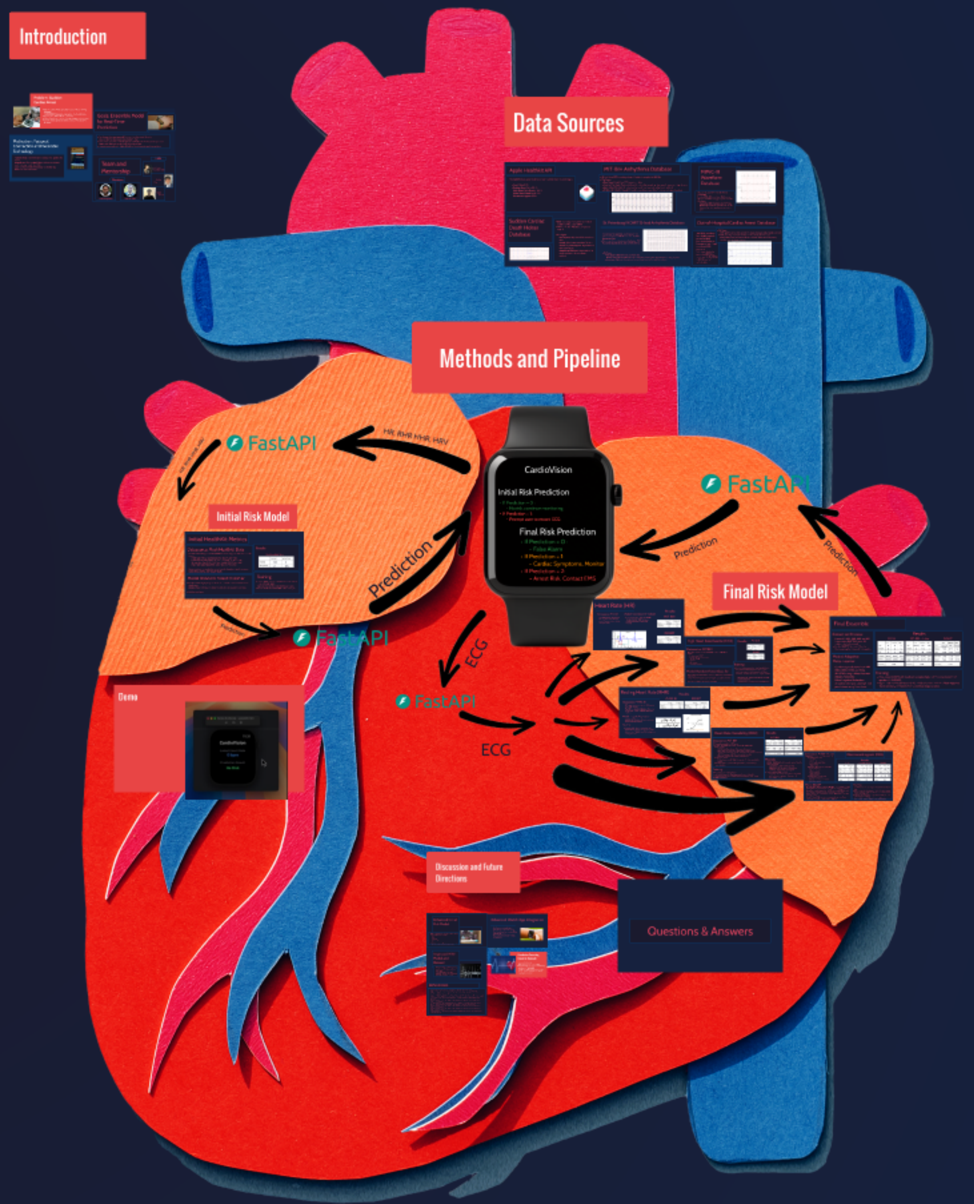
CardioVision: Predicting Cardiac Arrest Risk

An Ensemble Approach Utilizing Wearable Physiological Data



Introduction

Problem: Sudden Cardiac Arrest

- Sudden Cardiac Arrest (SCA) often occurs without warning symptoms.
- ~350,000 **SCA cases occur annually** in the United States.
- Survival rate is **below 10%** in most cases.
- Early risk prediction is critical for improving survival outcomes.
- Current clinical approaches often lack real-time, proactive detection.

Goals: Ensemble Model for Real-Time Prediction

- Build a **machine learning model** to predict cardiac arrest risk early.
- Classify risk into **three levels**: Low, Medium, High.
- Combine outputs from **multiple specialized models** for more accurate prediction.
- Enable **real-time risk** scoring using physiological data.
- Integrate seamlessly with **Apple Watch** for proactive **alerts and interventions.**

Motivation: Personal Connection and Wearable Technology

- A **personal loss** in our team due to cardiac arrest sparked the project.
- **Wearable tech** like the **Apple Watch** collects continuous health data (e.g., heart rate, ECG).
- Opportunities for **real-time cardiac risk monitoring**
- **Millions of users worldwide**

Team and Mentorship

LNV


Lukhsaan Elankumaran

Nirjhar Dey



Varun Gande

Mentors


Vineel Nagisetty

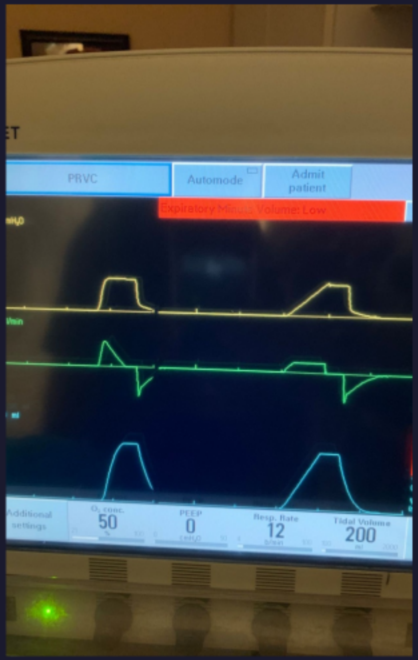

Jaemar Millar

Problem: Sudden Cardiac Arrest

- Sudden Cardiac Arrest (SCA) often occurs without warning symptoms.
- **~350,000 SCA cases occur annually** in the United States.
- Survival rate is **below 10%** in most cases.
- Early risk prediction is critical for improving survival outcomes.
- Current clinical approaches often lack real-time, proactive detection.

Motivation: Personal Connection and Wearable Technology

- A **personal loss** in our team due to cardiac arrest sparked the project.
- **Wearable tech** like the **Apple Watch** collects continuous health data (e.g., heart rate, ECG).
- Opportunities for **real-time cardiac risk monitoring**
- **Millions of users worldwide**



Goals: Ensemble Model for Real-Time Prediction

- Build a **machine learning model** to predict cardiac arrest risk early.
- Classify risk into **three levels**: Low, Medium, High.
- Combine outputs from **multiple specialized models** for more accurate prediction.
- Enable **real-time risk** scoring using physiological data.
- Integrate seamlessly with **Apple Watch** for proactive **alerts and interventions.**

