

Time series in Healthcare(ECG)

Naman Modi(202101421),^{*} Bhavya Shah(202101426),[†] and Vedant Shah(202101507)[‡]

Dhirubhai Ambani Institute of Information & Communication Technology,

Gandhinagar, Gujarat 382007, India

SC475, Time Series and Analysis

In this project, we will examine the time series data of heart rate(ECG) from the publicly available dataset of MIT-BIH Arrhythmia and PhysioNet mirrors worldwide over 4 years. Our goal would be to understand the provided time series data by examining different characteristics like autocorrelation, auto covariance and stationarity of the data. We would look into which time domain models i.e., AR, MA, or ARMA—fit the data the best.

I. INTRODUCTION

In this project, we set out to explore the Electrocardiogram signals using MIT-BIH Arrhythmia Database and use time series analysis to uncover the complex relationships of ECG waveforms. We pursue two objectives: to discover the temporal aspects of ECG data, and extract trends, seasonality and make it stationary to perform the prediction on the data.

II. DATA DESCRIPTION

The MIT-BIH Arrhythmia Database contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979.

The recordings were digitized at 360 ticks per second per channel with 11-bit resolution over a 10 mV range. Two or more cardiologists independently annotated each record; disagreements were resolved to obtain the computer-readable reference annotations for each beat (approximately 110,000 annotations in all) included with the database. [1]

III. DATA INTERPRETATION

Firstly, we started studying what are ECG signals and understood its features. They are [2]

- **Details :** We observed the patient, gender and data/time

- Patient Name : Aldomet, 69 M
- Patient Name : Inderal, 24 F
- Patient Name : Digoxin, 72 M
- Patient Name : Pronestyl, 51 F

- **Rate :** The normal resting heart rate for a healthy person is between 60 and 100 beats per minute. A resting heart rate of fewer than 60 beats per minute (bpm) is termed Bradycardia, while a rate exceeding 100 bpm is categorized as Tachycardia. If left untreated, both conditions can pose significant medical risks and require prompt attention.

The heart rate of all the observed patients lies between 60 and 100 bpm.

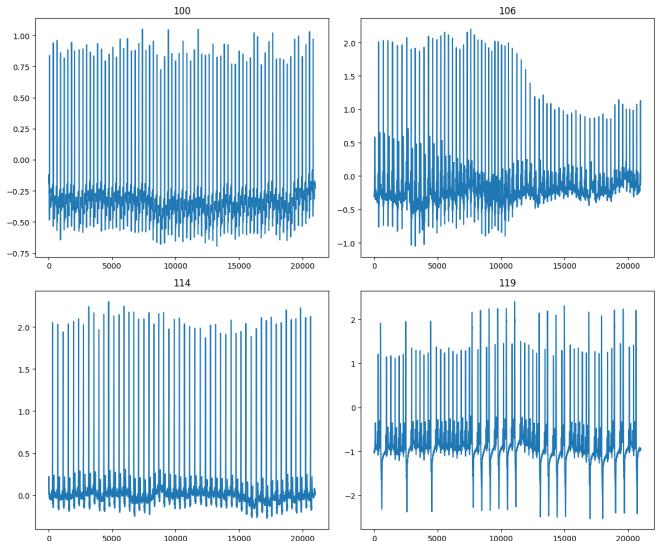


FIG. 1: ECG signals of the patients. Here, the Clock frequency is 360 Hz. y-axis is in μ V and x-axis is t(s) with 360ticks/sec

- **Rhythm :** The rhythm of the heart can be determined by various factors. We have tried to analyse and study some of them.

R-R interval: The RR interval refers to the distance between two consecutive R waves. The R wave represents the depolarization of the ventricles of the heart, which is the electrical signal that precedes each heartbeat. By measuring the time between R waves, clinicians can assess the regularity and rhythm of the heart, diagnose abnormalities such as arrhythmias, and determine the heart rate variability, which can provide insights into cardiac health and autonomic nervous system function.

