor / 3 s Yaseen)

Resample (CirCor ≈ 4 kHz; Yaseen 22.05 kHz or downsample)

CNN14 (AudioSet pretrained, 1-ch input) → $e\_cnn \in \mathbb{R}^{1024}$

PaSST-S (Spectrogram Transformer) → $e\_tr \in \mathbb{R}^{768\_1024}$

Concatenate → $z\_site \in \mathbb{R}^{512}$

Output: $z\_patient \in \mathbb{R}^{512}$

Add site/location embeddings (CirCor only)

Yaseen: Single-site embedding

CirCor: Attention pooling (MIL) across multiple heart sites

CirCor:
Murmur (3-class softmax)
Outcome (binary sigmoid)

Yaseen:
Disease-5 (Normal, AS, MS, MR, MVP)

- Losses: CE / BCE (weighted)
- Metrics: Weighted accuracy, AUROC, AUPRC

1 Input & Preprocessing   **2 Dual Encoders**   **3 Fusion & Representation**   **4 Heads & Outputs**