Team and Mentorship

LNV

Lukhsaan Elankumaran

Nirjhar Dey

Varun Gande

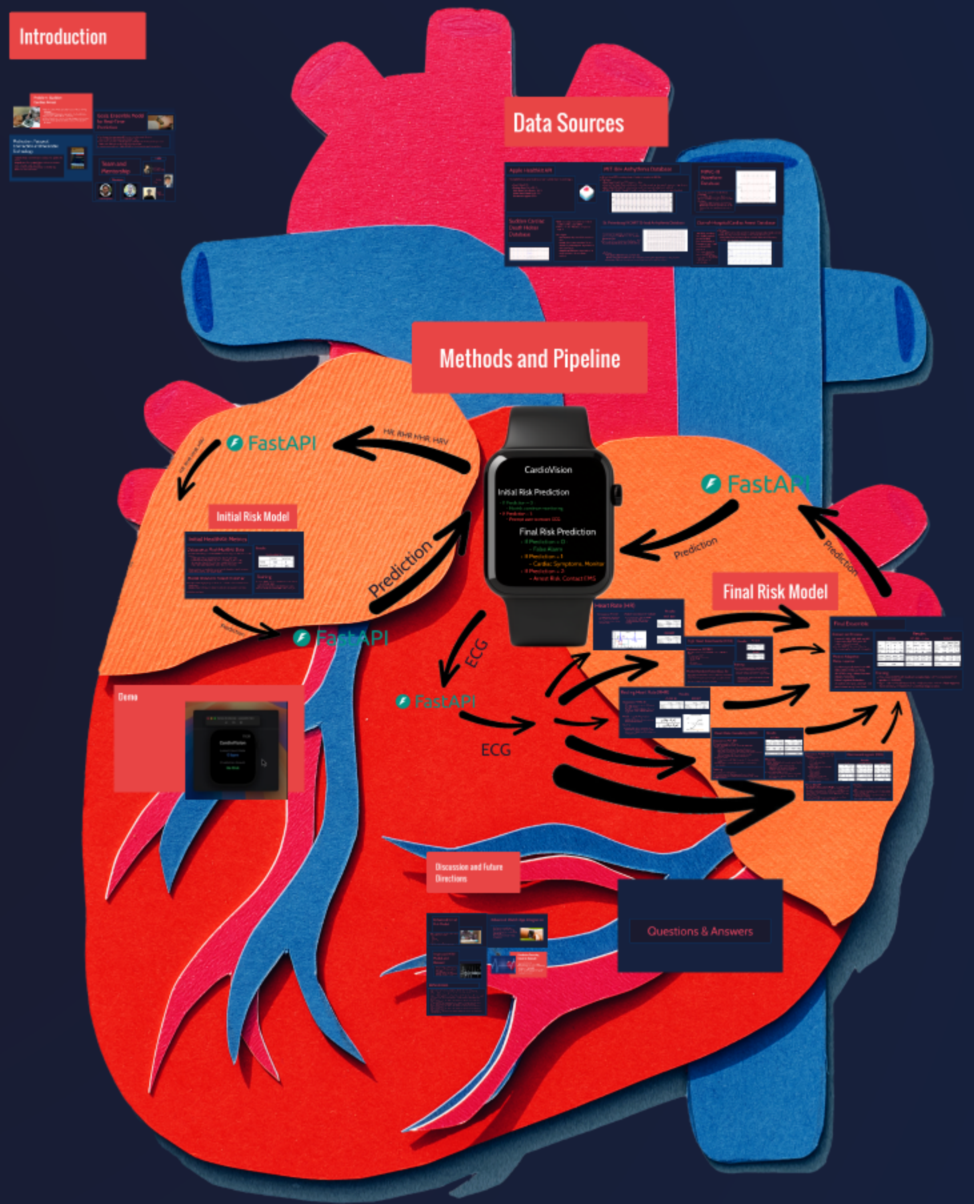
Mentors

Vineel Nagisetty

Jaemar Millar

CardioVision: Predicting Cardiac Arrest Risk

An Ensemble Approach Utilizing Wearable Physiological Data



Data Sources

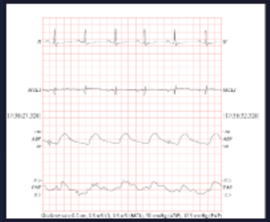
Apple Healthkit API

- Mock **JSON** Data generated based on healthkit metrics and ranges:
  - **Heart Rate** (HR)
  - **Resting Heart Rate** (RHR)
  - **High Heart Rate Events** (HHR)
  - **Heart Rate Variability** (HRV)
  - **Electrocardiogram** (ECG)

MIT-BIH Arrhythmia Database

- **48** half-hour **ECG** recordings from **47** patients sampled at **360 Hz**
- **File Types:**
  - **Raw Signal (.dat):** Raw ECG waveform data.
  - **Header File (.hea):** Metadata file with details like sampling rate, signal names, and signal gains.
  - **Annotation File (.atr):** Annotations for each beat (normal, arrhythmic, artifact).
  - **X-Wave Annotation (.xws):** Extended waveform signal data for advanced signal analysis.

MIMIC-III Waveform Database



- **67,000 ICU** waveform records linked to clinical data
- Includes **ECG**, **ABP**, respiration, **SpO_2** signals
- **48** hours of **ECG** recordings sampled at **125 Hz**
- **File Types:**
  - **Raw Signal (.dat):** Raw ECG waveform data.
  - **Header File (.hea):** Metadata file with details like sampling rate, signal names, and signal gains.

Sudden Cardiac Death Holter Database

- **ECG** recordings from patients at risk of **sudden cardiac death (SCD)**.
- **24-hour** Holter **ECG** data, sampled at **250 Hz**.
- **File Types:**
  - **Raw Signal (.dat):** Raw ECG waveform data.
  - **Header File (.hea):** Metadata file with details like sampling rate, signal names, and signal gains.
  - **Annotation File (.atr):** Annotations for each beat (normal, arrhythmic, artifact).

St. Petersburg INCART 12-lead Arrhythmia Database

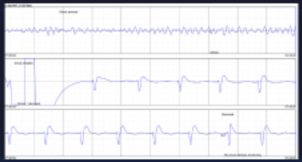
- **75** half-hour **12-lead ECG** recordings from **32** patients sampled at **257 Hz** with over **175,000** annotated beats
- Patient conditions include **atrial fibrillation**, **SVT**, **MI**, **bundle branch blocks**, and **ischemic events**.

- **File Types:**
  - **Raw Signal (.dat):** Raw ECG waveform data.
  - **Header File (.hea):** Metadata file with details like sampling rate, signal names, and signal gains.
  - **Annotation File (.atr):** Annotations for each beat (normal, arrhythmic, artifact).

Out-of-Hospital Cardiac Arrest Database

- **260 ECG** recordings from **OHCA** patients treated by **EMS**.
- Each record includes **9 sec pre-shock + 1 min** post-shock **ECG** segments.
- Each case labeled as successful **(ROEA)** or unsuccessful **(NoROEA)** defibrillation by 3 cardiologists.

- **File Types:**
  - **PDF (.pdf):** Scanned ECG waveform images (anonymized, printed, scanned).
  - **Text (.txt):** Digitized ECG waveforms with amplitude-time data.
  - **Excel (.xls):** Extracted feature values in tabular format (time, frequency, wavelet, and nonlinear metrics).



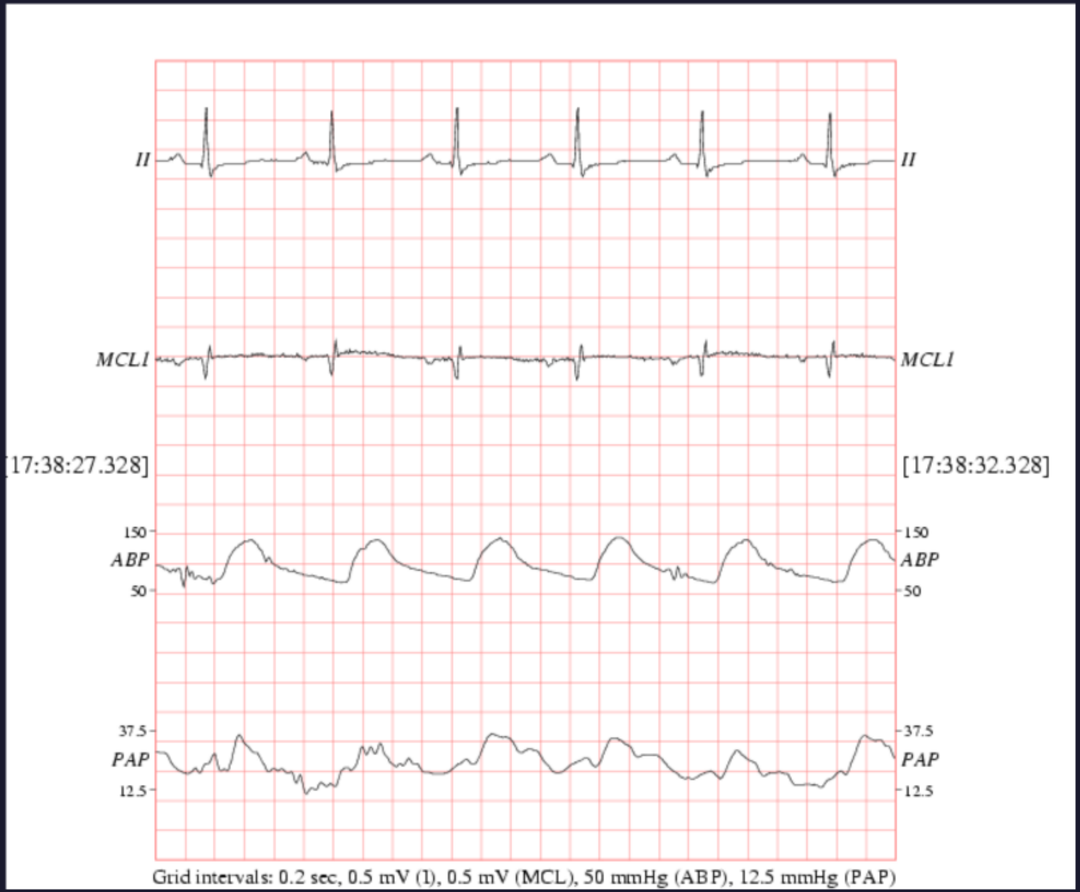
Apple Healthkit API

- Mock **JSON** Data generated based on healthkit metrics and ranges:
  - **Heart Rate** (HR) :
  - **Resting Heart Rate** (RHR)
  - **High Heart Rate Events** (HHR)
  - **Heart Rate Variability** (HRV)
  - **Electrocardiogram** (ECG)

MIT-BIH Arrhythmia Database

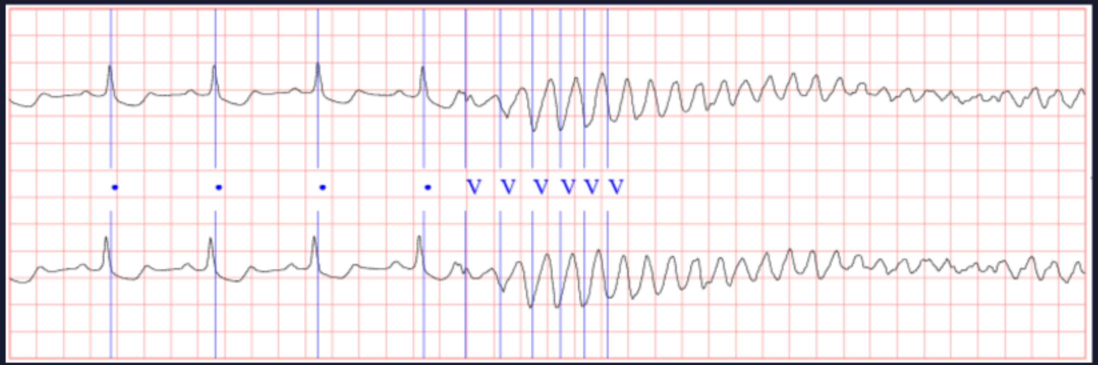
- **48** half-hour **ECG** recordings from 47 patients sampled at **360 Hz**
- **File Types**:
  - **Raw Signal (.dat):** Raw ECG waveform data.
  - **Header File (.hea):** Metadata file with details like sampling rate, signal names, and signal gains.
  - **Annotation File (.atr):** Annotations for each beat (normal, arrhythmic, artifact).
  - **X-Wave Annotation (.xws):** Extended waveform signal data for advanced signal analysis.

