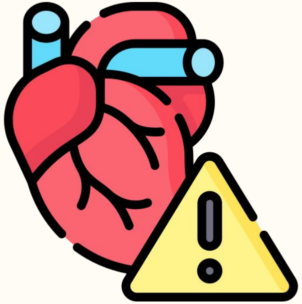
Automated Diagnosis of Cardiovascular Diseases

Using Multi-Phase Cardiac MRI and Convolutional Neural Networks

Instructor: Wade Schulz
Group Members:
Abbey Yuan; Ruoxi Teng; Yuchen Liu; Rong Sun

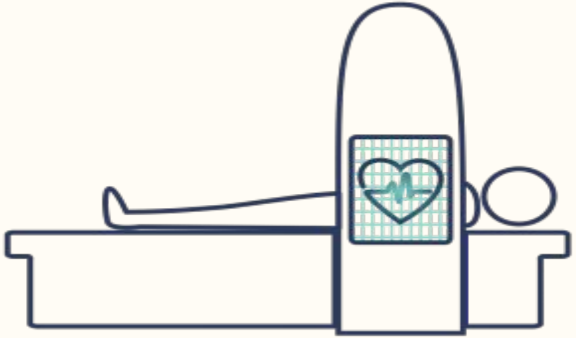


Introduction & Problem Specification

- **Cardiovascular disease** is the leading cause of death globally and **Cardiac MRI** is the gold standard for disease diagnosis
- **AI** offers a solution by automating diagnosis, reducing workload, and improving consistency
- Leveraging Cardiac MRI dataset from Kaggle, we explore **CNN-based model** for image classification

Project Objective:

*Build a reliable, interpretable, and clinically useful tool to support **cardiologists and radiologists** in diagnosing cardiac disease*



Method Overview

Data Preprocessing

Data Source: Automated Cardiac Diagnosis Challenge Dataset, with 150 patients

Inclusion Criteria: high-quality short-axis cine-MRI scans with clearly labeled end-diastolic (ED) and end-systolic (ES) phases

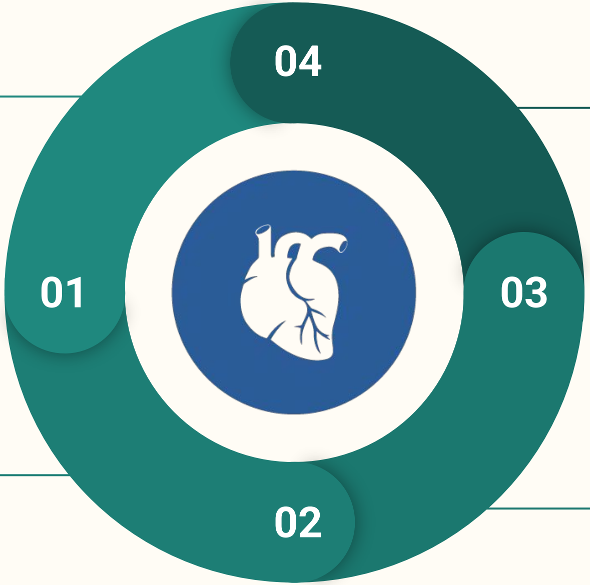
Exclusion Criteria:corrupted or incomplete files

