DATA ANALYTICS & MACHINE LEARNING

PREMANANDA INDIC, PH.D.

DEPARTMENT OF ELECTRICAL ENGINEERING

The University of Texas at TYLER Center for Health Informatics & Analytics

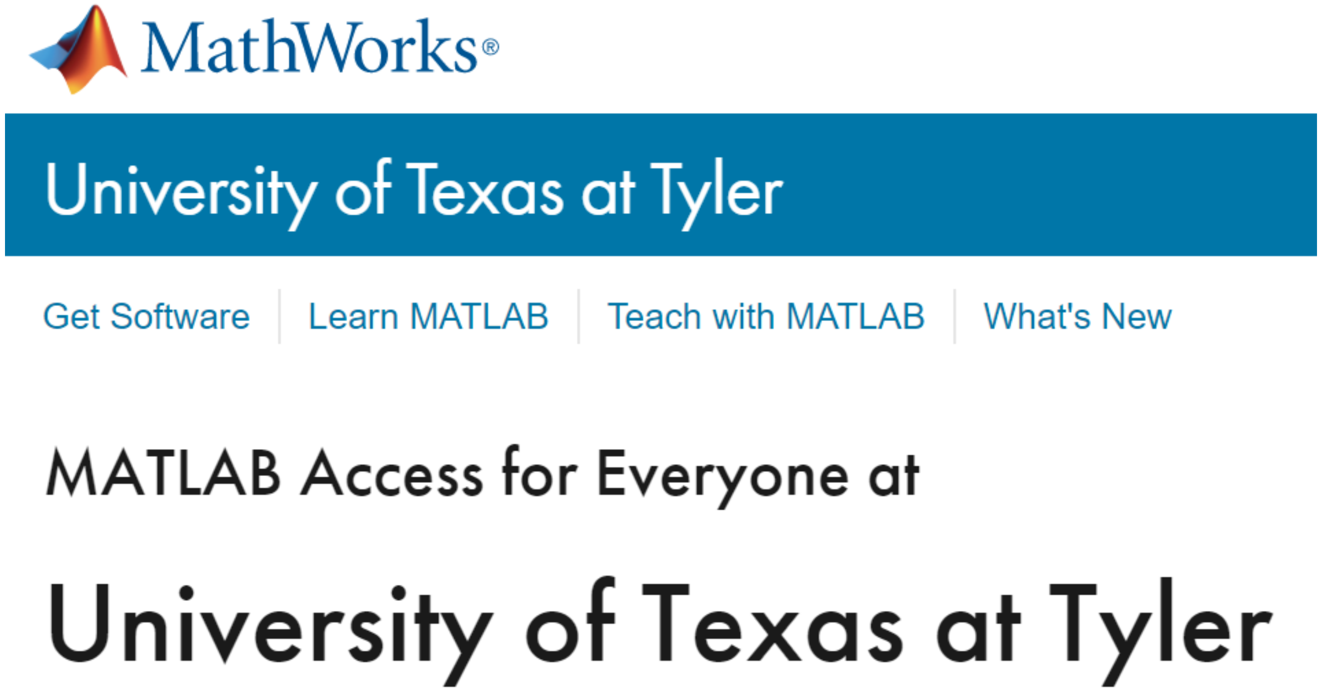
ORS Research Design & Data Analysis Lab

Office of Research and Scholarship

WORKSHOP SCHEDULE

- WEEK1: DATA ANALYTICS
- WEEK2: FEATURE EXTRACTION
- WEEK3: MACHINE LEARNING

ANALYSIS PLATFORM



https://www.mathworks.com/academia/tah-portal/university-of-texas-at-tyler-1108545.html

HYPOTHESIS

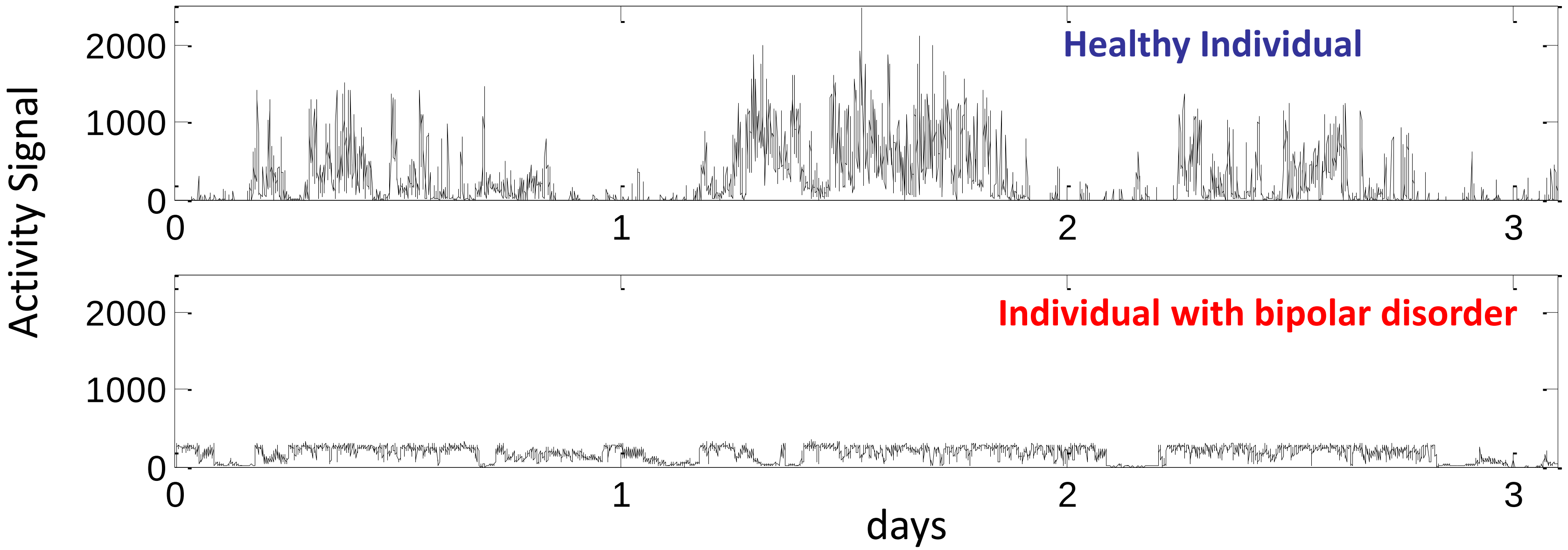
Scientific hypothesis, an idea that proposes a tentative explanation about a phenomenon or a narrow set of phenomena observed in the natural world. The two primary features of a scientific hypothesis are falsifiability and testability