^{*}Electronic address: 202101421@daiict.ac.in

[†]Electronic address: 202101426@daiict.ac.in

[‡]Electronic address: 202101507@daiict.ac.in

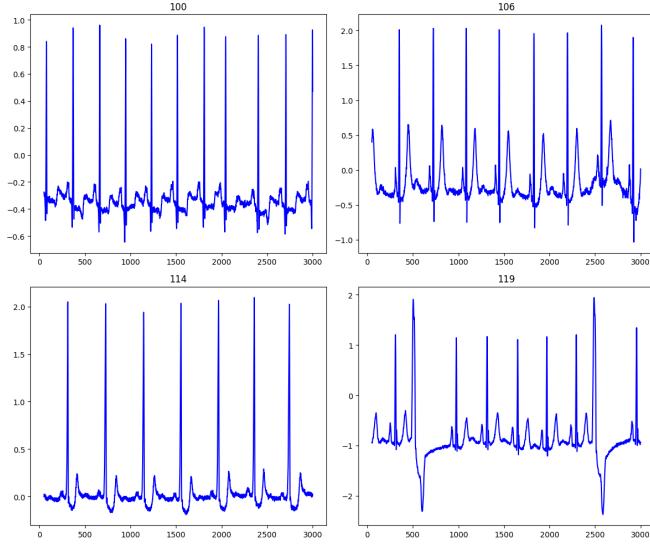


FIG. 2: R-R interval of patients. The R-R interval of first 3 patients are regular over the period and shows normal features but the patient 119 has irregularly irregular R-R interval with no clear pattern. The R-R interval is the difference between two peaks in the graph. y-axis is in μ V and x-axis is t(s) with 360ticks/sec

P-Wave: The P-wave is a signal that results from the atrial depolarization that stimulates atrial contraction. It signifies the start of the heartbeat and indicates the proper functioning of the SA node which is the heart's pacemaker. The duration of a normal P wave ranges from 60 to 120 milliseconds.

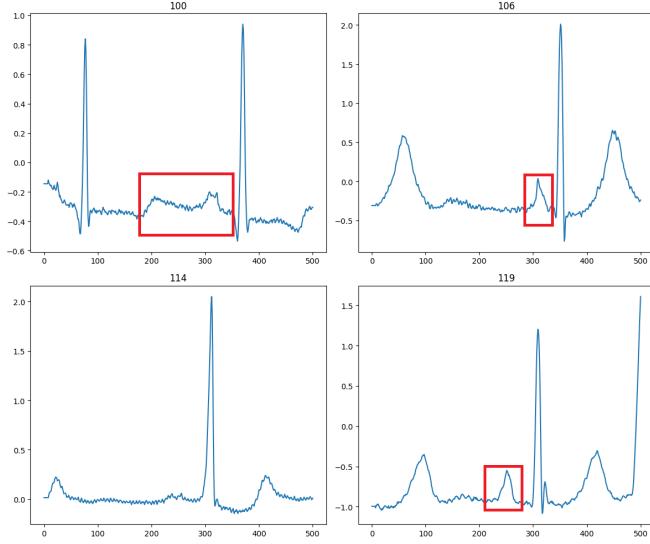


FIG. 3: P-waves of patients. The P-wave of the patient labeled 100 has an interval greater than 120ms (Dilated Left Atrium - P Mitrate), and for the patient labeled 114, the P-wave is missing (Atrial Fibrillation). Other patients have normal P-wave. y-axis is in μ V and x-axis is t(s) with 360ticks/sec

QRS interval: The QRS complex represents ventricular depolarization, which is the electrical signal that causes the ventricles (lower chambers of the heart) to contract. Its significance lies in indicating the initiation of ventricular contraction, which is essential for pumping blood to the lungs and the

rest of the body. The duration of the QRS complex is typically less than 0.12 seconds (120 milliseconds). A widened QRS complex may indicate bundle branch block, ventricular conduction delays, or ventricular hypertrophy.

