# Context-Aware Weighted KNN with Locally Adaptive k

Anieesh Saravanan

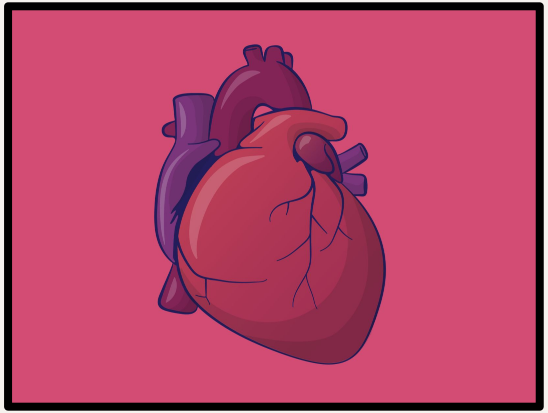
03/28/25

You know that feeling when you meet someone and your heart skips a beat?

Yeah, that's arrhythmia. You can die from that.

https://rusafu.com/quotes/tag/arrhythmia/

# Table of contents



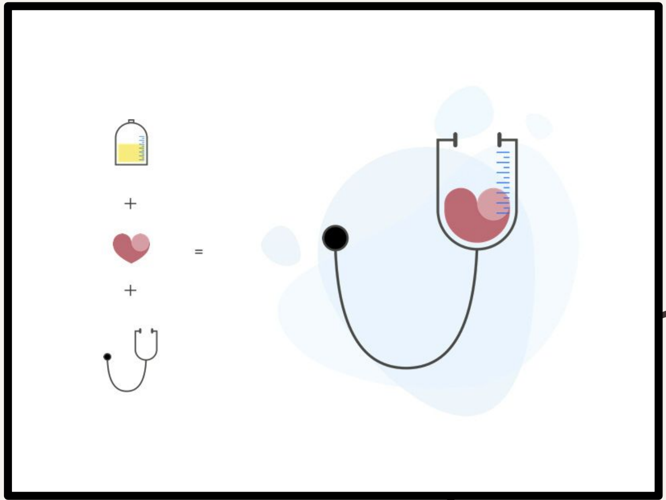
# 00
# Introduction

# Project Goal

- Develop a binary classification model for arrhythmia
- Diagnose patients with arrhythmia based on patient biometrics and ECG readings
- Streamline diagnosis process with supplementary outlook of data

# Arrhythmia Overview

- Arrhythmia refers to irregular heartbeats
- Antiarrhythmic medications/therapies exist
- No simple guidelines for diagnosis
- Widely varying clinical manifestations and minimal warning signs
- Even experienced cardiologists may misinterpret ECG patterns



# Common Myths

**Myth**: Arrhythmia always presents with severe, dramatic symptoms

**Fact:** Many arrhythmias can be **subtle or even asymptomatic**

**Myth**: Only unhealthy/elderly individuals develop arrhythmia

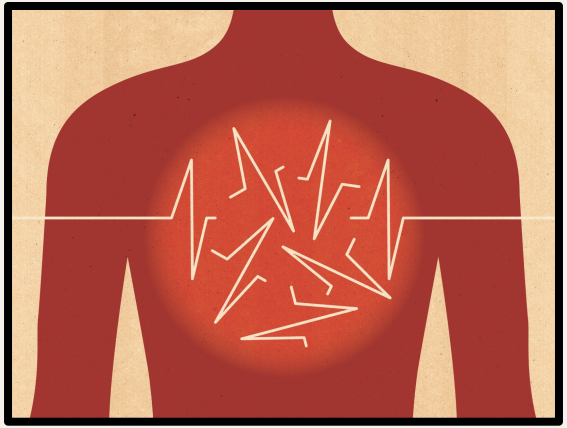
**Fact**: Arrhythmia can occur in healthy individuals (without heart conditions)