MIMIC-III Waveform Database



- **67,000 ICU** waveform records linked to clinical data
- Includes **ECG**, **ABP**, respiration, **SpO₂** signals
- 48 hours of **ECG** recordings sampled at **125 Hz**
- **File Types:**
  - **Raw Signal (.dat):** Raw ECG waveform data.
  - **Header File (.hea):** Metadata file with details like sampling rate, signal names, and signal gains.

Sudden Cardiac Death Holter Database

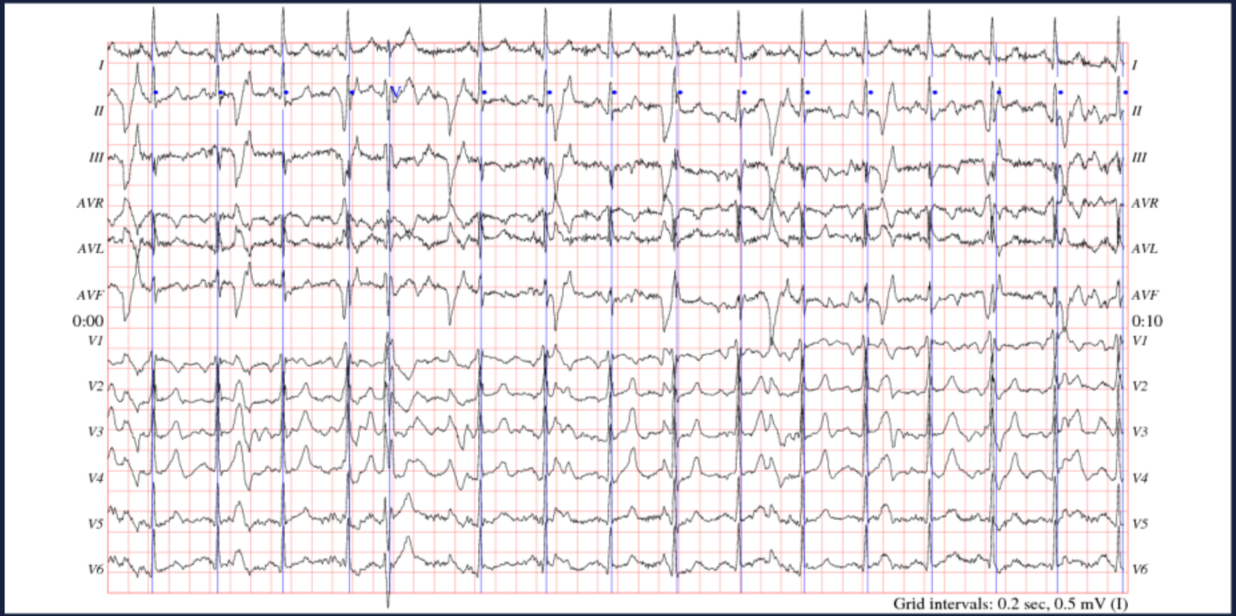


- **ECG** recordings from patients at risk of **sudden cardiac death (SCD).**
- 24-hour Holter **ECG** data, sampled at 250 Hz.

- **File Types:**
  - **Raw Signal (.dat):** Raw ECG waveform data.
  - **Header File (.hea):** Metadata file with details like sampling rate, signal names, and signal gains.
  - **Annotation File (.atr):** Annotations for each beat (normal, arrhythmic, artifact).

St. Petersburg INCART 12-lead Arrhythmia Database

- **75** half-hour **12-lead ECG** recordings from 32 patients sampled at **257 Hz** with over 175,000 annotated beats
- Patient conditions include **atrial fibrillation, SVT, MI, bundle branch blocks, and ischemic events.**

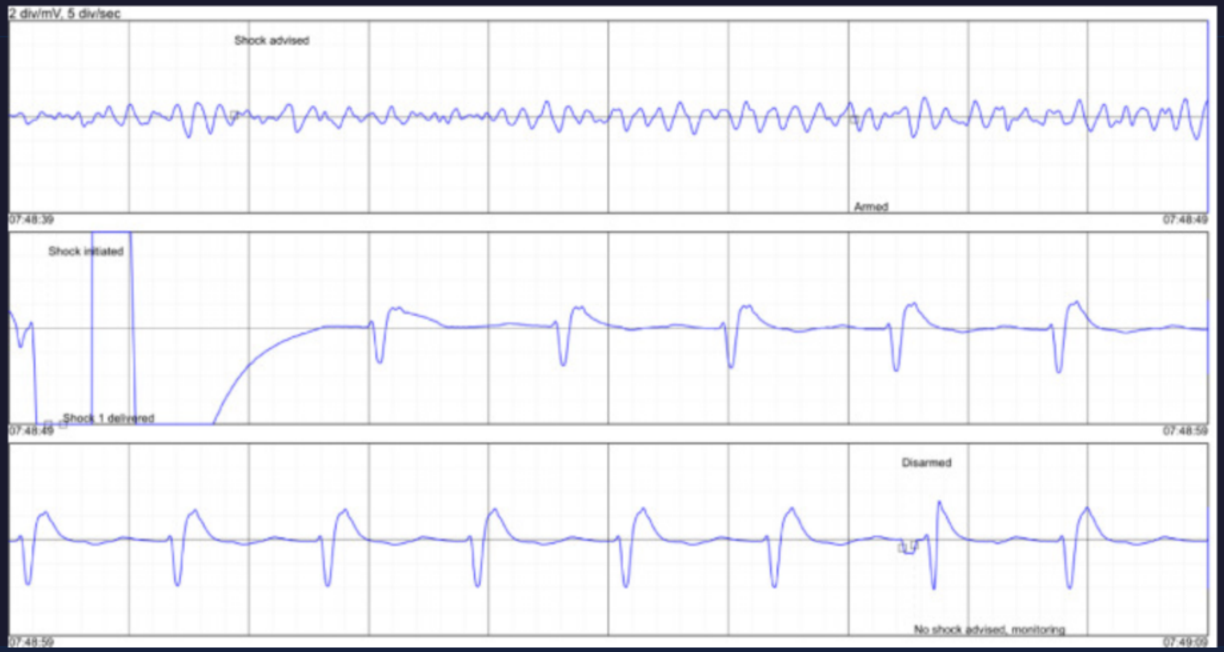


- **File Types:**
  - **Raw Signal (.dat):** Raw ECG waveform data.
  - **Header File (.hea):** Metadata file with details like sampling rate, signal names, and signal gains.
  - **Annotation File (.atr):** Annotations for each beat (normal, arrhythmic, artifact).

Out-of-Hospital Cardiac Arrest Database

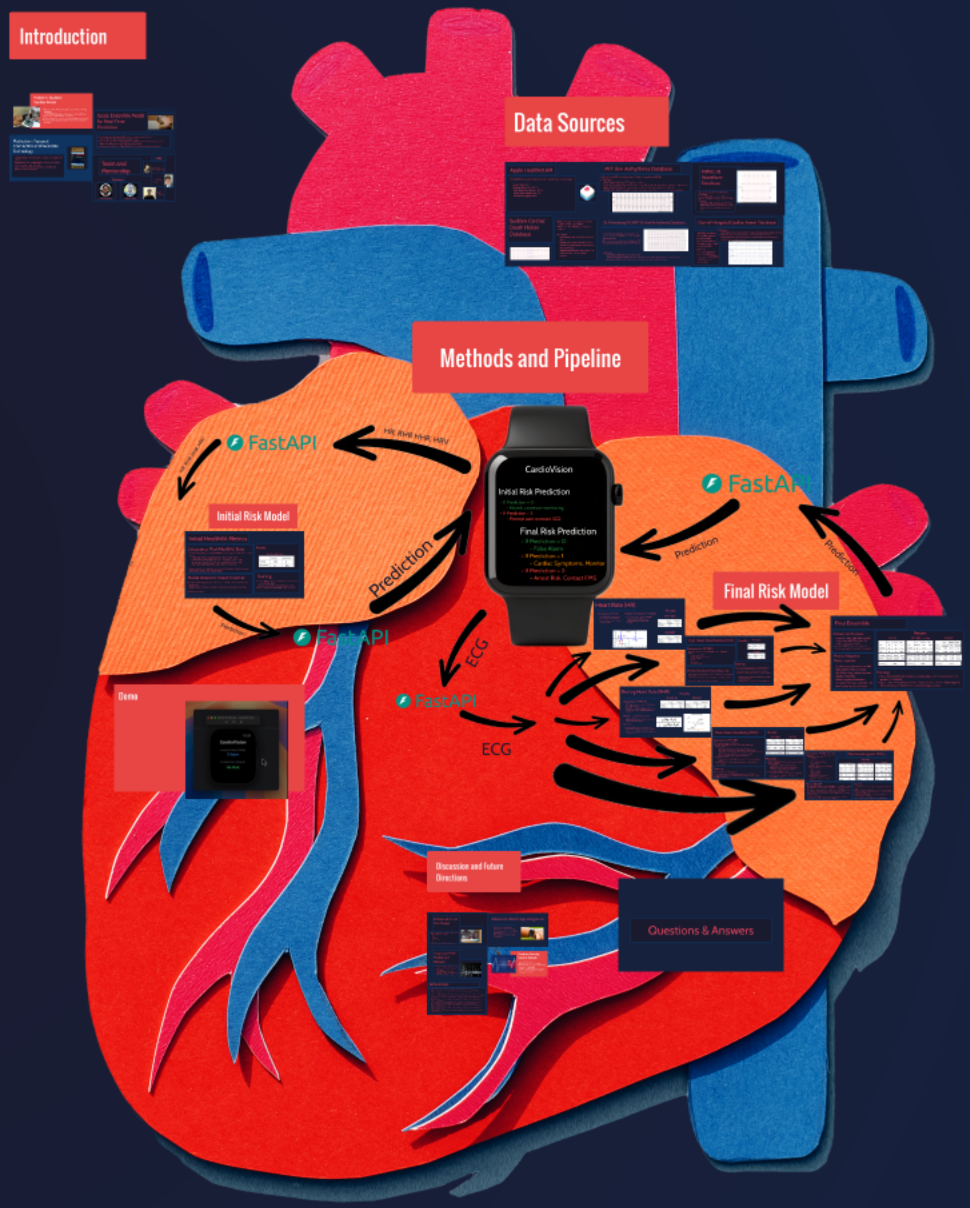
- **260 ECG** recordings from **OHCA** patients treated by **EMS**.
- Each record includes **9 sec pre-shock + 1 min** post-shock **ECG** segments.
- Each case labeled as successful **(ROEA)** or unsuccessful **(NoROEA)** defibrillation by 3 cardiologists.

