1. Introduction

Within the realm of cardiovascular diagnostics, the electrocardiogram (ECG) signal assumes a paramount role. Among the crucial elements for ECG analysis lies the RR interval, signifying the time duration between consecutive R-peaks within the cardiac cycle. A deeper comprehension of heart rate variability and autonomic regulation can be attained through the analysis of RR intervals, rendering them indispensable in assessing cardiac health. Consequently, it becomes imperative to discern the relationship between RR intervals and various demographic factors, particularly age and gender, to foster comprehensive cardiac analysis.

This master's thesis endeavors to delve into the distinctive characteristics of RR intervals present in ECG signals procured from a diverse age spectrum, ranging from one month to 55 years. The primary focus lies in undertaking a comparative analysis between two distinctive methodologies: Time-Domain Analysis and Frequency-Domain Analysis, utilizing the renowned Welch method. By conducting this meticulous examination, the study aims to gauge the effectiveness of these approaches in evaluating RR intervals based on age and gender.

1.1. Background and Significance

The electrocardiogram (ECG) is a crucial tool for cardio because it provides essential details about the electrical activity of the heart. The RR interval, which indicates the amount of time between sequential R-wave peaks and basically represents heart rate, is one of the key variables established from the ECG. With changes in these intervals (heart rate variability, HRV) being suggestive of the balance between sympathetic and parasympathetic effects on heart rate, the study of RR intervals has been frequently utilized to evaluate autonomic nervous system function [1], [2].

The utilization of HRV analysis has been found through a range of physiological and pathological circumstances, corresponding to its capacity as an invisible quantitative assessment of the sympathovagal balance [1]. Another study has shown that heart rate variability (HRV) is a valuable indicator of the possibility of mortality following a heart attack [2]. The studies mentioned previously highlight the significance of the use of HRV evaluation in both medical and research areas.

The analysis of Heart Rate Variability (HRV) can be conducted through the utilization of techniques based on time-domain and frequency-domain methodologies. The analysis of time domain offers metrics to quantify the level of variability in the RR intervals. On the contrary hand, frequency-domain analysis, represented through the Welch method, offers knowledge about the variation of power (variance) with relation to frequency [3]. Both approaches offer distinct benefits and can be used in combination to create an in-depth knowledge of HRV.

The majority of current research on HRV has concentrated on adult populations, thus resulting in a lack of comprehensive examination spanning a broad spectrum of ages, ranging from infants to adults. Moreover, an all-inclusive understanding of the effect of gender on HRV parameters remains elusive. Consequently, a notable knowledge deficit exists regarding the impact of age and gender on HRV parameters, which the present study endeavors to rectify.

The research project plans to use the RR interval dataset sourced from PhysioNet, which includes RR interval information obtained from a population of healthy individuals ranging from 1 month to 55 years of age [4]. The dataset presents a distinctive prospect to examine the impact of age and gender on HRV parameters, utilizing time-domain and frequency-domain analysis techniques.

The present study's results will make an important contribution to the current HRV literature by offering valuable insights into the influence of age and gender on HRV parameters. The establishment of reference ranges for HRV parameters that are specific to age and gender has the potential to enhance the medical efficacy of HRV analysis. Additionally, the results may offer valuable perspectives on the alterations in autonomic nervous system function among various age cohorts and among males and females.

The study's importance lies in its potential to offer a comprehensive comprehension of the behavior of the RR interval, thereby contributing to the enhancement of cardiovascular health evaluation. Using taking advantage of Welch's method, the present research aims to compare Time-Domain and Frequency-Domain methods with the goal to identify age-related structures in heart health and evaluate the possible impact of gender on variations in the RR interval [5]–[8]. The comparative analysis offered within it has the potential to make a valuable contribution to the subject of personalized healthcare. This is because of the fact that autonomic dysfunction has been identified in a significant proportion of patients with specific medical conditions, such as cirrhosis, influencing up to 60% of such individuals [8].

Additionally, the results of the study can provide insight into unaddressed orthostatic symptoms that manifest during the post-COVID period, potentially linked to autonomic dysregulation[5]. Furthermore, this research has the potential to make a valuable contribution to the advancement of smartphone applications for HRV measurement, an area that has undergone significant development over the past ten years [7].

The background and significance of the study have implications across multiple healthcare domains. In the field of sports medicine, knowledge of the RR interval's behavior can facilitate the monitoring of athletes' cardiovascular health and performance. The assessment of cardiovascular health in older adults within the field of geriatrics can be facilitated by considering the changes in the RR interval, which is known to be associated with the aging process [9].

Furthermore, the results of the study possess an opportunity to create significant

contributions regarding the advancement of unique diagnostic methods and treatments. A comprehension of the structure of the RR interval can facilitate the development of algorithms that focus on detecting arrhythmias or other cardiovascular pathologies. The RR interval can be affected by stress or anxiety, which implies that interventions aimed at managing these conditions could benefit from this information [10].

In summary, the study's background and significance originate in its capacity to offer enhanced comprehension of the behavior of the RR interval in ECG signals. The findings derived from this investigation have the potential to enhance the evaluation of cardiovascular health, facilitate customized medical care, and foster the creation of unique HRV measurement instruments.

1.2. Research Objectives

This research, entitled RR Interval Analysis in Electrocardiogram Signals of Individuals Aged 1 Month to 55 Years, seeks to perform a thorough investigation of RR intervals obtained from electrocardiogram signals across a broad spectrum of ages and genders. The study will utilize the RR interval dataset sourced from PhysioNet, comprising of RR interval data obtained from individuals who are in good health and aged between one month and 55 years [4]. The stated objectives of the research are as follows:

1. To perform an analysis of RR intervals in the time domain:
   The analysis in the time domain will yield metrics pertaining to the extent of fluctuation in the RR intervals. Different time-domain parameters, including SDNN, RMSSD, and pNN50, will be computed and compared among various age groups and genders.
2. To perform an analysis of RR intervals in the frequency domain:
   The analysis in the frequency domain produces insights into the variance distribution of power with regard to frequency. The study will involve the computation and comparison of several frequency-domain parameters, including but not limited to the total power, power in the low frequency (LF) and high frequency (HF) bands, and the LF/HF ratio. The analysis will be conducted across diverse age cohorts and between genders.
3. The objective of this study is to examine the impact of age and gender on HRV parameters:
   The study targets to investigate the variability of HRV parameters obtained from time-domain and frequency-domain analyses across different age groups and genders. The results have the potential to facilitate the establishment of HRV parameter reference ranges that are specific to age and gender.
4. To determine a relationship between the results and the previous literature:
   The study aims to conduct a comparative analysis of the results with the extant body of literature on the analysis of heart rate variability (HRV). The investigation seeks to

offer novel insights into the physiological alterations in autonomic nervous system activity in diverse age cohorts and gender categories.

The expected research has the potential to make an important contribution to the current corpus of knowledge on Heart Rate Variability (HRV), by furnishing valuable perspectives on the impact of age and gender on HRV parameters. The results have the potential to enhance the clinical efficacy of HRV examination, thereby facilitating the identification and treatment of diverse cardiovascular medical conditions.

1.3. Scope of the Work

The proposed master's thesis titled RR interval in the ECG signal from subjects between 1 month and 55 years aims to conduct a comprehensive analysis of RR intervals derived from ECG signals across a wide age range and between genders. The scope of work for this research is outlined as follows:

1. Data Collection
   The research is using the RR interval dataset sourced from PhysioNet, comprising of RR interval data obtained from a population of healthy individuals ranging from 1 month to 55 years of age. The present dataset presents an exceptional prospect to examine the impact of age and gender on HRV parameters, utilizing both time-domain and frequency-domain analysis techniques.
2. Data Preprocessing
   The RR interval data in its raw state is subject to preprocessing processes to remove any artifacts and irregularities. Ensuring the accuracy and reliability of the subsequent HRV analysis is of utmost importance, making this stage of the process essential [11].
3. Time-Domain Analysis
   The preprocessed RR interval data will be subjected to time-domain analysis, which will provide metrics that indicate the quantity of variability offered in the RR intervals. The study is to determine diverse time-domain parameters while performing comparisons among distinct age and gender groups [12].
4. Frequency-Domain Analysis
   The RR interval data that completed preprocessing will be analyzed in the frequency domain utilizing the Welch method. The present analysis seeks to clarify the relationship between power distribution, specifically variance, and frequency. The study seeks to determine diverse frequency-domain parameters and conduct a comparative analysis among distinct age groups and genders [13].
5. Statistical Analysis
   The HRV parameters obtained from the time-domain and frequency-domain analyses will be subjected to statistical analysis to determine the impact of age and gender.

The analytical approach will include descriptive statistics, correlation analysis, and regression analysis, as indicated by reference [14].

6. Interpretation of Results
   The interpretation of the statistical analysis outcomes will be conducted within the context of the current research related to the analysis of HRV. This research will explain the results in connection with the adjustments in autonomic nervous system activity at the physiological level, which vary across diverse age cohorts and between male and female individuals [15].
7. Thesis Writing
   The research process and findings will be documented in a master's thesis. The thesis will include an introduction, literature review, methodology, results, discussion, conclusion, and references.

The expected study will have the potential to make a substantial contribution to the current corpus of knowledge regarding HRV, furnishing valuable perspectives on the impact of age and gender on HRV parameters. The results have the potential to enhance the clinical effectiveness of HRV analysis, thus helping in the identification and treatment of diverse cardiovascular medical conditions.

2. State of Art

2.1. Overview of ECG Signal Analysis

An electrocardiogram (ECG) is an examination that analyzes the electrical impulses of the heart periodically. It is a non-invasive, comfortable implementation. That provides crucial data on the condition of the heart. The investigation of ECG signals is critical in the identification of a broad variety of cardiac problems, from modest rhythm irregularities to serious heart illnesses. It supports physicians in understanding the electrical activity of the heart, allowing them to identify anomalies and development appropriate medical methods.

The primary goal of researchers has been to identify and prevent life-threatening cardiovascular situations via ECG analysis. Traditional signal processing approaches, machine learning, and their subbranches, such as deep learning, are popular techniques for analyzing and categorizing ECG signals, with the goal of developing applications for the early identification and treatment of heart diseases and arrhythmias [16].

Data extraction, signal processing and denoising, identification of ECG fiducial points based on feature engineering, and ECG signal categorization are all phases in ECG signal analysis [16].

The evaluation of ECG signal quality after compression is an essential stage in the compression process. Compression makes signal archiving easier, speeds up signal transmission, and decreases energy use. Lossy compression, on the other hand, distorts the signals. As a result, compression performance must be expressed in terms of both compression efficiency and signal quality [17].

Because signals are highly polluted by noise and other abnormalities, cardiovascular signal processing is often highly computational work. An efficient method for detecting and localizing peak points in noisy ECG data is given. The developed approach has six phases and uses the Hilbert transform and a thresholding technique to find zones within the ECG signal that may contain a peak [18].

Wavelet-based approaches, such as signal denoising, wave identification, and heartbeat classification, play essential roles in ECG signal analysis. These approaches are appropriate for ECG analysis because they expose the fundamental mechanism of the ECG signals and the design principles that these methods may follow [18].

2.2. RR Interval and its Importance

The RR interval is the period between two successive R waves in the ECG signal, measured in milliseconds. The R wave is a component of the QRS complex, which depicts the electrical depolarization of the ventricles of the heart.

The RR interval is an important measure in determining heart health. It gives information about the heart's rhythm and pace, and its variability might reflect stress, fitness levels, and numerous health issues, such as cardiovascular disease.

In the area of biological data-enabled security, the RR interval has been used. An ECG-based authentication system suited for identity checks was presented in a 2019 article titled An Enhanced Machine Learning-Based Biometric Authentication System Using RR-Interval Framed Electrocardiograms by Song-Kyoo Kim et al. The approach reshapes supplied data via RR-interval slicing [19].

QT/RR hysteresis and QT/RR adaptation are two interconnected but distinct physiological processes that indicate how fast and how much the QT interval varies when the heart rate changes. The importance of correcting for QT/RR hysteresis in studies of drug-induced QTc interval changes was highlighted in a 2018 study titled Importance of QT/RR hysteresis correction in studies of drug-induced QTc interval changes by M. Malik et al [20].

The RR interval is also important in heartbeat categorization using an ECG. Yande Xiang et al published a study in 2018 titled ECG-Based Heartbeat Classification Using Two-Level Convolutional Neural Network and RR Interval Difference that proposed a method for patient-specific ECG beat classification that uses the difference between previous and post RR intervals as a dynamic feature [21].

In HIV patients, the RR interval and QTc interval are very essential. According to a 2019 study titled Prolonged QTc in HIV-Infected Patients: A Need for Routine ECG Screening by M. Myerson et al, HIV-infected patients may have clinically relevant longer QTc intervals on ECG, which is linked to the RR interval [22].

2.3. Time-Domain Analysis

Time-domain analysis is a technique for analyzing signals with a time dimension. It entails examining the shape of the waveform and determining significant elements such as the RR interval. This sort of analysis is especially valuable in ECG signal analysis, where the time and shape of distinct waveform components may provide important information about heart health.

Machine learning approaches are rapidly being applied in time-domain ECG data interpretation. A study published titled A Comparison of Physiological Signal Analysis Techniques and Classifiers for Automatic Emotional Evaluation of Audiovisual Contents by Adrián Colomer Granero et al. used various cutting-edge metrics to extract features from physiological signals, including ECG. These time-domain characteristics were used as inputs to numerous simple and complex classifiers [23].

Deep learning, a type of machine learning, has also been used to analyze ECG data in the time domain. A. Patané and M. Kwiatkowska's paper "Calibrating the Classifier: Siamese Neural Network Architecture for End-to-End Arousal Recognition from ECG"

offered a deep learning framework for arousal classification from ECG data. To extract features from ECG and assess time-domain variation patterns, the researchers created an end-to-end convolutional and recurrent neural network architecture [24].

Time-domain analysis is applied in domains other than heart health evaluation, such as seizure detection. D. Dash et al. published a review paper in 2022 titled "Review of Machine and Deep Learning Techniques in Epileptic Seizure Detection Using Physiological Signals and Sentiment Analysis" that discussed the effectiveness of time-domain features in seizure detection using machine learning techniques [25].

2.4. Frequency-Domain Analysis

Frequency-domain analysis is a technique for analyzing signals with a frequency component. It entails analyzing the spectral density of the signal and detecting essential aspects such as the power spectral density (PSD). This sort of analysis is very beneficial in ECG signal analysis, where the frequency components may provide important information about heart health.

Machine learning approaches are increasingly being applied in frequency-domain ECG data processing. A study published in 2016 titled A Comparison of Physiological Signal Analysis Techniques and Classifiers for Automatic Emotional Evaluation of Audiovisual Contents by Adrián Colomer Granero et al. used various cutting-edge metrics to extract features from physiological signals, including ECG. These frequency domain variables were used as inputs to numerous simple and sophisticated classifiers [26].

Frequency-Domain Analysis Using Deep Learning Deep learning, a type of machine learning, has also been used to analyze ECG data in the frequency domain. D. Dash et al. published a study in 2022 titled Review of Machine and Deep Learning Techniques in Epileptic Seizure Detection Using Physiological Signals and Sentiment Analysis, that discussed the effectiveness of frequency-domain features in seizure detection using machine learning techniques [27].