**Myth**: Abnormal heart rhythms are always obvious and easy to detect

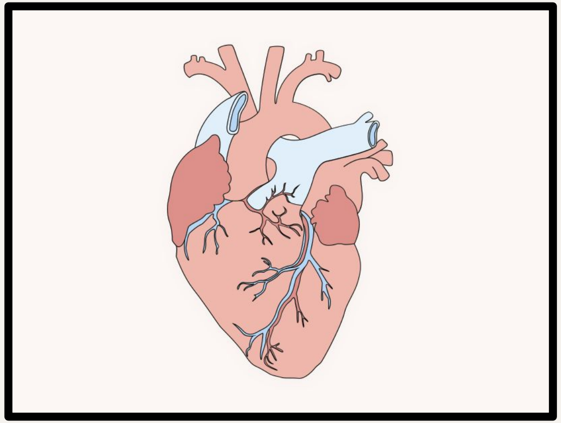
**Fact**: Many arrhythmias require **specialized diagnostic tools** (like extended ECG monitoring) that may not be apparent with a simple pulse check

# Impact

In 2021, arrhythmia was mentioned on at least **232,030** death certificates

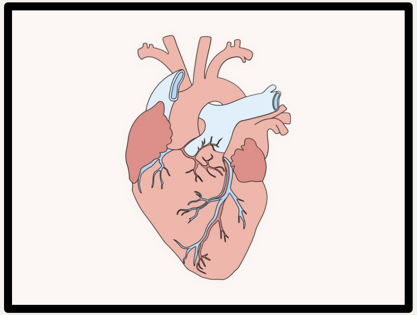


# 01
# Project Overview



# Dataset Overview

- **Quantitative/Categorical** dataset with **452 instances** and **279 attributes**
- Class attribute is the diagnosis of the patient (nonarrhythmia or type of arrhythmia)
- Sourced from UCI and originally compiled by H. Altay Guvenir
- <u>Missing values</u>: instances regarding vector angles (ECG electrode similarity)
- <u>Class distribution</u>: 54.20% nonarrhythmia, 45.80% arrhythmia



# Attributes

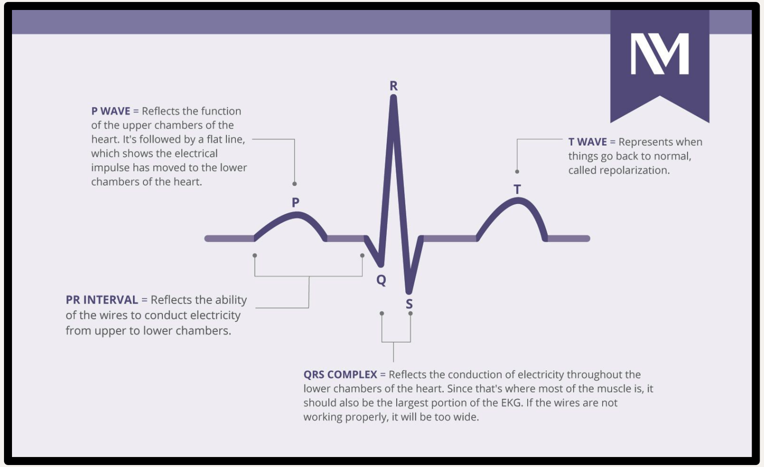
1. **class: Class Code 01-16**
   a. Code 01: Normal
   b. Code 02: Ischemic changes (Coronary Artery Disease)
   c. Code 03: Old Anterior Myocardial Infarction
   d. Code 04: Old Inferior Myocardial Infarction
   …
   p. Code 16: Others
2. age: Age in years, linear
3. sex: Sex (0 = male; 1 = female), nominal
4. height: Height in centimeters, linear
5. weight: Weight in kilograms, linear
6. QRSduration: Average of QRS duration in milliseconds, linear
7. PRinterval: Average duration between onset of P and Q waves in milliseconds, linear
8. Q-Tinterval: Average duration between onset of Q and offset of T waves in msec., linear
9. Tinterval: Average duration of T wave in milliseconds, linear
10. Pinterval: Average duration of P wave in milliseconds, linear