Source: https://www.britannica.com/science/scientific-hypothesis

FEATURE EXTRACTION

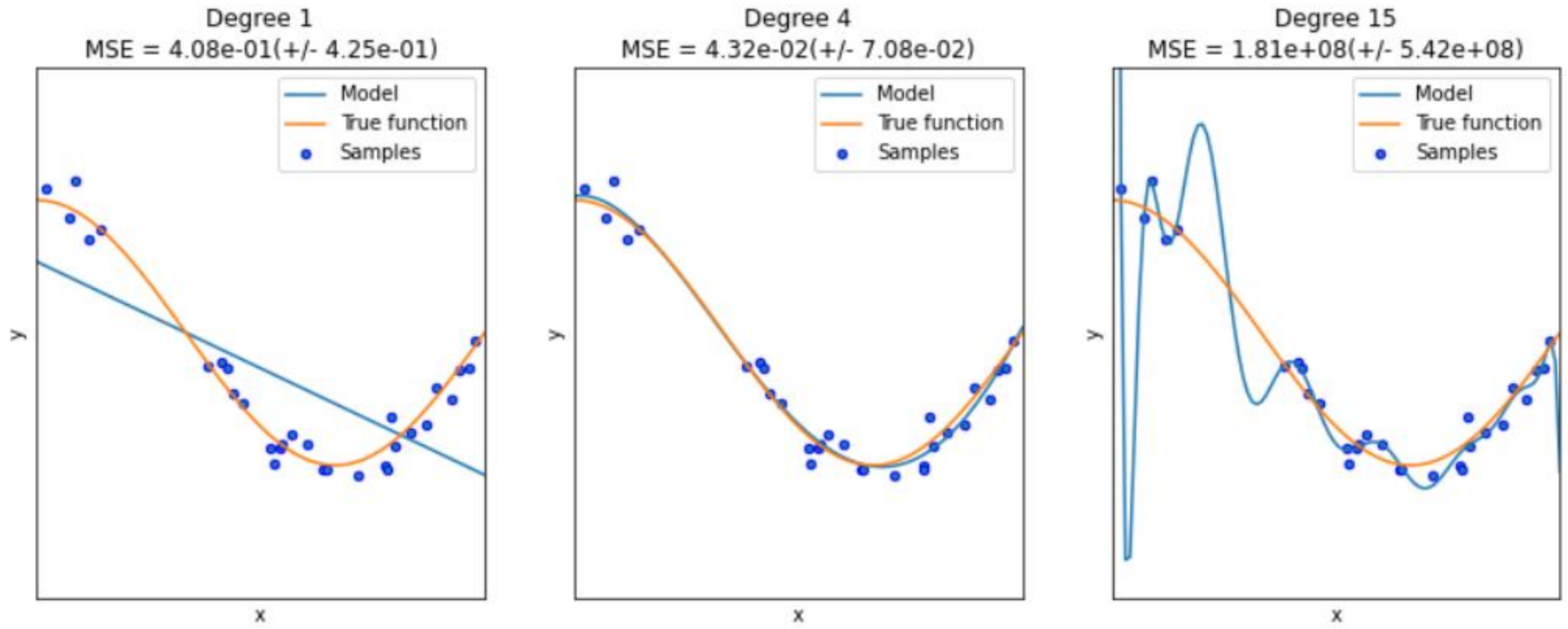
- Feature extraction is the process of converting raw data into useful information for machine learning algorithms to predict or classify

Philips Actiwatch 2



FEATURE EXTRACTION

- Avoid too many features (computational resources and overfitting)

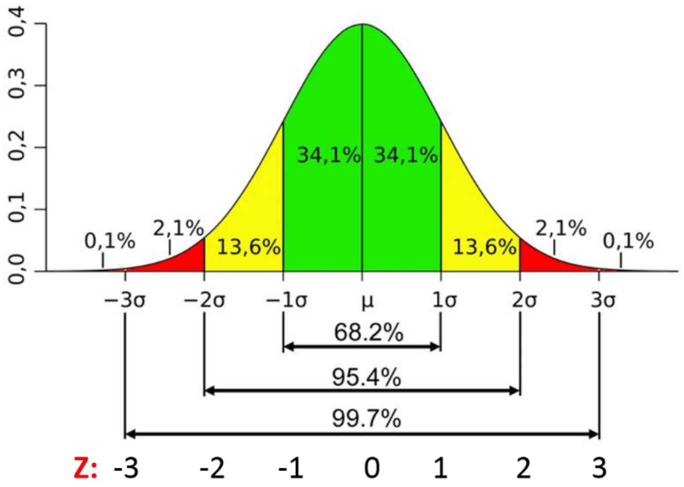


https://datascience.foundation/sciencewhitepaper/underfitting-and-overfitting-in-machine-learning

FEATURE EXTRACTION

- Statistical Features

(Mean, Standard Deviation, Mode, Skewness, Kurtosis)



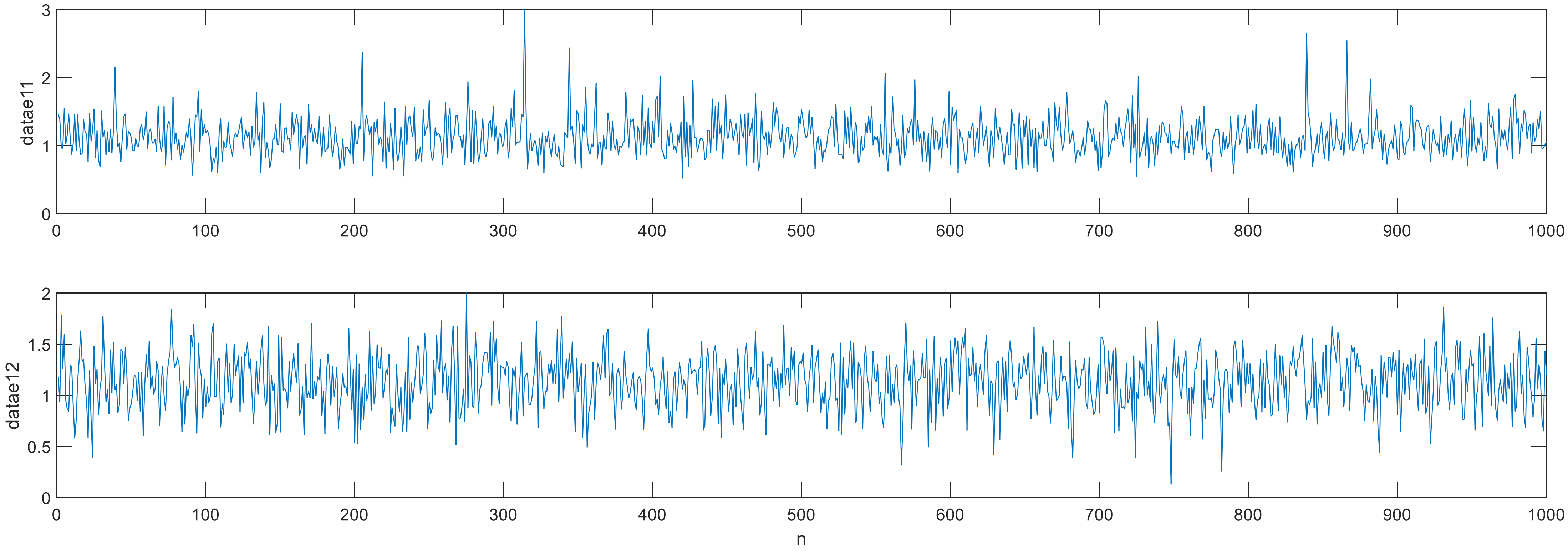
$\mu \rightarrow Mean$ (Mode and Median)