ECG signal improvement also makes use of frequency-domain analysis. G. Ramesh et al. presented a new ECG denoising approach based on noise reduction aspects in the frequency domain in their 2018 study titled "ECG Signal Enhancement through Subband Adaptive Soft Thresholding and EMD for Efficient Cardiac Arrhythmia Analysis" [28].

The Welch method is a prominent approach for estimating spectral density. It enhances power spectral density (PSD) estimation by minimizing noise. The approach entails segmenting the temporal signal into overlapping segments, calculating a periodogram for each segment, then averaging the periodograms to generate the PSD estimate. As a result, the Welch approach is especially effective in ECG signal processing, where precise PSD estimate is critical.

2.5. Polynomial Reggression

A study by Vacher et al. used a polynomial mixed-effects multilevel regression analysis approach to examine the effect of an asynchronous heart rate variability biofeedback (HRV-BFBasync) protocol on adolescent swimmers' cognitive appraisals and recovery-stress states during a six-week training period. The results suggested that the HRV-BFBasync protocol significantly predicts lower levels of biopsychosocial stress states and cognitive stress [29].

In a retrospective study by Perek et al., logistic regression was used to assess the prognostic role of ultra-short heart-rate variability (HRV) in myocarditis. The study found that HRV indices, specifically the root mean square of successive differences (RMSSD), may be prognostic indicators in myocarditis [30].

Zhou et al. used a logistic regression model to explore the relationship between HRV, the brain distribution of enlarged perivascular space (EPVS), and cognitive impairment in patients with EPVS. The study found that reduced HRV is involved in the pathophysiological mechanisms of the formation and development of BG-EPVS and is associated with cognitive impairment in patients with EPVS [31].

Lastly, a cross-sectional study by Niveatha et al. found that HRV indices were reduced in underweight, overweight, and obese groups compared to the normal weight BMI group. The study used second-order polynomial regression to show an inverted U-shaped relationship between BMI and HF log power in both genders [32].

These studies illustrate the diverse applications of Polynomial Regression models in HRV analysis, from sports training to disease prognosis and cognitive function assessment. However, more research is needed to further understand and optimize these models for HRV analysis.

3. Methodology

This research is divided into five stages: data selection, preprocessing, extraction, R-R interval, time-frequency analysis, and Frequency-Domain Analysis. The first stage is the selection of data. This stage involves the selection of the signal dataset obtained at physionet.org. The second stage is signal processing. The electrocardiogram signal is then analyzed for its characteristics and features. The captured signal characteristics are used as a reference for signal processing. This signal processing uses digital filters to determine the characteristics of the ECG signal. In the third stage, the filter normalization results are still processed to find R peaks, P peaks, T peaks, Q valleys, and S valleys, as well as obtain R-R intervals. In these three initial stages, the data has been processed from the ECG data source, and the R-R Interval data has been obtained. Then the data from the R-R Interval signal is processed. The R-R Interval signal data is then processed. These signals will then be evaluated using the Time-Frequency Analysis and Frequency-Domain Analysis methods in the next step. The findings of the two approaches will then be compared depending on the gender and age of the patients. Then the results of HRV calculations will be tested using polynomial regression modeling. Figure (3.1) shows the block diagram process of this research.

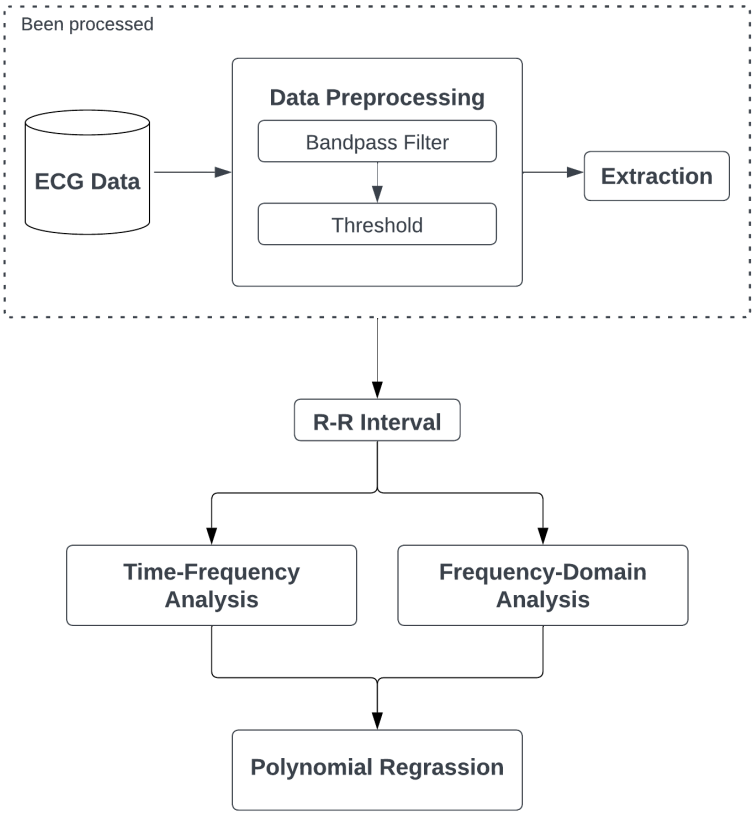

Figure 3.1. Block diagram

3.1. Data Collection

Data was obtained from the MIT-BIH arrhythmia Database. The RR Interval Database is a collection of RR interval data from healthy subjects. RR intervals are the time intervals between successive R waves in the QRS complex of the ECG of a heart. The R wave is the first upward deflection after the P wave, and it represents the depolarization of the ventricles.

The database contains the heart rate time series of 147 individuals, 72 males and 67 females (gender data on the other 8 individuals are unavailable). Of the 147 individuals, 71 were under one year of age, and 10 were over 18 years of age. Each time series was recorded using a two-lead ECG, for 24 hours, with a sampling frequency of 128 Hz. The RR intervals were then obtained from these ECG recordings. In the database, each individual is identified by an ID number, and the HRV time series of this individual is an ID.txt file. The files are text files (.txt, ANSI format) and contain RR intervals in milliseconds. Additional patient information is stored in the file patient-info.csv.

The text files contain the RR interval data, with each line representing one RR interval. The annotation files contain additional information about the RR intervals, such as the time of occurrence and the type of beat.

3.2. Preprocessing of ECG Signals

ECG signal preprocessing can help in the removal of contamination from an ECG signal. ECG contamination can be broadly classified into the following categories:

- Power line interference.
- Baseline wandering (e.g. due to subject movement).
- Electrode pop or contact noise (e.g. due to contraction between the electrode and the skin, causing an initial shift during electrode insertion).
- Patient-electrode motion artifacts (e.g. due to the lack of contact between ECG electrodes with the skin which causes impedance changes).
- Electromyographic (EMG) noise (e.g. muscle movement).

This data will be extracted so that it can be identified and classified later. It takes several stages of normalization or transformation to obtain the data. The preprocessing stage is depicted as a block diagram in Figure 3.2. The stages consist of data transformation, filtering using a bandpass filter, derivative, squaring, moving average, and thresholding. The explanation regarding preprocessing stages is explained below.

3.2.1. Bandpass Filter

A band-pass filter is a type of filter that allows only signals with frequencies between certain cutoff frequencies to pass through the filter, while signals with frequencies outside that range are blocked or thwarted. In other words, a band-pass filter signals with frequencies that lie within a certain range.

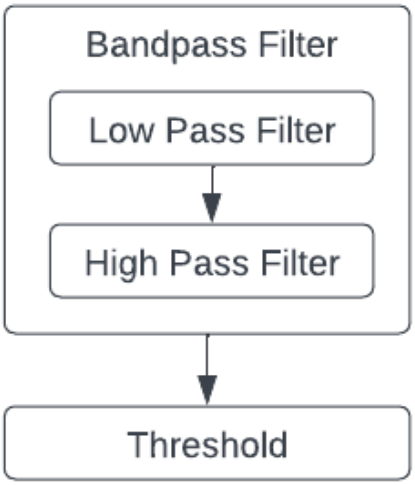

Figure 3.2. Preprocessing block diagram

Band-pass filters are very useful in separating certain frequency components from more complex signals. This filter is used to remove unwanted frequency components and isolate specific frequencies or desired frequency ranges.

Bandpass filter Reduces the influence of muscle noise, distraction, initial braking, and T-wave disturbance. Two filters are used. These filters are known as low-pass filters (LPF) and high-pass filters (HPF).

1. Low pass filter

   The transfer function of the second-order low-pass filter is according to equation 1.

$$H(z) = \frac{(1-z^{-6})^2}{(1-z^{-1})^2} \tag{1}$$

   So, the amplitude response is shown in equation 2.

$$|H(\omega T)| = \frac{(sin^2(3\omega T)}{(sin^2(\frac{\omega T}{2})} \tag{2}$$

   Where T is the sampling period. The difference equation of the filter is according to equation 3.

$$y(nT) = 2y(nT-T) - y(nT-2T) + x(nT-6T) + x(nT-12T) \tag{3}$$

2. High pass filter

   The high-pass filter design is based on reducing the output of a first order low-pass filter from an all-pass filter (i.e. sampled in the original signal). The transfer function

for this high-pass filter is shown in equation 4.

$$H(z) = \frac{(-1-32z^{-6}+z^{-32})^2}{(1-z^{-1})^2} \tag{4}$$

Then the amplitude of the response is in accordance with equation 5.

$$\left|H_{(\omega T)}\right| = \frac{[256+sin^2(16\omega T)]^{1/2}}{cos(\frac{\omega T}{2})} \tag{5}$$

The difference equation is shown in equation 6.

$$y(nT) = 32x(nT-16T) - [y(nT-T) + x(nT) - y(nT) - x(nT-32T) \tag{6}$$

3.2.2. Threshold

Threshold is a simple filtering method used to reduce the amount of noise in a signal. Where x represents the input, y represents the output, and λ represents the threshold.

$$threshold: \begin{cases} y = x \; if \; |x| > \lambda \\ y = 0 \; if \; |x| < \lambda \end{cases} \tag{7}$$

3.3. Extraction Data

The attributes of a typical ECG can be grouped under the signals associated with a normal sinus rhythm. The heart's regular rhythm, or pace, is regulated by the sinoatrial (SA) node. This node is strategically located in the right atrium, near the entrance of the superior vena cava. This rhythm, often referred to as the normal sinus rhythm, instigates an action potential that results in 60-100 heartbeats per minute when the body is at rest.

The defining characteristics of a normal sinus rhythm are as follows:

- Speed: 60 – 100 beats per minute.
- Rhythm: Regular.
- PR interval: 0.12 – 0.20 seconds.
- QRS interval: 0.05 – 0.11 seconds.
- RR interval: 0.6 – 1 seconds.

After normalizing, the calculation for the feature point search is performed. This search begins by detecting R peak, then looking for points P and T, followed by points Q and S. The ECG signal (shown in Figure 3.3) is used to detect heart defects.

R peak detection is accomplished by locating the highest point in a signal. R detection on normal signals typically uses an empirically determined threshold, whereas on arrhythmic signals, this function cannot be used for all signals, so R detection employs a separate algorithm.

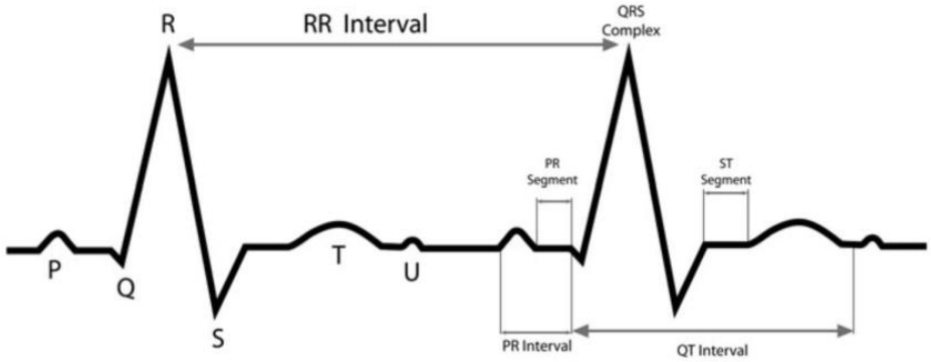

Figure 3.3. The ECG waves, segments, and intervals [33]

3.4. RR interval

After obtaining the R peak value, the R-R interval can be calculated by measuring the distance between the first and second R peaks, the second and third R peaks, and so on up to one minute. The measurement result is then divided by one minute of time.

$$RR_{int}(i) = \frac{\sum_{i=n}^{n}(R_{i+1} - R_i)}{60} \tag{8}$$

Where:

- $RR_{int}(i)$: The i peak R peak interval value
- R: Peak location value R.
- n: total peaks detected in one signal.
- i: repetition of the process for 1 location of the detected R peak.

3.5. Time-Domain Analysis

Measures in the time-domain capture the overall variability of heart rate, but they aren't particularly specific in quantifying the individual contributions of various regulatory mechanisms. Despite this, their broad sensitivity can be advantageous, such as in preliminary studies or when the focus isn't on specific neurophysiological mechanisms. Additionally, due to their simplicity in computation and interpretation, time-domain measures continue to be some of the most frequently reported indices of heart rate variability [34].

Time-Domain Analysis is a signal analysis technique that acts directly on time-series data. It is one of the most basic and widely used techniques across a wide range of industries, including but not limited to electronics, telecommunications, and signal processing.

Time-Domain Analysis in the context of ECG and RR interval analysis entails evaluating the RR interval across time. The RR interval is measured in milliseconds and represents one full heart cycle. We may get insights about the heart's function and perhaps uncover anomalies or abnormalities by examining these intervals over time.

Some of the primary statistical metrics employed in Time-Domain Analysis are as follows:

Table 3.1. Time-domain measures of HRV

Parameter	Unit	Description
MeanNN	ms	The mean of the RR intervals
SDNN	ms	The standard deviation of the RR intervals
RMSSD	ms	The square root of the mean of the squared successive RR intervals
pNN50	%	Percentage of Successive RR intervals that differ by more than 50ms
NN50		Number of consecutive RR intervals that differ by more than 50ms

3.5.1. Mean RR interval

The Mean RR interval is an approach utilized within the domain of heart rate variability (HRV) analysis. It indicates the mean value of all RR intervals within a particular time frame, see equation 14. The RR interval refers to the time duration between two consecutive R-waves of the QRS complex on the electrocardiogram (ECG), serving as a measure of the time elapsed between two successive heartbeats.

$$\overline{x} = \frac{1}{N} \times \sum_{t=1}^{N} x(t) \tag{9}$$

Where:

- $x(t)$: time interval NN at time
- N: number of samples

By performing calculations on the Mean RR interval, it is possible to obtain an approximation of the average heart rate within the specified temporal context. The cause for this occurrence is that it has a correlation between the heart rate and the RR interval. Specifically, a shorter RR interval is a sign of a heart rate that is greater, whereas a longer RR interval is indicative of a lower heart rate.

