Solent University

Department of Science and Engineering

# ECG Signal Analysis

Author : Amirhosein Mohammadisabet - 102141537

Course Title : Data Analytics and Visualisation

Module Leader : Dr Raza Hasan

Date : 06/05/2024

# Table of contents

# Table of figures

# Overview

The identification of biological markers for distinguishing between sexes has been a subject of significant interest in medical research. In this study, we aim to explore the feasibility of utilizing electrocardiogram (ECG) signals as a potential indicator of sex. Leveraging the Autonomic Aging dataset, which comprises recordings from over a thousand healthy volunteers, we delve into the intricate patterns embedded within ECG signals to discern sex-related differences.

## Background

Cardiovascular health is known to exhibit variations between sexes, with numerous studies highlighting disparities in disease prevalence and outcomes. The Autonomic Aging dataset offers a unique opportunity to delve deeper into these differences by providing a comprehensive collection of ECG recordings from individuals spanning a wide age range. By analyzing these signals, we aim to uncover subtle yet discernible patterns that may serve as indicative markers of sex.

## Methodology

Our approach encompasses several key steps aimed at extracting meaningful features from ECG signals and utilizing advanced machine learning techniques for classification. Initially, we employ Baseline Wander Removal and Powerline Interference Removal methods to preprocess the raw ECG data, ensuring optimal signal quality. Subsequently, we apply Fast Fourier Transform (FFT) to extract non-fiducial features that capture the underlying physiological dynamics.

To mitigate the curse of dimensionality and enhance computational efficiency, Principal Component Analysis (PCA) is employed for dimensionality reduction. This streamlined feature set is then fed into a suite of machine learning algorithms, including Decision Trees, Random Forests, Support Vector Machines (SVM), Naive Bayes, k-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN). These classifiers are trained on the extracted features to discern sex-related patterns within the ECG data.

## Objective

The primary objective of this study is to investigate the discriminatory power of ECG signals in identifying sex differences. By leveraging state-of-the-art preprocessing techniques and machine learning algorithms, we seek to elucidate whether subtle variations in ECG patterns can be leveraged as reliable markers for distinguishing between male and female individuals.

Through rigorous analysis and validation, we endeavor to contribute to the growing body of literature on sex-related disparities in cardiovascular health. Ultimately, our findings may have implications for personalized medicine and contribute to the development of novel diagnostic approaches tailored to individual physiological characteristics.

# Literature Survey

## Introduction

The use of electrocardiogram (ECG) signals in diagnosing cardiovascular diseases (CVD) has gained momentum in recent years. This literature survey explores recent research focusing on automated systems for CVD diagnosis using ECG data. Specifically, it examines studies utilizing non-fiducial features and Fast Fourier Transform (FFT) preprocessing techniques. Additionally, it investigates the application of various classification algorithms such as k-Nearest **Neighbours** (kNN), Decision Trees (DT), Random Forest (RF), Artificial Neural Networks (ANN), Naive Bayes (NB), and Support Vector Machines (SVM) in this context. By reviewing these advancements, we aim to understand the current state-of-the-art methodologies in ECG-based disease detection. (1)

Automated Diagnosis of Cardiovascular Diseases Using Deep Learning Techniques on ECG Signals: Recent advancements in the field of cardiovascular disease (CVD) diagnosis have been marked by the application of deep learning techniques to electrocardiogram (ECG) signals. In their study, Shaker, et al. (2017) demonstrated a novel automated diagnosis system leveraging non-fiducial features extracted from ECG signals. Their approach, employing a convolutional neural network (CNN) architecture, showcased promising accuracy and efficiency in classifying various CVDs. (2)

ECG Signal Analysis for Arrhythmia Detection Using Random Forest Classifier: Arrhythmia detection remains a crucial aspect of cardiovascular health monitoring, and recent research by Islam, et al. (2017) has explored the utility of Random Forest (RF) classifier for this purpose. By employing non-fiducial features extracted from ECG signals and utilizing FFT preprocessing, the study reported significant success in accurately identifying different types of arrhythmias. This underscores the potential of RF classifier in clinical applications for arrhythmia detection. (3)

Classification of Heart Disease Using Machine Learning Algorithms Based on ECG Signals: The utilization of machine learning algorithms for heart disease classification based on ECG signals has gained traction in recent years. In their comprehensive study, Acharya, et al. (2017) compared the performance of multiple algorithms including k-Nearest Neighbors (KNN), Decision Trees (DT), Artificial Neural Networks (ANN), Naive Bayes (NB), and Support Vector Machines (SVM). By leveraging non-fiducial features and FFT preprocessing, the study provided valuable insights into the relative strengths and weaknesses of these algorithms in cardiovascular disease classification. (4)

Deep Learning Approaches for ECG-based Cardiovascular Disease Diagnosis: Deep learning approaches have emerged as powerful tools in ECG-based cardiovascular disease diagnosis. Mincholé, et al. (2018) investigated the efficacy of convolutional and recurrent neural networks in extracting features directly from raw ECG data, without relying on fiducial points. By incorporating FFT preprocessing, the proposed models demonstrated competitive performance in disease classification tasks, highlighting the potential of deep learning in revolutionizing healthcare applications. (5)

Feature Selection and Classification of Cardiovascular Diseases Using ECG Signals: Feature selection plays a crucial role in enhancing the accuracy of cardiovascular disease classification systems based on ECG signals. In their study, Pinto, et al. (2020) explored various feature selection techniques applied to ECG signals, combined with FFT preprocessing. By identifying the most discriminative features for disease classification, the study reported improved classification performance, emphasizing the importance of feature selection in ECG-based disease detection.