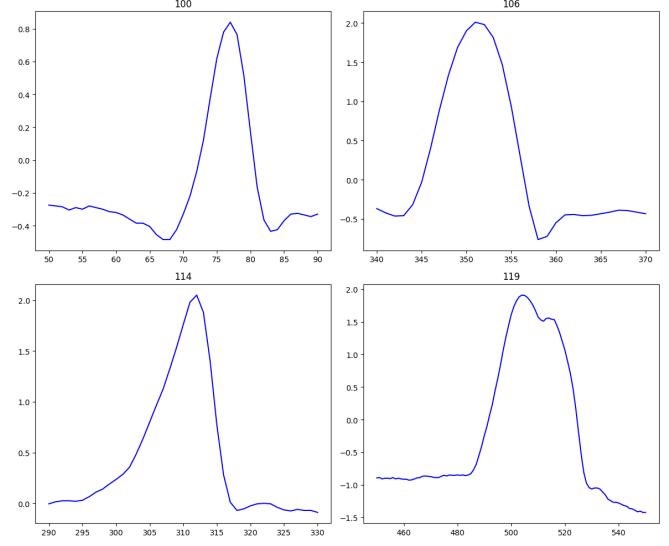


FIG. 4: This are the width of QRS interval of patient. The patient labelled 100,106,114 are having the interval length less than 120ms. The first patient has interval length of 55ms, second patient has length of 45ms, third patient has interval of 111ms whereas the last patient has widened QRS complex of approximately 145ms. y-axis is in μ V and x-axis is t(s) with 360ticks/sec

IV. DATA ANALYSIS

A. Frequency Distribution plots

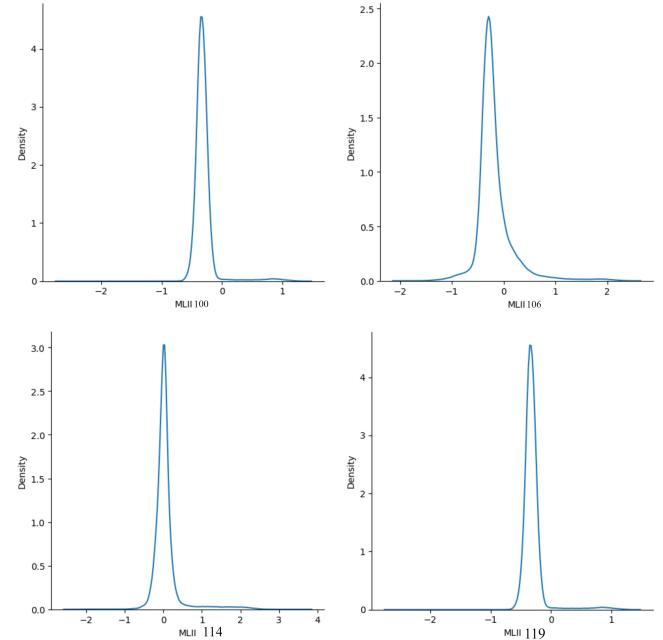


FIG. 5: The data of all the observed patients follow a normal distribution

B. Stationarity Test

To check the stationarity of our timeseries, we performed two different tests which are as follows:

1. KPSS Test

It is used to test for the null hypothesis that the data is stationary around a deterministic trend against the alternative of a unit root. If the p-value is less than a chosen significance level (e.g., 0.05), then the null hypothesis of stationarity is rejected, suggesting the presence of a unit root and non-stationarity.

KPSS Test Results:

- For data labeled 100: p-value: 0.01
- For data labeled 106: p-value: 0.01
- For data labeled 114: p-value: 0.01
- For data labeled 119: p-value: 0.01

The data is non-stationary due to a deterministic trend, as the p-value is less than 0.05 for all the datasets.

2. ADF Test

It tests the null hypothesis that a unit root is present in a time series sample. If the p-value is less than a chosen significance level (e.g., 0.05), then the null hypothesis of a unit root is rejected, suggesting stationarity.

ADF Test Results:

- For data labeled 100: p-value: 0.964080
- For data labeled 106: p-value: 0.964080
- For data labeled 114: p-value: 0.964080
- For data labeled 119: p-value: 0.964080

The data is non-stationary due to a deterministic trend, as the p-value is greater than 0.05 for all the datasets.

C. Detrending and removing seasonality

Now, we have started performing transformation operators such as differencing, moving averages, and polynomial fitting to remove the trend from the data. Following were the results for the trend, which we eventually subtracted from the original data to de-trend it.

