Real time prediction of heart(Cardiovascular) disease using smart watch data and Short term HRV(Heart Rate Variability) analysis.

Submitted By

Raj Bhavsar

17MCEC03

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2019

Real time prediction of heart(Cardiovascular) disease using smart watch data and Short term HRV(Heart Rate Variability) analysis.

Major Project

Submitted in fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering (CSE)

Submitted By

Raj Bhavsar

(17MCEC03)

Guided By

Dr. Sanjay Garg

NIRMA UNIVERSITY
INSTITUTE OF TECHNOLOGY
NAAC ACCREDITED 'A' GRADE

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

Certificate

This is to certify that the Major Project entitled **"Real time prediction of heart (Cardiovascular) disease using smart watch data and Short term HRV(Heart Rate Variability) analysis."** submitted by **RAJ BHAVSAR (17MCEC03)**, towards the fulfillment of the requirements for the degree of Master of Technology in Computer Science and Engineering of Nirma University is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination.

Dr. Sanjay Garg
Guide and Professor,
Department of CSE,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Priyanka Sharma
Professor,
Coordinator M.Tech CSE,
Institute of Technology,
Nirma University, Ahmedabad

Dr. Madhuri Bhavsar
Professor and Head,
Department of CSE,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Alka Mahajan
Director,
Institute of Technology,
Nirma University, Ahmedabad

Statement of Originality

I, **Raj Bhavsar**, Roll. No. **17MCEC03**, give undertaking that the Major Project entitled "**Real time prediction of heart (Cardiovascular) disease using smart watch data and Short term HRV(Heart Rate Variability) analysis.**" submitted by me, towards the fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering (CSE)** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made.It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

____

Signature of Student

Date:

Place:

Endorsed by

Dr. Sanjay Garg

____

(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Dr. Sanjay Garg**, Professor, Computer Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Madhuri Bhavsar**, Hon'ble Head of Computer Engineering/ Information Technology Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr. Alka Mahajan**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation he has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

- Raj Bhavsar

17MCEC03

Abstract

Nowadays the Health-care systems are shifted from patient care to monitored care. A motive of this project is the Real-time prediction of Cardiovascular disease using smart watch data and Short term Heart Rate Variability analysis. Therefore here we use classification and regression machine learning algorithms on that data set in which we have implemented the regression algorithms on data set which is created using MI Band 3 with time laps of 10 minutes. Performance matrix for different algorithms is shown and rectify that RNN and SVR algorithm are suitable and giving good results and with respect to data-set, what is the size of data-set matters a lot. Here as we can see the size of the data-set is small so the models are not giving best fit.

Abbreviations

BPM	Beats per minute
SVR	Support Vector Regression
NIH	National Institutes of Health
HRV	Heart Rate Variability)
AHA	American Heart Association
CHF	Chronic heart failure

–

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Objective

Nowadays the Health-care systems are shifted from patient care to monitored care. A motive of this project is the Real-time prediction of Cardiovascular disease using smart watch data and Short term Heart Rate Variability analysis. therefor in this, we cover two modules first is prediction using smart watch data and prediction using the hospital's data of heart rate and other related parameters like age, Respiratory rate, Temperature, Blood Pressure.

1.2 A brief about Heart disease and Facts about the Heart rate

As we know the heart is very important organ of the body. Heart rate is nothing but the number of time per minute that heart contacts. Heart-rate is a vital indicator of health for the Human body. Heart rate is varied as a result of physical activity. but the normal heart rate of a human is in between 60bpm to 100bpm while he/she is relaxed and that heart rate is called resting heart rate. More about heart rate and abnormal heart-related facts are mentioned in chapter 4.

1.3 A brief about Fitness belts and Machine learning

Smart watch and fitness belts are recorded the heart rate, number of steps, sleeping pattern and Energy expenditure of the user. Hence the recorded data can be used for predicting the chances of heart-related disease which leads to heart failure using machine learning techniques. we use both classification and regression techniques for prediction which is elaborated in chapter 5.

Chapter 2

Motivation

In 2.1 the First Heart attack among Indians that occurs before a particular Age by Asian Indian Heart Disease is shown.

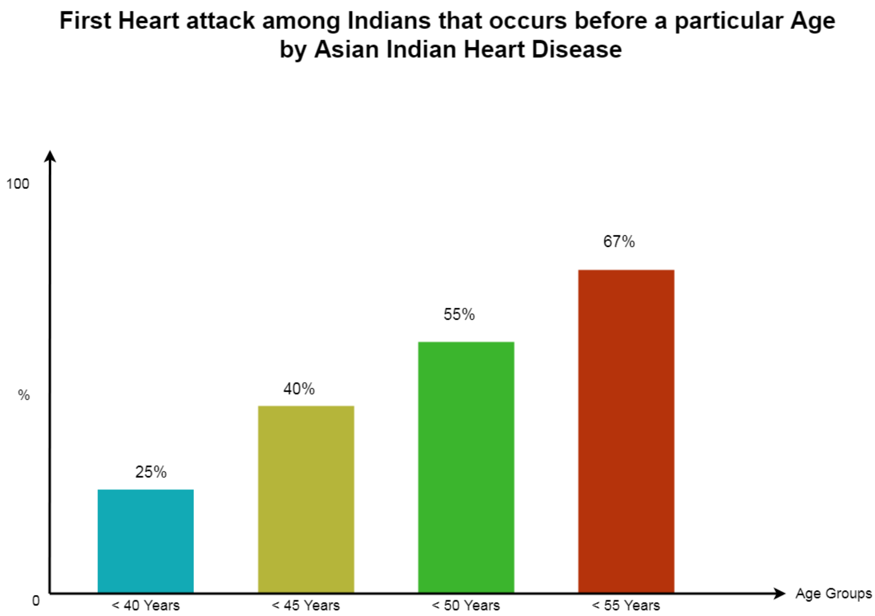

Figure 2.1: First Heart attack among Indians that occurs before a particular Age by Asian Indian Heart Disease

Nowadays the Health-care systems are shifted from patient care to monitored care. A motive of this project is the Real-time prediction of Cardiovascular disease using smart watch data and Short term Heart Rate Variability analysis.

Chapter 3

Literature Survey

3.1 Case-study 1

According to [1] Chronic disease management is the fastest growing, most expensive, most fractious problem facing health care provider and government across the world. and greater than 90 percent population over the globe suffer from heart-related health problems. Fitness belts are recorded the heart rate, number of steps, sleeping pattern and Energy expenditure of the user.

3.2 Case-study 2

As [2] Naive base algo is basically represented as an Bayes theorem considered with assumptions. Naive base algorithm is nothing but conditional probability based algorithm. In [2], System performance is observed on cleveland data-sets.

3.3 Case-study 3

In [3] they perform a study on 382 patient and they have used long and short term heart rate variability, shortness of breathing, Blood pressure as a parameter.

3.4 Case-study 4

In [4] and [5] also the Naive base, SVM, KNN, DT and Random forest are used for classification. [4] reviews that all machine learning algorithms mention above are having huge scope of predicting cardiovascular or heart related diseases.and from above algorithms each are performing well in some cases and very poor also in some cases which could be because of over-fitting. Similarly [6] also reviews the same classification algorithms.

Chapter 4

A brief about Heart disease and Facts on the Heart rate

4.1 Facts about Heart-rate

Heartrate is nothing but no of time per minute that heart contacts. The heart-rate is a vital indicator of health for the Human body. Heart rate is varied as physical activity. but normal heartrate of a human is in between 60 bpm to 100 bpm while he/she is relaxed and that heart rate is called resting heart rate. In the 4.1 Normal resting heart-rate of healthy people at different age according to the National Institute of Health-care (NIH) is represented. The heart-rates get slower as the person moves through childhood to adolescence [6], which we can see in the 4.1.