- **File Types:**
  - **PDF (.pdf):** Scanned ECG waveform images (anonymized, printed, scanned).
  - **Text (.txt):** Digitized ECG waveforms with amplitude-time data.
  - **Excel (.xls):** Extracted feature values in tabular format (time, frequency, wavelet, and nonlinear metrics).



CardioVision: Predicting Cardiac Arrest Risk

An Ensemble Approach Utilizing Wearable Physiological Data



Methods and Pipeline

RHR HHR, HRV
CardioVision
Initial Risk Prediction
If Prediction = 0 :
No risk, continue monitoring
If Prediction = 1:
Prompt user to record ECG:
Final Risk Prediction
If Prediction = 0 :
False Alarm
If Prediction = 1:
Cardiac Symptoms, Monitor
If Prediction = 2:
Arrest Risk, Contact EMS

HR, RHR HHR, HRV

FastAPI


HR, RHR HHR, HRV

Initial Risk Model

Initial HealthKit Metrics

Datasource: Mock Healthkit Data

- Mock HealthKit data generated as JSON files in two categories: "risk" and "no_risk"
- Metrics:
  - **Heart Rate (HR):** 60-100 bpm (**no risk**), 110-140 bpm (**risk**).
  - **Heart Rate Variability** (HRV): 55-100 ms (**no risk**), 10-40 ms (**risk**).
  - **Resting Heart Rate** (RHR): 55-85 bpm (**no risk**), 90-110 bpm (**risk**).
  - **High Heart Rate Events** (HHR): 0 (**no risk**), 1-5 events (**risk**).

- Ranges are based on established values from the **American Heart Association, Kubios HRV, WebMD, and Harvard Health**

Model: Random Forest Classifier

- **Random Forest**: Aggregates predictions from multiple decision trees for robust classification.
- **Decision Trees:** Individual tree classifiers that split data based on conditions.

- 100 Decision Trees

Results

Mock Healthkit Data

	Normal	Arrhythmia
Normal	100	0
Arrhythmia	0	100

Training:

- Loaded **JSON** data into a DataFrame with 4 features (**HR, HRV, RHR, HHR**).
- Split into training (80%) and testing (20%) sets.
- Trained **Random Forest model** on the training set.

Initial HealthKit Metrics

Datasource: Mock Healthkit Data

- Mock HealthKit data generated as JSON files in two categories: "risk" and "no_risk"
- Metrics:
  - **Heart Rate (HR):** 60-100 bpm (**no risk**), 110-140 bpm (**risk**).
  - **Heart Rate Variability** (HRV): 55-100 ms (**no risk**), 10-40 ms (**risk**).
  - **Resting Heart Rate** (RHR): 55-85 bpm (**no risk**), 90-110 bpm (**risk**).
  - **High Heart Rate Events** (HHR): 0 (**no risk**), 1-5 events (**risk**).
- Ranges are based on established values from the **American Heart Association, Kubios HRV, WebMD, and Harvard Health**

Results

Mock Healthkit Data

	Normal	Arrhythmia
Normal	100	0
Arrhythmia	0	100

Model: Random Forest Classifier

- **Random Forest**: Aggregates predictions from multiple decision trees for robust classification.
- **Decision Trees:** Individual tree classifiers that split data based on conditions.
- 100 Decision Trees

Training:

- Loaded **JSON** data into a DataFrame with 4 features (**HR, HRV, RHR, HHR**).
- Split into training (80%) and testing (20%) sets.
- Trained **Random Forest model** on the training set.


Prediction


FastAPI

Prediction

Initial Risk Prediction

- If Prediction = 0 :
  - No risk, continue monitoring
- If Prediction = 1:
  - Prompt user to record ECG:

ECG

FastAPI

ECG

Risk, Contact EMS

Final Risk Model

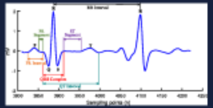
Heart Rate (HR)

Datasource: MIT-BIH

- HR Computation: Calculated as **HR = 60/RR** ; where **RR** is the interval between consecutive **R-peaks (seconds)**.

Model: Rule-Based Threshold

- Binary Classification: **High HR (1) if HR > 110 bpm, Normal HR (0)** otherwise.
- No training required as it is a fixed threshold



Results

MIT-BIH

	Normal	Arrhythmia
Normal	23063	0
Arrhythmia	0	12

INCART

	Normal	Arrhythmia
Normal	58951	0
Arrhythmia	0	78

High Heart Rate Events (HHR)

Datasource: MITBIH

- **RR intervals** calculated from **R-peaks in ECG**, converted to **HR (BPM)**
- **Features:**
  - Duration above 150 BPM
  - sustained high HR (all > 150 BPM),
  - max, min, avg HR
  - HR slope
  - spike frequency
  - HR standard deviation.

Model: Random Forest Classifier

- **Random Forest:** Aggregates predictions from multiple decision trees for robust classification.
- **Decision Trees:** Individual tree classifiers that split data based on conditions
- 300 Trees, max depth of 10, balanced class weights

Results

MIT-BIH

	Normal	Arrhythmia
Normal	59583	0
Arrhythmia	0	53117

INCART

	Normal	Arrhythmia
Normal	20674	0
Arrhythmia	0	14656

Training:

- Extracted **HR features** scaled using **StandardScaler**, balanced with SMOTE (Synthetic Minority Over-sampling - interpolating between minority class samples to introduce variation) and mapped to labels:
  - High HR event = High duration above threshold + sustained High HR
  - Normal event = Lower values or mixed patterns

Resting Heart Rate (RHR)

Datasource: MIMIC-III

- **ECG signals** processed to detect **RR** intervals using peak detection.
- **Mean HR** and **standard deviation** (filtered for low variation, **std < 5 for MIMIC, std < 15 for** INCART).

Model: Logistic Regression

- **Linear model** that predicts binary outcomes using the sigmoid function

$$\frac{e^{(\beta_0+\beta_1 x)}}{1+e^{(\beta_0+\beta_1 x)}}$$

Training:

- Extracted **RHR** values standardized with **StandardScaler**, labeled as **high** (1) or **normal** (0) based on a **75 BPM** threshold, and split into training/test sets.

Results

MIMIC-III

	Normal	Arrhythmia
Normal	96	0
Arrhythmia	0	218

INCART

	Normal	Arrhythmia
Normal	957	1
Arrhythmia	0	1187

Heart Rate Variability (HRV)

Datasource: MIT-BIH

- **RR intervals computed** from **ECG** signals
- Calculated **14 advanced HRV features** for each **30-beat** window:
  - **Time-Domain:**
    - RMSSD (Root Mean Square of Successive Differences)
    - SDNN (Standard Deviation of NN (Normal-Normal) Intervals
    - NN50 (Number of consecutive RR intervals differing by more than 50ms.
    - pNN50 (Percentage of NN50 relative to total intervals)
    - Mean, Min, Max, Skew abd Kurtosis of RR Intervals
  - **Frequency-Domain:**
    - LF (Low-Frequency), HF (High-Frequency), LF/HF ratio using Welch's method.
  - **Nonlinear**: SD1, SD2 (Short and Long term HRV), Shannon entropy (unpredictability)

Training:

- **HRV features** scaled with StandardScaler
- Models trained using optimized **hyperparameters** (RandomizedSearchCV).
- Ensemble model trained with **StackingClassifier** using predictions from **XGBoost**, **LightGBM**, and **CatBoost**.

Results

MIT-BIH

	Normal	Arrhythmia
Normal	18735	12491
Arrhythmia	2050	2298
Precision	0.87	0.22
Recall	0.60	0.53
Accuracy	0.62	

INCART

	Normal	Arrhythmia
Normal	25135	37154
Arrhythmia	4156	8895
Precision	0.90	0.15
Recall	0.15	0.68
Accuracy	0.46	

Models:

- **Gradient Boosting**: Combines weak models (trees) to improve accuracy.
  - **XGBoost:** Fast, depth-controlled trees.
  - **LightGBM:** Leaf-wise growth, fast training.
  - **CatBoost:** Categorical feature support.
- **Ensemble (Stacking Classifier):**
  - Combines predictions from XGBoost, LightGBM, and CatBoost using Logistic Regression for improved accuracy.

Electrocardiogram (ECG)

Datasource: MITBIH/Holter

- **ECG beat** segments from raw signals classified into
  - **Low Risk: Normal rhythms:**
    - N (Normal)
    - L (Left bundle branch block)
    - R (Right bundle branch block)
    - e (Atrial escape)
    - j (Nodal escape).
  - **Medium Risk: Supraventricular arrhythmias**
    - A (Atrial premature)
    - S (Supraventricular)
    - a (Aberrated atrial)
    - J (Nodal premature)
    - ? (Unclassifiable)
  - **High Risk: Ventricular arrhythmias:**
    - V (Premature ventricular)
    - F (Fusion of ventricular and normal)
    - E (Ventricular escape).