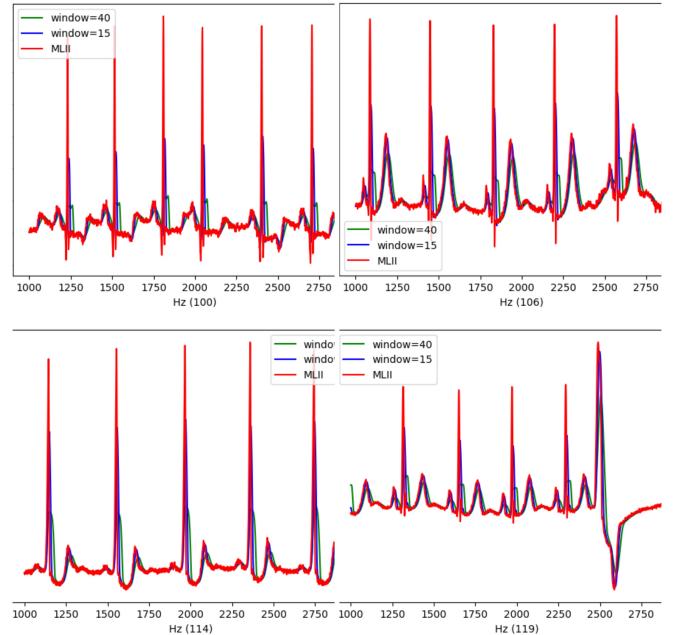


FIG. 6: Fitting Moving average process of order 15 and 40 on the four datasets.

KPSS Test Results for MA process:

- For MA(15) and MA(40) on data 100: p-value: 0.1 and 0.1
- For MA(15) and MA(40) on data 106: p-value: 0.1 and 0.1
- For MA(15) and MA(40) on data 114: p-value: 0.1 and 0.1
- For MA(15) and MA(40) on data 119: p-value: 0.1 and 0.1

ADF Test Results for MA process:

- For MA(15) and MA(40) on data 100: p-value: $2.61 * 10^{-8}$ and $5.13 * 10^{-10}$
- For MA(15) and MA(40) on data 106: p-value: 0.000004 and $1.49 * 10^{-11}$
- For MA(15) and MA(40) on data 114: p-value: $1.11 * 10^{-10}$ and $5.89 * 10^{-11}$
- For MA(15) and MA(40) on data 119: p-value: 0.000025 and $5.16 * 10^{-8}$

Both the tests show that Moving average processes are stationary for the datasets.

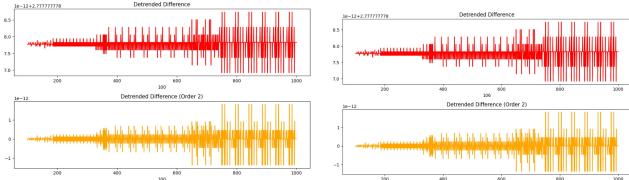


FIG. 7: The trend in the data can be seen in the plots of de-trended difference. The results are evident for orders 1 and 2, but the model does not give proper results for further orders.

Once the trend was removed, we used the seasonal decomposition library to remove the seasonal part from detrended data to obtain the residual on which all the remaining analyses will be performed.