Hence, the Mean RR interval offers an informative and uncomplicated assessment of the overall heart rate. This tool has the capability to evaluate the long-term functionality of the heart and can offer significant perspectives into an individual's cardiovascular well-being. For example, a consistently elevated heart rate (resulting in a decreased Mean RR interval) could potentially suggest the presence of stress or excessive physical strain. Conversely, a lower heart rate (corresponding to an increased Mean RR interval) may indicate favorable levels of physical fitness or a state of relaxation. Nevertheless, it is crucial to acknowledge that the assessment of the Mean RR interval necessitates a contextual approach, taking into account additional variables such as age, level of physical fitness, and overall health condition.

3.5.2. Standard Deviation of RR intervals (SDNN)

SDNN is a time-domain metric that measures the variability in intervals between normal sinus heartbeats, expressed in milliseconds (ms), see equation 15. "Normal" here implies that any abnormal beats, such as ectopic beats (heartbeats originating outside the sinoatrial node in the right atrium), have been excluded. While the normal recording interval for short-term tests is 5 minutes, some researchers propose ultra-short recording times ranging from 60 seconds to 240 seconds.

$$s_x = \sqrt{\frac{1}{N} \times \sum_{t=1}^{N} (x(t) - \overline{x})^2} \tag{10}$$

The Sympathetic Nervous System (SNS) and the Parasympathetic Nervous System (PNS) influence SDNN. It has a strong connection between overall power and the power in the Ultra-Low Frequency (ULF), Very-Low Frequency (VLF), and Low Frequency (LF) bands. However, the contribution of these bands to SDNN depends on the measurement conditions. When these bands have more power than the High Frequency (HF) band, they contribute more to SDNN.

During short-term resting recordings, the primary source of variation is parasympathetically mediated Respiratory Sinus Arrhythmia (RSA), especially with slow, paced breathing protocols. In contrast, during 24-hour recordings, LF band power significantly contributes to SDNN. SDNN is more accurately calculated over a 24-hour period than during shorter periods monitored during biofeedback sessions.

Longer recording periods provide a broader perspective on cardiac reactions to various environmental stimuli. They can index the heart's response to changing workloads, anticipatory central nervous activity, and circadian processes, including sleep-wake cycles. These 24-hour recordings reveal the SNS contribution to HRV.

SDNN, when recorded over a 24-hour period, is considered the "gold standard" for medical stratification of cardiac risk. SDNN values can predict both morbidity and mortality. Based on 24-hour monitoring, patients with SDNN values below 50 ms are classified as unhealthy, those with values between 50-100 ms have compromised health, and those with values above 100 ms are considered healthy. Heart attack survivors with higher 24-hour SDNN values had a greater probability of survival during a 31-month mean follow-up period. For instance, patients with SDNN values over 100 ms had a 5.3 times lower risk of mortality at follow-up than those with values under 50 ms. This raises the question of whether training patients to increase their SDNN to a higher category could potentially reduce their risk of mortality.

3.5.3. Root Mean Square of Successive Differences (RMSSD)

RMSSD is a time-domain measure used to estimate the vagally mediated changes reflected in HRV. It's calculated by taking the square root of the average of the square of

differences between successive normal heartbeats over a certain period, see equation 16. While the conventional minimum recording period is 5 minutes, some researchers have proposed ultra-short-term periods of 10 seconds, 30 seconds, and 60 seconds.

$$s_x = \sqrt{\frac{1}{N} \times \sum_{t=1}^{N} (x(t) - x(t-1))^2} \tag{11}$$

Where:

- $x(t)$: NN interval time at time t+1

RMSSD reflects the beat-to-beat variance in heart rate and is identical to the non-linear metric SD1, which reflects short-term HRV. It's strongly correlated with pNN50 and high-frequency (HF) power when measured over 24 hours. However, the correlation between RMSSD and the maximum and minimum heart rate is inconsistent and weak.

While RMSSD is correlated with HF power, the influence of respiration rate on this index is uncertain. It's less affected by respiration than Respiratory Sinus Arrhythmia (RSA) across several tasks. RMSSD is more influenced by the Parasympathetic Nervous System (PNS) than SDNN. Lower RMSSD values are correlated with higher scores on a risk inventory of sudden unexplained death in epilepsy.

NN50, pNN50, and RMSSD are calculated using the differences between successive NN intervals. Since their computation depends on NN interval differences, they primarily index high-frequency heart rate oscillations, are largely unaffected by trends in an extended time series, and are strongly correlated.

3.5.4. Percentage of Successive RR intervals that differ by more than 50ms (pNN50)

The pNN50 metric refers to the percentage of adjacent NN intervals (the time between successive heartbeats) that differ by more than 50 milliseconds. This metric requires a minimum of a 2-minute recording period, although some researchers have proposed ultra-short-term periods of 60 seconds.

The pNN50 is closely associated with Parasympathetic Nervous System (PNS) activity and is correlated with the RMSSD (Root Mean Square of Successive Differences) and High Frequency (HF) power. These correlations suggest that pNN50, RMSSD, and HF power can all provide insights into the activity of the PNS, which is the part of the nervous system responsible for rest and digestion.

However, the article notes that RMSSD is often preferred over pNN50 for assessing Respiratory Sinus Arrhythmia (RSA), especially in older subjects. RSA refers to variations in heart rate that occur during a breathing cycle, and it's a key factor contributing to HRV.

RMSSD may give an indicator that is more reliable than short-term SDNN measures in the setting of biofeedback, where short-term data are often employed. Biofeedback is a method that teaches you how to manage the functions of your body, such as your heart rate.

While SDNN is an important parameter in HRV study, additional metrics such as pNN50 and RMSSD also give vital insights into heart rate variability and nervous system function. The choice of metric, however, may be influenced by the unique situation, such as the duration of the recording period and the age of the subjects.

3.5.5. NN50

NN50 is the count of pairs of successive NN intervals that differ by more than 50 milliseconds. To accurately calculate NN50, a minimum of a 2-minute recording period (epoch) is required.

In simpler terms, NN50 is a measure that helps us understand the changes in heart rate over time. It counts how many times the gap between two heartbeats (NN intervals) changes by more than 50 milliseconds from one beat to the next within a 2-minute period. This statistic gives useful information on the ability of the heart for adjusting to changes, which is an important part of cardiovascular health.

A higher NN50 count usually correlates with heart wellness, since it reflects more variability in heart rate and, therefore, better adaptation of the heart to variations in the body's needs. A smaller NN50 count, on the other hand, may indicate less heart rate variability, which might be a symptom of possible health problems. However, this metric should be interpreted in the context of other HRV measures and individual health profiles for a comprehensive understanding.

These measurements give useful information about the functioning of the heart and may aid in the detection of probable cardiac problems or stress levels. A lower SDNN or RMSSD, for example, may imply less variability in the heart rate, which might be an indication of stress or cardiac disease.

It is essential to understand that Time-Domain Analysis is merely one method of evaluating ECG data. Other approaches, such as Frequency-Domain Analysis, give various insights and may supplement Time-Domain Analysis results.

3.6. Frequency-Domain Analysis

The frequency parts of a signal or dataset are analyzed using Frequency-Domain Analysis. It is an effective tool that is employed in a variety of domains including signal analysis, physics, and engineering. The frequency-Domain research provides insights into the patterns of activity which influence the variability of heart rate in the context of Heart Rate Variability (HRV) research.

The fundamental idea underlying Frequency-Domain Analysis is to analyze time-series data (like RR intervals in ECG) into a combination of sinusoidal components of various frequencies. This frequently happens through an equation known as the Fourier Transform. This process gives a signal spectrum that shows the power distribution of the signal across various frequencies.

In HRV analysis, the frequency spectrum is usually divided into several bands, each corresponding to different physiological mechanisms:

Table 3.2. Frequency-domain measures of HRV.

Parameter	Unit	Frequency range	Description
VLF	ms^2	0.0033–0.04 Hz	This band is thought to be influenced by various factors including thermoregulation and the renin-angiotensin system
LF	ms^2	0.04–0.15 Hz	This band is generally associated with both sympathetic and parasympathetic activity
HF	ms^2	0.15–0.4 Hz	This band is primarily associated with parasympathetic (vagal) activity and corresponds to the respiratory rhythm
LF/HF Ratio	%		Provide an estimate of the balance between the activities of the Sympathetic Nervous System (SNS) and the Parasympathetic Nervous System (PNS)

3.6.1. Very Low Frequency (VLF)

The Very Low Frequency (VLF) band, which ranges from 0.0033-0.04 Hz, is an important component of HRV research. A recording duration of at least 5 minutes is required to properly catch the oscillations in this band, while a 24-hour monitoring period may offer the most thorough data. A 5-minute sample might include anything from 0 to 12 complete periods of oscillation.

Low values in the VLF band and all 24-hour clinical HRV tests are indicative of a greater risk of unfavorable health consequences. However, compared to either Low Frequency (LF) or High Frequency (HF) power, VLF power shows a greater connection with all-cause mortality. This shows that the VLF rhythm could be very important for general health.

3.6.2. Low Frequency (LF)

The Low Frequency (LF) band, which ranges from 0.04 to 0.15 Hz, is an important component of HRV study. It is usually recorded for at least 2 minutes. This band was originally known as the baroreceptor range because it predominantly represents resting baroreceptor activity. Baroreceptors are blood pressure sensors found in the heart, aorta, and carotid sinus.

The parasympathetic and sympathetic nervous systems, as well as blood pressure modulation through baroreceptors, may all influence LF power. However, it is primarily driven by the PNS or by baroreflex activity alone. The SNS does not seem to produce rhythms much above 0.1 Hz, while the parasympathetic system can affect heart rhythms

down to 0.05 Hz (20-second rhythm). Under resting conditions, the LF band reflects baroreflex activity and not cardiac sympathetic innervation.

During periods of slow respiration rates, vagal activity, which is mediated by the vagus nerve, part of the PNS, can generate oscillations in heart rhythms that cross over into the LF band. Therefore, respiratory-related efferent vagally mediated influences are particularly present in the LF band when respiration rates are below 8.5 breaths per minute or 7-second periods, or when one sighs or takes a deep breath.

3.6.3. High Frequency (HF)

The High Frequency (HF) band, which ranges from 0.15 to 0.40 Hz, is normally recorded for at least one minute. The resting range may be changed to 0.24 to 1.04 Hz for newborns and children who have quicker breathing rates than adults. Because of its relationship with heart rate fluctuations associated with the respiratory cycle, the HF band is also known as the respiratory band. Respiratory Sinus Arrhythmia (RSA) refers to these periodic heart rate variations, which may or may not be a pure sign of cardiac vagal regulation.

The heart rate rises during breathing and falls when exhale. When we inhale, the cardiovascular center restricts the outflow of the vagus nerve, causing our heart rate to rise. During expiration, on the other hand, the cardiovascular center restores the vagus nerve's outflow, resulting in a drop in heart rate owing to the release of acetylcholine. Total blockage of the vagus nerve almost completely eliminates HF oscillations and reduces power in the Low Frequency (LF) range.

HF power is closely associated with the pNN50 and RMSSD time-domain measures. HF band power may increase during the night and decrease during the day. Lower HF power is linked with stress, panic, anxiety, or worry. The modulation of vagal tone is crucial for maintaining dynamic autonomic regulation, which is important for cardiovascular health. A deficiency in vagal inhibition is associated with increased morbidity.

3.6.4. LF/HF Ratio

This LF/HF ratio was initially established based on 24-hour records, in which both the PNS and SNS contribute to LF power, but PNS activity largely contributes to HF power. The goal was to calculate the balance of SNS and PNS activity.

The SNS generates LF power, whereas the PNS generates HF power, according to the underlying assumptions of the LF/HF ratio. A low LF/HF ratio in this paradigm shows parasympathetic dominance, which is related to energy conservation and social bonding activities. A high LF/HF ratio, on the other hand, indicates sympathetic system dominance, which is associated with fight-or-flight reactions or parasympathetic system withdrawal.

However, the LF/HF ratio as a measure of "sympatho-vagal balance" has been called into question. To begin with, LF power is not a clean measure of SNS activity since the PNS produces half of the fluctuation in this frequency range and unidentified causes provide the other half. Second, interactions between the PNS and SNS are complicated, non-linear,

and often non-reciprocal. Third, the effect of respiratory mechanics and resting heart rate generates ambiguity about the PNS and SNS contributions to the LF/HF ratio during the measurement time.

The LF/HF ratio is considered controversial because different processes seem to generate 24-hour and 5-minute values, and these values correlate poorly. Moreover, the SNS contribution to LF power varies significantly with testing conditions. For instance, when LF is calculated while sitting upright during resting conditions, the primary contributors are PNS activity and baroreflex activity, not SNS activity. Therefore, the interpretation of 5-minute resting baseline LF/HF ratios depends on specific measurement conditions.

3.6.5. Welch's Method in Frequency-Domain Analysis

Welch's method is a technique used in the Frequency-Domain Analysis to estimate the power spectral density (PSD) of a signal. The PSD is a measure of a signal's power content versus frequency, a fundamental tool in Frequency-Domain Analysis.

Welch's method involves dividing the time signal into overlapping segments, computing a modified periodogram of each segment, and then averaging these periodograms to produce the PSD estimate. This method reduces the noise in the PSD estimate and is widely used in HRV analysis.

The steps of Welch's method are as follows:

1. The time signal is divided into overlapping segments.
2. Each segment is windowed with a discrete window function (like a Hamming window).
3. The periodogram is computed for each windowed segment.
4. The periodograms are averaged to produce the PSD estimate.

By applying Welch's method in the Frequency-Domain Analysis of HRV, we can gain insights into the autonomic nervous system's activity and its influence on heart rate variability.

3.7. Polynomial Regression

Polynomial regression is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modeled as an nth degree polynomial. Polynomial regression can be used to model relationships between variables that aren't linear. It fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted E(y |x).

The general form of a polynomial regression is:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + ... + \epsilon \tag{12}$$

Where:

- y: typically represents the dependent variable, also known as the response or target variable. This is the variable that we want to predict or explain. In this case, y represents the HRV (Heart Rate Variability) indicators ('Mean RR', 'SDNN', 'RMSSD', 'NN50', 'pNN50', 'LF', 'HF', 'LF/HF Ratio').
- x: typically represents the independent variable(s), also known as the predictor(s) or feature(s). These are the variables that we use to predict or explain the dependent variable. In this case, x represents 'Age'.
- $\beta_0, \beta_1, \beta_2, \beta_3, ...$: are the regression coefficients, which represent the weight of each term. β_0 is the y-intercept of the regression line.
- ϵ: is the error term, which is a random variable that adds noise to the linear relationship.

The goal of polynomial regression is to model a non-linear relationship between the independent and dependent variables. It's important to note that while this method allows for a flexible, curve-linear relationship between the predictors and the response variable, it is still considered a form of multiple linear regression. This is because the regression function is linear in terms of the unknown parameters ($\beta_0, \beta_1, \beta_2, \beta_3, ...$).

The goodness of fit of the model can be determined by the coefficient of determination R Square and the adjusted coefficient of determination. The R Square value represents the proportion of the variance for the dependent variable that's explained by the independent variable. The adjusted R Square value adjusts the statistic based on the number of independent variables in the model.