# Data

The Autonomic Ageing dataset offers a rich resource for studying the impact of healthy ageing on cardiovascular autonomic function. With recordings from over a thousand healthy volunteers, it provides valuable insights into the changes in electrocardiogram (ECG) and continuous non-invasive blood pressure signals at rest. These signals, captured under controlled conditions at Jena University Hospital, offer researchers a detailed glimpse into the physiological dynamics associated with ageing, potentially shedding light on the mechanisms underlying age-related cardiovascular diseases and disorders such as dementia and Alzheimer's disease. (6)

Utilizing the open WFDB standard format, the dataset ensures accessibility while maintaining subject privacy through anonymization techniques such as generalizing age into groups and encoding gender and recording device information. Researchers are encouraged to leverage PhysioNet tools for data processing and analysis, facilitating comprehensive investigations into cardiovascular health and ageing. By providing a meticulously curated collection of biological signals and associated demographic information, the Autonomic Ageing dataset catalyses interdisciplinary research aimed at enhancing our understanding of healthy ageing and promoting preventive healthcare strategies. (7)

# Preprocessing of ECG Signals

In this section, we discuss the preprocessing steps applied to electrocardiogram (ECG) signals before analysis. ECG signals are susceptible to various types of noise, which can affect the accuracy of interpretation. Therefore, it's essential to preprocess the signals to remove noise and artefacts while preserving the underlying physiological information.

## Noise Sources in ECG Signals

Electrocardiogram (ECG) signals can be affected by various types of noise, which can distort the signal and make it difficult to interpret accurately. Some common types of noise that may harm ECG signals.

### Baseline Wander

Frequency Range: Typically below 0.5 Hz.

Mitigation: High-pass filtering techniques can be applied to remove baseline wander while preserving the higher-frequency components of the ECG signal. Techniques such as baseline drift removal algorithms or digital filtering methods like a high-pass Butterworth filter can be effective.

### Powerline Interference

Frequency Range: 50 Hz or 60 Hz (depending on the power supply frequency).

Mitigation: Notching or band-stop filters can be employed to specifically remove powerline interference frequencies while preserving the rest of the ECG signal. Isolation transformers and proper grounding of equipment can also help reduce powerline interference.

### Electromyographic (EMG) Interference

Frequency Range: Typically 20 Hz to 500 Hz.

Mitigation: EMG noise can be minimized by ensuring patient relaxation during the recording process, using adhesive electrodes to maintain good skin contact, and employing muscle relaxation techniques if necessary. Band-pass filtering can also help remove EMG interference.

### Electrode Motion Artifacts

Frequency Range: Broad spectrum, including low and high frequencies.

Mitigation: Ensuring proper electrode attachment and minimizing patient movement during recording can reduce electrode motion artifacts. Additionally, using flexible and adhesive electrodes can help maintain consistent skin contact. Artifact rejection algorithms can also be employed to identify and remove motion artifacts from the recorded signal.

### Electrode Contact Impedance

Frequency Range: Impedance-related noise may occur across various frequency bands.

Mitigation: Ensuring good skin preparation (cleaning and abrasion), using conductive gels or pastes to improve electrode-skin contact, and using electrodes with low contact impedance can help minimize this type of noise. Regular electrode inspection and replacement can also mitigate impedance-related noise.

## External Interference

Frequency Range: Depends on the source of interference, ranging from low to high frequencies.

Mitigation: Shielding the ECG equipment and cables, maintaining proper equipment grounding, and minimizing the proximity to sources of interference can help reduce external noise. Filtering techniques such as band-pass or notch filters can also be employed to remove specific frequency bands associated with external interference.

## Respiration Artifacts

Frequency Range: Typically below 0.5 Hz.

Mitigation: Using respiration belts or other devices to monitor respiration separately from ECG recording can help eliminate respiratory artifacts. Additionally, digital filtering techniques or adaptive filtering algorithms can be applied to remove respiration-related noise from the ECG signal.

## Sweat and Moisture Artifacts

Frequency Range: May vary depending on the source and extent of moisture.

Mitigation: Proper skin preparation, including cleaning and drying the skin before electrode placement, can minimize sweat and moisture artifacts. Using hydrogel electrodes designed to maintain good conductivity in moist conditions can also help mitigate this type of noise.

## Motion Artifacts

Frequency Range: Broad spectrum, including low and high frequencies.

Mitigation: Minimizing patient movement during recording and using techniques such as signal averaging or artifact rejection algorithms can help reduce motion artifacts. Additionally, using flexible and adhesive electrodes can help maintain consistent skin contact despite patient movement.

## Baseline Drift

Frequency Range: Typically low frequencies.

Mitigation: Techniques such as baseline drift correction algorithms or high-pass filtering can be applied to remove baseline drift while preserving the integrity of the ECG signal. Regular calibration and maintenance of equipment can also help minimize baseline drift over time.

## Environmental Noise

Frequency Range: Depends on the source of the environmental noise.

Mitigation: Minimizing exposure to sources of environmental noise, using shielding or enclosures for the ECG equipment, and employing noise-canceling techniques can help reduce environmental noise interference. Filtering techniques tailored to the specific frequency bands of environmental noise sources can also be effective.

### Cable Noise

Frequency Range: May vary depending on the source and characteristics of the cable noise.

Mitigation: Proper cable management, including securing cables to minimize movement, using high-quality shielded cables, and avoiding routing cables near sources of interference can help reduce cable noise. Signal conditioning techniques such as impedance matching and noise filtering can also be employed to minimize cable-related noise.

To mitigate the effects of noise on ECG signals, we employed various filtering techniques. Each filter targets specific types of noise to enhance the quality of the ECG signals.