$\sigma \rightarrow Standard\ Deviation$

https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module6-RandomError/PH717-Module6-RandomError5.html

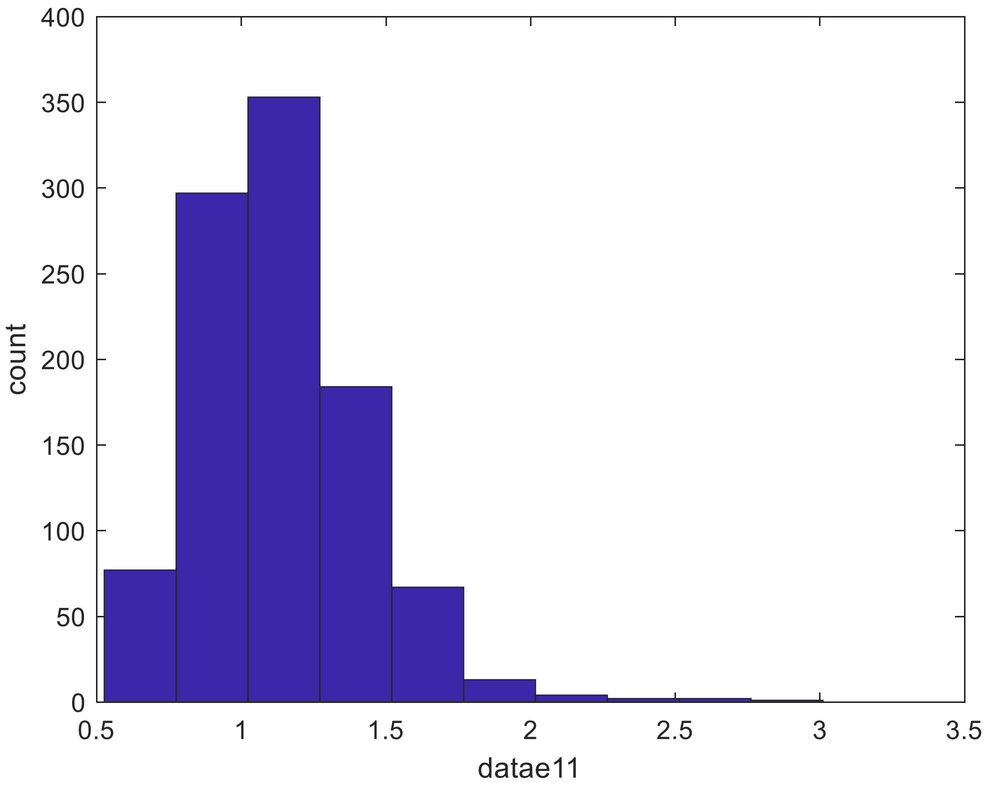
Exercise 1

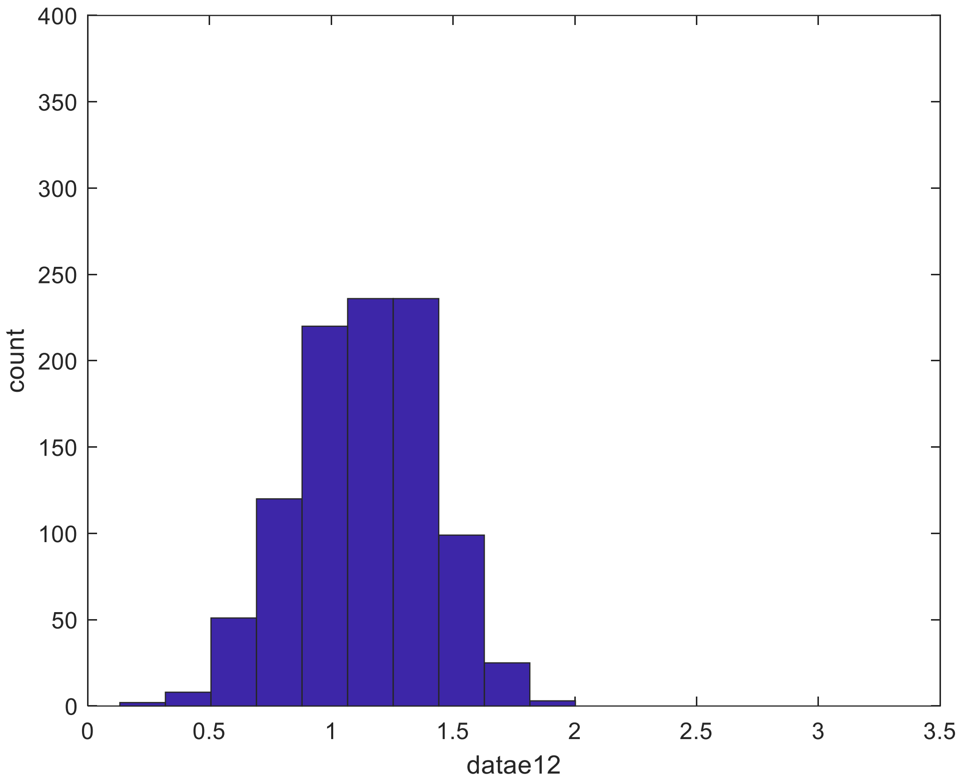
Check whether the given two data sets have same features