The significance of the regression coefficients is tested using the p-value. If the p-value is less than 0.05 ($p < 0.05$), then the regression coefficient is considered statistically significant. This means that there is strong evidence that the coefficient is not zero, and the independent variable does have an effect on the dependent variable. If the p-value is greater than 0.05 ($p > 0.05$), then the regression coefficient is not statistically significant. This means that we do not have enough evidence to conclude that the coefficient is different from zero, and the independent variable may not have an effect on the dependent variable

It's also important to check the assumptions of polynomial regression, which are similar to the assumptions of multiple linear regression:

1. Linearity: The relationship between the predictors and the response variable is linear in the parameters.
2. Independence: The residuals are independent.
3. Homoscedasticity: The variance of the residuals is constant across all levels of the predictors.
4. Normality: The residuals are normally distributed.

Violations of these assumptions can lead to inaccurate estimates of the regression

coefficients and incorrect inferences. Therefore, it's important to check these assumptions using appropriate diagnostic plots and tests..

4. Results and Discussion

This section explains the process of analyzing Heart Rate Variability (HRV) based on the RR interval signals pattern, which is next further studied utilizing the methods of time-domain analysis and frequency-domain analysis. The final results of this analysis are additionally addressed in this section. In this study, a comparison will be made between the computation results obtained from the two different approaches to data analysis, taking into consideration of the subject's age and gender.

4.1. Data Processing

The study data that were employed have been discussed in Chapter 3; from the data source, the database comprises the HRV time series of 147 people, including 72 men and 67 females (gender data on the remaining 8 persons is not accessible). 71 of the people were under the age of one, while just 10 of the individuals were more than 18 years old. Each individual has an ID number, and the ID.txt file for this individual is their HRV time series. Text files (.txt, ANSI format) that contain RR intervals expressed in milliseconds may be found in these folders. However, after checking back with the existing patient data, there are differences in the number of patients based on age and gender composition. RR interval data in text format processed using Jupyter notebook with Python programming language. As will be shown in Table 4.1.

Table 4.1. Data subjects

Age	Female	Male
0-1	35	37
1-5	12	14
5-10	5	7
10-20	5	6
20-55	5	4
Total:		130

There are 147 people listed in this table, 68 of whom are male and 62 of whom are female. The table contains 147 individuals, 68 males, and 62 females. The gender data for the remaining 17 people is not available. In addition, 72 people were less than one year old, while 9 people were older than 20 years old (the ages of the other 9 persons could not be determined).

Following this, testing of all previously collected interval RR signal data is performed. The collection of all of this data took place over an interval of about twenty-four hours. One of the RR interval signals taken from a male patient who was 53 years old may be shown in Figure 4.1.

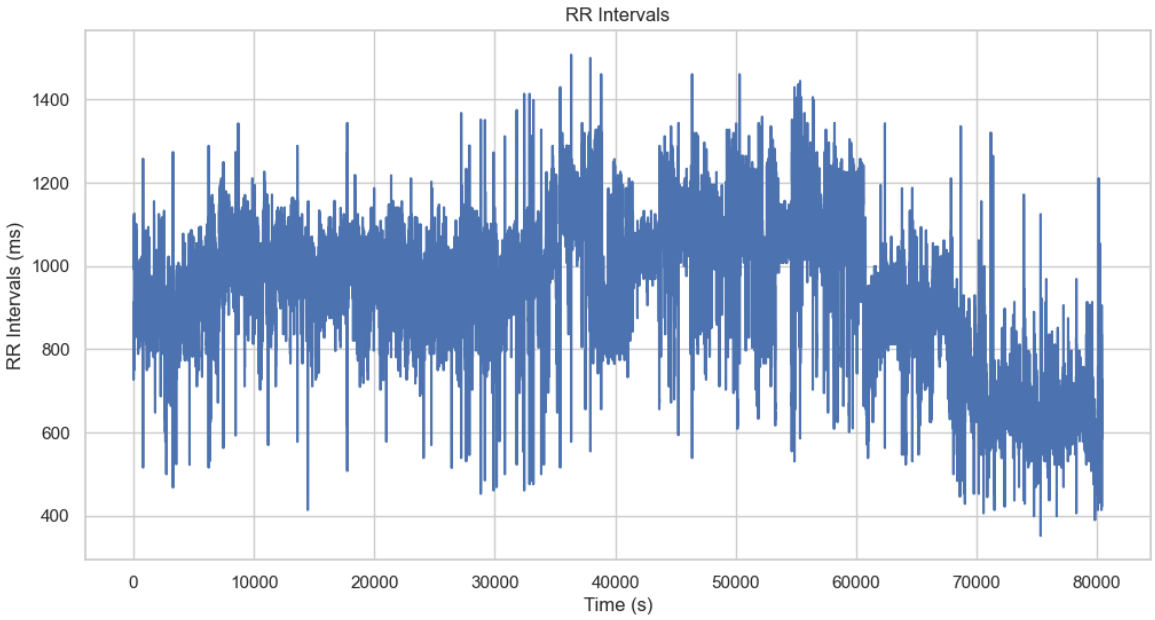

Figure 4.1. RR interval waves

4.2. Time-Domain Analysis

Table 4.2 shows the results of time domain analysis calculations and obtains several parameters (mean rr, SDNN, RMSSD, NN50 dan pNN50). Figure 4.2 shows a wave image from one of the subjects whose parameters have been found.

Table 4.2. Time domain measurement from several subjects

Age (year(s))	Gender	mean_rr (ms)	sdnn (ms)	rmssd (ms)	nn50	pNN50 (%)
5	M	716.4431265	179.5425117	107.3603399	50247	42.84764089
12	F	761.5967786	160.9319538	75.84080329	46788	41.56679489
14	M	863.8274956	183.0443175	93.01482797	44348	49.59627815
1	F	611.8550553	154.2087883	123.8322828	43134	31.65751695
12		860.2435174	155.6728649	79.06369599	42587	42.93563738
5	F	601.3167574	162.4778607	91.02855067	40663	28.67387809
8	F	691.0414067	133.6178885	68.5300412	40339	33.38602619
10		924.1608593	249.9287158	154.1951707	39462	43.97223182
7	M	643.8212291	206.2752889	97.59757677	39170	29.32105697
10	M	884.2574462	177.4593635	80.9987416	37572	44.67379285

Heart Rate Variability is a physiological phenomenon that denotes the temporal variation in the intervals between successive cardiac contractions. The non-invasive assessment is related to the evaluation of the autonomic nervous system's performance, alongside a specific focus on the relationship between sympathetic and parasympathetic activities. The influence of gender and age on cardiovascular variability has been estab-

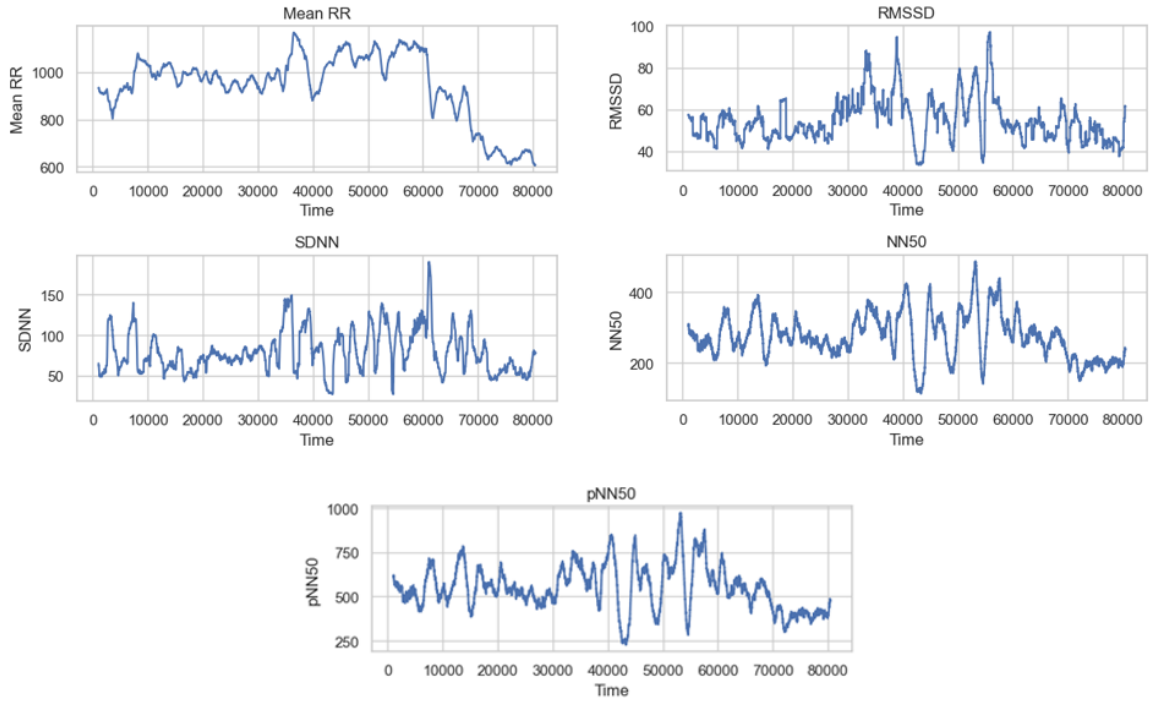

Figure 4.2. Waves of mean RR, SDNN, RMSSD, NN50 and pNN50 from one of the subjects

lished, with several studies indicating that males may exhibit higher levels of HRV in comparison to females.

In this research, an investigation has been done to investigate the association between four heart rate variability (HRV) measures, namely mean RR, SDNN, RMSSD, NN50, and pNN50, and age. Subjects were split into five categories of age: '0-1', '1-5', '5-10', '10-20', and '20-55'. Following that, the average value of each HRV measure was calculated for each age category. Differences in the aforementioned measures were observed between genders within each age category, indicating that gender may exert an influence on HRV. Nevertheless, the variability of gender differences in HRV appears to be contingent upon age groups and assessment methods, indicating complicated interactions between gender and HRV.

4.2.1. Mean RR

Mean RR is the average time between heartbeats. Observation of differences in Mean RR between genders within each age category, suggests that gender may influence Mean RR. Females tend to have a lower Mean RR than males in the '0-1', '1-5', and '5-10' age categories, suggesting a faster heart rate in females than in males in these age groups. However, in the '10-20' and '20-55' age categories, males tend to have a lower Mean RR than females, indicating a faster heart rate in males than in females in these age groups. It was found that the mean RR decreases as the age category increases from '0-1' to '20-55'.

This suggests that heart rate tends to increase (i.e., RR interval decreases) with age. Figure 4.3 shows a comparison diagram of mean RR for age and gender.

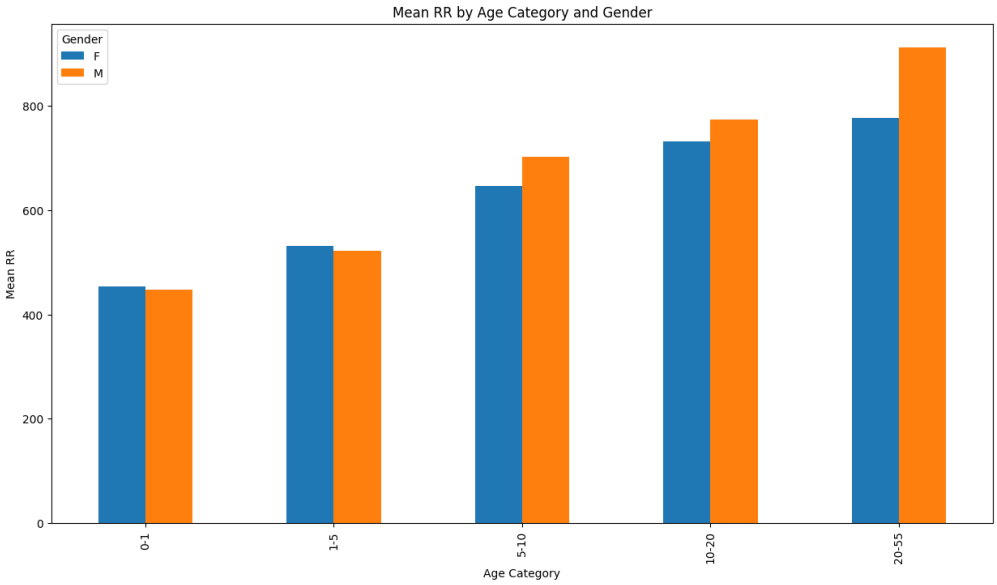

Figure 4.3. Comparision of mean RR

4.2.2. SDNN

Standard Deviation of NN intervals (SDNN) is a metric used to quantify the variability in heart rate variability (HRV), providing an assessment of the overall HRV. The given value denotes the statistical measure of variability, specifically the standard deviation, across all NN intervals. More significant heart rate variability is indicated by higher values of SDNN. Variations in the mean SDNN were observed between genders within each age category, indicating a potential influence of gender on SDNN. The results obtained indicate that there is a decrease in the mean SDNN as the age category grows, which is similar to the overall pattern of declining heart rate variability observed in older people. Figure 4.4 shows a comparison diagram of SDNN for age and gender.

4.2.3. RMSSD

Root Mean Square of Successive Differences (RMSSD) is a gauge for heart rate fluctuations that last for just a brief period of time. The mean squared differences of subsequent NN intervals are represented by its square root. Heart rate variability is larger when RMSSD values are higher. We discovered gender disparities in mean RMSSD, much like SDNN. We saw a decline in mean RMSSD with increasing age category, much as SDNN. This shows that HRV declines as people become older. Figure 4.5 shows a comparison diagram of RMSSD for age and gender.