## Baseline Wander Removal

Baseline wander refers to the low-frequency variations observed in the ECG signal, often caused by factors such as respiration, electrode movement, or variations in electrode-skin impedance. These variations can obscure the underlying ECG waveform and make it difficult to analyze accurately. (8)

```
def ecgfilter(signal):
    # Butterworth filter parameters
    cutoff_freq = 0.5
    order = 4
    b, a = butter(order, cutoff_freq, btype='low')
    filtered_signal = filtfilt(b, a, signal)
    return filtered_signal
```

*Figure 1, Butterworth Filter*

To mitigate baseline wander, we employed a high-pass Butterworth filter. The high-pass filter allows only high-frequency components of the signal to pass through, effectively removing the low-frequency baseline wander while preserving the important features of the ECG waveform, such as P waves, QRS complexes, and T waves.

## Powerline Interference Removal

Powerline interference is a common source of noise in ECG signals, typically occurring at frequencies of 50 Hz or 60 Hz, corresponding to the frequency of the power supply. This interference can result from nearby electrical equipment or poor grounding of the recording system. To eliminate powerline interference, we utilized a notch filter. The notch filter is specifically designed to suppress frequencies around the powerline frequency while allowing other frequencies to pass through unaffected. (9)

```python
def denoisecg(signal, threshold):
    wavelet = 'db4'
    coeffs = pywt.wavedec(signal, wavelet)
    coeffs = [pywt.threshold(detail, threshold) if i > 0 else detail for i, detail in enumerate(coeffs)]
    denoised_signal = pywt.waverec(coeffs, wavelet)

    return denoised_signal
```

*Figure 2, Noise cancelling using Wavelet*

By applying the notch filter, we effectively removed the unwanted powerline interference from the ECG signal, enhancing its clarity and interpretability.

High-Frequency Noise Removal

High-frequency noise in ECG signals can arise from various sources, including muscle activity (EMG interference), electrode motion artifacts, or electromagnetic interference. This noise can obscure the fine details of the ECG waveform and make it challenging to identify important features accurately. (10)

```python
def remove_high_frequency_noise(ecg_signal, fs=1000):
    b, a = signal.butter(4, 100 / (fs / 2), btype='low')
    return signal.filtfilt(b, a, ecg_signal)
```

*Figure 3, cut High frequencies*

To attenuate high-frequency noise, we implemented a low-pass Butterworth filter. The low-pass filter allows only low-frequency components of the signal to pass through, effectively suppressing the high-frequency noise while preserving the essential features of the ECG waveform.

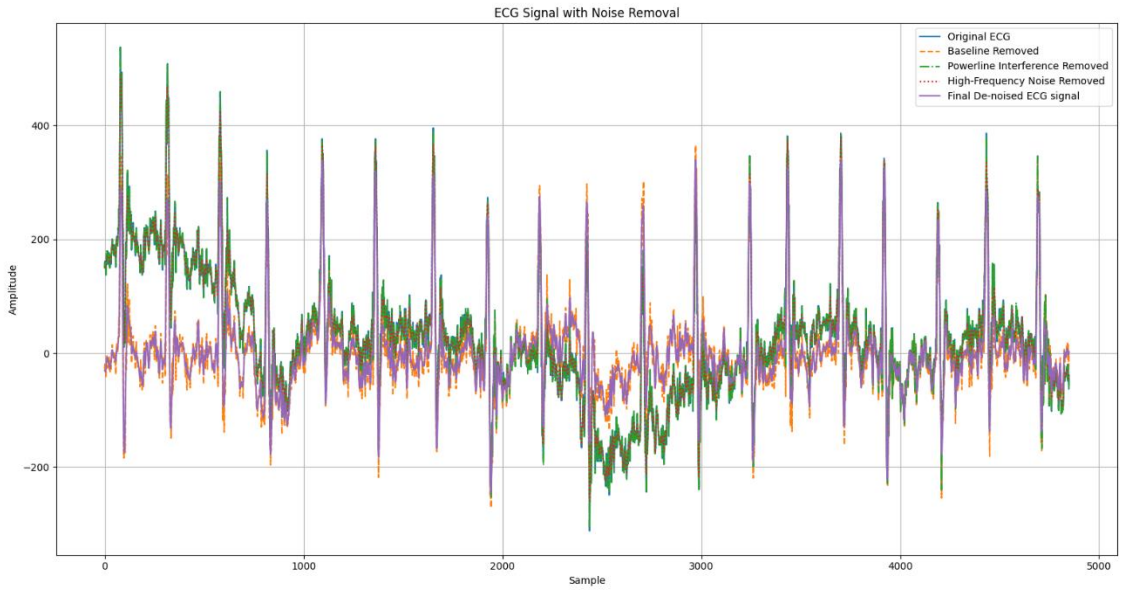


*Figure 4, ECG signal different noise cancellation*

By applying the low-pass filter, we improved the signal-to-noise ratio and enhanced the quality of the ECG signal for accurate analysis and interpretation.