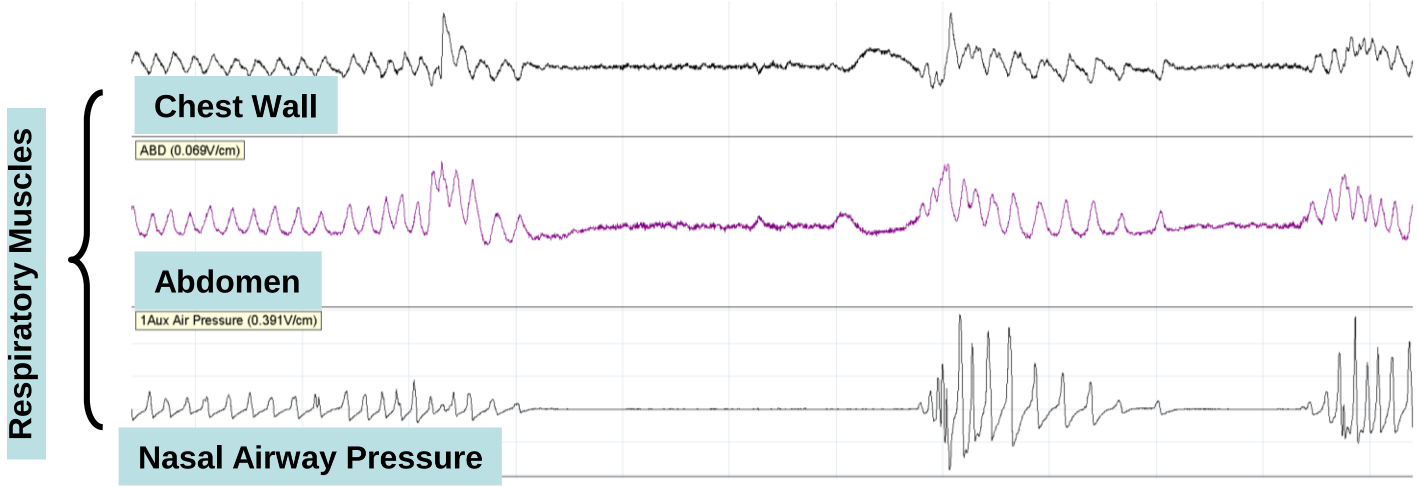
Exercise 1

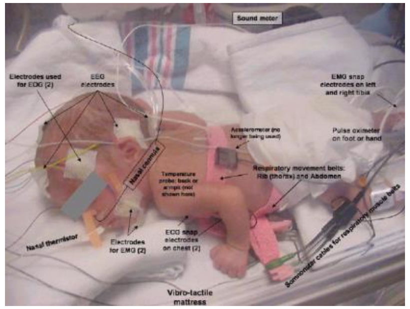
Check whether the given two data sets have same features

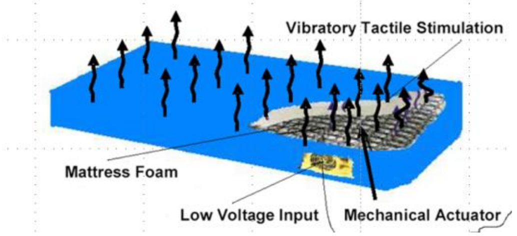




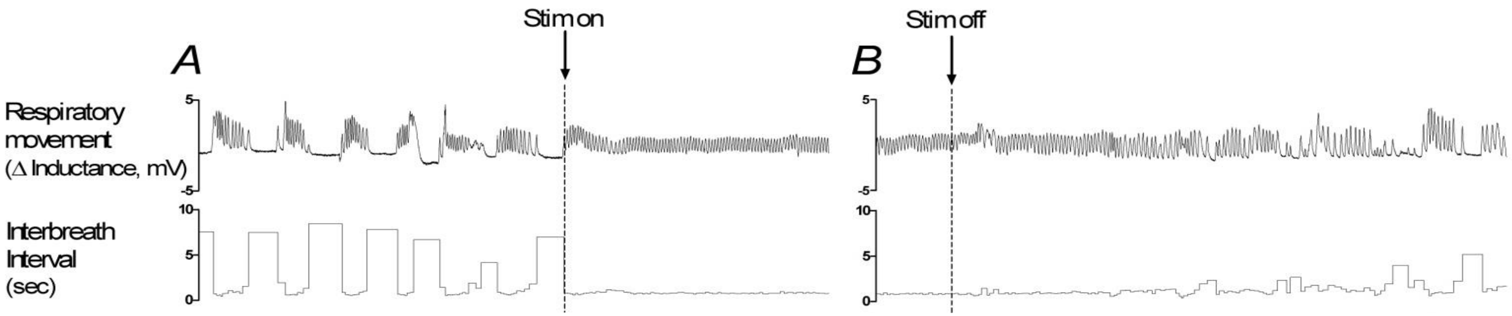
Exercise 1


Respiratory Muscles
Chest Wall
ABD (0.069V/cm)
Abdomen
1Aux Air Pressure (0.391V/cm)
Nasal Airway Pressure

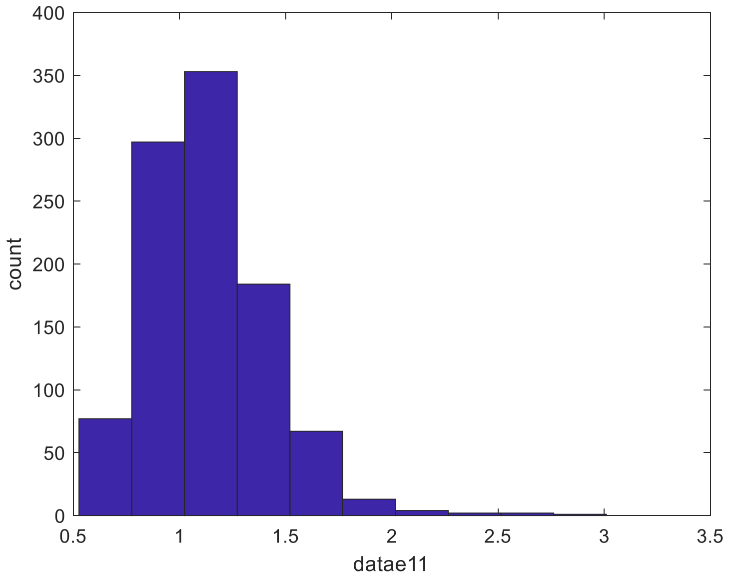



Vibratory Tactile Stimulation
Mattress Foam
Low Voltage Input
Mechanical Actuator