…

279. chV6_QRSTA: Of channel V6

# Related Work

- **Signal Pre-Processing and Fixed-Feature Extraction**
  - Classification of Arrhythmia using KNN-Classifier
  - Heart Arrhythmia Classification Using Machine Learning Algorithms
- **Classifier Comparisons with Feature Selection**
  - Classification of Arrhythmia
- **Algorithmic Approaches with Feature Projection**
  - A Supervised Machine Learning Algorithm for Arrhythmia Analysis
  - Identifying Best Feature Subset for Cardiac Arrhythmia Classification




# 02 Preprocessing

# Missing Value Handling

Imputed using the mean for numerical features and mode for categorical features.

```
numeric_pipeline = Pipeline(

    [("imputer", SimpleImputer(strategy = "mean")),
     ("scaler", StandardScaler())]

)

categorical_pipeline = Pipeline(

    [("imputer", SimpleImputer(strategy = "most_frequent")),
     ("onehot", OneHotEncoder(handle_unknown = "ignore", drop = "if_binary"))]

)
```

scikit-learn pipelines were used to streamline preprocessing

# Feature Encoding

Used one-hot encoding on all categorical features

(necessary for distance computations)

```
categorical_pipeline = Pipeline(

    [("imputer", SimpleImputer(strategy = "most_frequent")),
     ("onehot", OneHotEncoder(handle_unknown = "ignore", drop = "if_binary"))]

)
```

Encodings using scikit-learn functionality

# Dataset Splitting

- Training Set: 80% of data for model training
- Testing Set: 20% for unbiased evaluation of model performance

```
x_train, x_test, y_train, y_test = train_test_split(
    *arrays: x_scaled, y_binary.values, test_size = 0.2, random_state = 42, stratify = y_binary.values
)
```

Split with stratified distribution using scikit-learn functionality

# Feature Selection

```
pca = PCA(n_components = 0.95, random_state = 42)
x_pca = pca.fit_transform(x_processed)
```

Preserves 95% of the variance while avoiding the curse of dimensionality

Eliminates all redundancy from correlated features.

# Converting Dataset to OvR

```python
y_binary = y.apply(lambda x: 0 if str(x).strip() == "1" else 1)
```

Class Code 1 is separated from Class Codes 2 - 16

**class: Class Codes 01-16**

a. Code 01: Normal
b. Code 02: Ischemic changes (Coronary Artery Disease)
c. Code 03: Old Anterior Myocardial Infarction
…
p. Code 16: Others

# 03
# Analysis


Overview
Conditions
My Heart Condition
My Schedule
My Body Condition

# Feature Weighting

- **Mutual Information (MI):**
    - Captures nonlinear dependencies between features and target
- **Absolute Pearson Correlation:**
    - Quantifies linear relationships

$$w = \alpha \times \frac{MI}{\Sigma MI + \epsilon} + (1 - \alpha) \times \frac{|Corr|}{\Sigma |Corr| + \epsilon}$$

# Neighborhood Size

$$k = k_{min} + \frac{local\ density - minimum\ density}{maximum\ density - minimum\ density + \epsilon} \times (k_{max} - k_{min})$$

$k_{min}$ defaults to 3 and $k_{max}$ defaults to 15

```
self.training_densities = np.median(distances[:, 1:], axis = 1)
self.min_density = np.min(self.training_densities)
self.max_density = np.max(self.training_densities)
```

“distances” variable contains distances from *r* nearest neighbors

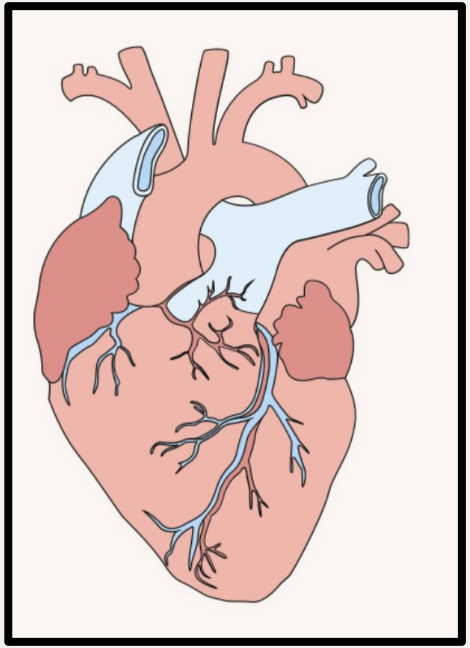
# Distance Metrics

- **Weighted Euclidean Distance:**
  - Incorporates feature weights directly
- **Local Mahalanobis Distance:**
  - Uses a locally computed covariance matrix to account for interdependencies in features.