Results

OHCA

	Low	Medium	High
Low	0	0	0
Medium	0	0	0
High	80	39	3455
Precision	0.00	0.00	0.99
Recall	0.00	0.00	0.97
Accuracy	0.97		

MIT-BIH + Holter

	Low	Medium	High
Low	24177	75	254
Medium	0	475	1
High	170	104	23650
Precision	0.99	0.73	0.99
Recall	0.99	0.99	0.99
Accuracy	0.99		

INCART

	Low	Medium	High
Low	91664	43100	18832
Medium	545	1355	59
High	3473	1001	15751
Precision	0.96	0.03	0.45
Recall	0.60	0.69	0.78
Accuracy	0.62		

Model: BiLSTM

- **Bidirectional Long Short-Term Memory (BiLSTM)** - a type of **recurrent neural network (RNN)** that learns from both past and future contexts in sequential data, enhancing pattern recognition - with 3 hidden layers (128 hidden units each), followed by dropout (0.3) and a fully connected output layer.
- Captures bidirectional temporal dependencies in **ECG** waveforms for robust pattern recognition.
- ,Takes 250-sample **ECG** beat segments as input and outputs multiclass classification (0: **Low Risk, 1: Medium Risk, 2: High Risk**).

Training:

- **CrossEntropyLoss** with class weights ensures all risk levels are learned equally
- Adam optimizer with a learning rate of **0.0005** for efficient training
- Medium-risk beats are augmented **10x** with random transformations to ensure balanced training
- **30 epochs** (full training passes through data), monitoring validation accuracy to prevent overfitting
- Trained in small groups of **32 beats** for faster, stable learning

Final Ensemble

Extraction Process:

- Extracts **HR**, **HRV**, **RHR**, **HHR**, and **ECG** features using **BiLSTM** (3 layers, 128 hidden units, bidirectional).
- Applies **SMOTE** (Synthetic Minority Oversampling) for balanced training data.

Model: Adaptive Meta-Learner

- Combines predictions from **HR**, **HRV**, **RHR**, **HHR**, and **ECG** (**BiLSTM**) using a **Meta-Learner** (**Neural Network**).
- **False Negative Reduction**: Adaptive voting adjusts high-risk predictions using Focal Loss.

Results

OHCA

	Low	Medium	High
Low	0	0	0
Medium	0	0	0
High	31	19	3524
Precision	0.00	0.00	0.99
Recall	0.00	0.00	0.98
Accuracy	0.98		

MIT-BIH + Holter

	Low	Medium	High
Low	24200	60	246
Medium	0	475	1
High	62	38	23824
Precision	0.99	0.89	0.99
Recall	0.99	0.99	0.99
Accuracy	0.99		

INCART

	Low	Medium	High
Low	105679	29417	18500
Medium	500	1400	59
High	2499	1001	16725
Precision	0.96	0.05	0.47
Recall	0.69	0.71	0.83
Accuracy	0.67		

Training:

- Fine-tunes **BiLSTM** with feedback samples (high-risk FN corrections) for **25 epochs** (LR: **0.0001**).
- Optimized with **Focal Loss** (·=0.95, ·=2.0) to further minimize **false negatives**.
- Tracks accuracy, confusion matrix, and false negative rates.

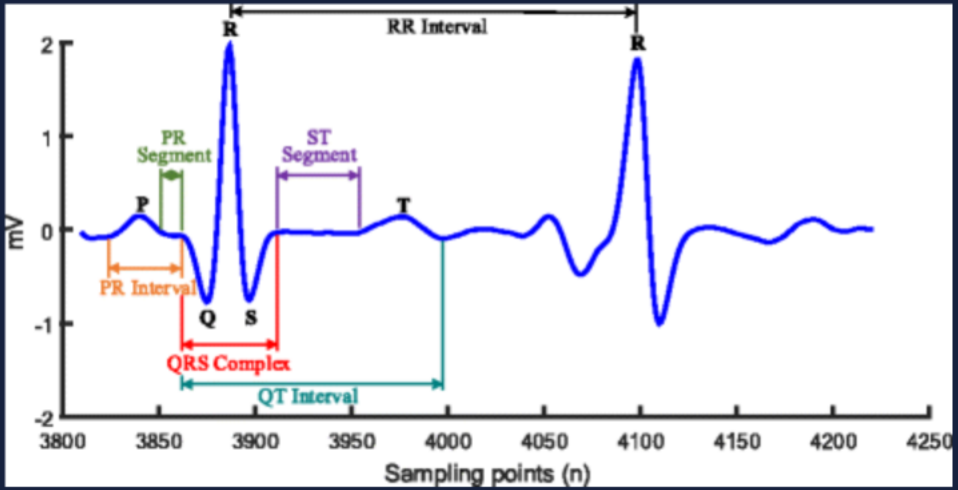
Heart Rate (HR)

Datasource: MIT-BIH

- HR Computation: Calculated as **HR = 60/RR :** where **RR** is the interval between consecutive **R-peaks (seconds).**

Model: Rule-Based Threshold

- Binary Classification: **High HR (1) if HR > 110 bpm**, **Normal HR (0)** otherwise.
- No training required as it is a fixed threshold



Results

MIT-BIH

	Normal	Arrhythmia
Normal	23063	0
Arrhythmia	0	12

INCART

	Normal	Arrhythmia
Normal	58951	0
Arrhythmia	0	78

High Heart Rate Events (HHR)

Datasource: MITBIH

- **RR intervals** calculated from **R-peaks in ECG**, converted to **HR (BPM)**.
- **Features:**
  - Duration above 150 BPM
  - sustained high HR (all > 150 BPM),
  - max, min, avg HR,
  - HR slope
  - spike frequency
  - HR standard deviation.

Model: Random Forest Classifier

- **Random Forest:** Aggregates predictions from multiple decision trees for robust classification.
- **Decision Trees:** Individual tree classifiers that split data based on conditions.

- 300 Trees, max depth of 10, balanced class weights

Results

MIT-BIH

	Normal	Arrhythmia
Normal	59583	0
Arrhythmia	0	53117

INCART

	Normal	Arrhythmia
Normal	20674	0
Arrhythmia	0	14656

Training:

- Extracted **HR features** scaled using **StandardScaler**, balanced with SMOTE (Synthetic Minority Over-sampling - interpolating between minority class samples to introduce variation) and mapped to labels:

  - High HR event = High duration above threshold + sustained High HR
  - Normal event = Lower values or mixed patterns

Resting Heart Rate (RHR)

Datasource: MIMIC-III

- **ECG signals** processed to detect **RR** intervals using peak detection.
- **Mean HR** and **standard deviation** (filtered for low variation, **std < 5 for MIMIC, std < 15** for INCART).

Model: Logistic Regression

- **Linear model** that predicts binary outcomes using the sigmoid function

$$\frac{e^{(\beta_0+\beta_1 x)}}{1+e^{(\beta_0+\beta_1 x)}}$$

Training:

- Extracted **RHR** values standardized with **StandardScaler**, labeled as **high (1)** or **normal (0)** based on a **75 BPM** threshold, and split into training/test sets.

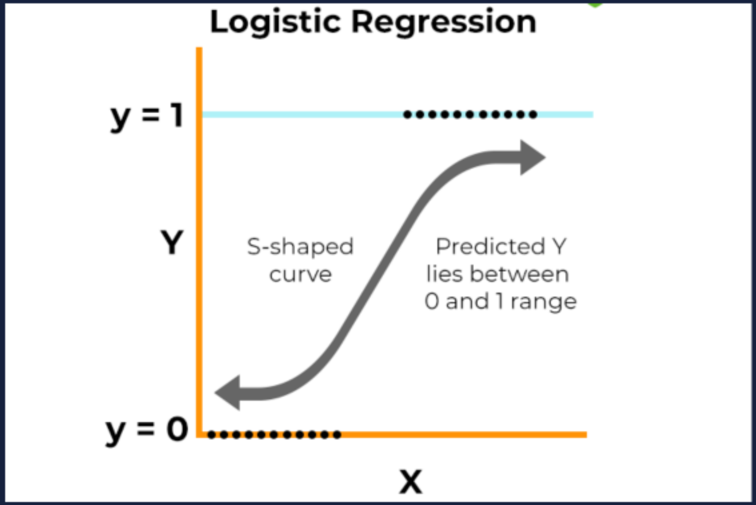
Results

MIMIC-III

	Normal	Arrhythmia
Normal	96	0
Arrhythmia	0	218

INCART

	Normal	Arrhythmia
Normal	957	1
Arrhythmia	0	1187



Heart Rate Variability (HRV)

Datasource: MIT-BIH

- **RR intervals computed** from **ECG** signals
- Calculated **14 advanced HRV features** for each **30-beat** window:
  - **Time-Domain:**
    - RMSSD (Root Mean Square of Successive Differences)
    - SDNN (Standard Deviation of NN (Normal-Normal) Intervals
    - NN50 (Number of consecutive RR intervals differing by more than 50ms,
    - pNN50 (Percentage of NN50 relative to total intervals)
    - Mean, Min, Max, Skew abd Kurtosis of RR Intervals
  - **Frequency-Domain:**
    - LF (Low-Frequency), HF (High-Frequency), LF/HF ratio using Welch's method.
  - **Nonlinear**: SD1, SD2 (Short and Long term HRV), Shannon entropy (unpredictability)

Training:

- **HRV features** scaled with StandardScaler.
- Models trained using optimized **hyperparameters** (RandomizedSearchCV).
- Ensemble model trained with **StackingClassifier** using predictions from **XGBoost**, **LightGBM**, and **CatBoost**.

Results

MIT-BIH

	Normal	Arrhythmia
Normal	18735	12491
Arrhythmia	2050	2298

Precision	0.87	0.22
Recall	0.60	0.53

Accuracy	0.62

INCART

	Normal	Arrhythmia
Normal	25135	37154
Arrhythmia	4156	8895

Precision	0.90	0.15
Recall	0.15	0.68

Accuracy	0.46

Models:

- **Gradient Boosting:** Combines weak models (trees) to improve accuracy.
  - **XGBoost:** Fast, depth-controlled trees.
  - **LightGBM:** Leaf-wise growth, fast training.
  - **CatBoost:** Categorical feature support.
- **Ensemble (Stacking Classifier):**
  - Combines predictions from XGBoost, LightGBM, and CatBoost using Logistic Regression for improved accuracy.