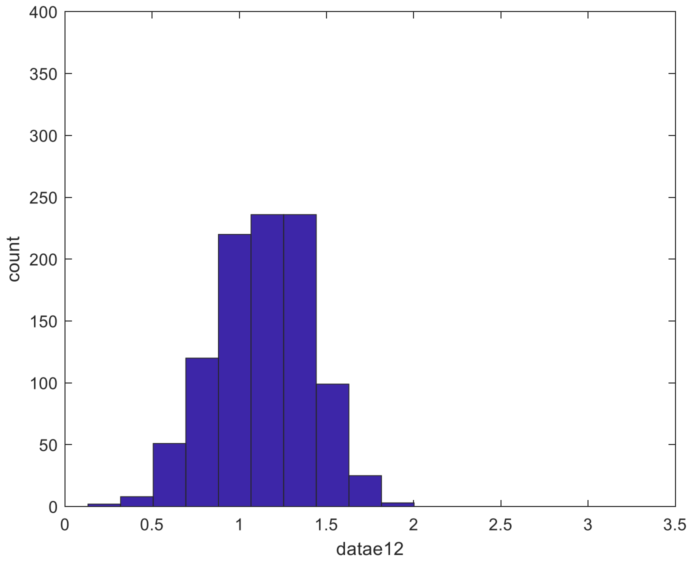
Exercise 1



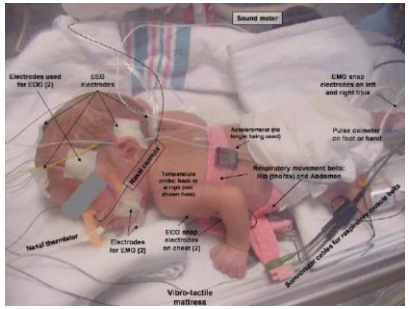
Stim on
Stim off
A
B
Respiratory movement (Δ Inductance, mV)
Interbreath Interval (sec)
5
-5
10
5
0

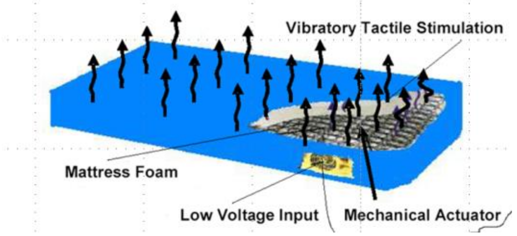


count
datae11
0
50
100
150
200
250
300
350
400
0.5
1
1.5
2
2.5
3
3.5



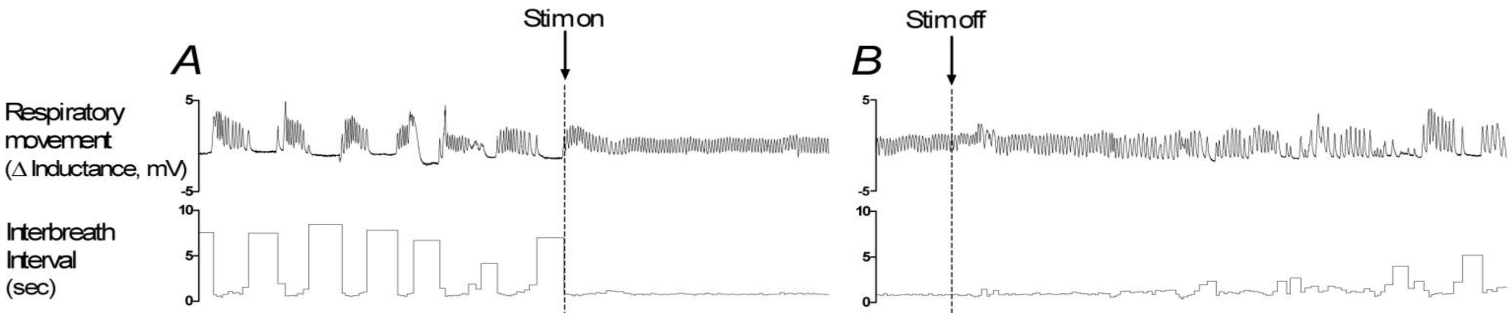
count
datae12
0
50
100
150
200
250
300
350
400
0
0.5
1
1.5
2
2.5
3
3.5

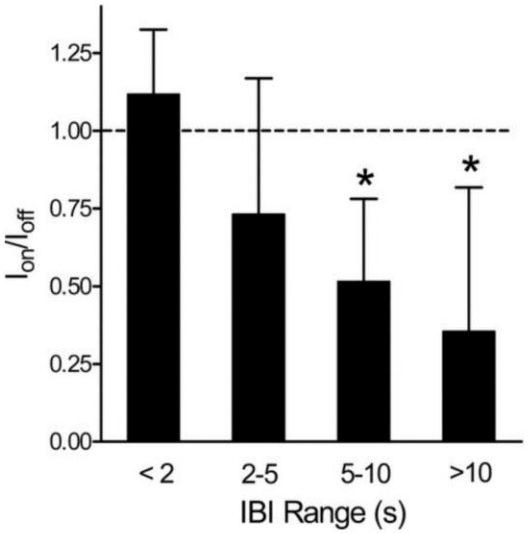


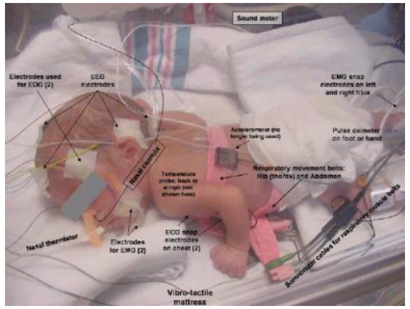


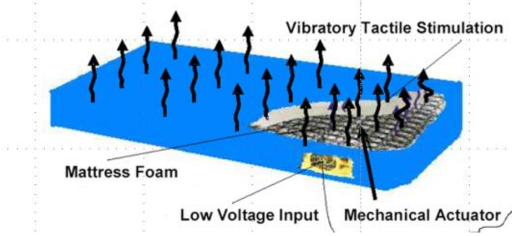
Vibratory Tactile Stimulation
Mattress Foam
Low Voltage Input
Mechanical Actuator

Exercise 1





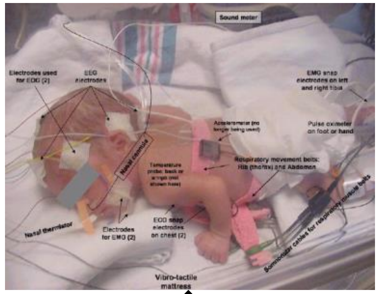


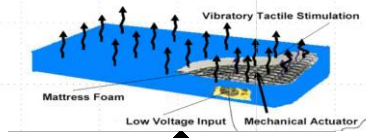


Bloch-Salisbury E, Indic P, Bednarek F, and Paydarfar D, J Appl Physiol., 2009, 107: 1017-1027