Age	**Normal heart-rate(bpm)**
Up - 1 month	between 70 - 190 bpm
From 1 - 11 months	between 80 - 160 bpm
From 1 - 2 years	between 80 - 130 bpm
From 3 - 4 years	between 80 - 120 bpm
From 5 - 6 years	between 75 - 115 bpm
From 7 - 9 years	between 70 - 110 bpm
Over 10 years	between 60 - 100 bpm

Table 4.1: Normal heartrate during resting at different ages From NIH

In the 4.2 heart rates of healthy people during exercise according to the NIH is represented. The AHA states the max hr while exercise should be nearly equals to 220 bpm minus age of the person. In this 4.2 exertion of 50 to 85 percent and at 100 percent with

Age (y)	**Target heart-rate at 50 - 85 percent exertion (bpm)**	**Average maximum heart-rate at 100 percent exertion (bpm)**
20	between 100 to 170	200
30	between 95 to 162	190
35	between 93 to 157	185
40	between 90 to 153	180
45	between 88 to 149	175
50	between 85 to 145	170
55	between 83 to 140	165
60	between 80 to 136	160
65	between 78 to 132	155
70	between 75 to 128	150

Table 4.2: Heart rates during exercise according to the National Institute of Healthcare(NIH)

Normal heart-rate	Between 60 to 100 bpm
Slow heart rate (Bradycardia)	Slower than 60 bpm
Fast heart rate (Tachycardia)	faster than 100 bpm

Table 4.3: Heart rate classification

different age is shown.

In the 4.3 the HR classification is represented. as we can see here if the heart rate is between 60 to 100 while resting or relaxing it's considered as a normal or resting heart rate. but according to experts, it should be 70 to 90. Slower than 60 is known as Bradycardia and faster than 100 is known as Tachycardia.

4.2 Abnormal heart-rate rhythms

The speed of heart rate is not only the factor for considering health but the heart rhythm is also important. The heartbeat rhythm should be a steady rhythm and the gap between beats should be there. The electrical system is there in the muscle which tells that when to beat and push the blood in the body. If there is some fault in that electrical-system it can lead to the abnormal heart rate rhythm.

Chapter 5

Algorithm and Techniques used

In this we are having two modules, first we use regression for prediction of heart rate and in second we use classification of bradycardia and tachycardia[7]. Therefore here this problem is considered as both classification and regression problem. In which we use regression for simple fitness belt recorded data-set and classification for hospital's data-set and smart-watch data with more number of parameters such as age, Respiratory rate, Temperature, Blood Pressure. In fig 5.2 this both modules are represented.

5.1 Basic Block Diagram

In 5.1 the basic flow of the system is represented. The block diagram is simple and self explanatory.

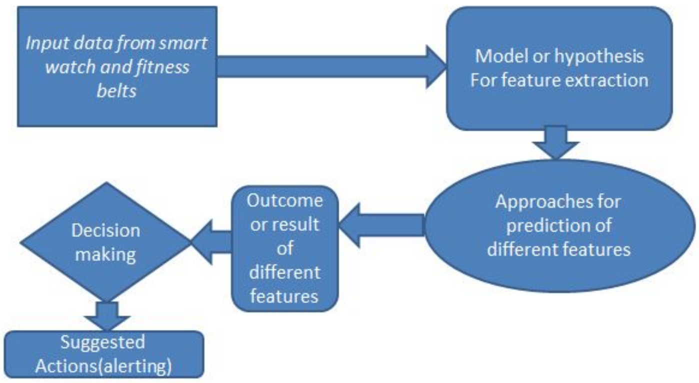

Figure 5.1: Basic Block Diagram of the system

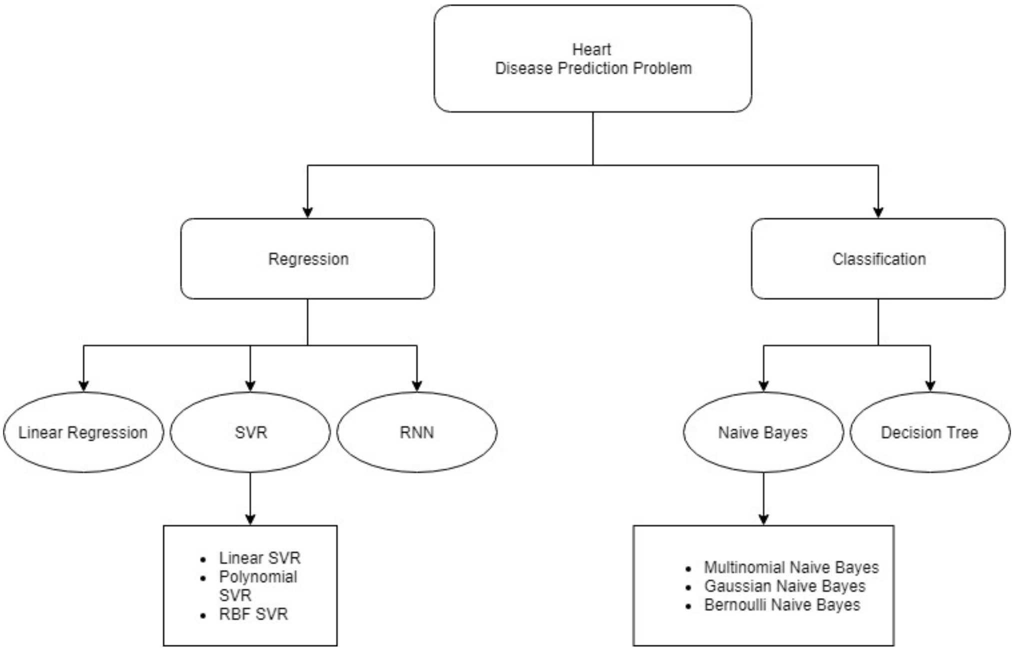

Figure 5.2: Machine Learning Techniques For Prediction

5.2 Module I: Considered problem as a Regression

In this we use heart rate data which is recorded through smart watch and we apply regression machine learning algorithm which can predict the heart rate of user and here we use following techniques using sklearn library. and we plot the graph of capture data of heart rate and predicted heart rate with time.

5.2.1 Linear-Regression

Regression is nothing but the method of making the model which can predict the target value independently. we can use for forecast and for find cause and effect relation between variables. Simple linear regression is analysis, where no of independent variable is only 1 and in that linear relation is there between the independent variable X and dependent variable Y. It is shown in fig **??** in which the red color line in the graph is refereed as the best fit straight line.