Electrocardiogram (ECG)

Datasource: MITBIH/Holter

- **ECG beat** segments from raw signals classified into:
  - **Low Risk: Normal rhythms:**
    - N (Normal)
    - **L** (Left bundle branch block)
    - **R** (Right bundle branch block)
    - **e** (Atrial escape)
    - j (Nodal escape).
  - **Medium Risk: Supraventricular arrhythmias**
    - **A** (Atrial premature)
    - S (Supraventricular)
    - **a** (Aberrated atrial)
    - **J** (Nodal premature)
    - **? (Unclassifiable**).
  - **High Risk: Ventricular arrhythmias:**
    - **V** (Premature ventricular)
    - **F** (Fusion of ventricular and normal)
    - **E** (Ventricular escape).

Results

OHCA

	Low	Medium	High
Low	0	0	0
Medium	0	0	0
High	80	39	3455

Precision	0.00	0.00	0.99
Recall	0.00	0.00	0.97

Accuracy	0.97

MIT-BIH + Holter

	Low	Medium	High
Low	24177	75	254
Medium	0	475	1
High	170	104	23650

Precision	0.99	0.73	0.99
Recall	0.99	0.99	0.99

Accuracy	0.99

INCART

	Low	Medium	High
Low	91664	43100	18832
Medium	545	1355	59
High	3473	1001	15751

Precision	0.96	0.03	0.45
Recall	0.60	0.69	0.78

Accuracy	0.62

Model: BiLSTM

- **Bidirectional Long Short-Term Memory (BiLSTM)** - a type of **recurrent neural network (RNN)** that learns from both past and future contests in sequential data, enhancing pattern recognition - with 3 hidden layers (128 hidden units each), followed by dropout (0.3) and a fully connected output layer.
- Captures bidirectional temporal dependencies in **ECG** waveforms for robust pattern recognition.
- ,Takes 250-sample **ECG** beat segments as input and outputs multiclass classification **(0: Low Risk, 1: Medium Risk, 2: High Risk).**

Training:

- **CrossEntropyLoss** with class weights ensures all risk levels are learned equally
- Adam optimizer with a learning rate of **0.0005** for efficient training.
- Medium-risk beats are augmented **10x** with random transformations to ensure balanced training.
- **30 epochs** (full training passes through data), monitoring validation accuracy to prevent overfitting
- Trained in small groups of **32 beats** for faster, stable learning

Final Risk Model

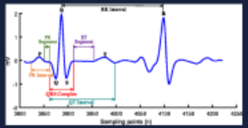
Heart Rate (HR)

Datasource: MIT-BIH

- HR Computation: Calculated as **HR = 60/RR** : where **RR** is the interval between consecutive **R-peaks (seconds)**.

Model: Rule-Based Threshold

- Binary Classification: **High HR (1) if HR > 110 bpm**, **Normal HR (0)** otherwise.
- No training required as it is a fixed threshold



Results

MIT-BIH

	Normal	Arrhythmia
Normal	23063	0
Arrhythmia	0	12

INCART

	Normal	Arrhythmia
Normal	58951	0
Arrhythmia	0	78

High Heart Rate Events (HHR)

Datasource: MITBIH

- **RR intervals** calculated from **R-peaks in ECG**, converted to **HR (BPM)**.
- **Features:**
  - Duration above 150 BPM
  - sustained high HR (all > 150 BPM),
  - max, min, avg HR,
  - HR slope
  - spike frequency
  - HR standard deviation.

Model: Random Forest Classifier

- **Random Forest:** Aggregates predictions from multiple decision trees for robust classification.
- **Decision Trees:** Individual tree classifiers that split data based on conditions.
- 300 Trees, max depth of 10, balanced class weights

Results

MIT-BIH

	Normal	Arrhythmia
Normal	59583	0
Arrhythmia	0	53117

INCART

	Normal	Arrhythmia
Normal	20674	0
Arrhythmia	0	14656

Training:

- Extracted **HR features** scaled using **StandardScaler**, balanced with SMOTE (Synthetic Minority Over-sampling - interpolating between minority class samples to introduce variation) and mapped to labels:
  - High HR event = High duration above threshold + sustained High HR
  - Normal event = Lower values or mixed patterns

Resting Heart Rate (RHR)

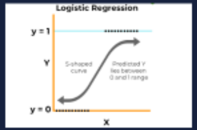
Datasource: MIMIC-III

- **ECG signals** processed to detect **RR** intervals using peak detection.
- **Mean HR** and **standard deviation** (filtered for low variation, **std < 5 for MIMIC, std < 15** for INCART).

Model: Logistic Regression

- **Linear model** that predicts binary outcomes using the sigmoid function

$$\frac{e^{(\beta_0+\beta_1 x)}}{1+e^{(\beta_0+\beta_1 x)}}$$



Training:

- Extracted **RHR** values standardized with **StandardScaler**, labeled as **high (1)** or **normal (0)** based on a **75 BPM** threshold, and split into training/test sets.

Results

MIMIC-III

	Normal	Arrhythmia
Normal	96	0
Arrhythmia	0	218

INCART

	Normal	Arrhythmia
Normal	957	1
Arrhythmia	0	1187

Heart Rate Variability (HRV)

Datasource: MIT-BIH

- **RR intervals computed** from **ECG** signals
- Calculated **14 advanced HRV features** for each **30-beat** window:
  - **Time-Domain:**
    - RMSSD (Root Mean Square of Successive Differences)
    - SDNN (Standard Deviation of NN [Normal-Normal] Intervals
    - NN50 (Number of consecutive RR intervals differing by more than 50ms.
    - pNN50 (Percentage of NN50 relative to total intervals)
    - Mean, Min, Max, Skew abd Kurtosis of RR Intervals
  - **Frequency-Domain:**
    - LF (Low-Frequency), HF (High-Frequency), LF/HF ratio using Welch's method.
  - **Nonlinear:** SD1, SD2 (Short and Long term HRV), Shannon entropy (unpredictability)

Training:

- **HRV features** scaled with StandardScaler.
- Models trained using optimized **hyperparameters** (RandomizedSearchCV).
- Ensemble model trained with **StackingClassifier** using predictions from **XGBoost**, **LightGBM**, and **CatBoost**.

Results

MIT-BIH

	Normal	Arrhythmia
Normal	18735	12491
Arrhythmia	2050	2298
Precision	0.87	0.22
Recall	0.60	0.53
Accuracy	0.62	

INCART

	Normal	Arrhythmia
Normal	25135	37154
Arrhythmia	4156	8895
Precision	0.90	0.15
Recall	0.15	0.68
Accuracy	0.46	

Models:

- **Gradient Boosting:** Combines weak models (trees) to improve accuracy.
  - **XGBoost:** Fast, depth-controlled trees.
  - **LightGBM:** Leaf-wise growth, fast training.
  - **CatBoost:** Categorical feature support.
- **Ensemble (Stacking Classifier):**
  - Combines predictions from XGBoost, LightGBM, and CatBoost using Logistic Regression for improved accuracy.

Final Ensemble

Extraction Process:

- Extracts **HR**, **HRV**, **RHR**, **HHR**, and **ECG** features using **BiLSTM** (3 layers, 128 hidden units, bidirectional).
- Applies **SMOTE** (Synthetic Minority Oversampling) for balanced training data.

Model: Adaptive Meta-Learner

- Combines predictions from **HR**, **HRV**, **RHR**, **HHR**, and **ECG** (**BiLSTM**) using a **Meta-Learner** (**Neural Network**).
- **False Negative Reduction**: Adaptive voting adjusts high-risk predictions using Focal Loss.

Results

OHCA

	Low	Medium	High
Low	0	0	0
Medium	0	0	0
High	31	19	3524
Precision	0.00	0.00	0.99
Recall	0.00	0.00	0.98
Accuracy	0.98		

MIT-BIH + Holter

	Low	Medium	High
Low	24200	60	246
Medium	0	475	1
High	62	38	23824
Precision	0.99	0.89	0.99
Recall	0.99	0.99	0.99
Accuracy	0.99		

Training:

- Fine-tunes **BiLSTM** with feedback samples (high-risk FN c
- **epochs** (LR: **0.0001**).
- Optimized with **Focal Loss** (·=0.95, ·=2.0) to further minim
- Tracks accuracy, confusion matrix, and false negative rates.

Electrocardiogram (ECG)

Datasource: MITBIH/Holter

- **ECG beat** segments from raw signals classified into:
  - **Low Risk: Normal rhythms:**
    - N (Normal)
    - **L** (Left bundle branch block)
    - **R** (Right bundle branch block)
    - **e** (Atrial escape)
    - j (Nodal escape).
  - **Medium Risk: Supraventricular arrhythmias**
    - **A** (Atrial premature)
    - S (Supraventricular)
    - a (Aberrated atrial)
    - J (Nodal premature)
    - ? (Unclassifiable).
  - **High Risk: Ventricular arrhythmias:**
    - **V** (Premature ventricular)
    - **F** (Fusion of ventricular and normal)
    - **E** (Ventricular escape).