Exercise 1

Big Data





Signal Processing → Machine Learning → Evaluation & Prediction

WO Patent 2,013,033,433
WO2014182792 A1

Indic P, Paydarfar D, Barbieri R. IEEE Trans. Biomed. Eng 2013, 60(10):2858-66

Gee AH, Barbieri R, Paydarfar D, Indic P. IEEE EMBC Conf. 2015, 5855-5858

BIOMEDICAL DATA

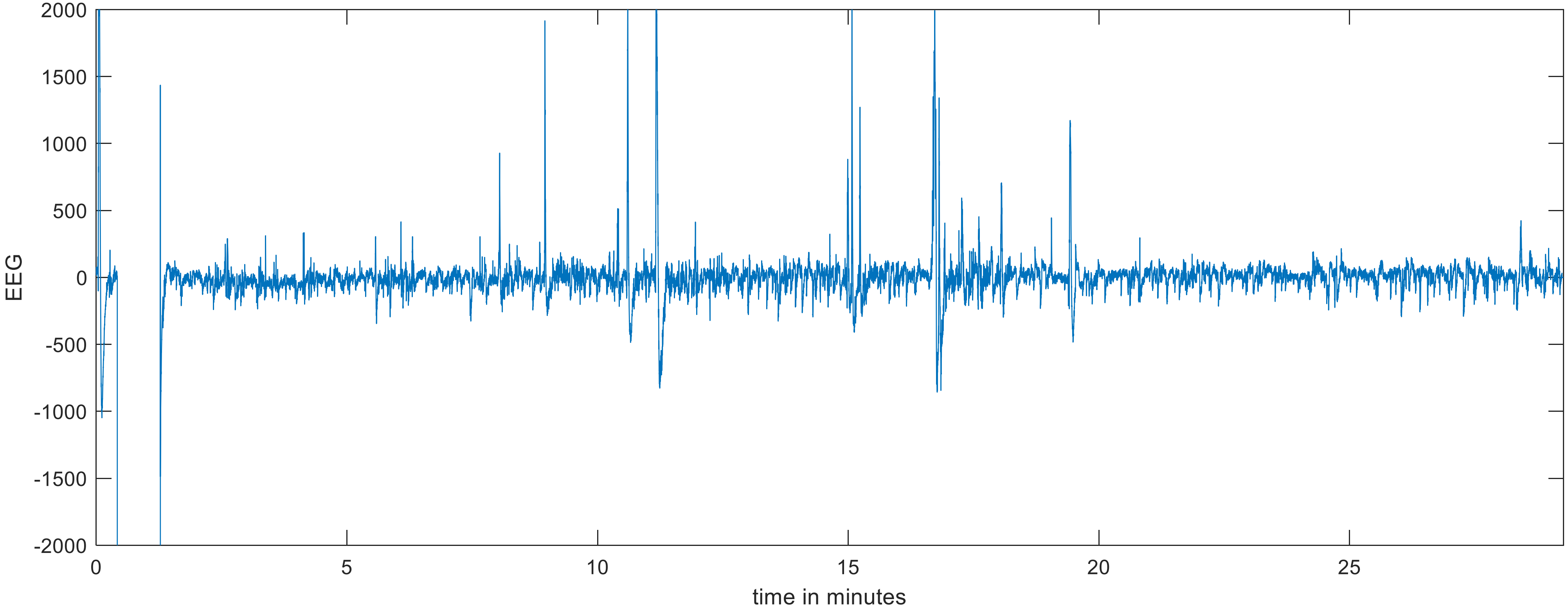
- LINEAR VS NONLINEAR
- DETERMINISTIC VS STOCHASTIC
- STATIONARY VS NONSTATIONARY

Biomedical data are nonlinear, nonstationary and deterministic / stochastic in nature

Analytical tools are applicable only for linear, deterministic/stochastic and stationary