Model Development

5-class multi-head CNN & binary classification

ReLU Activation & Max Pooling

Adam Optimizer and BCE Loss



MLOps Integration

Hyperparameters, training metrics, and model artifacts

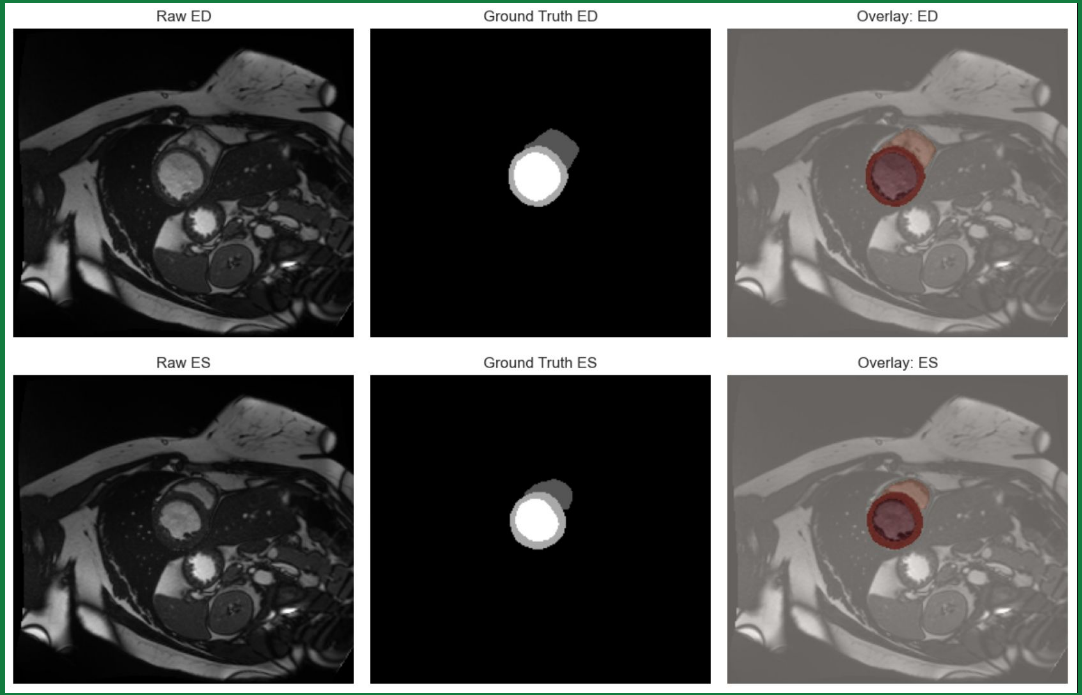
Robustness, transparency and reproducibility

Model Evaluation

Performance: accuracy, precision, recall, F1-score, confusion matrices, and AUC metrics (ROC-AUC and PR-AUC)

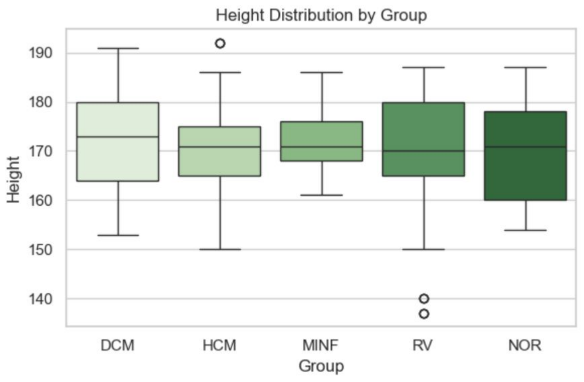
Interpretability: Grad-CAM and saliency maps