Results

OHCA

	Low	Medium	High
Low	0	0	0
Medium	0	0	0
High	80	39	3455
Precision	0.00	0.00	0.99
Recall	0.00	0.00	0.97
Accuracy	0.97		

MIT-BIH + Holter

	Low	Medium	High
Low	24177	75	254
Medium	0	475	1
High	170	104	23650
Precision	0.99	0.73	0.99
Recall	0.99	0.99	0.99
Accuracy	0.99		

INCART

	Low	Medium	High
Low	91664	45100	18832
Medium	545	1355	59
High	3473	1001	15751
Precision	0.96	0.03	0.45
Recall	0.60	0.69	0.78
Accuracy	0.62		

Model: BiLSTM

- **Bidirectional Long Short-Term Memory (BiLSTM)** - a type of **recurrent neural network (RNN)** that learns from both past and future contests in sequential data, enhancing pattern recognition - with 3 hidden layers (128 hidden units each), followed by dropout (0.3) and a fully connected output layer.
- Captures bidirectional temporal dependencies in **ECG** waveforms for robust pattern recognition.
- .Takes 250-sample **ECG** beat segments as input and outputs multiclass classification **(0: Low Risk, 1: Medium Risk, 2: High Risk)**.

Training:

- **CrossEntropyLoss** with class weights ensures all risk levels are learned equally
- Adam optimizer with a learning rate of **0.0005** for efficient training.
- Medium-risk beats are augmented **10x** with random transformations to ensure balanced training.
- **30 epochs** (full training passes through data), monitoring validation accuracy to prevent overfitting
- Trained in small groups of **32 beats** for faster, stable learning

Final Ensemble

Extraction Process:

- Extracts **HR**, **HRV**, **RHR**, **HHR**, and **ECG** features using **BiLSTM** (3 layers, 128 hidden units, bidirectional).
- Applies **SMOTE** (Synthetic Minority Oversampling) for balanced training data.

Model: Adaptive Meta-Learner

- Combines predictions from **HR**, **HRV**, **RHR**, **HHR**, and **ECG** (**BiLSTM**) using a **Meta-Learner** (**Neural Network**).
- **False Negative Reduction**: Adaptive voting adjusts high-risk predictions using Focal Loss.

Results

OHCA

	Low	Medium	High
Low	0	0	0
Medium	0	0	0
High	31	19	3524

Precision	0.00	0.00	0.99
Recall	0.00	0.00	0.98

Accuracy	0.98

MIT-BIH + Holter

	Low	Medium	High
Low	24200	60	246
Medium	0	475	1
High	62	38	23824

Precision	0.99	0.89	0.99
Recall	0.99	0.99	0.99

Accuracy	0.99

INCART

	Low	Medium	High
Low	105679	29417	18500
Medium	500	1400	59
High	2499	1001	16725

Precision	0.96	0.05	0.47
Recall	0.69	0.71	0.83

Accuracy	0.67

Training:

- Fine-tunes **BiLSTM** with feedback samples (high-risk FN corrections) for **25 epochs** (LR: **0.0001**).
- Optimized with **Focal Loss** (·=0.95, ·=2.0) to further minimize **false negatives**.
- Tracks accuracy, confusion matrix, and false negative rates.


Prediction

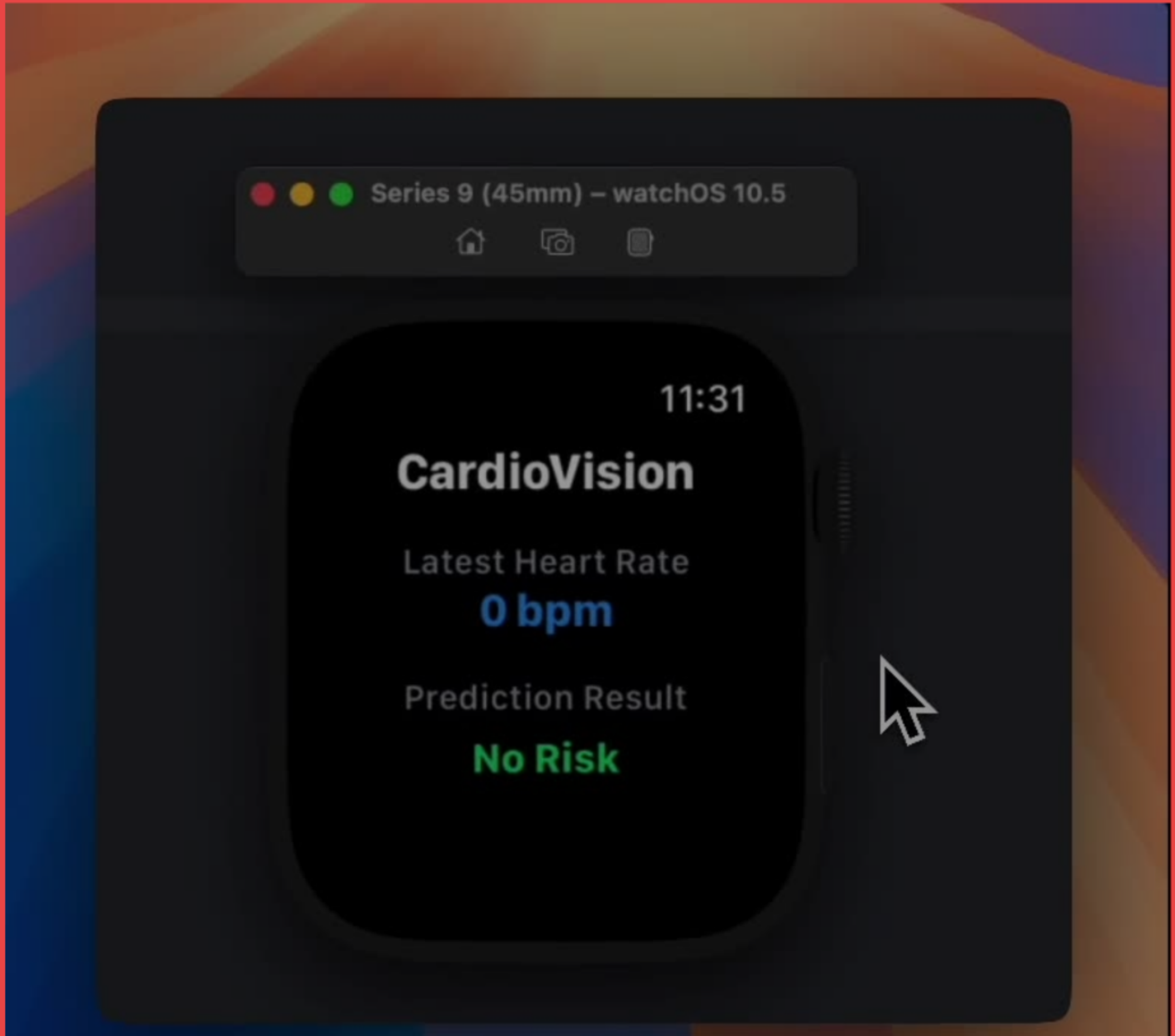
FastAPI

Prediction

Final Risk Prediction

- If Prediction = 0 :
  - False Alarm
- If Prediction = 1:
  - Cardiac Symptoms, Monitor
- If Prediction = 2:
  - Arrest Risk, Contact EMS

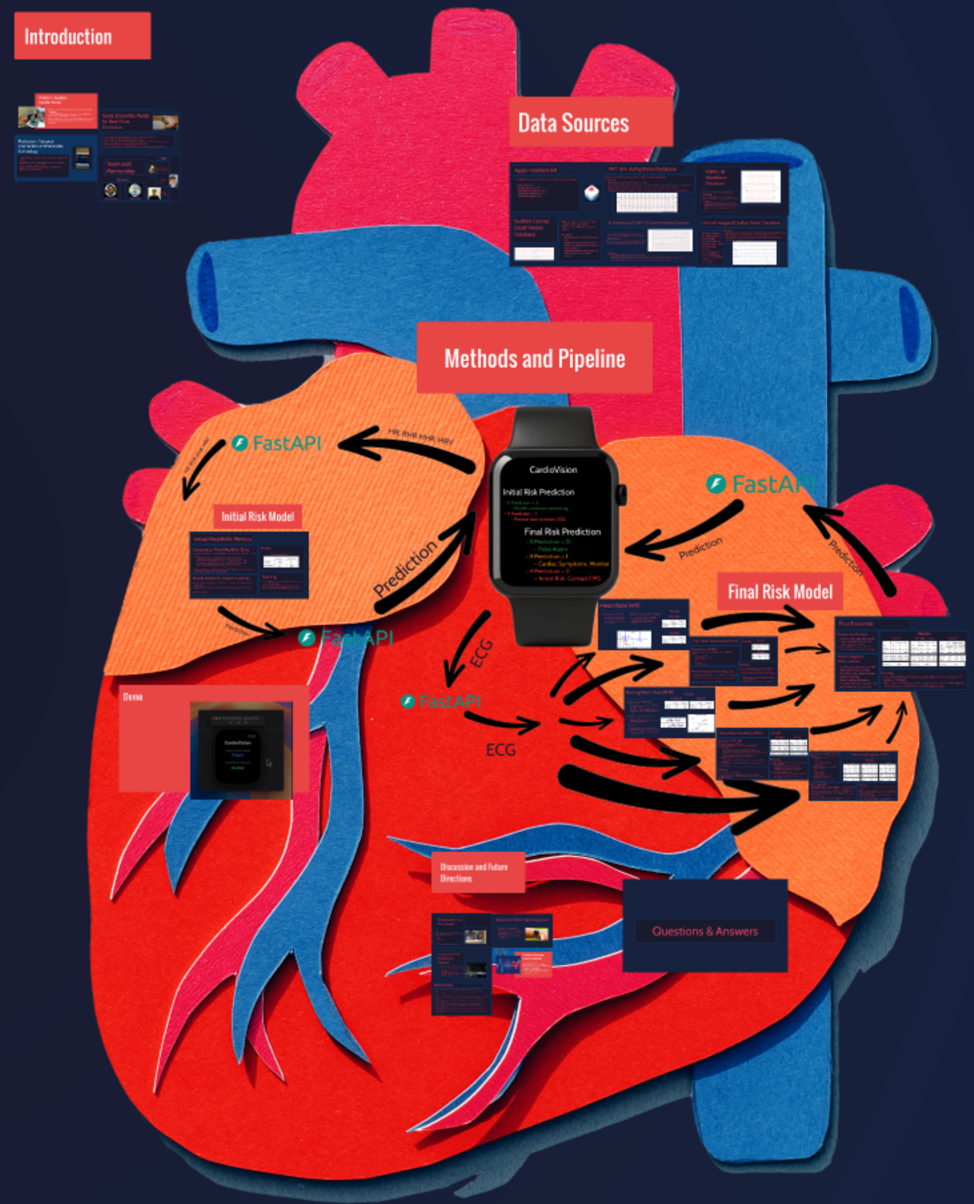
Demo


Series 9 (45mm) – watchOS 10.5
11:31
CardioVision
Latest Heart Rate
0 bpm
Prediction Result
No Risk