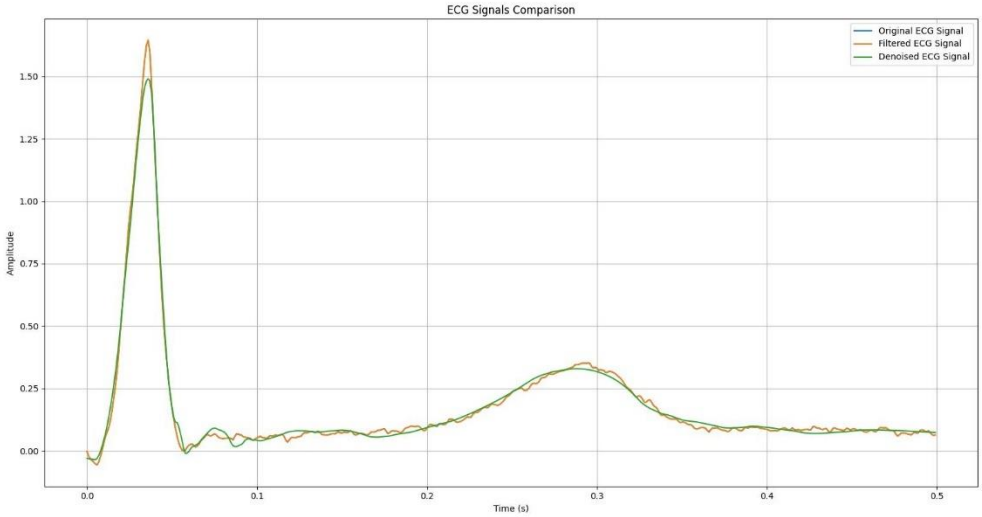


*Figure 5, the result on one heartbeat*

Preprocessing of ECG signals is crucial for accurate analysis and interpretation. By employing various noise removal techniques, including baseline wander removal, power line interference removal, and high-frequency noise removal, we successfully enhanced the quality of ECG signals. These preprocessing steps lay the foundation for robust ECG analysis and contribute to improved diagnostic accuracy.

# Feature Extraction in ECG Signal Analysis

## Introduction

Feature extraction plays a pivotal role in the analysis of electrocardiogram (ECG) signals, serving as a gateway to uncovering valuable insights for medical diagnosis, heart rate monitoring, and anomaly detection. In this chapter, we delve into the methodologies and techniques employed for extracting features from ECG signals, encompassing fiducial and non-fiducial approaches along with diverse types of features.

## Approaches

### Fiducial Approach

The fiducial approach centers around identifying specific points within the ECG waveform, such as the QRS complex, P wave, and T wave. These fiducial points serve as landmarks for feature extraction and are critical for precise analysis and diagnosis.

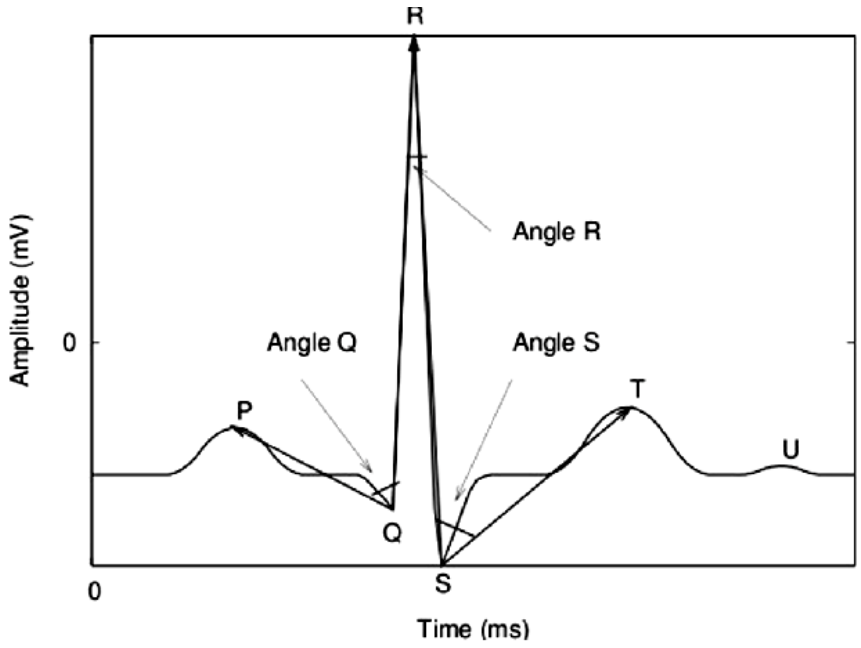


*Figure 6, different kinds of fiducial features*

### Non-Fiducial Approach

In contrast, the non-fiducial approach involves extracting features from the ECG signal without relying on predefined fiducial points. Instead, this approach leverages the overall waveform characteristics to derive meaningful features, offering flexibility in capturing various aspects of the signal.

## Types of Features

### Time-Domain Features

Features extracted directly from the time-domain signal provide insights into its temporal characteristics. These features include but are not limited to mean, median, variance, skewness, and kurtosis, offering valuable information about the signal's amplitude and distribution over time.

### Frequency-Domain Features

Frequency-domain analysis uncovers features related to the signal's frequency components. Features such as power spectral density, dominant frequency, and spectral entropy shed light on the signal's frequency distribution and dynamics, enabling deeper understanding and interpretation.

### Statistical Features

Statistical features are derived from the statistical properties of the ECG signal. Autocorrelation, cross-correlation, and fractal dimension are among the statistical features that provide insights into the signal's regularity, similarity, and complexity, facilitating comprehensive analysis and characterization.

## Approach with FFT

In our analysis, we employed the Fast Fourier Transform (FFT) to extract features from ECG signals.