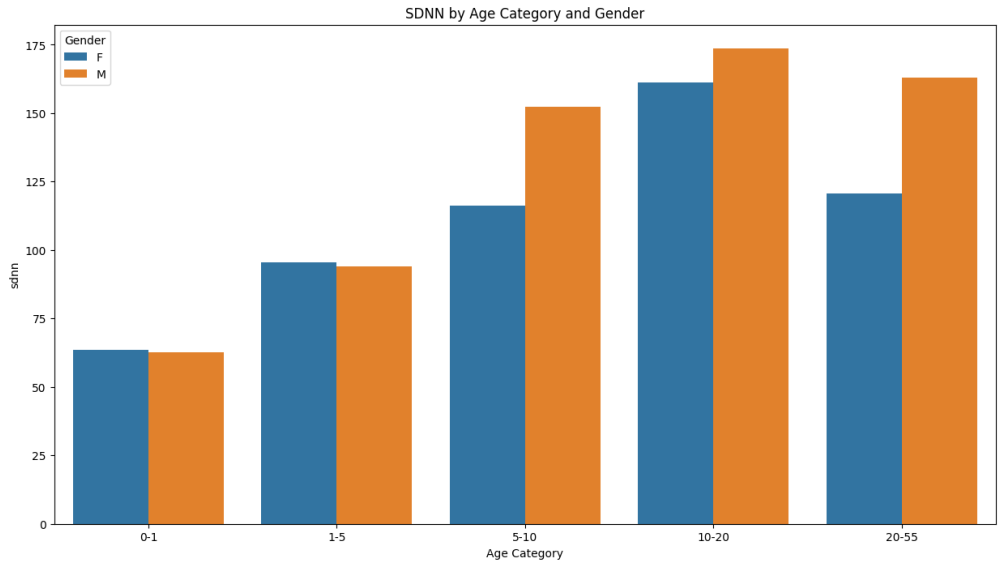

Figure 4.4. Comparision of SDNN

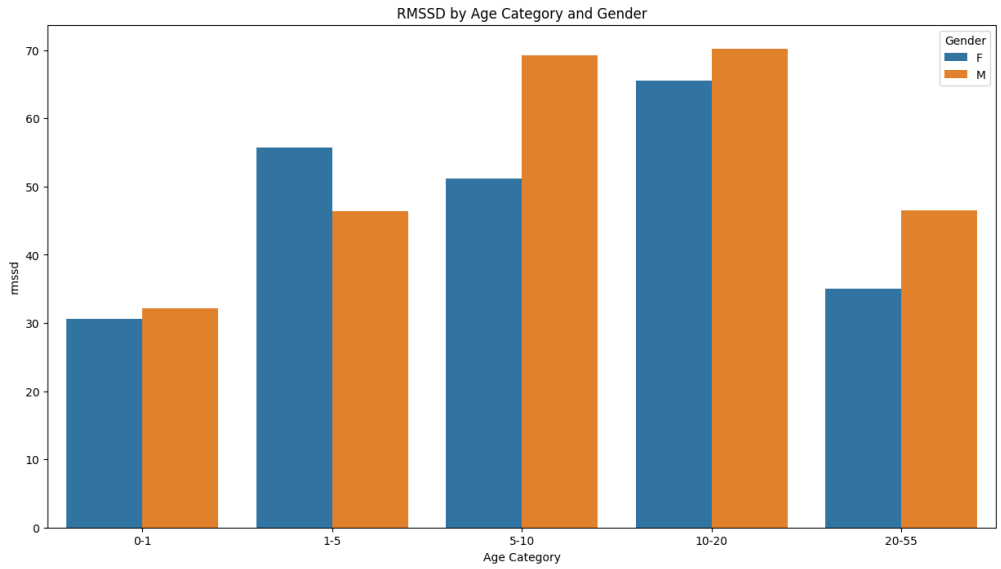

Figure 4.5. Comparison of RMSSD

4.2.4. NN50

Number of consecutive NN intervals greater than 50 ms (NN50) is another HRV measurement. Heart rate variability is higher when the NN50 value is higher. Our investigation found that the mean NN50 varied across gender. Analysis showed that the mean NN50 decreased with increasing age groups. This is consistent with the typical pattern of declining HRV with aging. Figure 4.6 shows a comparison diagram of NN50 for age and gender.

4.2.5. pNN50

The percentage of consecutive NN intervals greater than 50 ms (pNN50) is a measurement of the percentage of succeeding NN intervals that vary by more than 50 ms. It

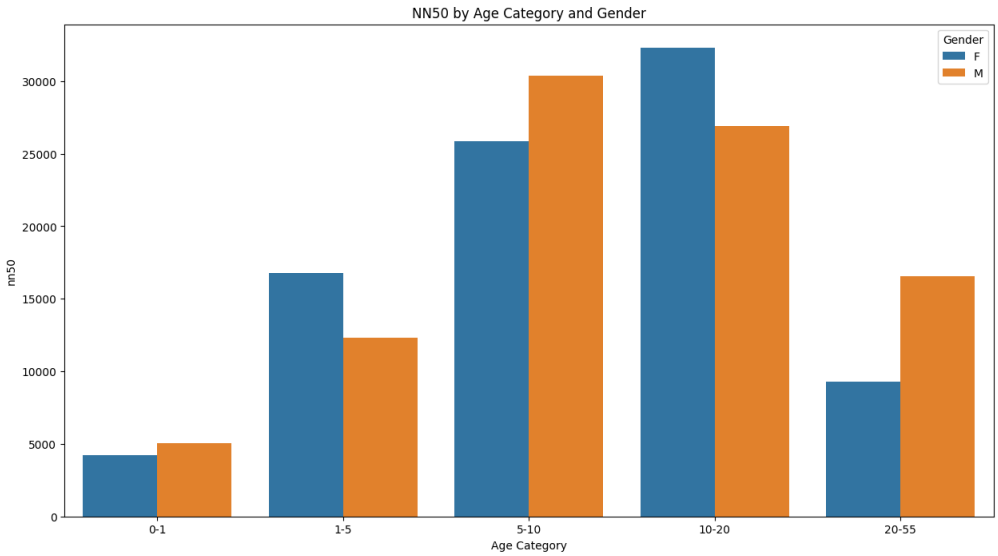

Figure 4.6. Comparison of NN50

is calculated as the proportion of the NN50 count divided by the entire number of NN intervals. Greater heart rate variability is indicated by higher pNN50 values. We observed variations in mean pNN50 between the sexes. With each age group up, the mean pNN50 was shown to be decreasing. This follows a decrease in HRV with age. Figure 4.7 shows a comparison diagram of pNN50 for age and gender.

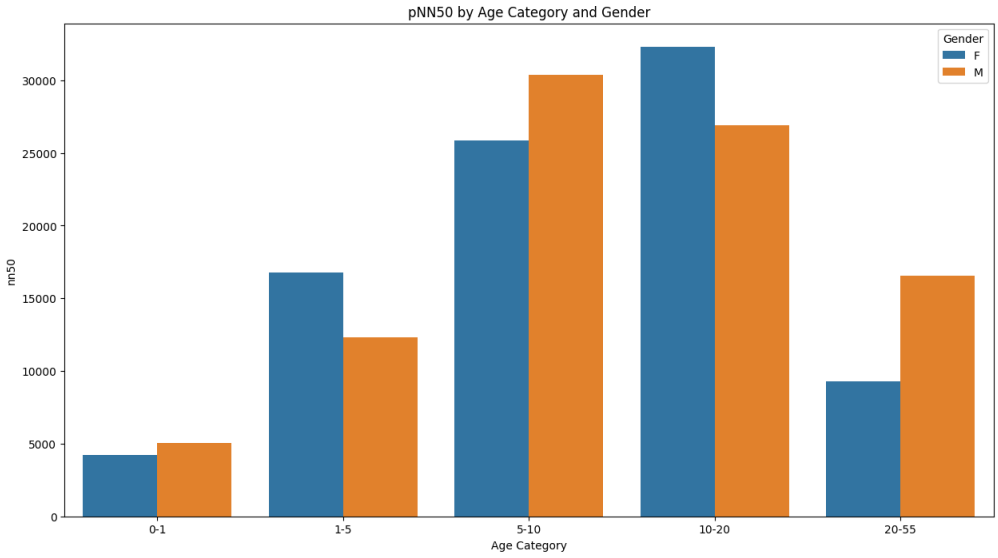

Figure 4.7. Comparison of pNN50

4.3. Frequency-Domain Analysis

The time-frequency HRV indices typically include the maximum and minimum energy of the time windows in the heart rate signal, the standard deviation between energy of

time windows, the total energy of signal in different frequency bands (e.g., energy of VLF, LF, or HF), and the mean energy of each band (the result obtained by separating overall energy by band length), shown in Figure 4.8.

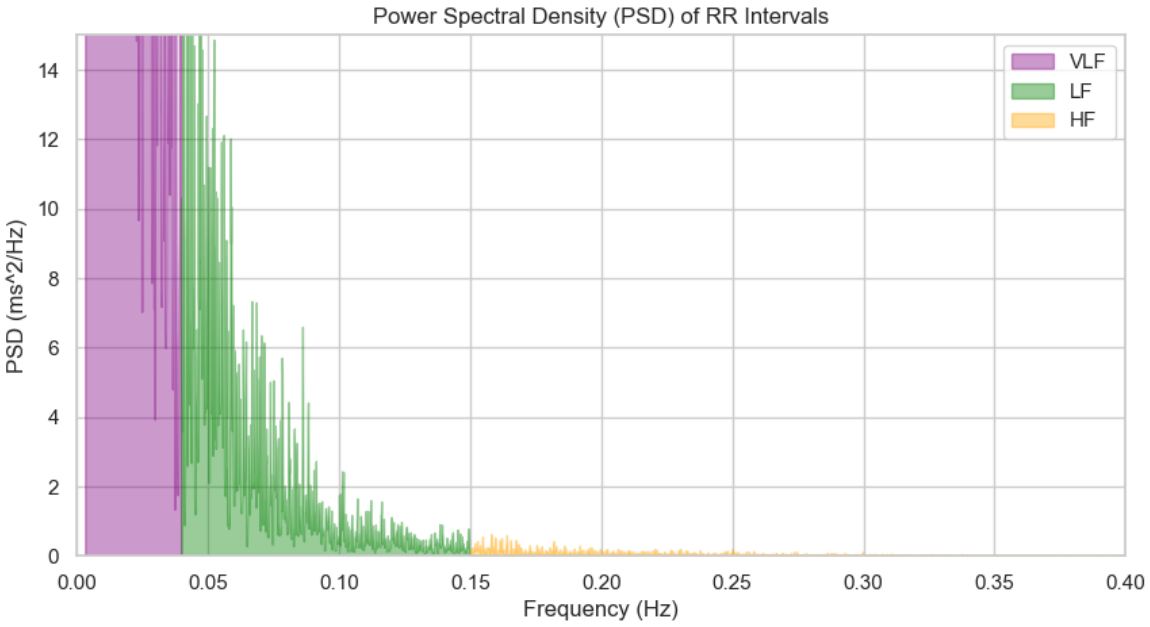

Figure 4.8. PSD of RR interval

After that, all of the patient data is processed, and results for frequency bands such as LF, HF, and VLF as well as the LF/HF ratio are obtained. Table 4.4 displays the values of the frequency bands for each and every patient's record.

Table 4.3. Frequency domain measurement from several subjects

Age (year(s))	Gender	VLF (ms^2)	LF (ms^2)	HF (ms^2)	LF/HF ratio
0.25	F	2094.029632	0.037955098	0.000804095	47.20223988
0.25	F	1966.688206	0.044587297	0.000937745	47.54735158
0.17	M	1644.504546	0.042234868	0.001048715	40.27298136
0.42	M	2229.718792	0.039308276	0.000898568	43.7454557
		2017.706748	0.056025584	0.001918674	29.20016387
0.33	F	2474.556519	0.043205352	0.001601005	26.98638815
0.083	M	1700.543455	0.039340244	0.001213243	32.4256908
		2182.99359	0.047686568	0.000750727	63.52052628
0.083	M	1999.042664	0.045830972	0.001705806	26.86763059
0.083	F	1845.663404	0.056450527	0.00145429	38.81654855

The HF (High Frequency), LF (Low Frequency), VLF (Very Low Frequency), and LF/HF ratio are essential components of heart rate variability and provide valuable insight into the autonomic nervous system's functioning. The LF component is frequently associated with

both sympathetic and parasympathetic activity, whereas the HF component is primarily associated with parasympathetic activity. The VLF component is less well understood but is believed to reflect very rapid variations in heart rate. The LF/HF ratio is frequently employed as a measure of the equilibrium between sympathetic and parasympathetic activity.

This analysis revealed distinct patterns in these HRV components between males and females across different age categories. Age is a key factor that influences HRV, reflecting changes in autonomic function over the lifespan.

4.3.1. VLF

VLF values show a similar trend to LF, with values increasing with age, peaking in the 10-20 years category, and then decreasing. VLF is thought to reflect both sympathetic and parasympathetic activity, and its interpretation can be complex. Furthermore, the values show a similar trend to LF, with values increasing with age, peaking in the 10-20 years category, and then decreasing. VLF is thought to reflect both sympathetic and parasympathetic activity, and its interpretation can be complex. Figure 4.9 shows a comparison diagram of VLF for age and gender.

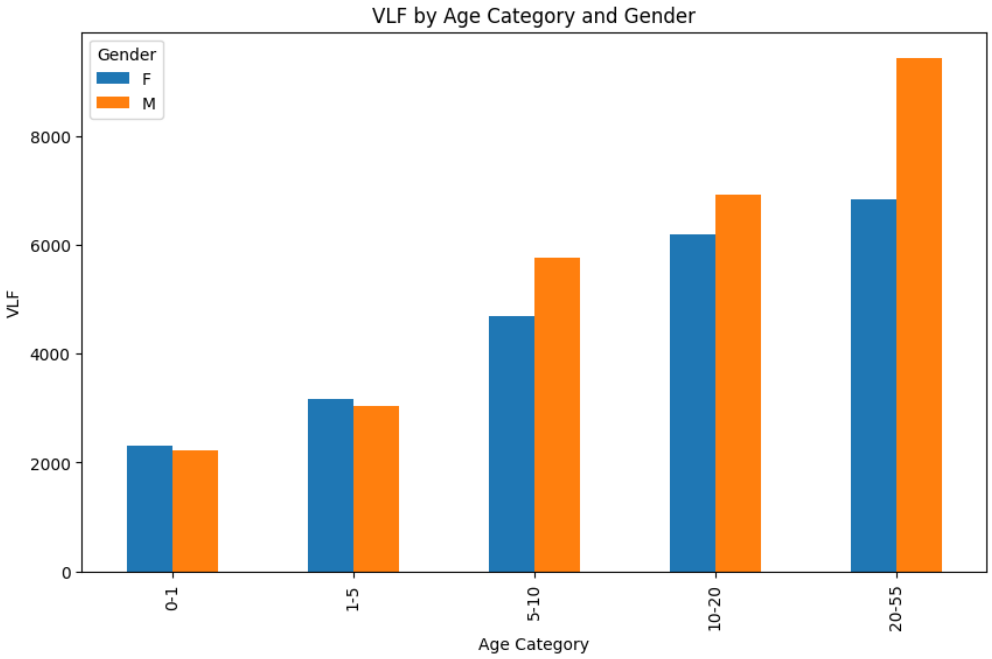

Figure 4.9. Comparision of VLF

4.3.2. LF

Low-Frequency values tend to increase with age, suggesting an increase in sympathetic and/or parasympathetic activity with age. However, males generally have higher LF values than females, particularly in the 10-20 and 20-55 age categories. This could suggest a higher level of sympathetic as well as parasympathetic activity in males compared to

females in these age groups. Moreover, LF values tend to increase with age, with the highest values observed in the 20-55 age category. This suggests that sympathetic as well as parasympathetic activity, as reflected by LF, may increase with age. This could be due to physiological changes associated with aging, such as changes in cardiovascular function and autonomic control. Figure 4.10 shows a comparison diagram of LF for age and gender.