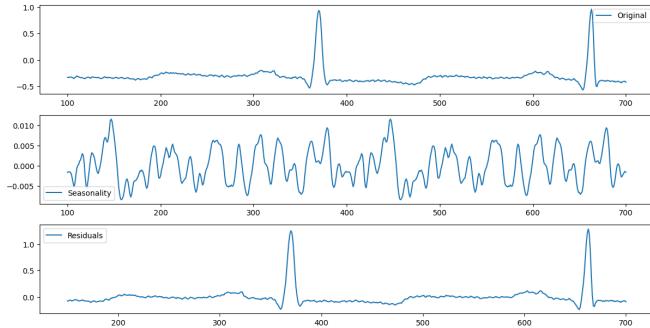


FIG. 8: The decomposition for the detrended data for patient labeled 100

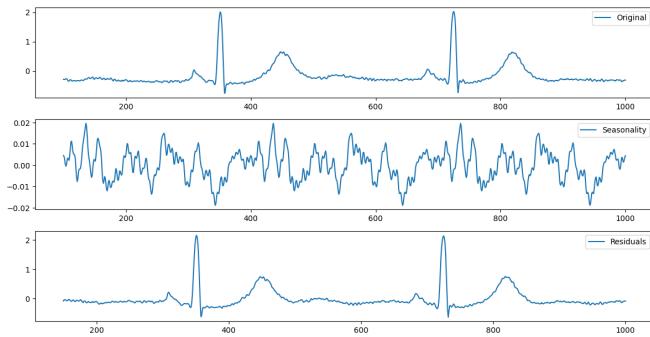


FIG. 9: The decomposition for the detrended data for patient labeled 106

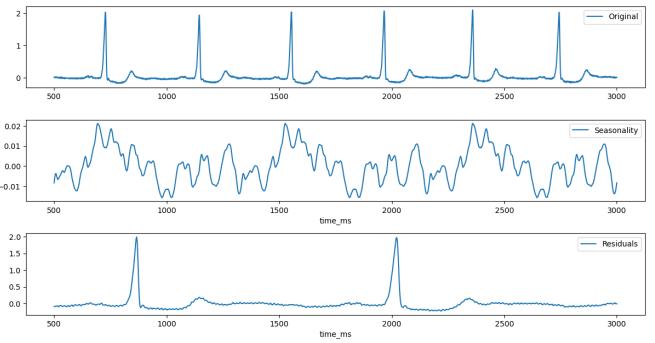


FIG. 10: The decomposition for the detrended data for patient labeled 114

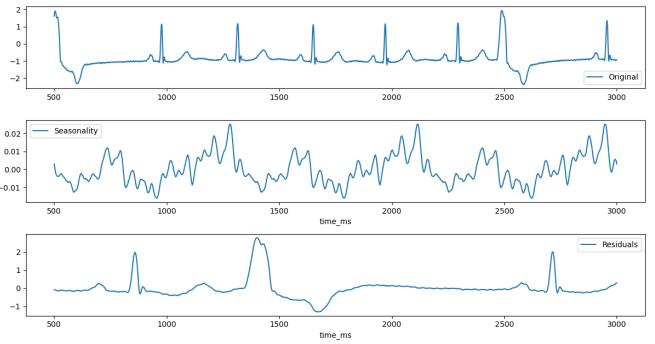


FIG. 11: The decomposition for the detrended data for patient labeled 119

The graphs show the seasonality in the detrended data with 300 units between two R-R intervals. Even though the graphs' magnitude of seasonality was insignificant, similar patterns existed in each patient.

D. Residuals

We obtained these final results after removing the trend and seasonality from the data and making it stationary.