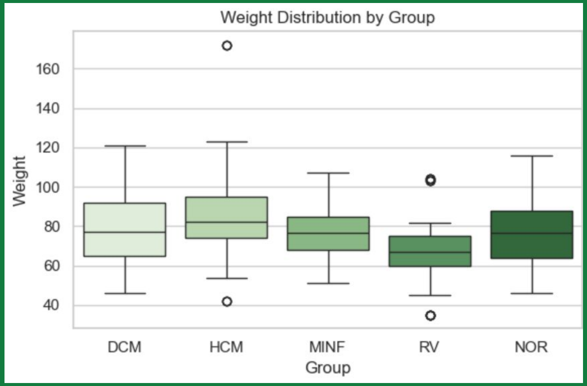
EDA

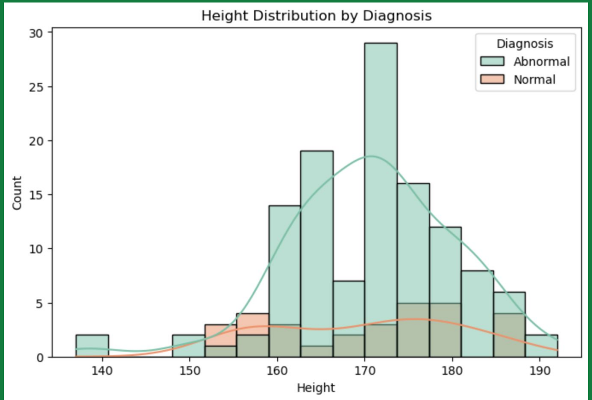


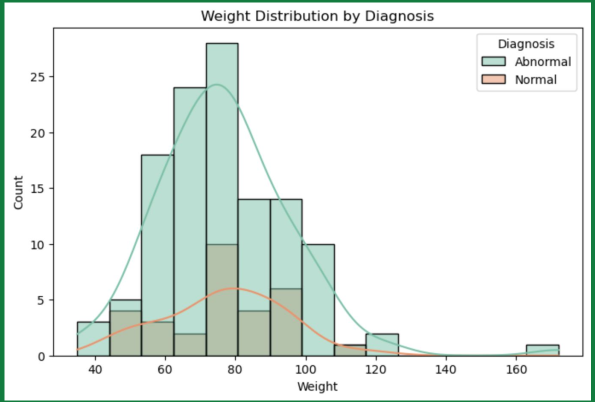

```
database/
├── training/
│   ├── patient001/
│   │   ├── info.cfg
│   │   ├── patient001_frame01.nii
│   │   ├── patient001_frame01_gt.nii
│   │   ├── ...
├── testing/
│   ├── patient101/
│   │   ├── info.cfg
│   │   ├── patient101_frame01.nii
│   │   ├── patient101_frame01_gt.nii
│   │   ├── ...
```

EDA


Height Distribution by Group
Height
Group
DCM
HCM
MINF
RV
NOR
190
180
170
160
150
140


Weight Distribution by Group
Weight
Group
DCM
HCM
MINF
RV
NOR
160
140
120
100
80
60
40


Height Distribution by Diagnosis
Diagnosis
Abnormal
Normal
Count
Height
0
5
10
15
20
25
30
140
150
160
170
180
190


Weight Distribution by Diagnosis
Diagnosis
Abnormal
Normal
Count
Weight
0
5
10
15
20
25
40
60
80
100
120
140
160

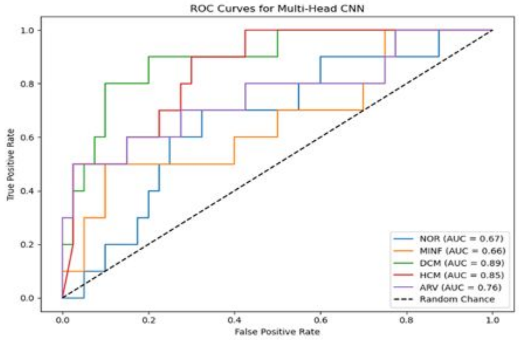
Model 1 Development and Performance

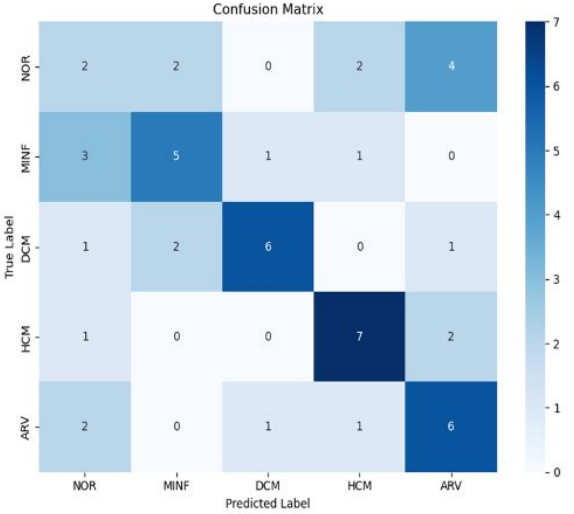
Model 1: Simple CNN Multi-Head (5-Class Classification)