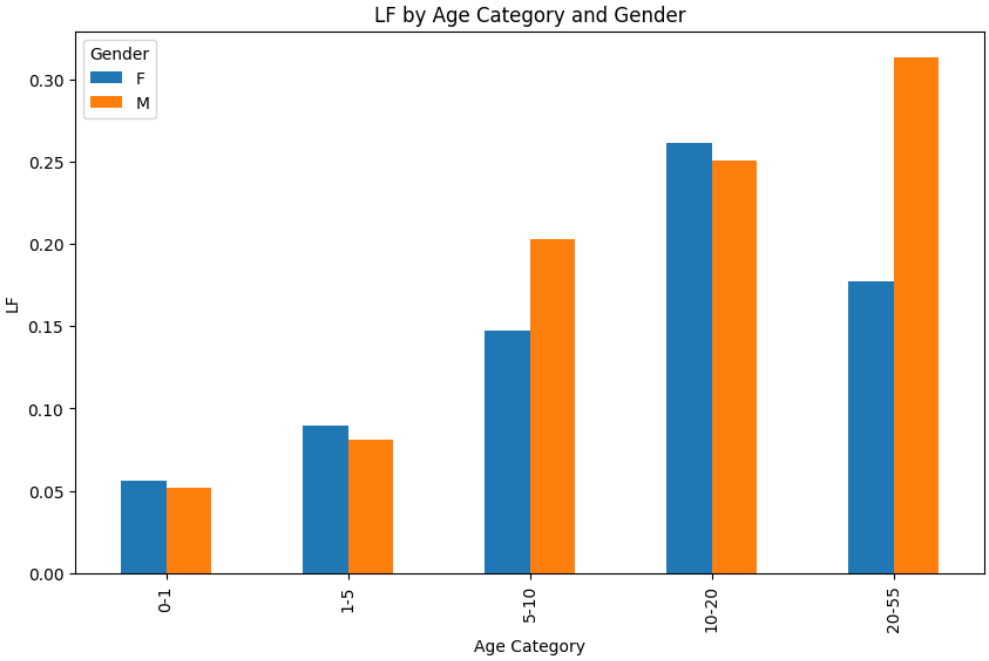

Figure 4.10. Comparision of LF

4.3.3. HF

High-Frequency values are generally higher in females than in males across all age categories, suggesting a higher level of parasympathetic activity in females. The difference is most pronounced in the 10-20 and 20-55 age categories, which could reflect gender differences in autonomic function during adolescence and adulthood. Moreover, HF values show a more complex pattern with age. They tend to be highest in the 1-5 and 20-55 age groups, suggesting that parasympathetic activity, as reflected by HF, may vary across different stages of life. This could reflect changes in autonomic function associated with development and aging. Figure 4.10 shows a comparison diagram of HF for age and gender.

4.3.4. LF/HF ratio

The LF/HF ratio is generally higher in males than in females across all age categories, suggesting a greater dominance of sympathetic activity relative to parasympathetic activity in males. This could reflect gender differences in the balance of autonomic function, with males potentially having a higher stress response compared to females. Moreover, the LF/HF ratio tends to increase with age, suggesting that the balance between sympathetic

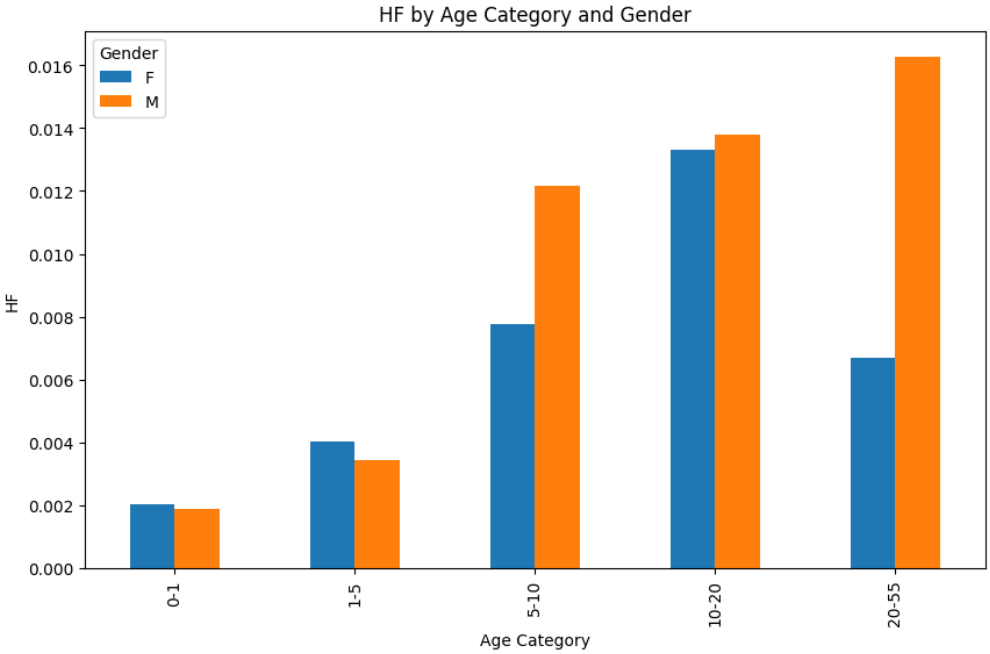

Figure 4.11. Comparision of HF

and parasympathetic activity may shift towards more sympathetic activity as individuals get older. This could reflect age-related changes in autonomic function, with a potential increase in the stress response and a decrease in the relaxation response with age. Figure 4.12 shows a comparison diagram of LF/RF ratio for age and gender.

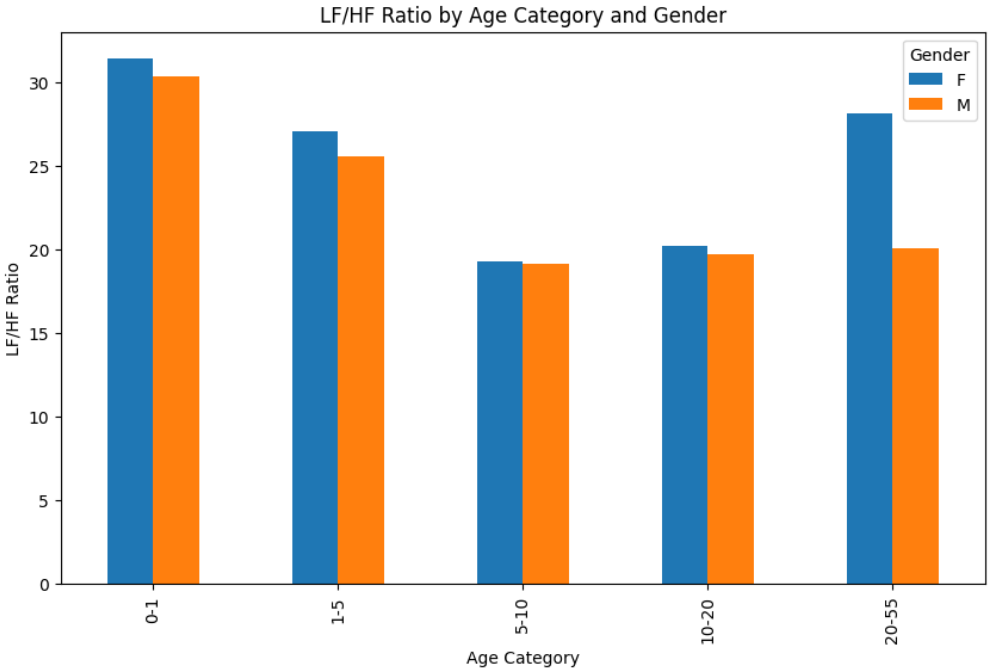

Figure 4.12. Comparision of LF/RF ratio

4.4. Polynomial Regression Test

The data analysis entails the utilization of a polynomial regression model, which is a variant of regression analysis, a statistical technique employed to establish a mathematical representation of the relationship between a dependent variable and one or more independent variables. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted E(y|x).

The application of Polynomial Regression to analyze the relationship between HRV variables and the independent variable 'Age'. The polynomial used possesses a degree of 7. The determination of the polynomial degree is established by evaluating the test outcomes across a range of degrees, specifically from 1 to 10. Between these options, the degree of 7 was selected due to its achieving the highest R-square value throughout all variables.

4.4.1. Mean RR

The polynomial equation is derived from the model. The equation possesses the capability of predicting the 'Mean RR' value for any specified 'Age' parameter. The scatter plot is in Figure 4.13

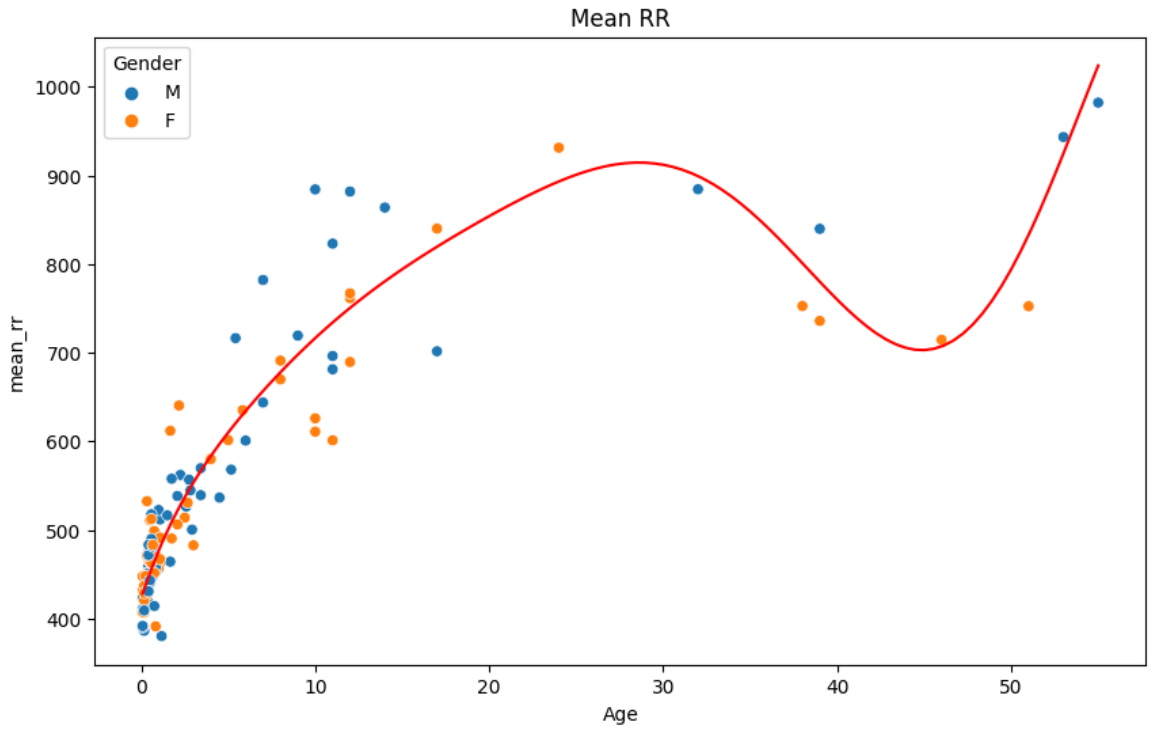

Figure 4.13. Polynomial regression of mean RR

The corrected polynomial regression line is in equation 13:

$$y = 423.32 + 60.49x - 8.10x^2 + 1.02x^3 - 0.08x^4 + 0.004x^5 - 0.0001x^6 + 0.000002x^7 \quad (13)$$

The R-squared value and p-value of the model are calculated. The coefficient of determination, indicated as R-squared, exhibits a value of 0.776047, which relatively high score. This demonstrates that the model explains a large proportion (approximately 77.60%) of the variability observed in the 'SDNN' variable. The obtained p-value is below the threshold of 0.05 ($p < 0.05$), indicating that the observed results possess statistical significance.

4.4.2. SDNN

The polynomial equation is derived from the model. The equation possesses the capability of predicting the 'SDNN' value for any specified 'Age' parameter. The scatter plot is in Figure 4.14.

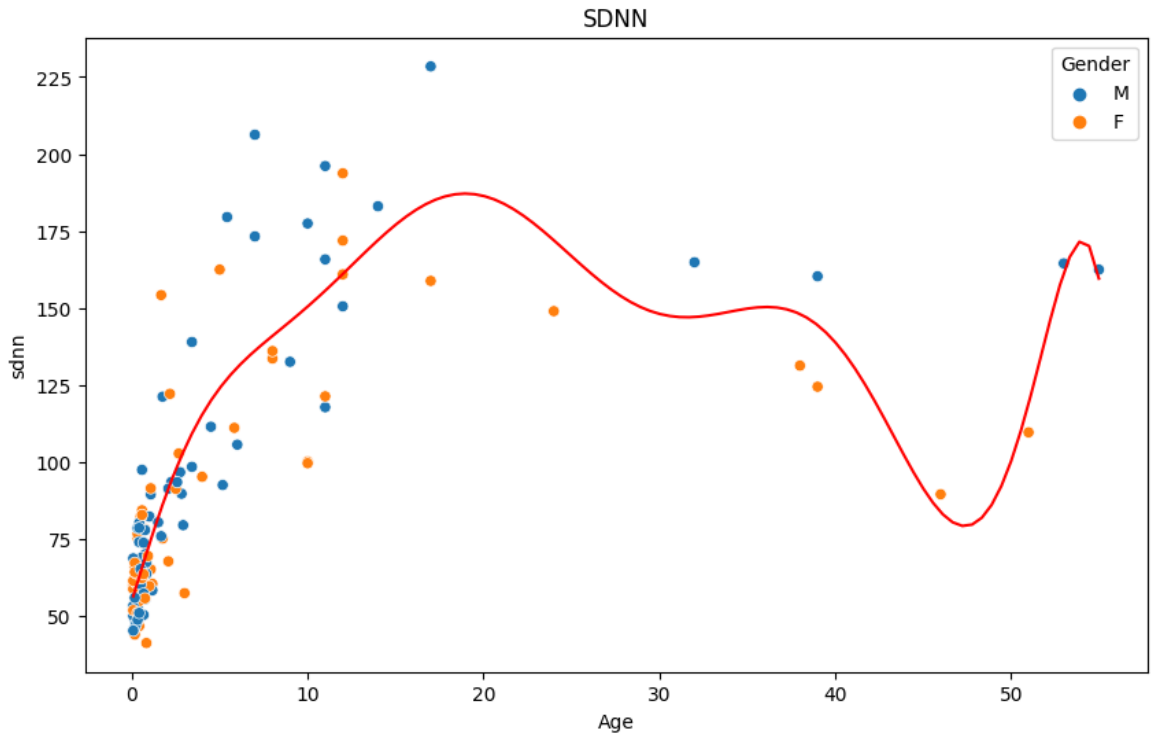

Figure 4.14. Polynomial regression of SDNN

The corrected polynomial regression line is in equation 14:

$$y = 54.90 + 18.57x + 0.33x^2 - 0.55x^3 + 0.08x^4 - 0.005x^5 + 0.0002x^6 - 0.000004x^7 \quad (14)$$

The R-squared value and p-value of the model are calculated. The coefficient of determination, indicated as R-squared, exhibits a value of 0.890831, which high score. This demonstrates that the model explains a large proportion (approximately 89.08%) of the variability observed in the 'SDNN' variable. The obtained p-value is below the threshold of 0.05 ($p < 0.05$), indicating that the observed results possess statistical significance.

4.4.3. RMSSD

The polynomial equation is derived from the model. The equation possesses the capability of predicting the 'RMSSD' value for any specified 'Age' parameter. The scatter plot is in Figure 4.15.