Linear Regression Equation:

y = a0 + a1 * x

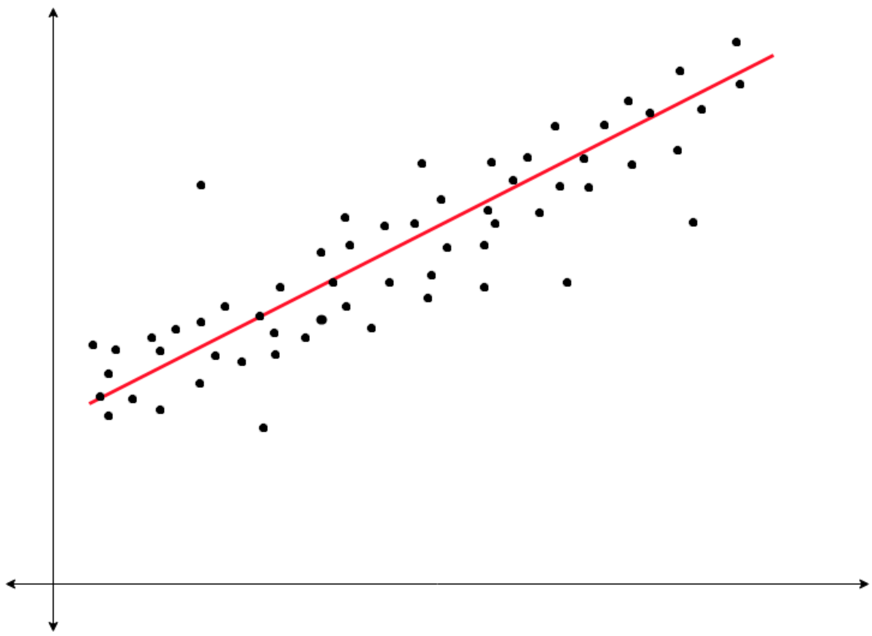

Figure 5.3: Linear Regression

$$minimize \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

$$J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

Cost Function:

The main aim of the linear regression is to find the best value of a0 and a1. and then after that we use that value for predict the value of y using that linear regression equation.

Gradient Descent

Gradient Descent is a method for updating the value of a0 and a1 to reduce the cost function. In this we simply start with some values for a0 and a1 and then we change the value of a0 and a1 iteratively to reduce the cost.

5.2.2 SVR

SVR is nothing but the regression but here difference is, we try to minimize the error rate in simple regression but SVR tries to fit the error withing some particular threshold. SVM is used for classification in which some terms are defined as kernel, Hyper plane Boundary line and Support vectors. In which kernel is nothing but the function which is used for map to a lower dimensional data to higher dimensional data. Hyper-plane is separation line between the data and classes in svm. although in svr this line is help us to predict the value. Boundary lines are nothing but two lines used for create a margin over the hyper plane. Points which are closest to the boundary line are called Support vectors. We are using three variation of regression in svr as below.this variation is based on kernel function. Linear and polynomial kernels are less time consuming and provides less accuracy than rbf kernels.

Linear SVR

In this Linear SVR the kernel function is linear function and the accuracy is less in this compare to rbf and it is less time consuming for decreasing the dimension of data.

Polynomial SVR

In this Polynomial SVR the kernel function is polynomial function with degree 2 or 3 and the accuracy is less in this compare to rbf and it is less time consuming for decreasing the dimension of data.

RBF SVR

In this rbf SVR the kernel function is radial base function and the accuracy is more. but it is time consuming for decreasing the dimension of data as compare to linear and polynomial function.

5.3 Module II: Considered problem as a Classification

In this we use classification of bradycardia and tachycardia. therefore here this problem is considered as classification problem. In which we use classification for hospital's data-set and smart-watch data with more number of parameters such as age, Respiratory rate, Temperature, Blood Pressure.

5.3.1 Naive Bayes Classifier

The Conditional probability is used in this algorithm which is based on Bayes' theorem with some approximation. It assumes that the existence of a class is independent of another class. The performance of the system is Observed on Cleveland data sets and the classification techniques are then applied to predict group membership for data instances. In 5.4 the equation is represented.

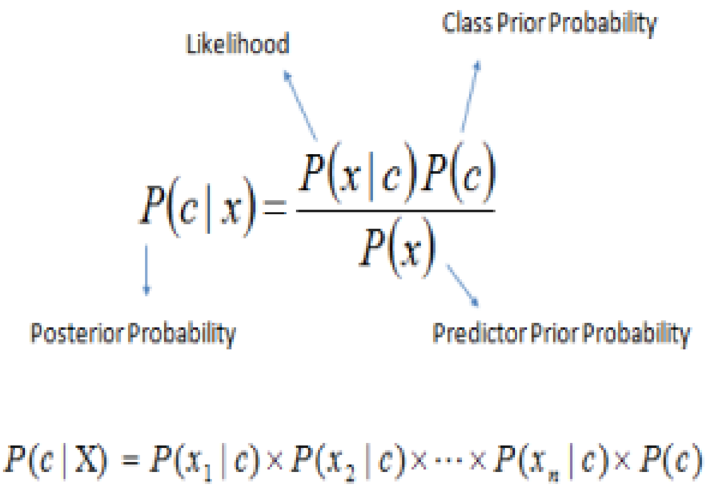

Figure 5.4: Naive Bayes Equation

5.3.2 Decision Trees

Tree like structure is there in Decision Tree algorithm. In this algorithm partitioning is performed on the data set and it will separate data-set into smaller subsets.Entropy and Information gain are calculated in this algorithm and based on that the classes are defined..In 5.5 the equation and structure of DT are represented.

5.3.3 Support Vector Machines

Support Vector Machine (SVM) is a method makes use of hyperplanes. As represented in [4] Support Vector Machine (SVM) can be used for heart failure prediction. The author has shown that they get the highest accuracy of 77.63 percent using this model. Similarly [8] also representing the svm used for classification.

$$Entropy(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

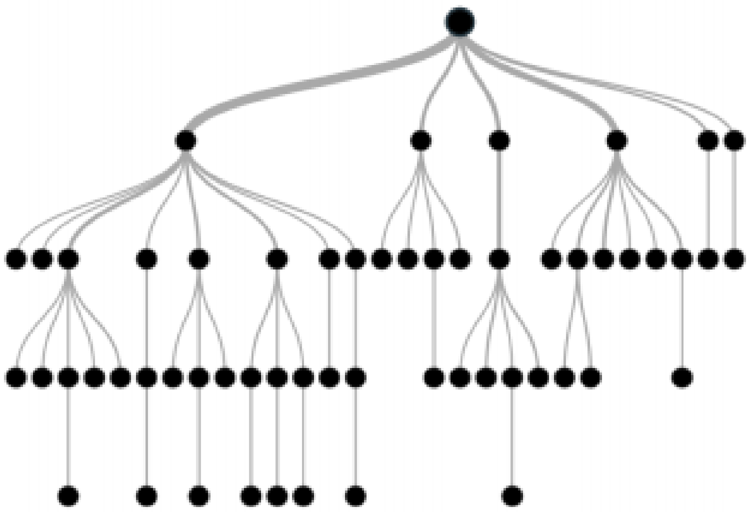

Figure 5.5: Decision tree

5.4 Using Long short-term memory (Recurrent neural network)

In this we use heart rate data which is recorded through smartwatch and we apply LSTM and here we are using keras and tensorflow for implementing this technique. we are using Google Colabfor implementation.

Chapter 6

Experimental Evaluation

In this we represented the prediction using regression module-I. And plot the graph of time vs heart rate.

6.1 Data-set Description

For this data set is created using MI band-3 fitness belt with time laps 10 minutes. from 10:40 AM 22-11-2018 to 12:10 AM 23-11-2018 total 82 rows which is represented in 6.1. In that ID and Heart rate we use for prediction here ID is considered as integer and Heart Rate data type is float. Here we need to use ID because, we can not use directly time. Machine learning models can not accept input as a time we need to convert it to some int number only then only we can apply that algorithms.

ID	**TIME**	**DATE**	**HeartRate**
1	10:40 AM	11/22/2018	61
2	10:50 AM	11/22/2018	87
3	11:00 AM	11/22/2018	104
4	11:10 AM	11/22/2018	77
5	11:20 AM	11/22/2018	83

Table 6.1: Structure of Heart rate data-set

6.2 Libraries Used

- **From sklearn.svm SVR** is used, here we use three kernel function in svr such as liner, polynomial and radial base function.And from this library we train the model and predict the heart rate.

