1.1 Background

Cardiovascular diseases (CVDs) are the foremost cause of mortality worldwide, claiming approximately 17.9 million lives annually; about 85% are due to heart attacks and strokes [1]. The burden is especially high in low- to middle-income countries, including Indonesia, where access to timely diagnosis and treatment is limited. Arrhythmias—irregular heart rhythms—pose substantial risk by disrupting blood flow and causing fatigue, dizziness, fainting, myocardial infarction, or sudden cardiac death. They are also indicators of broader systemic events such as stroke and heart attack; individuals with atrial fibrillation (AF) have about a five-fold increased risk of stroke [2][3]. Diagnostic challenges persist due to limited availability of affordable, user-friendly tools in resource-constrained settings. Valvular diseases such as mitral regurgitation (MR), mitral stenosis (MS), and mitral valve prolapse (MVP) often produce systolic murmurs detectable in heart sounds, yet differentiating among them requires advanced auscultation or imaging (e.g., 12-lead ECG) and related evaluations in acute heart failure [4][5]. Structural abnormalities can precipitate AF, ventricular tachycardia (VT), and premature ventricular contractions (PVCs), underscoring the need for methods that capture both rhythmic and valvular abnormalities from physiological signals such as PCG and ECG [6].

1.2 Problem Formulation

Given phonocardiogram (PCG) recordings, develop a method that classifies heart sounds into *Normal* vs. *Abnormal* with high sensitivity and specificity, robust to real-world noise, and feasible for deployment in resource-limited environments. Specifically:

Q1: Can a hybrid CNN–GRU model using PCG signals achieve accurate binary arrhythmia detection comparable to or better than strong baselines?

Q2: Which training strategies (e.g., augmentation, class-weighted loss) most improve performance under class imbalance and noise?

Q3: Is the approach suitable for eventual deployment on low-cost/edge hardware?

1.3 Objectives and Benefits

Objectives.

- Design a hybrid CNN–GRU architecture that leverages time–frequency PCG features for binary arrhythmia detection [9].
- Employ augmentation and class-weighted loss to handle real-world variability and class imbalance.
- Benchmark against CNN and CNN–LSTM baselines, reporting accuracy, precision, recall,

and F1-score.

Benefits.

- Provide an accurate, scalable screening aid for earlier detection and triage in clinics with limited resources.
- Reduce reliance on costlier modalities by leveraging PCG signals; PCG offers acoustic information complementary to ECG's electrical activity [10].

1.4 Problem Limitation

- Scope restricted to *binary* classification (Normal vs. Abnormal); specific arrhythmia subtypes (e.g., AF, VT, PVC) are not individually classified.
- Evaluation uses publicly available PCG datasets; no prospective or in-clinic patient data are included.
- Results may reflect dataset/sensor domain characteristics; external validation across devices/environments is needed.
- ECG is referenced for context but not integrated; multimodal fusion is deferred to future work.

1.5 Research Methods

Recent advances show strong potential for automated recognition of valvular and rhythm disorders using deep learning on heart sounds: an automated VHD detector based on PCG and deep models [7], and attention-based networks that improve explainability for heart-sound classification [8]. Building on this, a hybrid CNN–GRU model is adopted to capture both spatial/spectral and temporal characteristics from PCG recordings [9]; PCG contributes acoustic cues complementary to ECG's electrical signals [10]. Concretely:

1. **Data:** Public PCG recordings spanning Normal and valve-related abnormalities (e.g., MR, MS, MVP).
2. **Preprocessing:** Time–frequency representations (e.g., spectrograms) to expose joint temporal–spectral patterns.
3. **Model:** **CNN** layers learn local spectral–spatial features; **GRU** layers model temporal dependencies across cardiac cycles.
4. **Training:** Class-weighted loss for imbalance; augmentation (noise, gain, time shift, speed perturbation, spectral filtering) for robustness.

5. **Evaluation:** Accuracy, precision, recall, and F1-score from the confusion matrix, with comparisons to CNN and CNN–LSTM baselines.