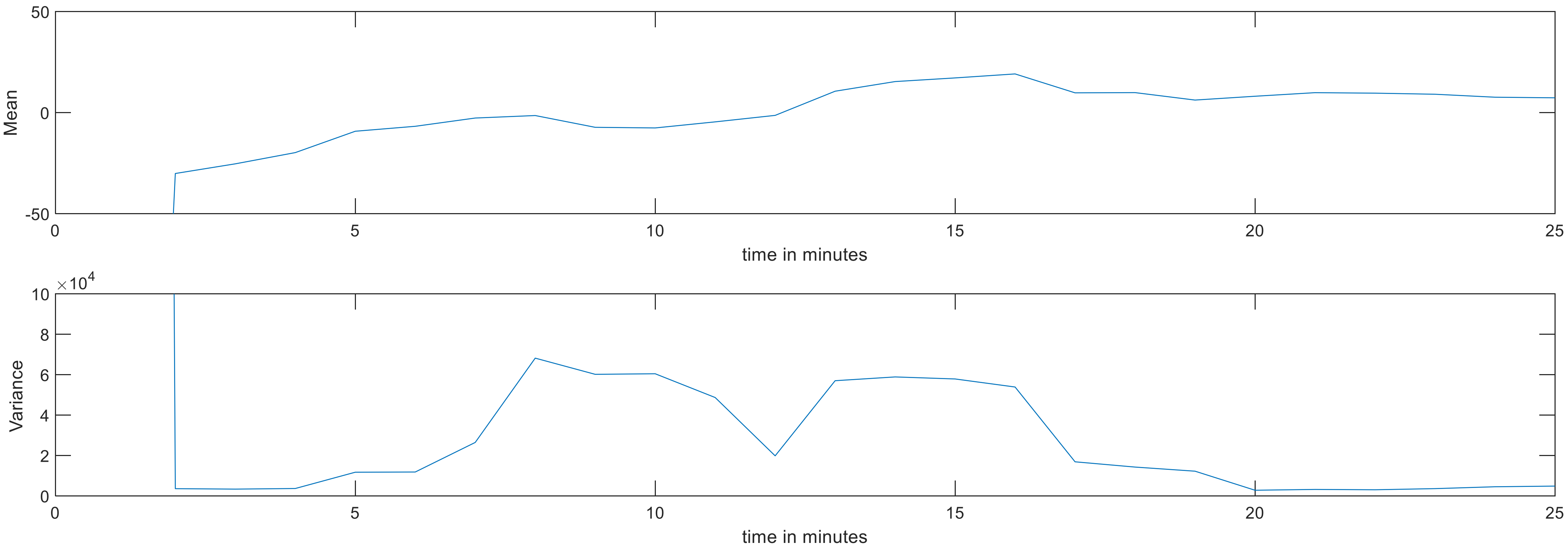
Exercise 2

Sliding Window Method: A few number of files and only one channel



Exercise 2

Sliding Window Method: A few number of files and only one channel



Exercise 3

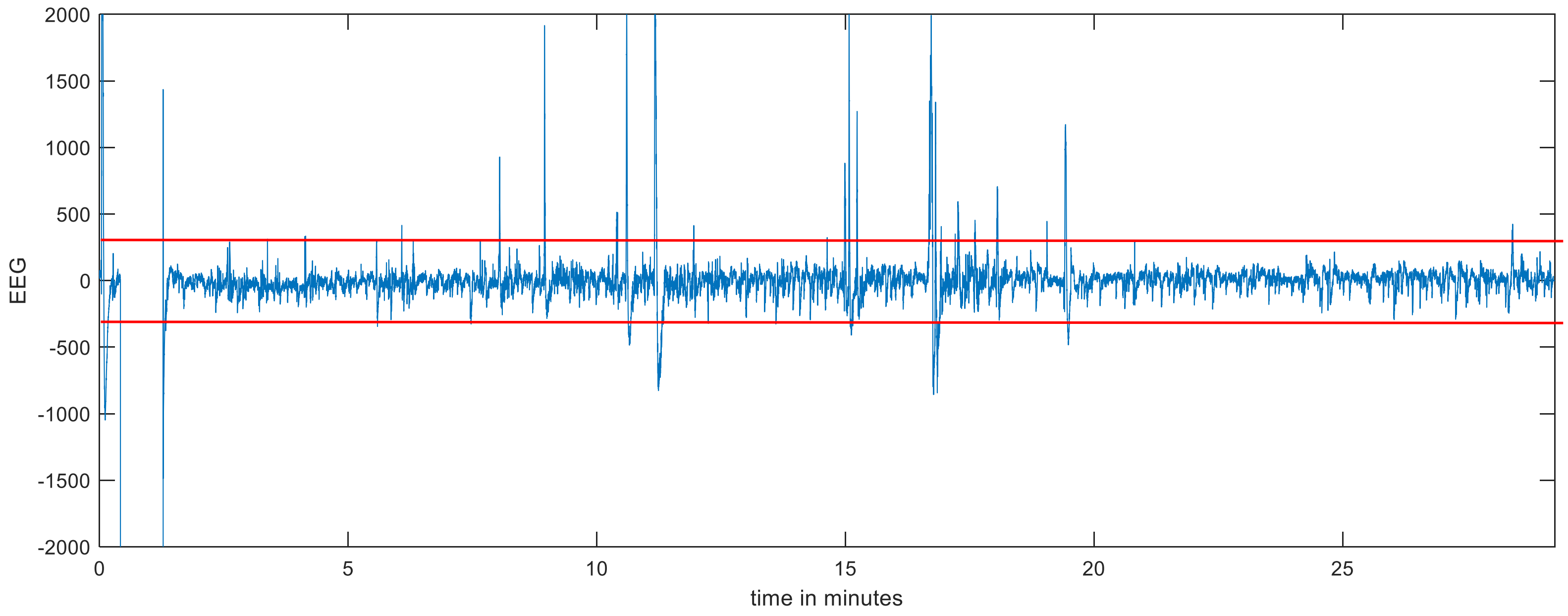
Sliding Window Method: A few number of files and multiple channels in each file

Exercise 4

Sliding Window Method: Several number of files in a folder

PREPROCESSING

- IDENTIFY OUTLIERS
- IDENTIFY NOISE



FEATURE EXTRACTION

- Statistical Features

(Mean, Standard Deviation, Mode, Skewness, Kurtosis)

- Spectral Features (linear)

(Frequency (Rate), Amplitude, Phase, Coherence, Spectrum Entropy)

- Nonlinear Features

(Detrended Fluctuation Coefficient, Multiscale Entropy, Mutual Information)

FEATURE EXTRACTION

- Statistical Features

(Mean, Standard Deviation, Mode, Skewness, Kurtosis)

MATLAB functions:

mean, sd, mode, skew, kurt

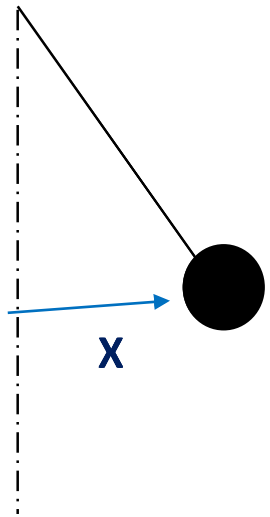
Works on two dimension arrays

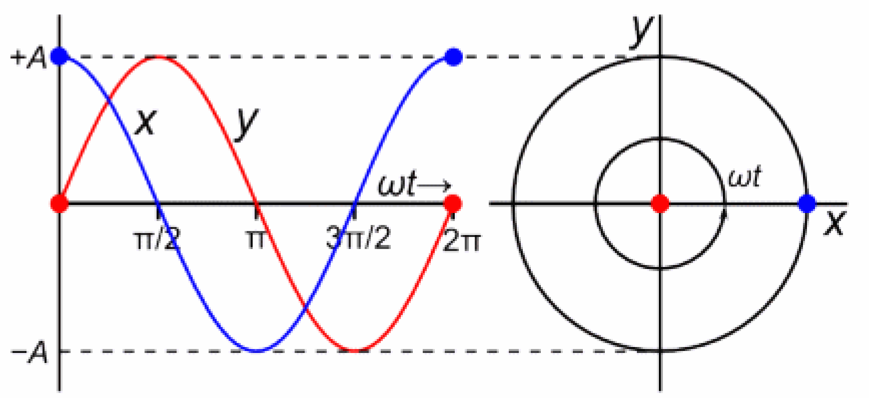
FEATURE EXTRACTION

- Spectral Features (linear)

(Frequency (Rate), Amplitude, Phase, Coherence, Spectrum Entropy)

Transformation of data in time to a new variable (example: Fourier Transform)





https://commons.wikimedia.org/wiki/File:Simple_harmonic_motion_animation_2.gif

FEATURE EXTRACTION

- Nonlinear Features

(Detrended Fluctuation Coefficient, Multiscale Entropy, Mutual Information)

-Complicated and may be useful

HYPOTHESIS

Scientific hypothesis, an idea that proposes a tentative explanation about a phenomenon or a narrow set of phenomena observed in the natural world. The two primary features of a scientific hypothesis are falsifiability and testability

Whatever features you select, and whatever conclusion you reach, always think, so what ?

Source: https://www.britannica.com/science/scientific-hypothesis

FEATURE EXTRACTION

Statistical Models are for inference (Linear Regression, Logistic Regression,……..)

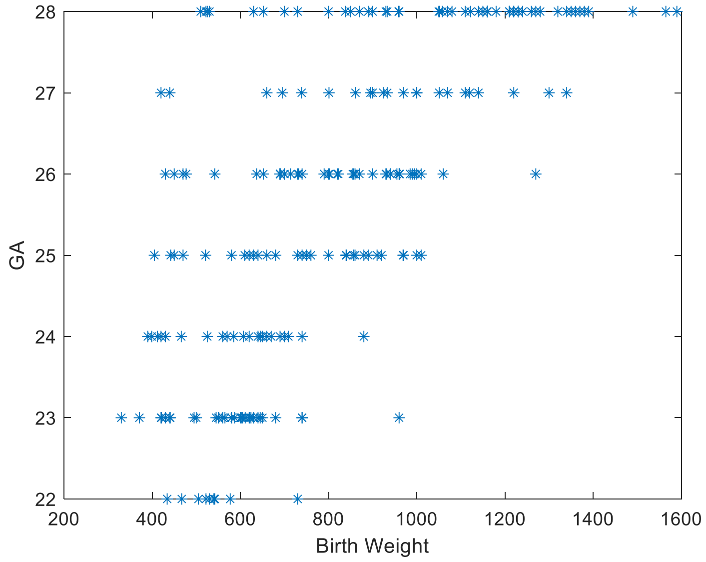
Machine Learning Models are for classification / prediction (Linear Regression, Support Vector Machine……)

FEATURE EXTRACTION

Statistical Models are for inference (Linear Regression, Logistic Regression,……..)