- **From sklearn.linearmodel LinearRegression** is used, and from this library we train the model and predict the heart rate.

- **from keras.models import Sequential**

- **from keras.layers import Dense**

- **from keras.layers import LSTM**

6.3 Implementation

Here as shown in 6.1 we use three variation of svr and Linear regression for prediction of heart rate in bpm. here p1,p2,p3 and p4 are corresponding output and it is in the array form for ID 61 to 80, which is shown in resultant graph 6.5.

```
from sklearn.linear_model import LinearRegression
from sklearn.svm import SVR

svr_lin = SVR(kernel='linear', C=1.0)
svr_poly = SVR(kernel='poly', C=1.0, degree=3, gamma='scale')
svr_rbf = SVR(kernel='rbf', C=1.0,degree=4, gamma='scale')

svr_lin.fit(TrTime, TrHr)
svr_poly.fit(TrTime, TrHr)
svr_rbf.fit(TrTime, TrHr)
reg = LinearRegression().fit(TrTime, TrHr)

p1 = svr_lin.predict(TestTime)
p2 = svr_poly.predict(TestTime)
p3 = svr_rbf.predict(TestTime)
p4 = reg.predict(TestTime)

```

Figure 6.1: Code snippet 1

```
model = Sequential()
model.add(LSTM(4, input_shape=(1, look_back)))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
model.fit(trainX, trainY, epochs=100, batch_size=1, verbose=2)
```

Figure 6.2: Code snippet 2

6.4 Results

Result is shown in graph 6.3 and 6.5. In which 6.3 represents the time(10 minutes) vs heart rate(bpm) of dataset. and 6.5 represents the plot for predicted values of heart rate using different models. Here We consider this problem as regression and we use regression algorithms on our data set therefore for comparing those algorithm we need to use Mean Square Error or R^2 Error. And because of our data set is very small so we are getting more error as compare to expected error. And here naive bayes or DT are not applicable.

6.4.1 Root Mean Square Error Using Different techniques:

RMSE SVR LINEAR = 16.845161885851955

RMSE SVR POLYNOMIAL = 19.677817526491435

RMSE SVR RBF = 16.56222142724984

RMSE LINER REGRESSION = 16.855963061692016

RMSE LSTM = 14.81

As we can see in the results the Long short-term memory (LSTM), which is an artificial recurrent neural network (RNN) architecture is giving the lowest value of Root mean square error.

6.4.2 Average relative oxygen consumption during different situations are compared:

Here as we can see during different physical situations the relative oxygen consumption in Millitres per kilogram per minute is different. As we can see in figures if we compare the relative oxygen consumption during Seated to Cycling stage 6 is much more higher. Data-set of average relative oxygen consumption is very small so model can not probably giving best fit, which we can see in following figures. As shown in fig 6.6

average relative oxygen consumption during Supine, fig 6.7 average relative oxygen consumption during Seated, fig 6.8 average relative oxygen consumption during Standing, fig 6.9 average relative oxygen consumption during Treadmill Stage1, fig 6.10 average relative oxygen consumption during Treadmill Stage2, fig 6.11 average relative oxygen consumption during Cycle Stage1, fig 6.12 average relative oxygen consumption during Cycle Stage2, fig 6.13 average relative oxygen consumption during Cycle Stage6 are there and for prediction we have compared different techniques which is represented in graph.