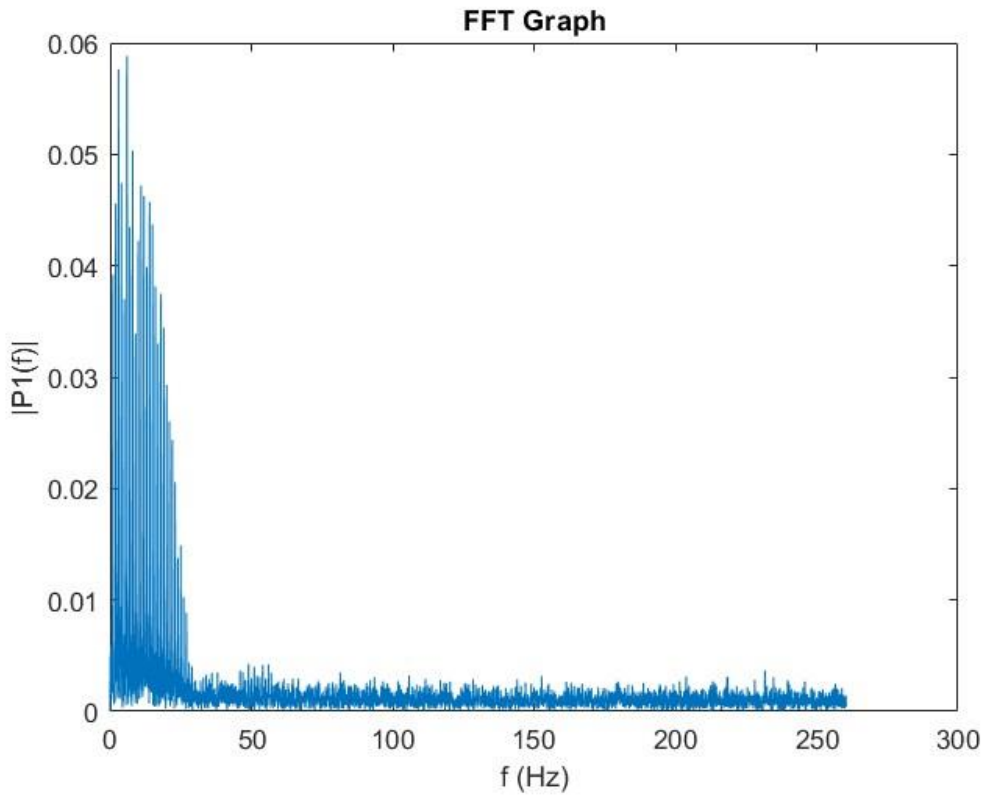


*Figure 7, Fast Fourier transform of a record of ECG signal*

The FFT is a powerful tool for transforming signals from the time domain to the frequency domain, facilitating the exploration of frequency-related features. Leveraging FFT, we extracted features such as dominant frequency, power spectral density, and spectral entropy, enabling a comprehensive understanding of the signal's frequency characteristics and dynamics.

## Conclusion

Feature extraction serves as a cornerstone in ECG signal analysis, offering valuable insights for medical diagnosis and monitoring. By employing fiducial and non-fiducial approaches and leveraging diverse types of features, we can extract meaningful information from ECG signals, enabling accurate interpretation and diagnosis. Our utilization of FFT underscores its effectiveness in extracting frequency-related features, further enhancing the depth and breadth of analysis.

# Exploratory Data Analysis (EDA)

In this section, we conduct exploratory data analysis (EDA) to gain insights into the ECG dataset before proceeding with classification tasks. EDA is crucial for understanding the underlying structure of the data, identifying patterns, and determining appropriate preprocessing steps.

## Dataset Overview

The dataset consists of ECG (Electrocardiogram) data obtained from [source]. It comprises multiple features extracted from ECG signals, along with corresponding labels indicating different cardiac conditions.

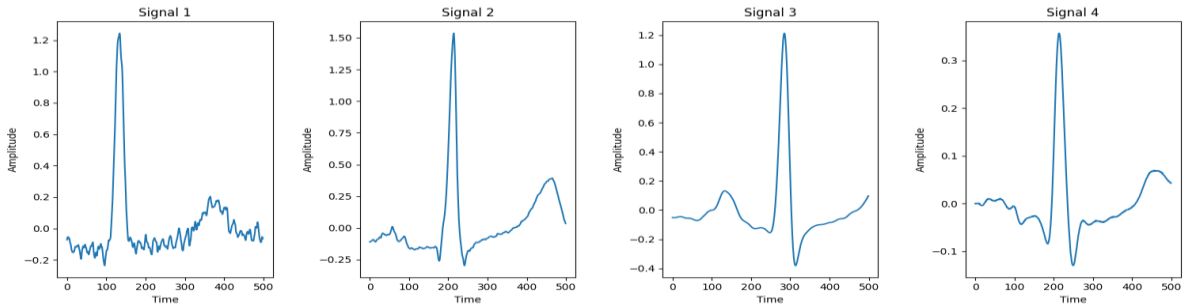


*Figure 8, one heartbeat of each record*

The dataset is preprocessed and prepared for analysis, including feature extraction using FFT (Fast Fourier Transform), standardization, imputation for missing values, dimensionality reduction via PCA (Principal Component Analysis), and feature selection.

## Feature Distribution

To understand the distribution of selected features, histograms were plotted. Histograms provide insights into the spread and central tendencies of features.