- **Task:** Multi-class classification for NOR (Normal), MINF (Systolic Heart Failure with Infarction), DCM (Dilated Cardiomyopathy), HCM (Hypertrophic Cardiomyopathy), and ARV (Abnormal Right Ventricle)
- **Accuracy:** 52.0% (2.6× better than random)
  NOR Class: Poor performance (F1: 0.211)
- DCM & HCM Classes: Best performance (F1: 0.667 for both)
- **AUCs:** DCM (0.89), HCM (0.85), ARV (0.76), MINF (0.66), NOR (0.67)

Tabel : Model 1 Classification Report

	precision	recall	f1-score	support
NOR	0.222	0.2	0.211	10
MINF	0.556	0.5	0.526	10
DCM	0.75	0.6	0.667	10
HCM	0.636	0.7	0.667	10
ARV	0.462	0.6	0.522	10
accuracy	0.52	0.52	0.52	0.52
macro avg	0.525	0.52	0.518	50
weighted avg	0.525	0.52	0.518	50





Promising initial results despite challenges; limited data and high inter-class variability likely constrained performance—highlighting future potential with expanded datasets and more complex architectures.

Model 2-4 Development and Performance

Model 2: Simple Binary CNN

- **Task**: Collapsed to binary NOR vs. ABN classification
- **Accuracy**: 74.0%
- **NOR Class**: Poor detection (F1: 0.316)
- **ABN Class**: High F1 (0.840)
- **AUC**: ROC (0.721); strong PR (0.910)
- **Takeaway**: Good at detecting ABN but biased toward over-predicting abnormalities

Model 3: Deep Binary CNN

- **Enhancement**: Increased depth with more layers
- **Accuracy**: 76.0%
- **NOR Class**: Improved F1 (0.571)
- **ABN Class**: Precision improved (0.938), F1 stable (0.833)
- **AUC**: ROC (0.862); PR (0.966)
- **Takeaway**: Better balance between classes and higher overall discrimination

Model 4: Improved Deep Binary CNN

- **Enhancement**: Further increased depth by adding more hidden layers than Model 3
- **Accuracy**: 86.0%
- **NOR Class**: Best so far (F1: 0.632)
- **ABN Class**: Outstanding performance (F1: 0.914)
- **AUC**: ROC (0.865); PR (0.949)
- **Takeaway**: Best overall model with strong, balanced performance

Progressively deeper binary CNN models (simplifying to NOR vs. ABN) achieved higher accuracy and better class balance, with Model 4 reaching 86% accuracy and the most robust, balanced performance.

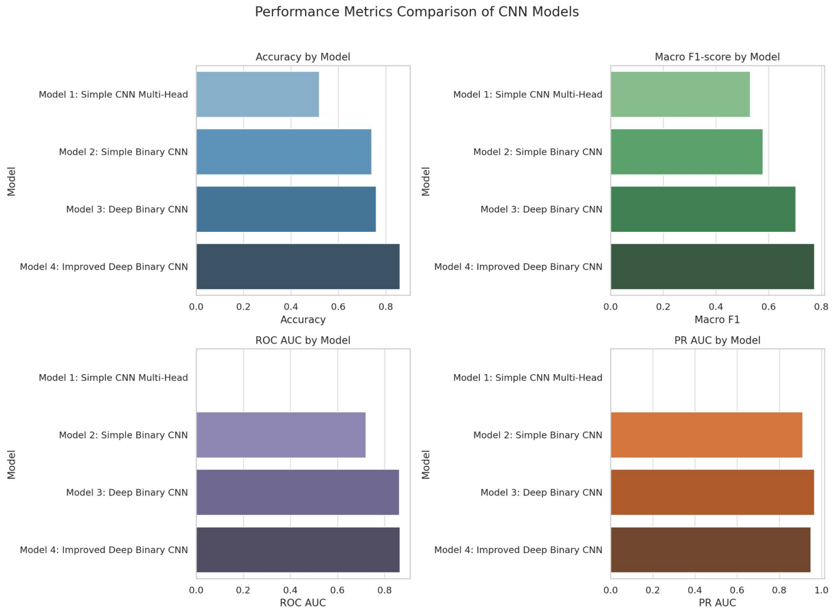
Model Performance Comparison

- **Reframed the Task**: Shifted from 5-class to binary classification (Normal vs. Abnormal), reducing confusion and boosting performance.