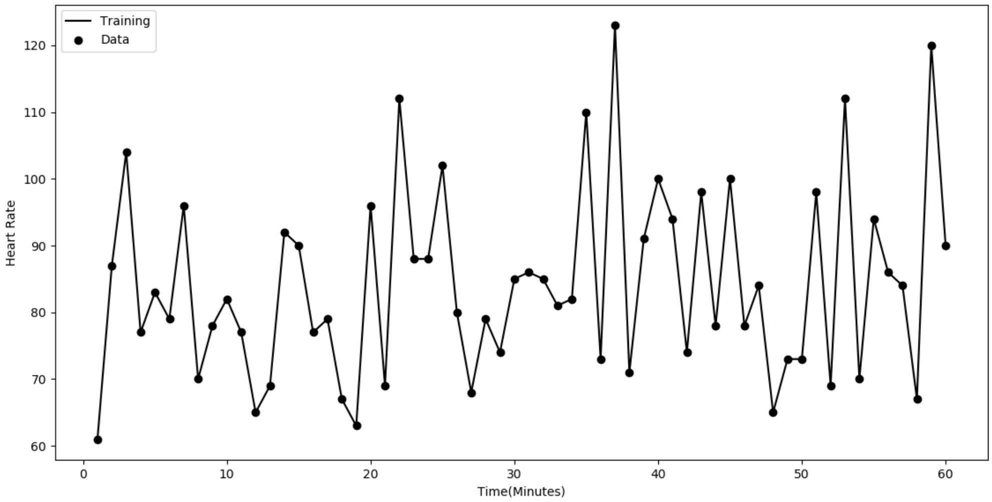

Figure 6.3: Time vs Heart rate using training data

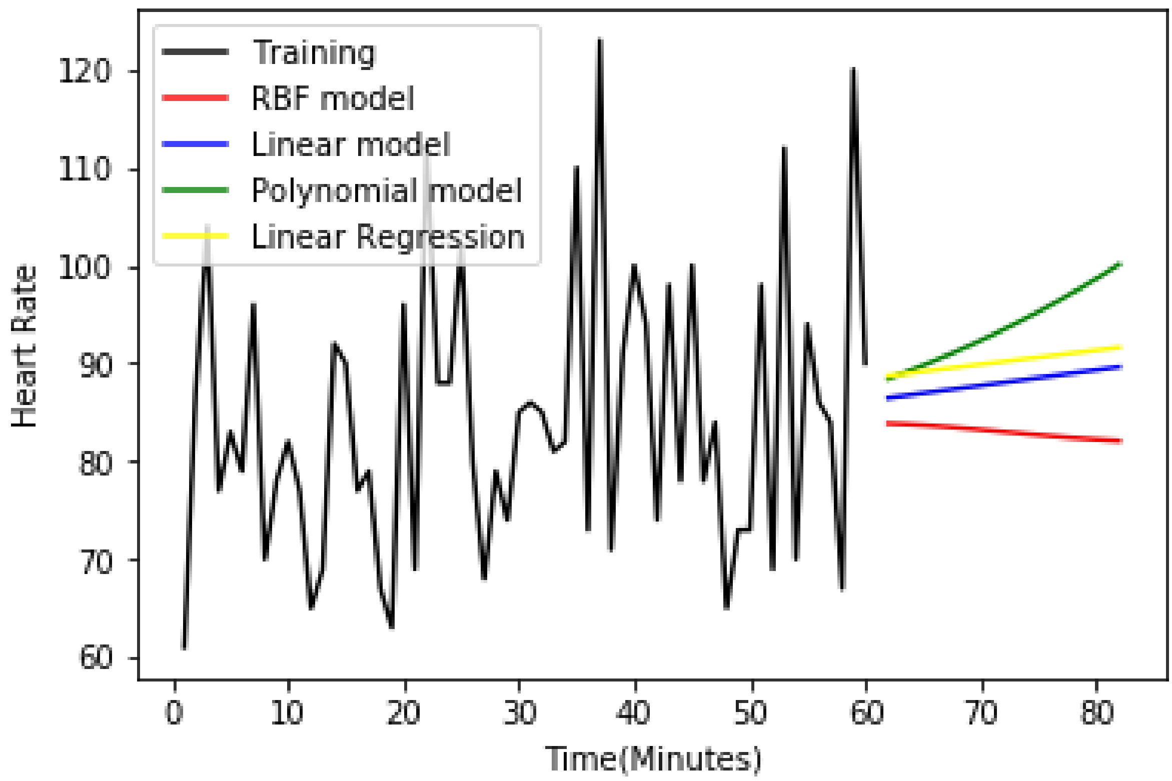

Figure 6.4: Resultant Graph

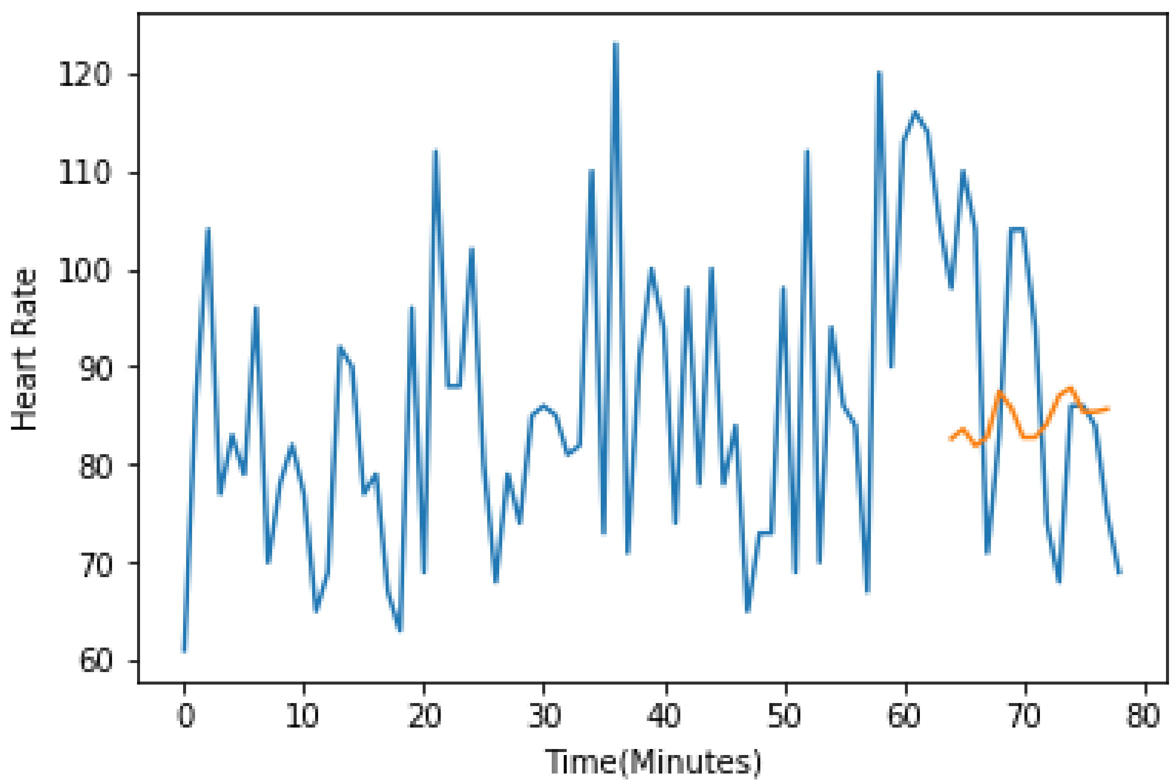

Figure 6.5: Resultant Graph

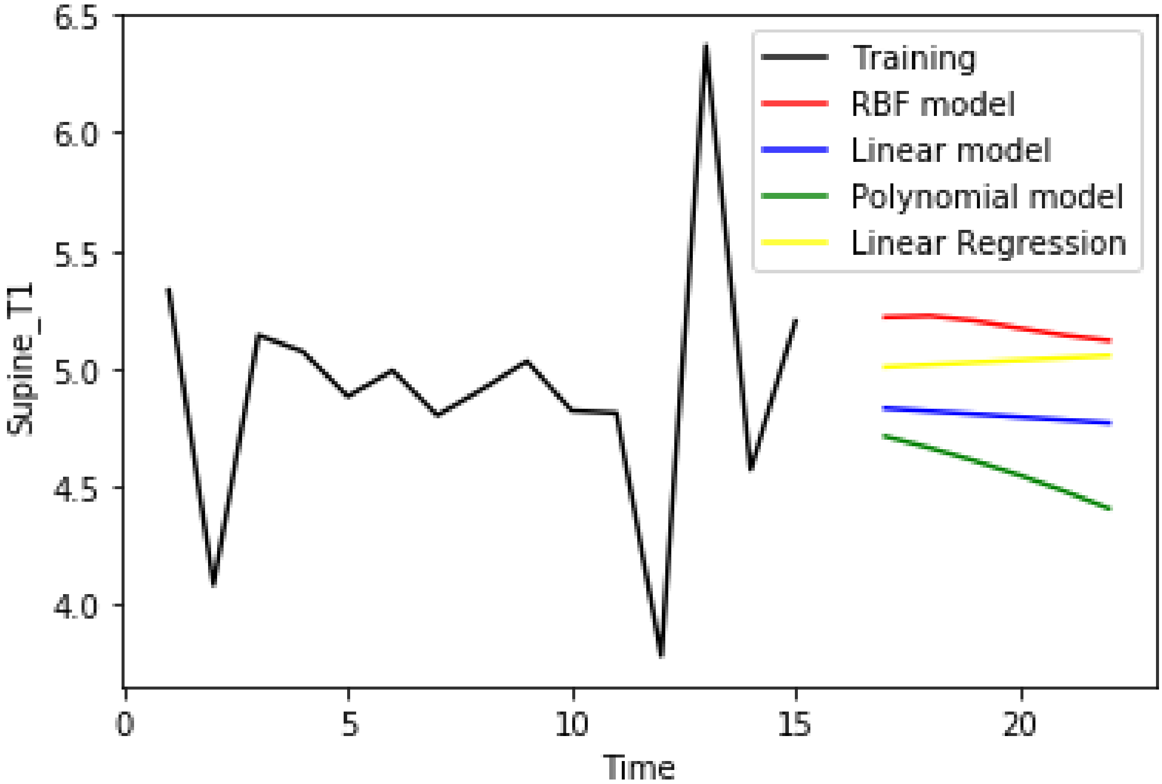

Figure 6.6: Average relative oxygen consumption during supine rest

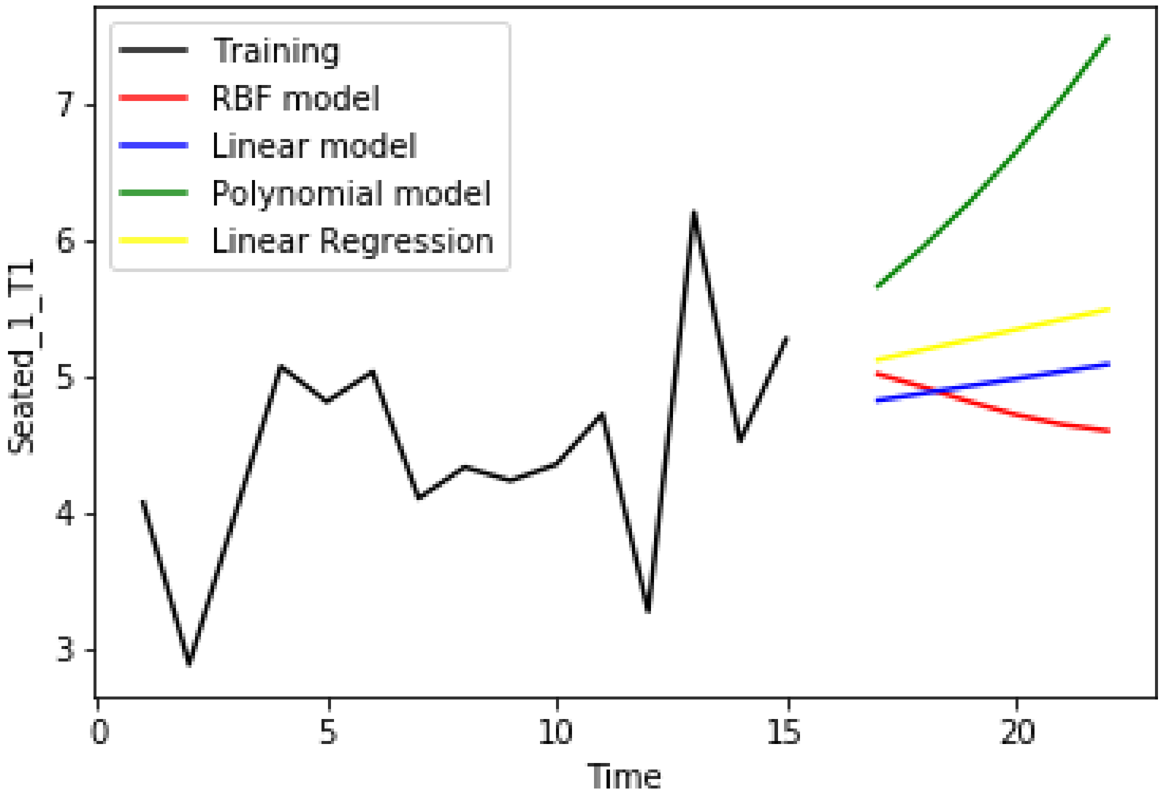

Figure 6.7: Average relative oxygen consumption during seated rest

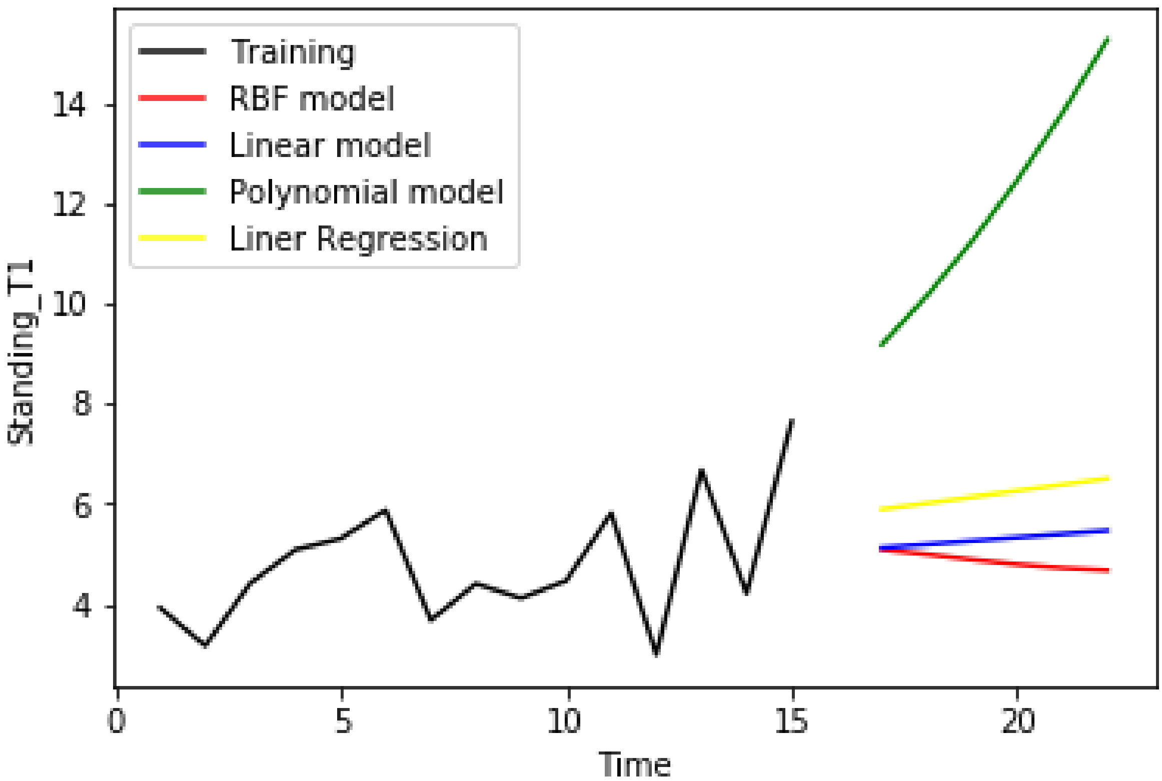

Figure 6.8: Average relative oxygen consumption during standing rest

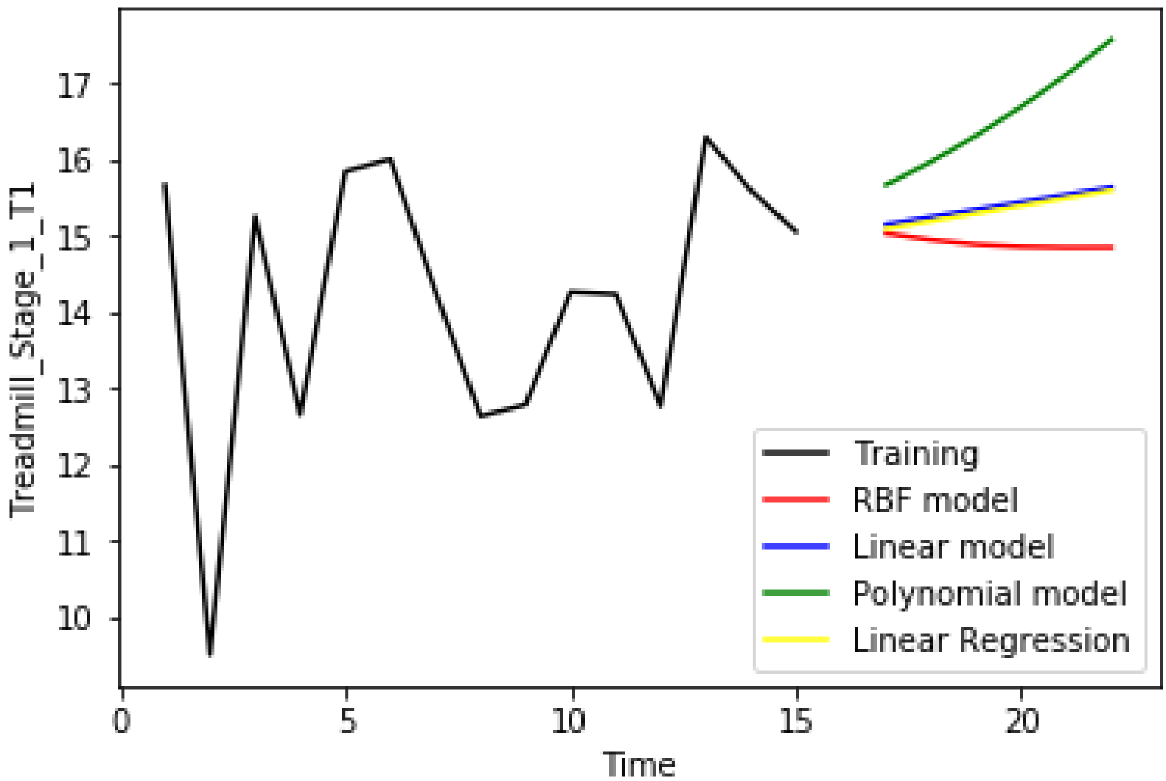

Figure 6.9: Average relative oxygen consumption during Treadmill Stage 1 rest

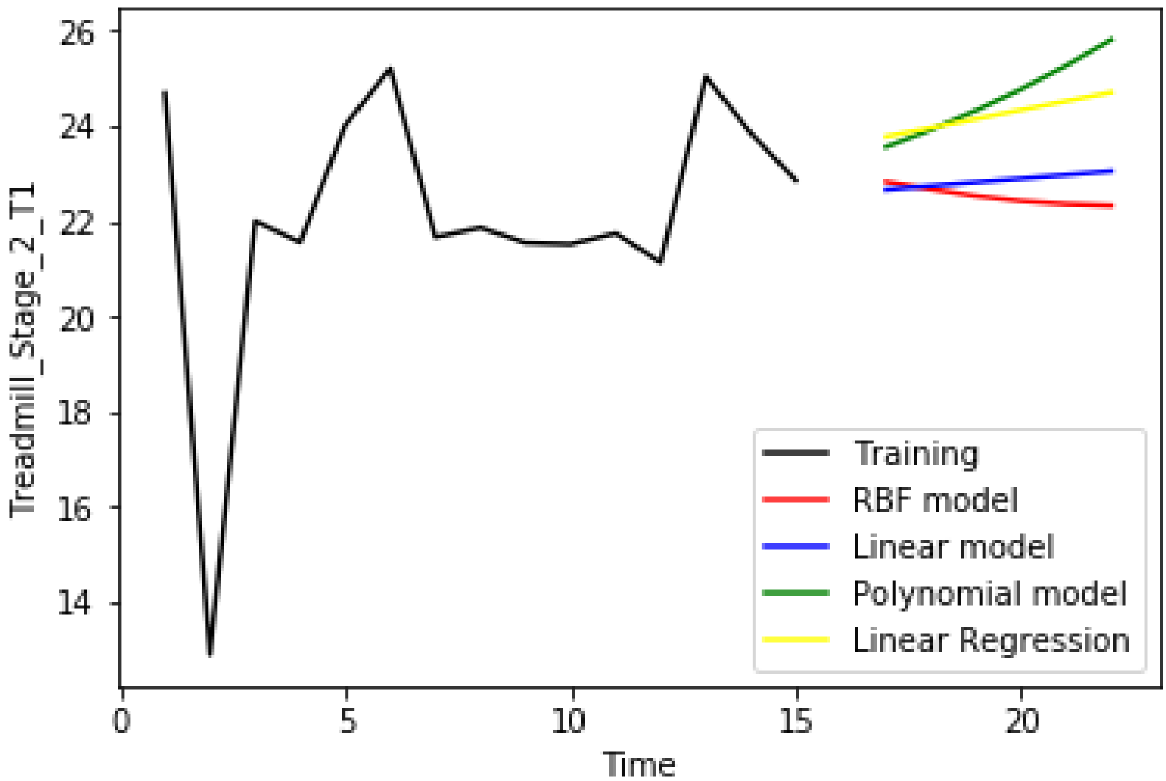

Figure 6.10: Average relative oxygen consumption during Treadmill Stage 2 rest

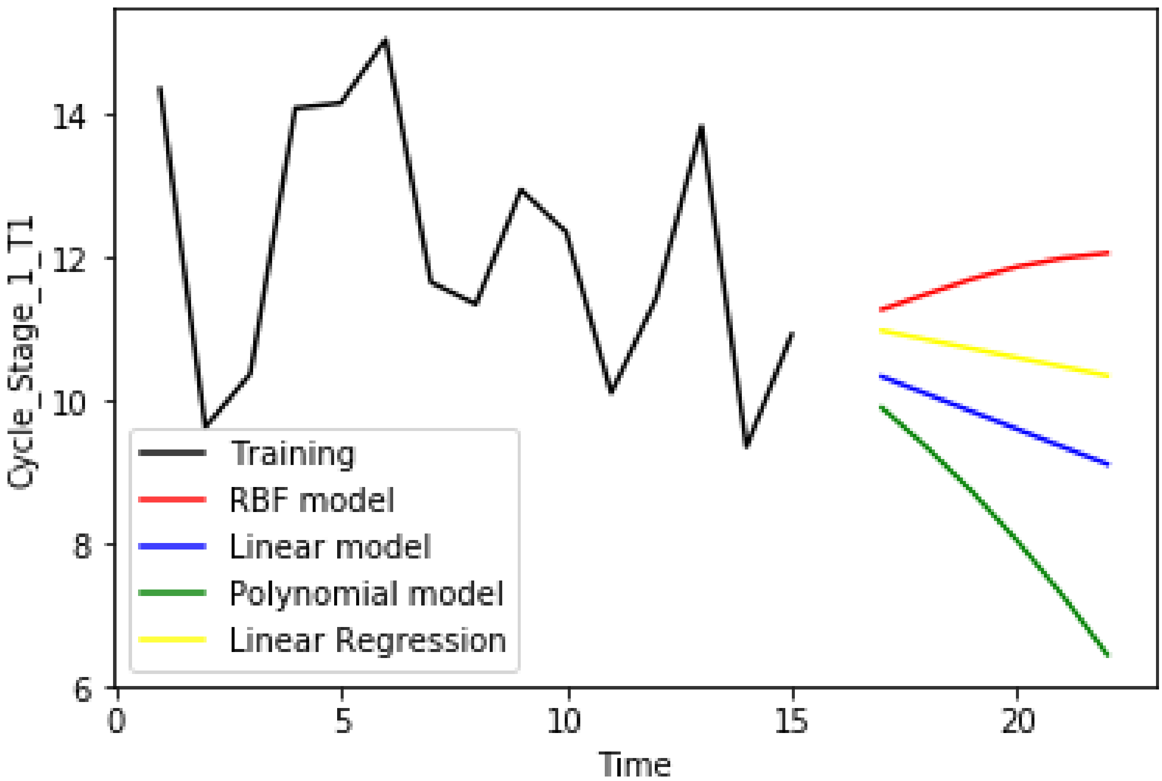

Figure 6.11: Average relative oxygen consumption during Cycle Stage 1 rest

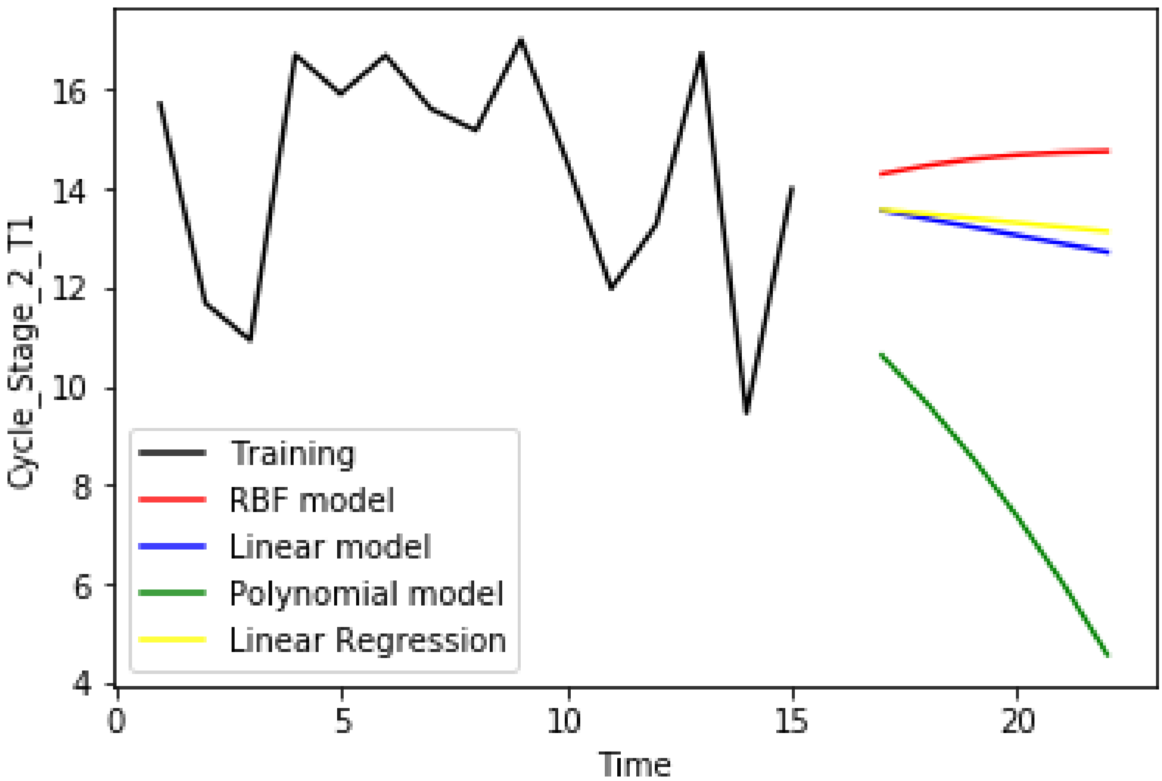