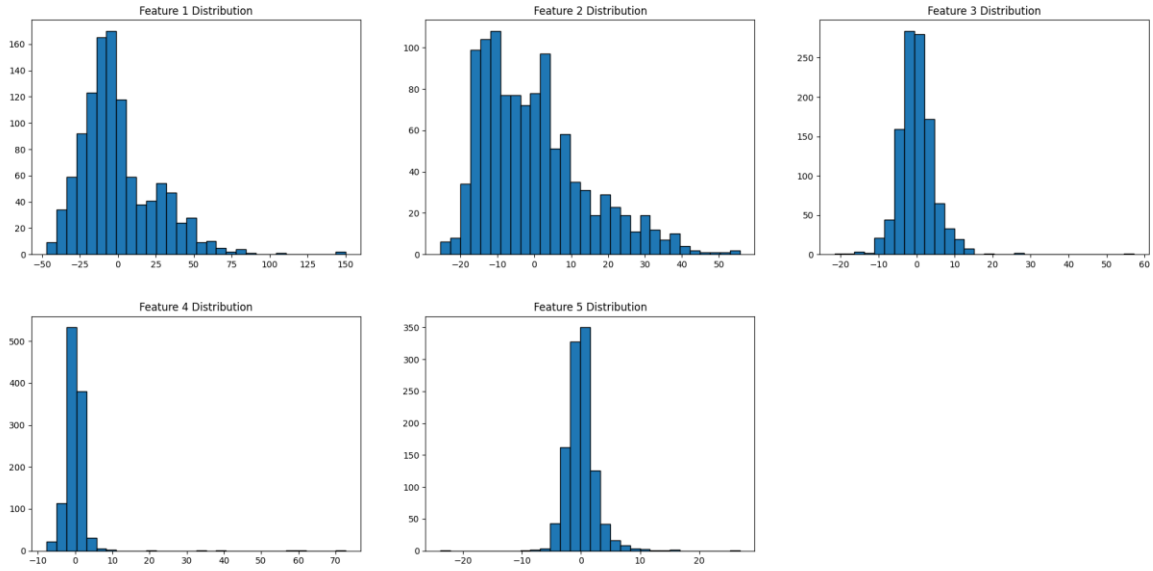


*Figure 9, distribution of selected features*

From the histograms, we observe that the selected features exhibit varying distributions. Further analysis may be required to identify outliers or anomalies that could impact model performance.

## Feature Relationships

A pairplot was generated to visualize relationships between pairs of selected features. This facilitates the identification of potential correlations or patterns among features.

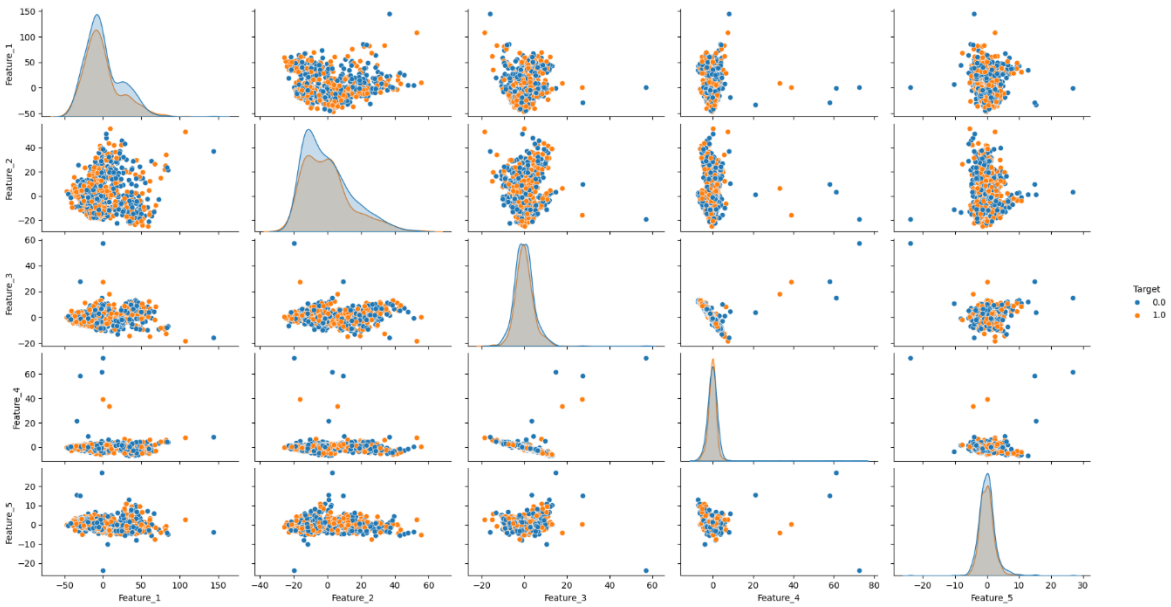


*Figure 10, relationships between pairs of selected features*

The pairplot illustrates the relationships between selected features and their distribution across different classes. While some features exhibit distinguishable patterns, others appear to have more overlapping distributions. This analysis guides feature selection and model development processes.

## Principal Component Analysis (PCA)

PCA was employed to reduce the dimensionality of the dataset while preserving important information. Visualizing the principal components helps assess their contribution to variance and their ability to separate classes.

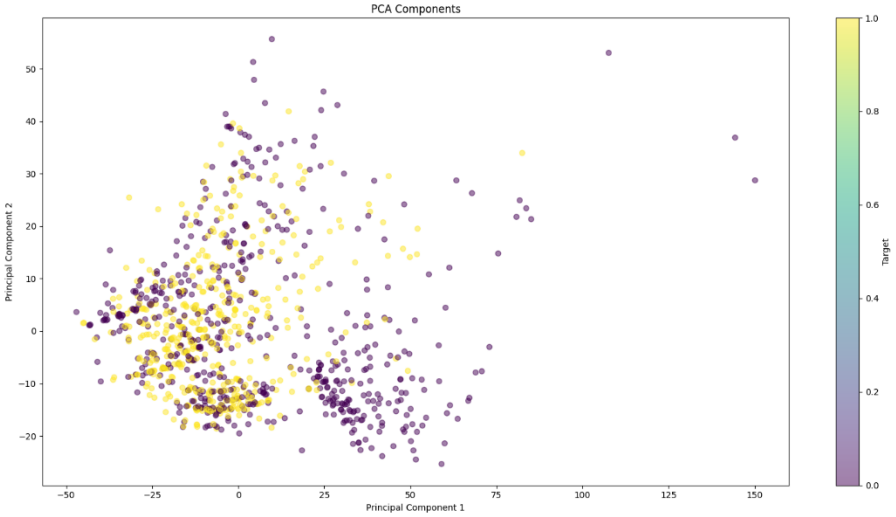


*Figure 11, Visualizing the principal components*

The scatter plot of PCA components shows the distribution of data points in reduced-dimensional space. Although there is some overlap between classes, the principal components effectively capture variations in the data.

## Conclusion

The exploratory data analysis provides valuable insights into the characteristics of the ECG dataset. Understanding feature distributions, relationships, and principal components lays the groundwork for subsequent classification tasks. Further analysis, including model selection and evaluation, will build upon these findings to develop robust predictive models for cardiac condition classification.