$$d = \beta \times d_{euclidean} + (1 - \beta) \times d_{mahalanobis}$$

# Grid Search


Accuracy Heatmap for Different Alpha and Beta Values
0.0 0.6703 0.6813 0.6923 0.7033 0.7143 0.6813 0.6813 0.6923 0.7033 0.7033 0.6923
0.1 0.6703 0.6813 0.6923 0.7033 0.7143 0.6813 0.6813 0.7033 0.7033 0.7033 0.6923
0.2 0.6703 0.6813 0.6923 0.7033 0.7143 0.6813 0.6813 0.7033 0.7033 0.7033 0.6923
0.3 0.6703 0.6923 0.6923 0.7033 0.7143 0.6813 0.6813 0.7033 0.7033 0.7033 0.6923
0.4 0.6703 0.6923 0.6923 0.6923 0.7143 0.6813 0.6813 0.7033 0.7033 0.7033 0.6923
0.5 0.6703 0.6923 0.6923 0.6923 0.7143 0.6813 0.6813 0.7033 0.7033 0.7143 0.6923
0.6 0.6703 0.6923 0.6923 0.6923 0.7143 0.6923 0.6813 0.7033 0.7033 0.7143 0.6923
0.7 0.6703 0.6923 0.6923 0.6923 0.7143 0.6923 0.6813 0.7033 0.7033 0.7143 0.6923
0.8 0.6813 0.6813 0.6923 0.6923 0.7033 0.6923 0.6813 0.7033 0.6923 0.7143 0.6923
0.9 0.6703 0.6813 0.6813 0.6923 0.7033 0.6813 0.6813 0.6813 0.7033 0.7033 0.6923
1.0 0.6703 0.6813 0.6813 0.6593 0.6593 0.6484 0.6484 0.6593 0.6703 0.6813 0.6923
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Alpha
0.71
0.70
0.69
0.68
0.67
0.66
0.65



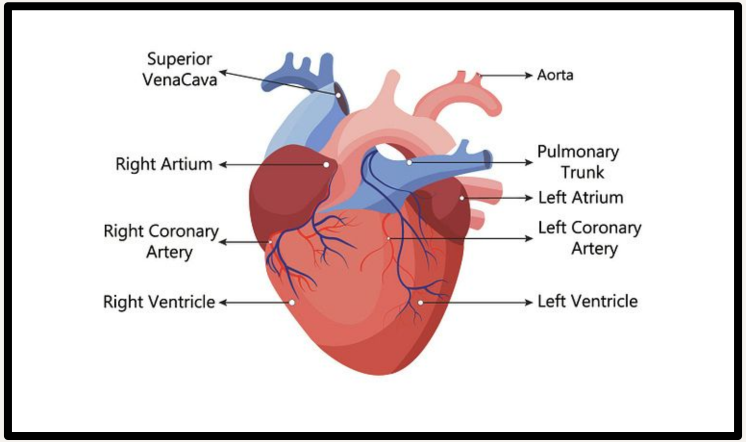
# 04 Conclusion

# Results

| | Accuracy (%) | Precision (%) | Recall (%) | F1 Score (%) |
|---|---|---|---|---|
| Default Parameters | 68.13 | 84.21 | 38.10 | 52.46 |
| Optimized Parameters | 71.43 | 86.00 | 40.50 | 55.07 |

# Future Work

- Hyperparameter tuning using cross-validation
- Application to other high-dimensional medical datasets
- Hybrid models that integrate adaptive KNN



# Sources

Liebman J. (2002). Nonarrhythmia clinical electrocardiography--can it be returned to clinical viability? *Journal of electrocardiology*, *35*(4), 293–298. https://doi.org/10.1054/jelc.2002.36273