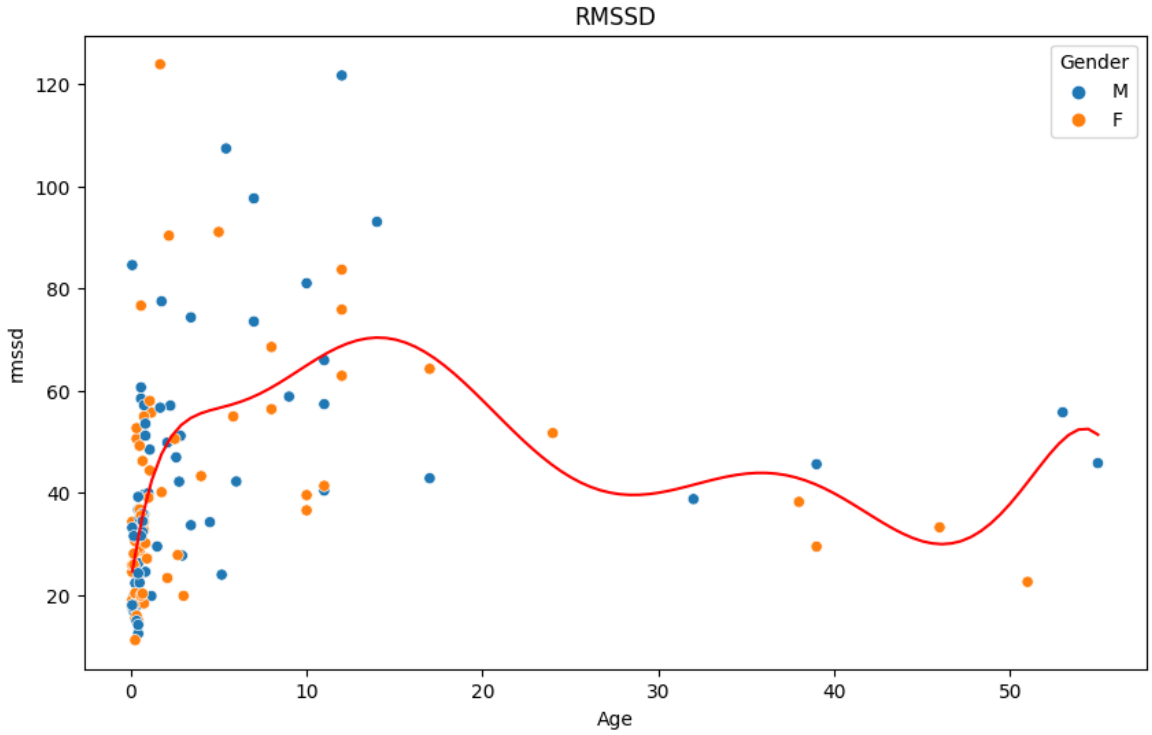

Figure 4.15. Polynomial regression of RMSSD

The corrected polynomial regression line is in equation 15:

$$y = 22.77 + 23.07x - 6.54x^2 + 0.95x^3 - 0.07x^4 + 0.003x^5 - 0.000087x^6 + 0.000001x^7 \quad (15)$$

The R-squared value and p-value of the model are calculated. The coefficient of determination, indicated as R-squared, exhibits a value of 0.347998, which relatively low score. This demonstrates that the model explains a weak proportion (approximately 34.80%) of the variability observed in the 'RMSDD' variable. The obtained p-value is below the threshold of 0.05 ($p < 0.05$), indicating that the observed results possess statistical significance.

4.4.4. pNN50

The polynomial equation is derived from the model. The equation possesses the capability of predicting the 'pNN50' value for any specified 'Age' parameter. The scatter plot is in Figure 4.16.

The corrected polynomial regression line is in equation 16:

$$y = 1.50 + 2.83x + 0.78x^2 - 0.22x^3 + 0.02x^4 - 0.001x^5 + 0.000057x^6 - 0.000001x^7 \quad (16)$$

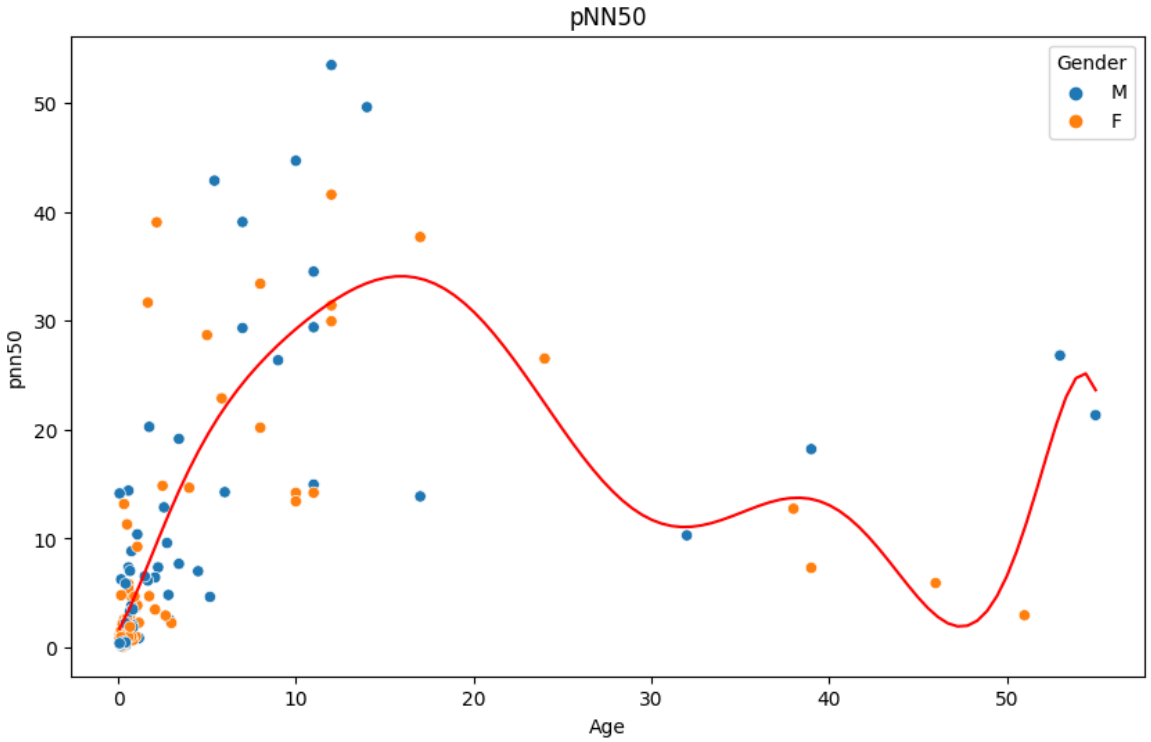

Figure 4.16. Polynomial regression of pNN50

The R-squared value and p-value of the model are calculated. The coefficient of determination, indicated as R-squared, exhibits a value of 0.646472, which relatively high score. This demonstrates that the model explains a strong proportion (approximately 64.65%) of the variability observed in the 'pNN50' variable. The obtained p-value is below the threshold of 0.05 ($p < 0.05$), indicating that the observed results possess statistical significance.

4.4.5. NN50

The polynomial equation is derived from the model. The equation possesses the capability of predicting the 'NN50' value for any specified 'Age' parameter. The scatter plot is in Figure 4.17.

The corrected polynomial regression line is in equation 17:

$$y = y = 2959.99 + 3974.87x + 726.47x^2 - 231.45x^3 + 25.42x^4 - 1.50x^5 + 0.05x^6 - 0.001x^7 \quad (17)$$

The R-squared value and p-value of the model are calculated. The coefficient of determination, indicated as R-squared, exhibits a value of 0.576404, which moderate score. This demonstrates that the model explains a moderate proportion (approximately 57.64%) of the variability observed in the 'NN50' variable. The obtained p-value is below the threshold of 0.05 ($p < 0.05$), indicating that the observed results possess statistical significance.

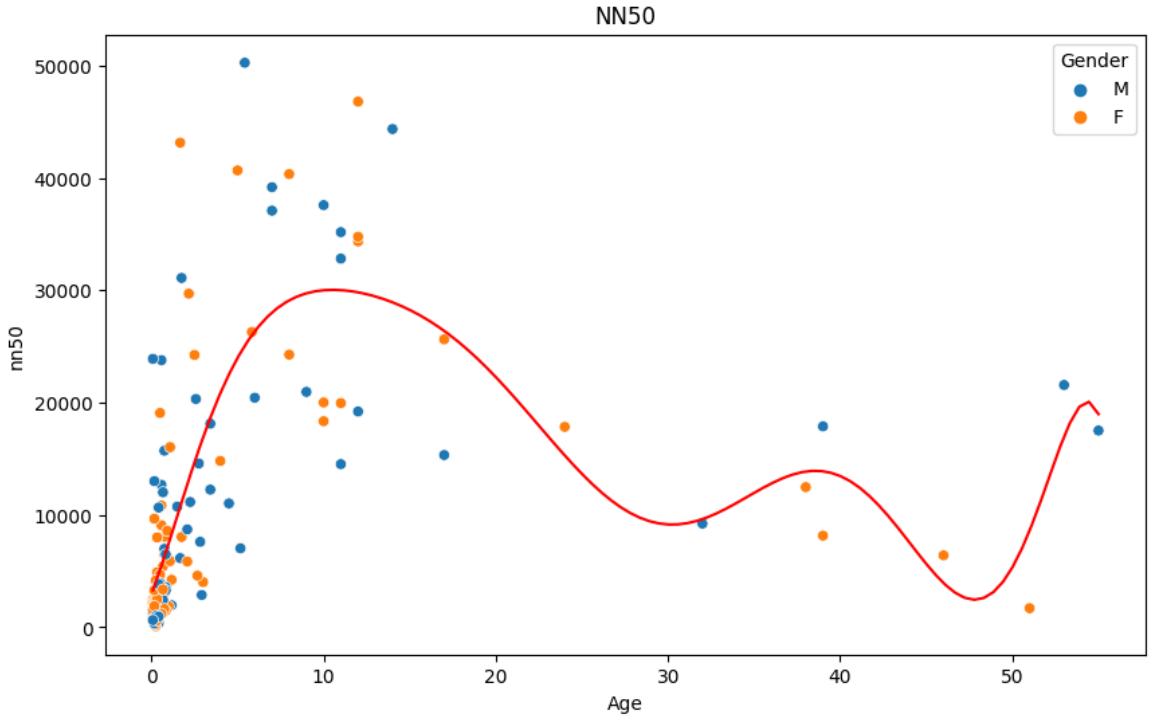

Figure 4.17. Polynomial regression of NN50

4.4.6. VLF

The polynomial equation is derived from the model. The equation possesses the capability of predicting the 'VLF' value for any specified 'Age' parameter. The scatter plot is in Figure 4.18.

The corrected polynomial regression line is in equation 18:

$$y = 2052.99 + 449.52x + 20.23x^2 - 7.01x^3 + 0.72x^4 - 0.039x^5 + 0.001x^6 - 0.000023x^7 \quad (18)$$

The R-squared value and p-value of the model are calculated. The coefficient of determination, indicated as R-squared, exhibits a value of 0.890367, which high score. This demonstrates that the model explains a very strong proportion (approximately 89.04%) of the variability observed in the 'VLF' variable. The obtained p-value is below the threshold of 0.05 ($p < 0.05$), indicating that the observed results possess statistical significance.

4.4.7. LF

The polynomial equation is derived from the model. The equation possesses the capability of predicting the 'LF' value for any specified 'Age' parameter. The scatter plot is in Figure 4.19.

The corrected polynomial regression line is in equation 19:

$$y = 0.05 + 0.02x - 0.0006x^2 + 0.0003x^3 - 0.00003x^4 + 0.000002x^5 - 0.00x^6 + 0.00x^7 \quad (19)$$

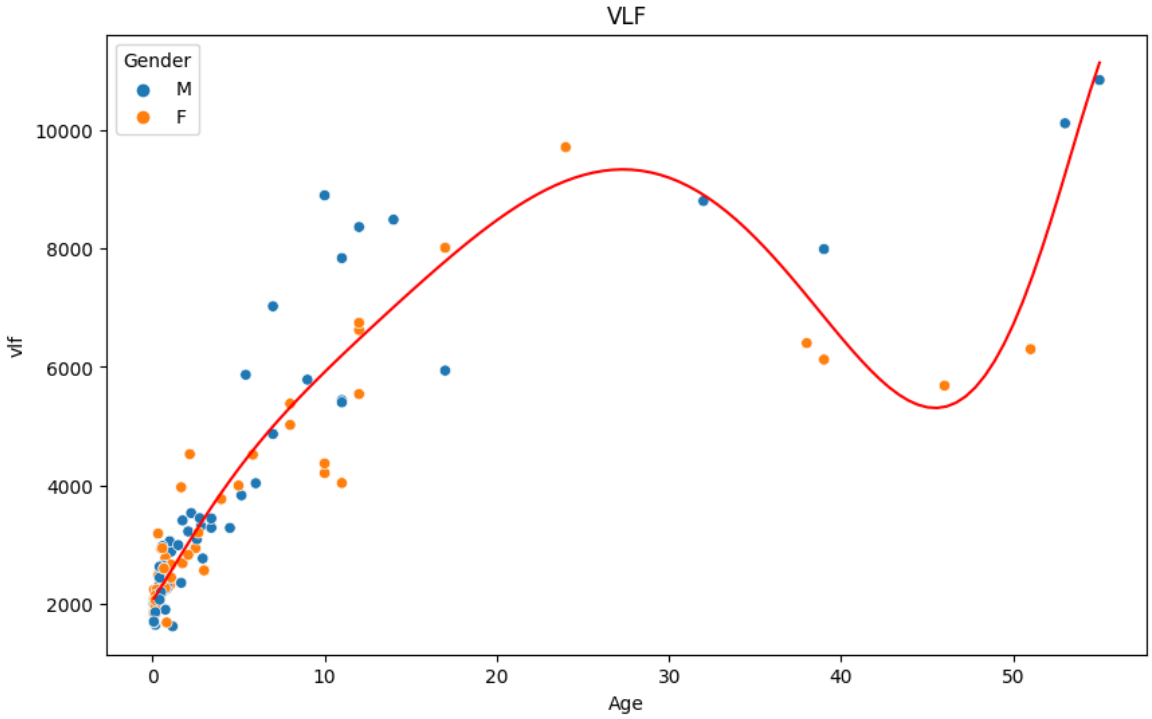

Figure 4.18. Polynomial regression of VLF

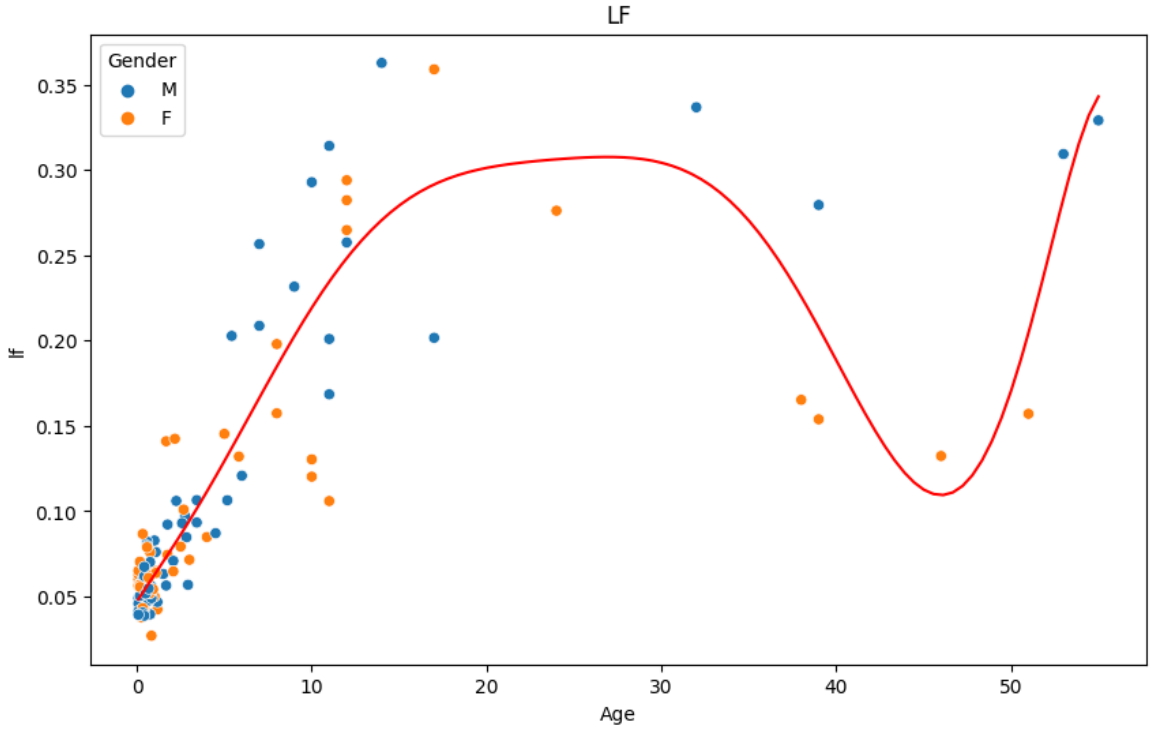

Figure 4.19. Polynomial regression of LF

The R-squared value and p-value of the model are calculated. The coefficient of determination, indicated as R-squared, exhibits a value of 0.840870, which high score. This

demonstrates that the model explains a very strong proportion (approximately 84.09%) of the variability observed in the 'LF' variable. The obtained p-value is below the threshold of 0.05 ($p < 0.05$), indicating that the observed results possess statistical significance.