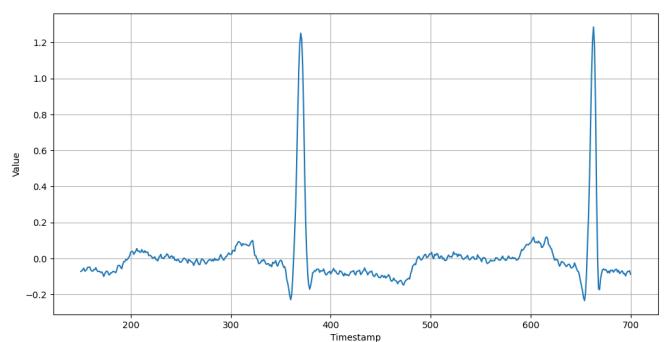


FIG. 12: Residual for patient labeled 100

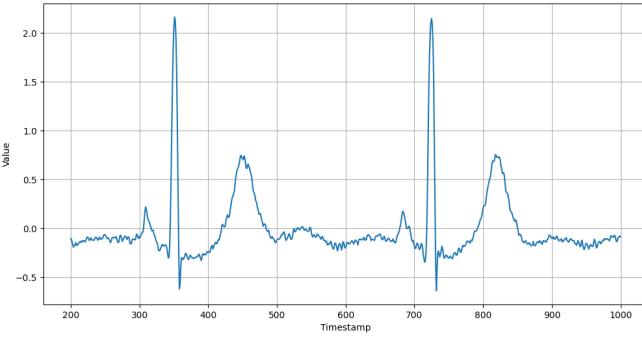


FIG. 13: Residual for patient labeled 106

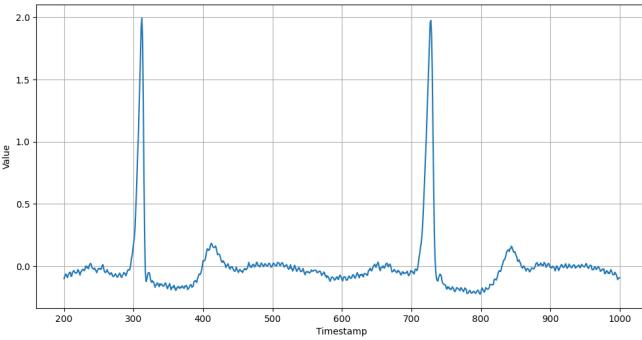


FIG. 14: Residual for patient labeled 114

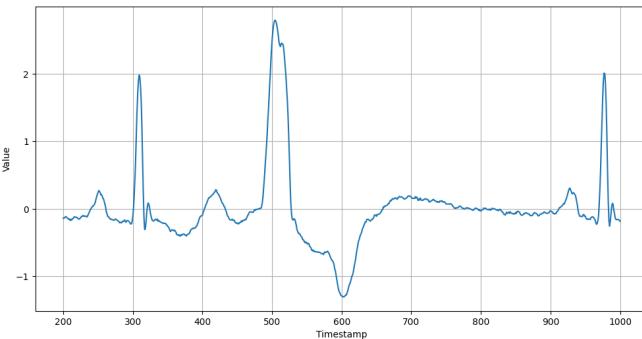


FIG. 15: Residual for patient labeled 119

E. Autocorrelation function

The ACF quantifies the relationship between a time series and its past observations at different lags. The ACF provides valuable information about the temporal structure of a time series, including seasonality and trend.

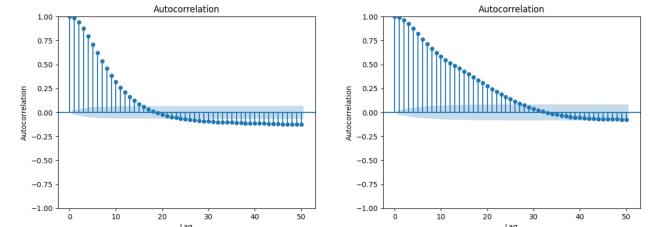
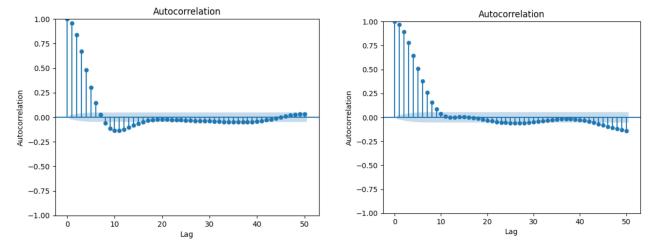


FIG. 16: autocorrelation function of residual for all the observed patients

According to the graphs, it can be seen that the autocorrelation function of residual depends upon lag and, hence, has become weak stationary.

V. MODEL FITTING AND VISUALIZATION

ARIMA models with different parameter combinations were applied to ECG data. These models were fitted on the residuals of each dataset. The visualizations can be seen below:

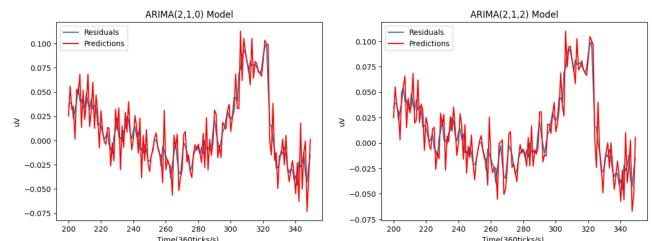
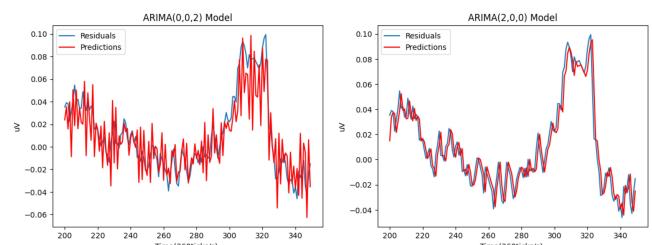