Figure 6.12: Average relative oxygen consumption during Cycle Stage 2 rest

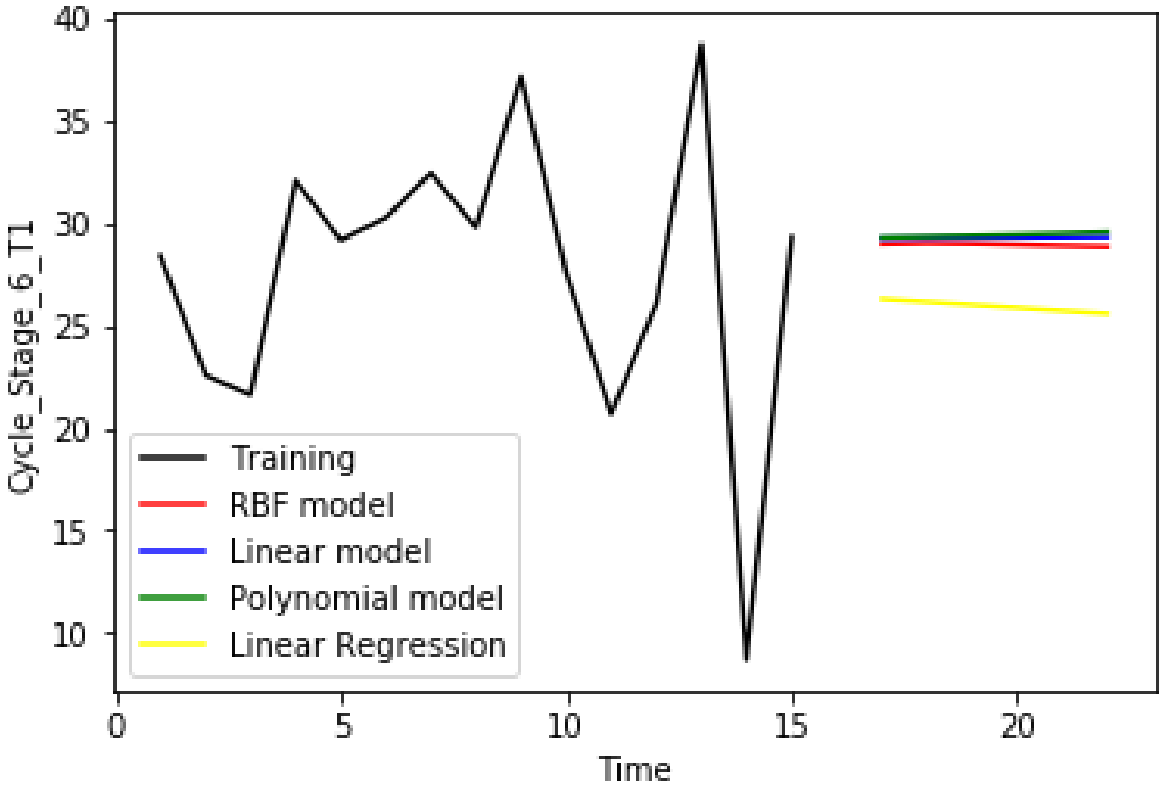

Figure 6.13: Average relative oxygen consumption during Cycle Stage 6 rest

Chapter 7

Conclusion and Future Scope

Health-care systems are shifted from patient care to monitored care. A motive of this project is the Real-time prediction of Cardiovascular disease using smart watch data and Short term Heart Rate Variability analysis. Therefore here we use classification and regression machine learning algorithms on that data set in which till now we have implemented the regression algorithms on data set which is created using MI Band 3 as shown in 6.1 . Performance matrix for different algorithms is shown and rectify that RNN and SVR algorithm are suitable and giving good results and with respect to data-set, what is the size of data-set matters a lot. Here as we can see the size of the data-set is small so the models are not giving best fit. But in future we can extend this problem with considering more parameters and larger data-set it can give best accuracy compared to this results.