Series 9 (45mm) – watchOS 10.5
11:31
CardioVision
Latest Heart Rate
0 bpm
Prediction Result
No Risk

CardioVision: Predicting Cardiac Arrest Risk

An Ensemble Approach Utilizing Wearable Physiological Data



Discussion and Future Directions

Enhanced Initial Risk Model

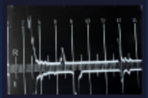
- **Expanded Metrics**: Integrate additional Apple Watch metrics to refine initial risk assessment such as:
  - HRV
  - VO2 Max
  - Blood Oxygen
  - Respiratory Rate
  - Stress
- Design to track and account for previous inputs over time to identify trends and make more informed predictions

Advanced Watch App Integration

- Deployment on **Apple Watch**
- Reliable and private data storage
- Real-time data streaming with **Kafka** and **FastAPI** server
- Have user get notification if **initial** risk identified

Improved HRV Model and Dataset

- **HRV model** struggles with distinguishing arrhythmias due to **limited arrhythmia-specific data**.
- Use the **PhysioNet MIT-BIH Atrial Fibrillation Database**, which is specifically designed for arrhythmia detection.



Conclusion: Promising Ensemble Approach

- Accurate Risk Prediction
- **Real-Time** Assessment without **ECG**
- **Modular Design**: Each model **independently trainable** and **replaceable**
- **Future Directions**: Further model training and optimization, Apple Watch and other smart watch integration

References

Enhanced Initial Risk Model

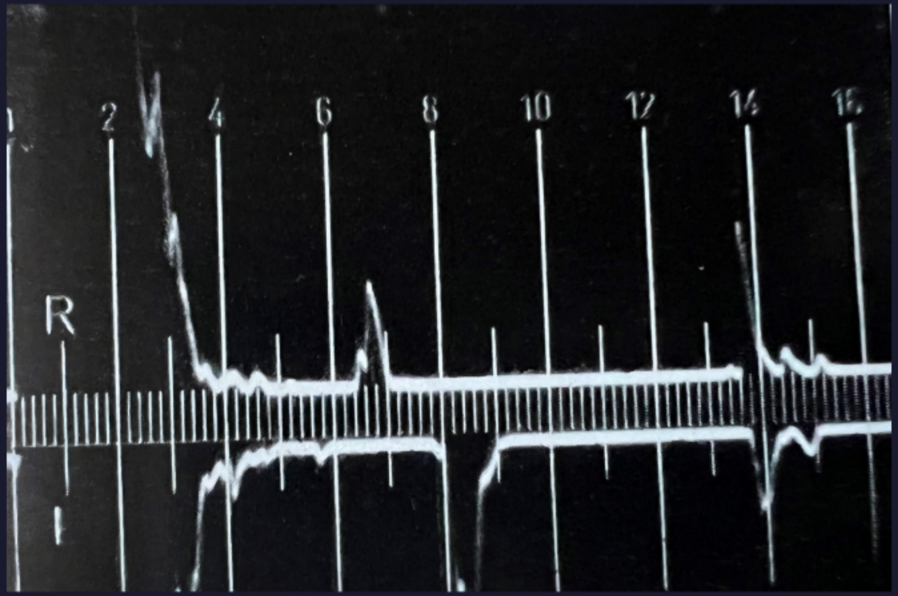
- **Expanded Metrics**: Integrate additional Apple Watch metrics to refine initial risk assessment such as:
  - HRV
  - VO2 Max
  - Blood Oxygen
  - Respiratory Rate
  - Stress
- Design to track and account for previous inputs over time to identify trends and make more informed predictions

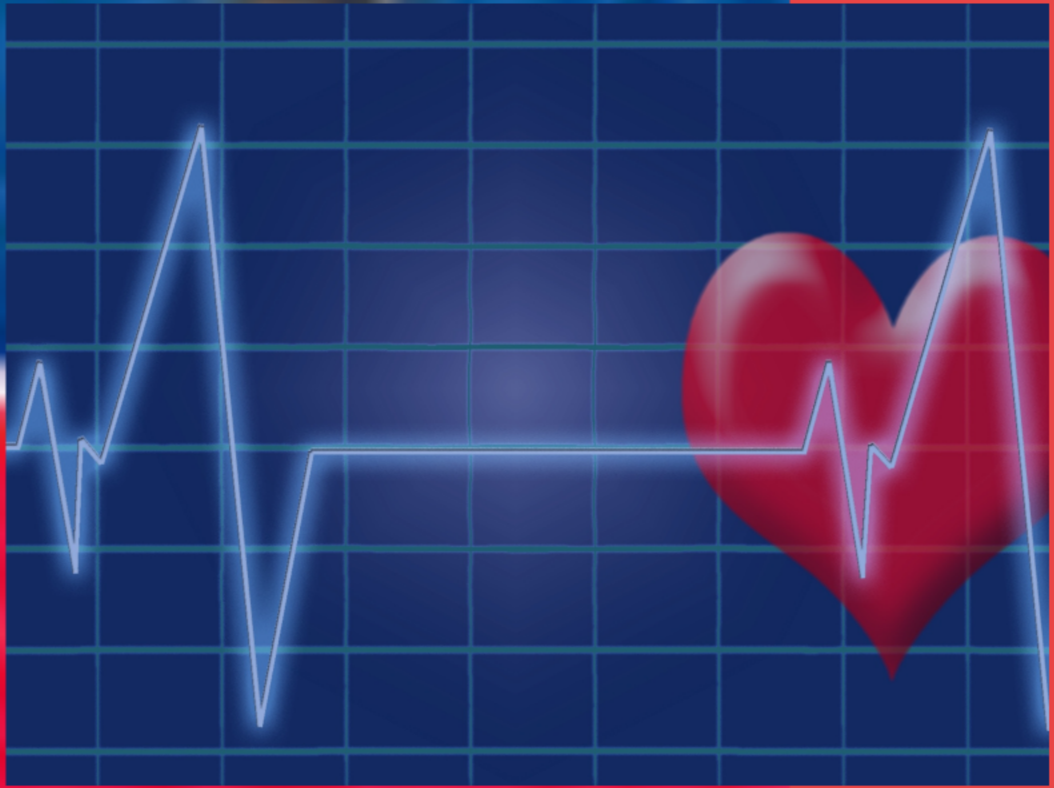
Advanced Watch App Integration

- Deployment on **Apple Watch**
- Reliable and private data storage
- Real-time data streaming with **Kafka** and **FastAPI** server
- Have user get notification if **initial risk** identified

Improved HRV Model and Dataset

- **HRV model** struggles with distinguishing arrhythmias due to **limited arrhythmia-specific data.**
- Use the **PhysioNet MIT-BIH Atrial Fibrillation Database,** which is specifically designed for arrhythmia detection.





Conclusion: Promising Ensemble Approach

- Accurate Risk Prediction
- **Real-Time** Assessment without **ECG**
- **Modular Design**: Each model **independently trainable** and **replaceable**
- **Future Directions**: Further model training and optimization, Apple Watch and other smart watch integration

References

- Apple HealthKit Documentation: Apple Developer. Available at: https://developer.apple.com/documentation/healthkit
- OHCA Database: "Out-of-Hospital Cardiac Arrest Registry," PMC. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC7753135/
- INCART Database: "The St. Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database," PhysioNet. Available at: https://archive.physionet.org/physiobank/database/incartdb/
- Holter Database (SDDB): "Sudden Cardiac Death Holter Database," PhysioNet. Available at: https://archive.physionet.org/physiobank/database/sddb/
- MIMIC-III Waveform Database: "MIMIC-III Waveform Database (Version 1.0)," PhysioNet. Available at: https://physionet.org/content/mimic3wdb/1.0/
- MIT-BIH Arrhythmia Database: "MIT-BIH Arrhythmia Database (Version 1.0.0)," PhysioNet. Available at: https://www.physionet.org/content/mitdb/1.0.0/
- American Heart Association. (n.d.). Tachycardia (Fast Heart Rate). Retrieved from https://www.heart.org/en/health-topics/arrhythmia/about-arrhythmia/tachycardia--fast-heart-rate
- Kubios. (n.d.). Heart Rate Variability (HRV) Normal Range. Retrieved from https://www.kubios.com/blog/heart-rate-variability-normal-range/
- WebMD. (n.d.). What Is Heart Rate Variability?. Retrieved from https://www.webmd.com/heart/what-is-heart-rate-variability
- American Heart Association. (n.d.). Tachycardia (Fast Heart Rate). Retrieved from https://www.heart.org/en/health-topics/arrhythmia/about-arrhythmia/tachycardia--fast-heart-rate
- American Heart Association. (n.d.). Tachycardia (Fast Heart Rate). Retrieved from https://www.heart.org/en/health-topics/arrhythmia/about-arrhythmia/tachycardia--fast-heart-rate
- Harvard Health Publishing. (2019, March). What your heart rate is telling you. Retrieved from https://www.health.harvard.edu/heart-health/what-your-heart-rate-is-telling-you

Questions & Answers

CardioVision: Predicting Cardiac Arrest Risk

An Ensemble Approach Utilizing Wearable Physiological Data