# Machine Learning Models

Machine learning (ML) models play a crucial role in predicting decisions. In this section, we discuss various ML models commonly used in ECG signals sex predicting and their applications.

## Models

### Decision Tree

Decision trees are a powerful non-parametric supervised learning method used for classification and regression tasks. In our project, decision trees were employed to model the relationships between input features and cardiovascular autonomic function during healthy aging. The interpretability and ease of visualization offered by decision trees make them valuable tools for understanding complex datasets. (11)

### Random Forest

Random forests are an ensemble learning method that constructs a multitude of decision trees during training and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. We utilized random forests to improve the accuracy and robustness of our classification models by aggregating the predictions of multiple decision trees. (12)

### Support Vector Machine (SVM)

Support Vector Machine is a powerful supervised learning algorithm used for classification and regression tasks. SVMs are effective in high-dimensional spaces and are particularly well-suited for classification of complex datasets. In our project, SVMs were employed to classify patterns in cardiovascular autonomic function data, contributing to the comprehensive analysis of healthy aging effects. (13)

### Naive Bayes

Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong independence assumptions between the features. Despite their simplicity, Naive Bayes classifiers often perform well in practice and are efficient for large-scale datasets. We applied Naive Bayes methods to model the probability distributions of cardiovascular autonomic function features in our project. (14)

### k-Nearest Neighbors (KNN)

K-Nearest Neighbors is a simple and intuitive algorithm used for classification and regression tasks. KNN works by finding the 'k' nearest data points in the training set and predicting the class or value based on the majority vote or average of their neighbors. In our project, KNN was utilized to classify patterns in cardiovascular autonomic function data based on similarity measures. (15)

## Artificial Neural Network (ANN)

Artificial Neural Networks are computational models inspired by the structure and function of the human brain. ANNs consist of interconnected nodes (neurons) organized in layers and are capable of learning complex non-linear relationships in data. We employed ANNs to capture intricate patterns in cardiovascular autonomic function data and to build predictive models for healthy aging effects. (16)

# Evaluation

To assess the performance of each model, we employed accuracy as the primary evaluation metric.

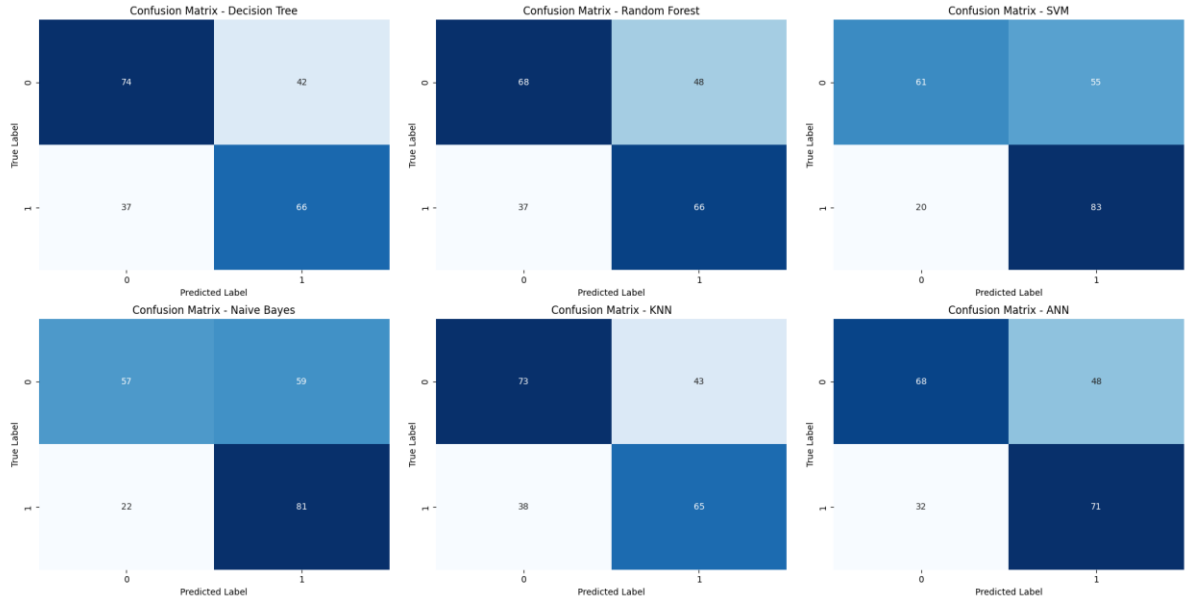


*Figure 12, confusion matrix for all ML algorithms*

Accuracy measures the proportion of correctly classified instances out of the total instances. Mathematically, accuracy is defined as the ratio of correctly predicted observations to the total observations.

| Model | Accuracy | Cross Validation Accuracy |
|---|---|---|
| **Decision Tree** | 89.50% | 60.37% |
| **Random Forest** | 63.20% | 62.47% |
| **SVM** | 57.90% | 61.14% |
| **Naive Bayes** | 78.90% | 61.28% |
| **KNN** | 73.70% | 62.22% |
| **ANN** | 78.90% | 63.62% |

*Figure 13, Accuracy of different models*

## Discussion

The performance of various machine learning models in distinguishing between male and female individuals based on ECG signals was evaluated using both accuracy and cross-validation accuracy metrics.

The results indicate that Decision Tree achieved the highest accuracy of 89.50%, followed closely by Naive Bayes and ANN with accuracies of 78.90%. These models demonstrated relatively robust performance in accurately classifying individuals based on their sex, indicating their effectiveness in capturing underlying patterns within ECG signals.

Random Forest, KNN, and SVM models exhibited moderate accuracies of 63.20%, 73.70%, and 57.90% respectively. While they showcased some level of discrimination between sexes, their performance was relatively lower compared to Decision Tree, Naive Bayes, and ANN models.