FIG. 17: Results of data 100

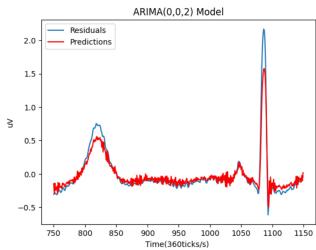


FIG. 18: Results of data 106

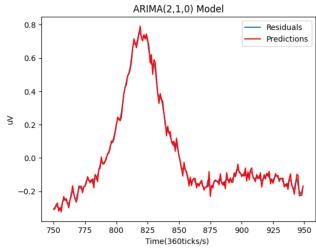
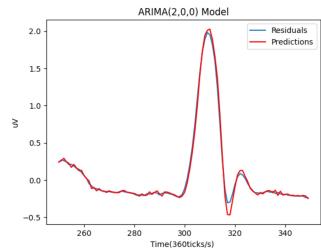
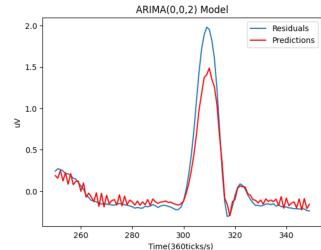
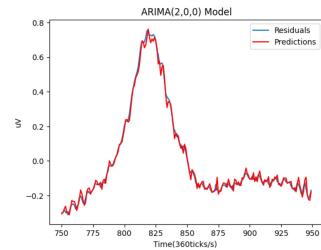


FIG. 19: Results of data 114

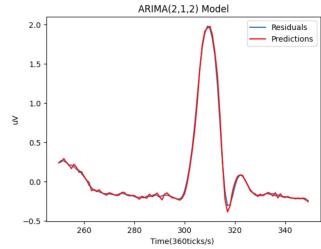
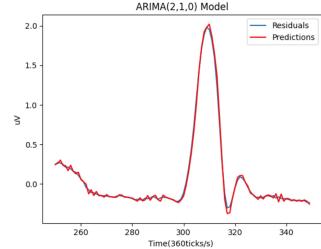
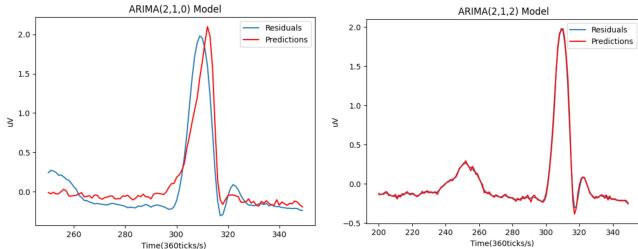
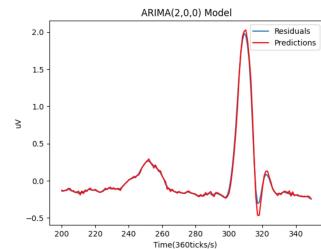


FIG. 20: Results of data 119

As seen in the above plots, ARIMA(2,1,2) model generally fits well to all the residuals.

VI. CONCLUSIONS

Throughout this time series project, we went deeply into the dynamics of the MIT-BIH Arrhythmia Database using sophisticated analytical methods to reveal significant information about cardiac electrophysiology. By examining stationarity, deterministic trends, and seasonality, we developed models that could model the fine details of ECG signals effectively. We used the auto-correlation function, stationarity tests and various transformations to dive deep into the data and study it using these analytical techniques.

Video link: [Video](#)

-
- [1] <https://archive.physionet.org/physiobank/database/html/mitdbdir/intro.htm>
[2] <https://youtu.be/u1m3HKW1VqU?si=8MKTms5D452LX8kX>.