4.4.8. HF

The polynomial equation is derived from the model. The equation possesses the capability of predicting the 'HF' value for any specified 'Age' parameter. The scatter plot is in Figure 4.20.

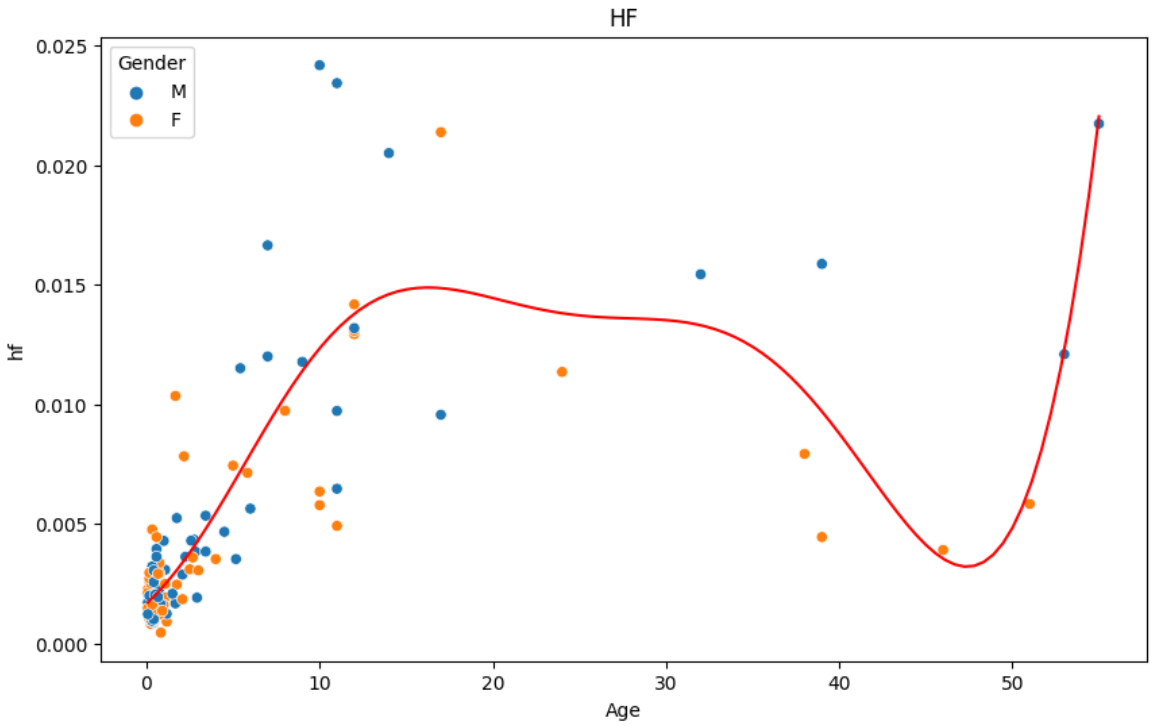

Figure 4.20. Polynomial regression of HF

The corrected polynomial regression line is in equation 20:

$$y = 0.001 + 0.0007x + 0.00004x^2 + 0.00001x^3 - 0.000002x^4 + 0.00x^5 - 0.00x^6 + 0.00x^7 \quad (20)$$

The R-squared value and p-value of the model are calculated. The coefficient of determination, indicated as R-squared, exhibits a value of 0.730289, which relatively high score. This demonstrates that the model explains a strong proportion (approximately 73.03%) of the variability observed in the 'HF' variable. The obtained p-value is below the threshold of 0.05 ($p < 0.05$), indicating that the observed results possess statistical significance.

4.4.9. LF/HF ratio

The polynomial equation is derived from the model. The equation possesses the capability of predicting the 'LF/HF ratio' value for any specified 'Age' parameter. The scatter plot is in Figure 4.21.

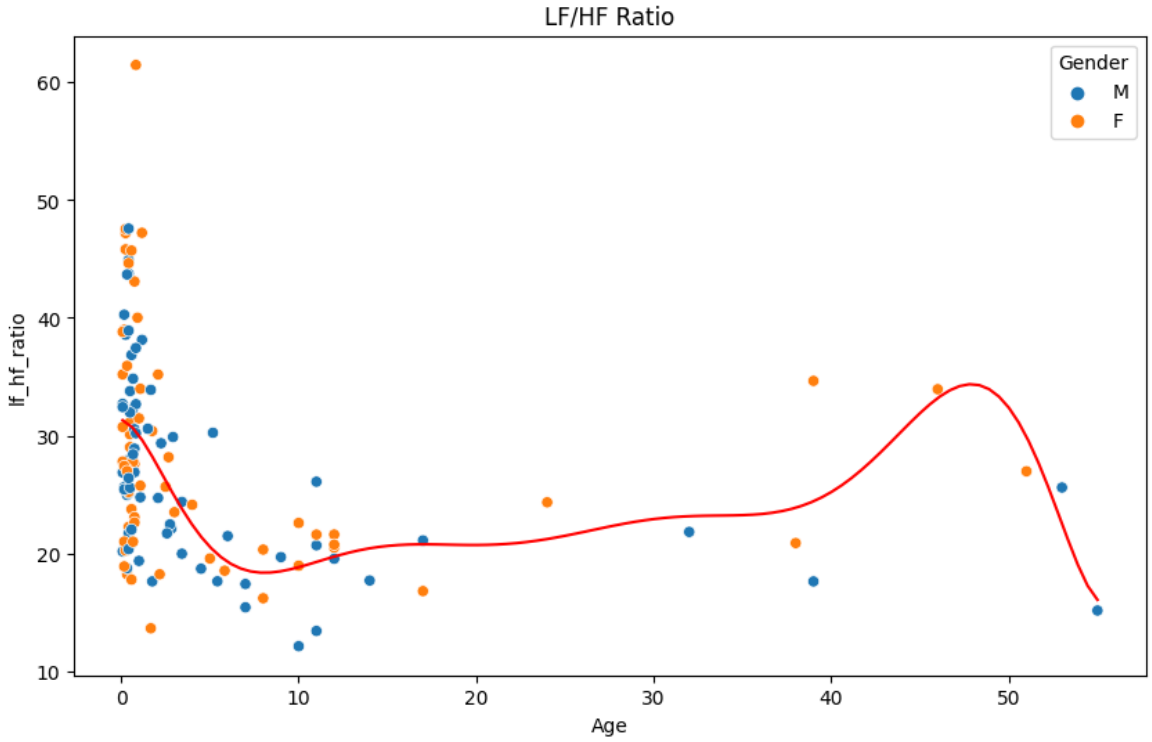

Figure 4.21. Polynomial regression of LF/HF ratio

The corrected polynomial regression line is in equation 21:

$$y = 31.34 - 0.34x - 1.24x^2 + 0.28x^3 - 0.03x^4 + 0.001x^5 - 0.00005 * x^6 + 0.000001x^7 \quad (21)$$

The R-squared value and p-value of the model are calculated. The coefficient of determination, indicated as R-squared, exhibits a value of 0.277415, which relatively low score. This demonstrates that the model explains a weak proportion (approximately 27.74%) of the variability observed in the 'LF/HF ratio' variable. The obtained p-value is below the threshold of 0.05 ($p < 0.05$), indicating that the observed results possess statistical significance.

4.4.10. Conclusion of polynomial regression test results

The Polynomial Regression models of degree 9 were able to fit the data well for most of the HRV indicators ('Mean RR', 'SDNN', 'RMSSD', 'NN50', 'pNN50', 'LF', 'HF'), with high R-squared values indicating that a significant proportion of the variance in these indicators can be explained by 'Age'. This suggests that 'Age' is a significant predictor of these HRV indicators. The conclusion of the results can be seen in Table 4.4.

However, for the 'LF/HF Ratio' HRV indicator, the Polynomial Regression model of degree 7 had a lower R-squared value, indicating that 'Age' explains a smaller proportion of the variance in 'LF/HF Ratio'. This suggests that 'Age' is a less significant predictor of

Table 4.4. Modeling results

HRV parameter	Model	R-square	P-value
Mean RR	7	0.776047	statistical significance
SDNN	7	0.890831	statistical significance
RMSSD	7	0.347998	statistical significance
pNN50	7	0.646472	statistical significance
NN50	7	0.576404	statistical significance
VLF	7	0.890367	statistical significance
LF	7	0.840870	statistical significance
HF	7	0.730289	statistical significance
LF/HF ratio	7	0.277415	statistical significance

'LF/HF Ratio', and other variables or a different type of model may be needed to better predict this indicator.

While the models were statistically significant with p-values less than 0.05, statistical significance does not necessarily imply practical significance. It's important to also consider the effect sizes and the practical implications of the findings.

5. Conclusion

5.1. Summary of Findings

5.1.1. HRV Analysis

Based on the comprehensive analysis and visualizations conducted on the HRV parameters with respect to gender and age, the following conclusions can be drawn:

1. Gender Differences in HRV Parameters: There are significant differences in HRV parameters between males and females. These differences may reflect variations in the autonomic nervous system activity between the two genders. Females generally exhibit higher HRV parameters, indicating higher parasympathetic activity. This could be attributed to hormonal differences between males and females, which might influence the autonomic regulation of heart rate.
2. Age-related Changes in HRV Parameters: HRV parameters tend to decrease with increasing age, indicating a decline in heart rate variability and parasympathetic activity. However, the LF/HF ratio, an indicator of the balance between sympathetic and parasympathetic activity, tends to increase with age, suggesting a relative increase in sympathetic activity compared to parasympathetic activity. This could be due to the physiological changes that occur with aging, including a decrease in the elasticity of the cardiovascular system and changes in the autonomic regulation of heart rate.

HRV parameters are important indicators of cardiovascular health. Lower HRV parameters are associated with various cardiovascular diseases and conditions, including hypertension, coronary artery disease, and heart failure. Therefore, the observed gender differences and age-related changes in HRV parameters could have significant implications for cardiovascular health and disease risk.

5.1.2. Polynomial Regression

The Polynomial Regression models of degree 7 were able to fit the data well for most of the HRV indicators ('Mean RR', 'SDNN', 'RMSSD', 'NN50', 'pNN50', 'LF', 'HF'), with high R-squared values indicating that a significant proportion of the variance in these indicators can be explained by 'Age'. This suggests that 'Age' is a significant predictor of these HRV indicators.

However, for the 'LF/HF Ratio' HRV indicator, the Polynomial Regression model of degree 7 had a lower R-squared value, indicating that age explains a smaller proportion of the variance in 'LF/HF Ratio'. This suggests that age is a less significant predictor of 'LF/HF Ratio', and other variables or a different type of model may be needed to better predict this indicator.

Given these findings, further research is recommended to improve the models and gain a deeper understanding of the HRV indicators. This could involve incorporating more variables into the models, trying different model types, using cross-validation for model

selection, performing a longitudinal analysis, expanding the dataset, doing a deep dive into the 'LF/HF Ratio' indicator, and collaborating with domain experts.

By pursuing these research directions, we can develop more accurate and robust models for predicting HRV indicators, which could have important implications for understanding heart health and guiding interventions to improve cardiovascular outcomes.

5.2. Limitations of the Study

This analysis has several limitations that should be considered in future research. The analysis was conducted using a single dataset, and as such, the implications of the findings may not apply to broader populations. Moreover, the analysis did not take into account other variables that could influence heart rate variability parameters, such as physical activity level, tension and lifestyle factors. Further investigation is warranted to examine the impact of these variables on HRV indicators, as well as their correlation with gender and age.

The use of a 7th-degree polynomial for regression analysis can lead to overfitting. Overfitting occurs when the model fits the training data too closely and may not perform well on new, unseen data. It's always a good idea to validate the model using a separate test dataset or cross-validation techniques to ensure that the model generalizes well to new data.

The analysis outcomes can be influenced by the quality of the data as well. The presence of errors, outliers, or missing values within the dataset has the potential to impact the precision and reliability of the models. Conducting data cleaning and preprocessing procedures prior to analysis is crucial in order to guarantee the accuracy and reliability of the data.

The size of the dataset can also have an influence on the outcomes of the analysis. Lacking dataset size could limit the models' capability to accurately represent the connections among variables. However, in the case that the dataset is of substantial terms of size, the execution of the analysis may incur significant computational expenses.

While the models were statistically significant with p-values less than 0.05, statistical significance does not necessarily imply practical significance. It's important also to consider the effect sizes and the practical implications of the findings.

5.3. Suggestions for Future Research

The results of this analysis underscore the significance of including gender and age differences in investigations on the relationship between heart rate variability and cardiovascular health. Further research should strive to provide a deeper knowledge of the mechanisms that support these disparities and their potential impacts on the possibility of cardiovascular disease and the development of treatment methodologies.

While Polynomial Regression models were used in this analysis, other types of models might be more appropriate for certain indicators. For example machine learning models (such as decision trees, random forests, or neural networks) could be explored. Comparing the performance of different model types could help identify the most suitable model for each HRV indicator.

Enhancing the dataset by including additional observations or diversifying the populations under study has the potential to enhance the generalizability of the results. This endeavor may entail the acquisition of additional data or the integration of the existing dataset with other similar datasets.