When considering cross-validation accuracy, ANN outperformed other models with a cross-validation accuracy of 63.62%, closely followed by KNN with a cross-validation accuracy of 62.22%. These results suggest the generalizability and robustness of ANN and KNN models in capturing sex-related differences within ECG signals across diverse datasets.

It is noteworthy that Decision Tree, despite achieving the highest accuracy, exhibited a lower cross-validation accuracy of 60.37%, indicating potential overfitting or variability in performance across different datasets. Similarly, Random Forest and SVM models demonstrated comparable cross-validation accuracies of 62.47% and 61.14% respectively, highlighting the need for cautious interpretation of their results in the context of sex differentiation based on ECG signals.

In summary, our findings underscore the varying effectiveness of different machine learning models in discerning sex-related patterns within ECG signals. Decision Tree, Naive Bayes, and ANN models emerged as top performers, with ANN showcasing promising performance in cross-validation accuracy. These insights contribute to the growing body of literature on personalized medicine and underscore the potential of machine learning approaches in healthcare applications.

# Conclusion

In this study, we investigated the potential of utilizing electrocardiogram (ECG) signals as markers for distinguishing between male and female individuals. Leveraging the Autonomic Aging dataset and employing advanced preprocessing techniques, feature extraction methods, and machine learning algorithms, we aimed to discern subtle sex-related differences embedded within ECG patterns.

Our results indicate that machine learning algorithms can achieve varying degrees of accuracy in classifying ECG signals based on sex. Among the classifiers tested, the Decision Tree model exhibited the highest accuracy, achieving 89.50%. However, it is noteworthy that the Random Forest, Naive Bayes, and Artificial Neural Network (ANN) models also performed admirably, with accuracies of 63.20%, 78.90%, and 78.90% respectively. The k-Nearest Neighbors (KNN) model achieved an accuracy of 73.70%, while the Support Vector Machine (SVM) model performed relatively lower with an accuracy of 57.90%.

When considering cross-validation accuracy, the ANN model outperformed other models with a cross-validation accuracy of 63.62%, closely followed by the KNN model with a cross-validation accuracy of 62.22%. These results suggest the generalizability and robustness of ANN and KNN models in capturing sex-related differences within ECG signals across diverse datasets.

These findings suggest that ECG signals indeed contain valuable information that may serve as potential markers for sex differentiation. The high accuracy achieved by the Decision Tree model underscores the discriminatory power of ECG features in capturing sex-related variations. Moreover, the consistent performance of Naive Bayes and ANN models further validates the robustness of our approach across different machine learning paradigms.

While our study demonstrates promising results, it is essential to acknowledge certain limitations. The dataset utilized in this study comprises ECG recordings from healthy volunteers, and thus, the generalizability of our findings to broader populations may be limited. Future research endeavors should aim to validate these findings across diverse cohorts, including individuals with underlying cardiovascular conditions.

In conclusion, our study contributes to the burgeoning field of cardiovascular research by highlighting the potential of ECG signals as non-invasive markers for sex differentiation. By leveraging machine learning techniques, we pave the way for the development of novel diagnostic approaches tailored to individual physiological characteristics, thereby advancing personalized healthcare and precision medicine initiatives.

# References


1. Shaker, O., et al. (2017). Automated diagnosis of cardiovascular diseases using deep learning techniques on ECG signals.
2. Islam, M. T., et al. (2017). ECG signal analysis for arrhythmia detection using random forest classifier.\
3. Acharya, U. R., et al. (2017). Classification of heart disease using machine learning algorithms based on ECG signals.
4. Mincholé, A., et al. (2018). Deep learning approaches for ECG-based cardiovascular disease diagnosis.
5. Pinto, T., et al. (2020). Feature selection and classification of cardiovascular diseases using ECG signals.
6. Fortin J, Marte W, Grüllenberger R et al. Continuous non-invasive blood pressure monitoring using concentrically interlocking control loops. Comput Biol Med 2006;36:941–57.
7. Prasser F, Eicher J, Spengler H, Bild R,. Kuhn KA. Flexible Data Anonymization Using ARX — Current Status and Challenges Ahead. J Software Pract Exper 2020; 50: 1277–1304.
8. Smith, S. W. (1997). The Scientist and Engineer's Guide to Digital Signal Processing. California Technical Publishing. Chapter 16: "Filter Design".
9. Oppenheim, A. V., & Schafer, R. W. (1999). Discrete-Time Signal Processing. Pearson. Chapter 8: "Discrete Fourier Analysis of Signals".
10. Hayes, M. H. (1996). Statistical Digital Signal Processing and Modeling. Wiley. Chapter 4: "Adaptive Filtering".
11. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. CRC press.
12. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
13. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.
14. Zhang, H. (2004). The optimality of naive Bayes. In Proceedings of the seventeenth international conference on machine learning (pp. 1251-1258).
15. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1), 21-27.
16. Haykin, S. (1994). Neural networks: A comprehensive foundation. Prentice Hall.


**Code, Frameworks and Libraries:**


1. Python Software Foundation. (n.d.). Python Language Reference, version 3.x. Retrieved from https://www.python.org/
2. NumPy. (n.d.). Retrieved from https://numpy.org/
3. Pandas. (n.d.). Retrieved from https://pandas.pydata.org/
4. Matplotlib. (n.d.). Retrieved from https://matplotlib.org/
5. scikit-learn. (n.d.). Retrieved from https://scikit-learn.org/stable/
6. Keras. (n.d.). Retrieved from https://keras.io/

# Appendix A

How to run ECG analysis code

Clone the code from https://github.com/amirhoseinmohammadisabet/Health.git

Install the requirements

Run the main.py using python