- **Simple Binary CNN:** Highly sensitive to abnormalities but over-predicted them—poor at detecting normal cases.

- **Deep Binary CNN:** Added layers improved balance—better at recognizing both normal and abnormal cases.

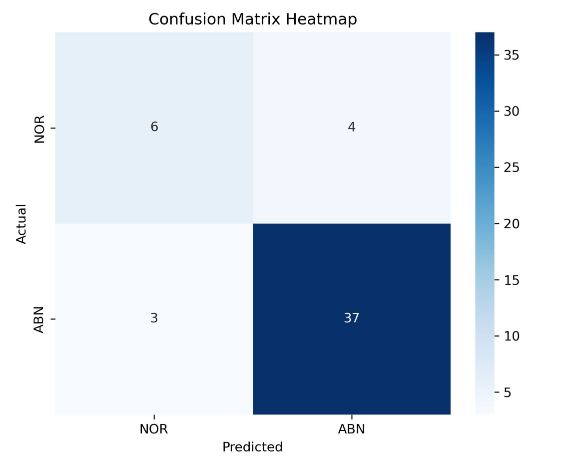
- **Improved Deep Binary CNN:** Best overall accuracy and trade-off between sensitivity and specificity—most clinically applicable.

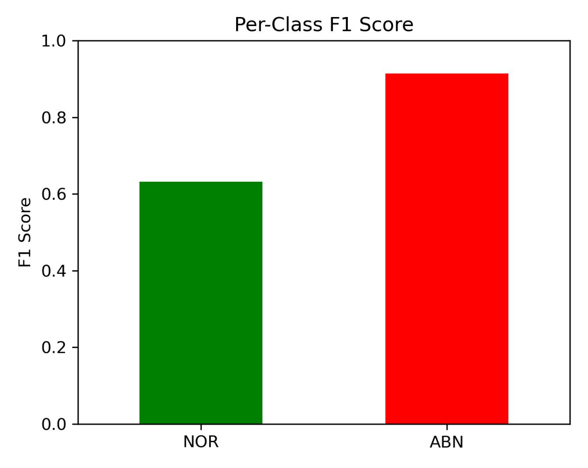


Binary models outperformed the multi-class model due to reduced complexity, better handling of class imbalance, improved feature learning in deeper architectures.

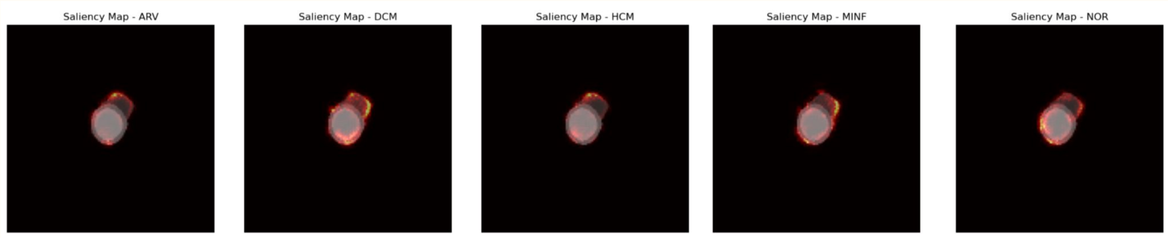
Bias Assessment

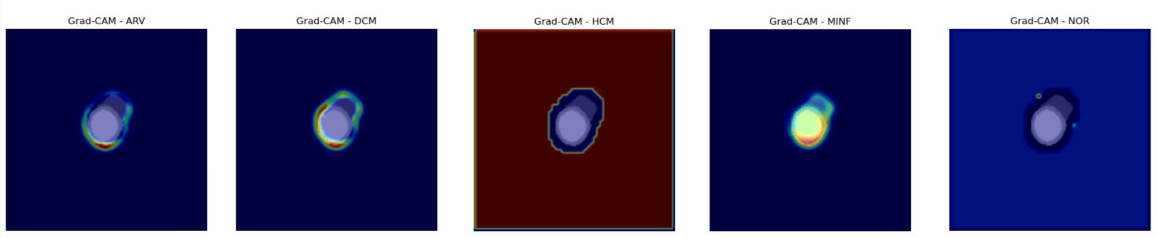
Bias was assessed using per-class F1 scores and confusion matrices, focusing on Model 4. While the model performed well overall, it showed class-specific disparities: 37 out of 40 abnormal cases were correctly classified, compared to just 6 out of 10 normal cases, indicating a tendency to over-predict abnormalities. The F1-score for the NOR class was 0.63 versus 0.92 for ABN, reflecting imbalance. This bias may result from class imbalance, inconsistent labeling, or scanner variability.