7.1 Publications

[1] R.Bhavsar and S.Garg, ”Real time prediction of heart(Cardiovascular) disease using smart watch data and Short term HRV(Heart Rate Variability) analysis,” in CISCT 2019, IEEE-International Conference on Innovative Sustainable Computational Technologies, (Communicated)

Bibliography

[1] N. S. Rajliwall, G. Chetty, and R. Davey, "Chronic disease risk monitoring based on an innovative predictive modelling framework," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–8, Nov 2017.

[2] B. Gnaneswar and M. R. E. Jebarani, "A review on prediction and diagnosis of heart failure," in *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pp. 1–3, March 2017.

[3] G. D. Pinna, R. Maestri, S. Capomolla, O. Febo, A. Mortara, P. G. Riccardi, and M. T. L. Rovere, "Determinant role of short-term heart rate variability in the prediction of mortality in patients with chronic heart failure," in *Computers in Cardiology 2000. Vol.27 (Cat. 00CH37163)*, pp. 735–738, Sept 2000.

[4] V. V. Ramalingam, A. Dandapath, and M. K. Raja, "Heart disease prediction using machine learning techniques : a survey," *International Journal of Engineering Technology*, vol. 7, no. 2.8, pp. 684–687, 2018.

[5] G. Guidi, M. C. Pettenati, P. Melillo, and E. Iadanza, "A machine learning system to improve heart failure patient assistance," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, pp. 1750–1756, Nov 2014.

[6] A. Hazra, S. Kumar Mandal, A. Gupta, A. Mukherjee, and A. Mukherjee, "Heart disease diagnosis and prediction using machine learning and data mining techniques: A review," *Advances in Computational Sciences and Technology*, vol. 10, pp. 2137–2159, 07 2017.

[7] M. Daniel Masethe, Hlaudi Masethe, "Prediction of heart disease using classification algorithms.," 2014.

[8] Mythili, D. Mukherji, N. Padalia, and A. N. Naidu, "A heart disease prediction model using svm-decision trees-logistic regression (sdl)," 2013.