University of Rochester Medical Center. (n.d.). *Arrhythmias*. Health Encyclopedia. https://www.urmc.rochester.edu/encyclopedia/content?ContentTypeID=85&ContentID=P00195

Hiremani, V., Devadas, R. M., L, H., S, S., & Pasha, A. (2023). Heart arrhythmia classification using machine learning algorithms. In *Proceedings of the 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC)* (pp. 1–4). IEEE. https://doi.org/10.1109/ICAIHC59020.2023.10431441

# Sources

Samad, S., Khan, S. A., Haq, A., & Riaz, A. (2014). Classification of arrhythmia. *International Journal of Electrical Energy*, 2(1), 57–61. https://doi.org/10.12720/ijoee.2.1.57-61

Guvenir, H. A., Acar, B., Demiroz, G., & Cekin, A. (1997). A supervised machine learning algorithm for arrhythmia analysis. In *Computers in Cardiology 1997* (pp. 433–436). IEEE. https://doi.org/10.1109/CIC.1997.647926

Niazi, K. A. K., Khan, S. A., Shaukat, A., & Akhtar, M. (2015). Identifying best feature subset for cardiac arrhythmia classification. In *Proceedings of the 2015 Science and Information Conference (SAI)* (pp. 494–499). IEEE. https://doi.org/10.1109/SAI.2015.7237188

# Sources


mutual_info_classif. (n.d.). scikit-learn. Retrieved March 27, 2025, https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_classif.html

StandardScaler. (n.d.). scikit-learn. Retrieved March 27, 2025, https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

NearestNeighbors. (n.d.). scikit-learn. Retrieved March 27, 2025, from https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.NearestNeighbors.html

# Sources

SimpleImputer. (n.d.). scikit-learn. Retrieved March 27, 2025, from https://scikit-learn.org/stable/modules/generated/sklearn.impute.SimpleImputer.html

ColumnTransformer. (n.d.). scikit-learn. Retrieved March 27, 2025, from https://scikit-learn.org/stable/modules/generated/sklearn.compose.ColumnTransformer.html

OneHotEncoder. (n.d.). scikit-learn. Retrieved March 27, 2025, from https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html

Pipeline. (n.d.). scikit-learn. Retrieved March 27, 2025, from https://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html

# Sources


PCA. (n.d.). scikit-learn. Retrieved March 27, 2025, from https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

Rusafu.com. (n.d.). Arrhythmia Quotes. Retrieved March 28, 2025, from https://rusafu.com/quotes/tag/arrhythmia/

Corvino, G. (n.d.). Beating Heart [Illustration]. Dribbble. Retrieved March 28, 2025, from https://dribbble.com/shots/11260618-Beating-Heart

Punkichi. (n.d.). Pumping Heart [Illustration]. Dribbble. Retrieved March 28, 2025, from

https://dribbble.com/shots/1963962-Pumping-Heart

Sandhya. (n.d.). Doctor's Day Dribbble [Illustration]. Dribbble. Retrieved March 28, 2025, from https://dribbble.com/shots/4770653-Doctor-s-Day-Dribbble

# Sources

Zukauskaite, K. (n.d.). Premium Stock: Medical, Health [Illustration]. Dribbble. Retrieved March 28, 2025, from https://dribbble.com/shots/10994835-Premium-Stock-Medical-Health

Chantale. (n.d.). Heart [Illustration]. Dribbble. Retrieved March 28, 2025, from https://dribbble.com/shots/5288643-Heart

Northwestern Medicine. (n.d.). How to Read an EKG (Infographic). Retrieved March 28, 2025, from https://www.nm.org/healthbeat/healthy-tips/how-to-read-an-ekg-infographic

Aakash, M. R. H. (n.d.). Human heart. The heart with a venous system. Anatomy. Flat vector [Illustration]. Dribbble. Retrieved March 28, 2025, from https://dribbble.com/shots/22623314-Human-heart-The-heart-with-a-venous-system-Anatomy-Flat-vecto

# Thank you!