Interpretability/ Explainability





GradCAM Plot:

- Visualizes the anatomic regions contributing most to class-specific predictions
- The model focused on relevant myocardial regions, varying by disease type—indicating clinically meaningful attention

Saliency Plot:

- Highlight pixel-level sensitivity to input changes
- Showed consistent edge-focused patterns, suggesting predictions were driven by structural contours in the MRI slices

Implementation & Dissemination Strategy

Implementation Plan

- Workflow Integration

- Inference triggered during routine cardiac cine-MRI scans
- Operates in batch mode (no real-time requirement)

- Standardized Pipeline

- Mid-slice selection, ED/ES phase stacking, etc

- Production Architecture

- Combining MRI processing, disease classification, and Grad-CAM interpretability

- Model Characteristics

- Modular, continuously updatable, and designed for clinical interpretability.

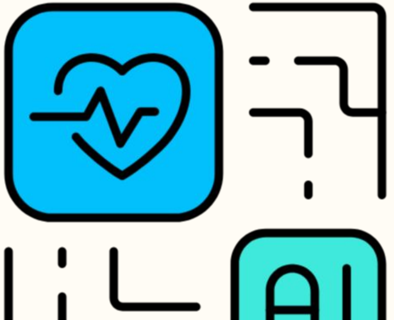
Dissemination Strategy

- Stakeholder Engagement

- Cardiologists/Radiologists/AI researchers

- Dissemination Channels

- Peer-reviewed journals
- AI/Healthcare conferences
- Clinical implementation bodies



References

1. Vaduganathan, M., Mensah, G. A., Turco, J. V., Fuster, V., & Roth, G. A. (2022). The Global Burden of Cardiovascular Diseases and Risk: A Compass for Future Health. Journal of the American College of Cardiology, 80(25), 2361–2371. https://doi.org/10.1016/j.jacc.2022.11.005

2. Taylor, C. J., Hartshorne-Evans, N., Satchithananda, D., & Hobbs, F. R. (2021). FASTer diagnosis: Time to BEAT heart failure. BJGP open, 5(3), BJGPO.2021.0006. https://doi.org/10.3399/BJGPO.2021.0006

3. Wang, Y. J., Yang, K., Wen, Y., Wang, P., Hu, Y., Lai, Y., Wang, Y., Zhao, K., Tang, S., Zhang, A., Zhan, H., Lu, M., Chen, X., Yang, S., Dong, Z., Wang, Y., Liu, H., Zhao, L., Huang, L., Li, Y., … Zhao, S. (2024). Screening and diagnosis of cardiovascular disease using artificial intelligence-enabled cardiac magnetic resonance imaging. Nature medicine, 30(5), 1471–1480. https://doi.org/10.1038/s41591-024-02971-2

4. https://www.kaggle.com/datasets/samdazel/automated-cardiac-diagnosis-challenge miccai17?resource=download

Group Work Allocation

All group members actively contributed to the project, participated in group meetings, and wrote according sections of the report.

Ruoxi Teng: Data augmentation and model tuning, literature review and method development

Yuchen Liu: EDA, initial model tuning, implementation and dissemination strategies

Abbey Yuan: Developed Model 1, including training, parameters tuning, model improvements and visualized interpretability of the model

Rong Sun: Developed and finalized Models 2, 3, and 4 for binary classification, including training, parameters tuning, model improvements, interpretability methods, and performance comparison, etc. (codes/analyses); conducted bias assessment (codes/analyses)

Questions?