Machine Learning Models are for classification / prediction (Linear Regression, Support Vector Machine……)

GA = 0.0047*Birth Weight + 21.78



r = 0.69 ***p*** < 0.05

Project 1: Hypertension

Physiol Genomics 42: 23–41, 2010.
First published March 30, 2010; doi:10.1152/physiolgenomics.00027.2010.

CALL FOR PAPERS: | *Computational Modeling of Physiological Systems*

Identifying physiological origins of baroreflex dysfunction in salt-sensitive hypertension in the Dahl SS rat

Scott M. Bugenhagen, Allen W. Cowley, Jr., and Daniel A. Beard
Department of Physiology, Medical College of Wisconsin, Milwaukee, Wisconsin

Submitted 3 February 2010; accepted in final form 25 March 2010

Bugenhagen SM, Cowley AW Jr, Beard DA. Identifying physiological origins of baroreflex dysfunction in salt-sensitive hypertension in the Dahl SS rat. *Physiol Genomics* 42: 23–41, 2010. First published March 30, 2010; doi:10.1152/physiolgenomics.00027.2010.—Salt-sensitive hypertension is known to be associated with dysfunction of the baroreflex control system in the Dahl salt-sensitive (SS) rat. However, neither the physiological mechanisms nor the genomic regions underlying the baroreflex dysfunction seen in this rat model are definitively known. Here, we have adopted a mathematical modeling approach to investigate the physiological and genetic origins of baroreflex dysfunction in the Dahl SS rat. We have developed a computational model of the overall baroreflex heart rate control system based on known physiological mechanisms to analyze telemetry-based blood pressure and heart rate data from two genetic strains of rat, the SS and consomic SS.13^{BN}, on low- and high-salt diets. With this approach, physiological parameters are estimated, unmeasured physiological variables related to the baroreflex control system are predicted, and differences in these quantities between the two strains of rat on low- and high-salt diets are detected. Specific findings include: a significant selective impairment in sympathetic gain with high-salt diet in SS rats and a protection from this impairment in SS.13^{BN} rats, elevated sympathetic and parasympathetic offsets with high-salt diet in both strains, and an elevated sympathetic tone with high-salt diet in SS but not SS.13^{BN} rats. In conclusion, we have

left unidentified because of these interactions. Thus, these types of measurements become diminishingly informative with an increased degree of genetic nonlinearity.

It seems, then, that more detailed phenotypic measurements are required to understand the underlying etiology and to make sense of the genetics associated with this complex disease. Of course, this is not always possible; many measurements of interest are either inaccessible or simply not practical to obtain. In addition, many of these measurements are operating-point dependent and are influenced to a high degree by physiologic state. Methods of obtaining these measurements often require invasive techniques that introduce stressors (surgical, pharmacological, etc.) that may themselves alter physiological state and therefore the observed measurements. Thus, differences detected in such experimental measurements may not always indicate differences in underlying physiology but can rather indicate differences in confounding variables related to experimental conditions and/or methods.

Mechanistic mathematical models offer a powerful complement to laboratory measurements (5). By accounting for the

Paper 4

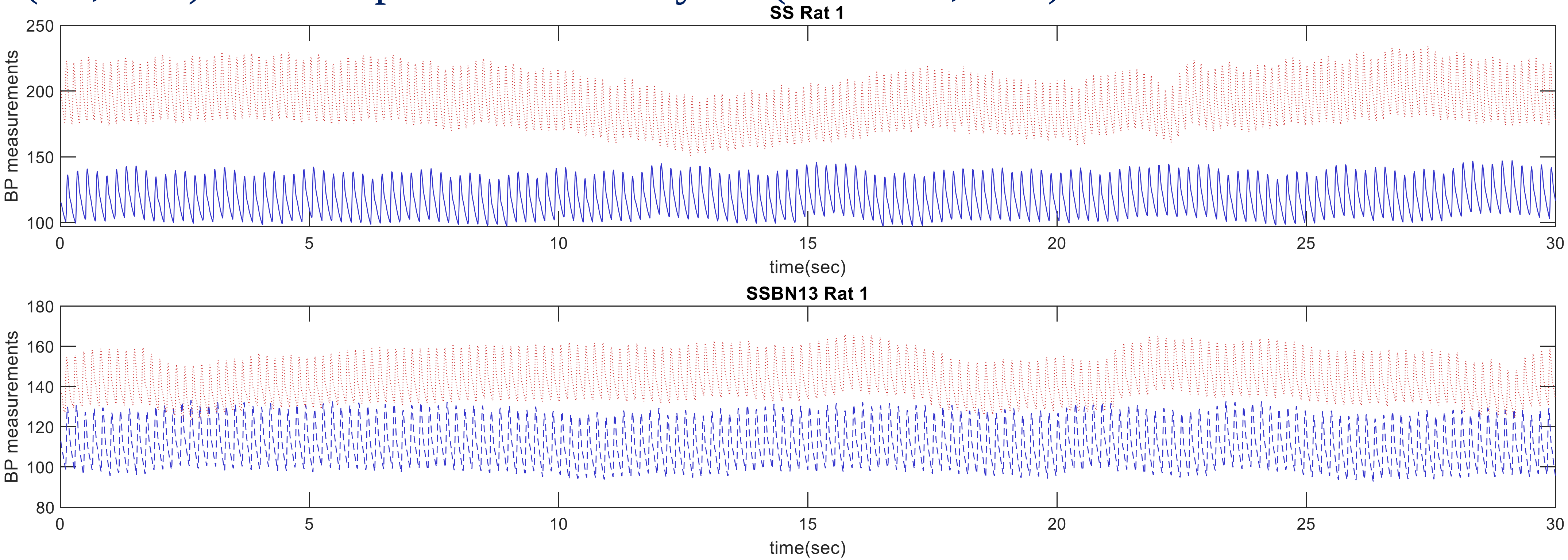
Project 1: Hypertension

Hypothesis: To test the hypothesis that high and low level of salt contents can identify dysfunction in baroreflex mechanisms to indicate hypertension

Paper 4

Project 1: Hypertension

Give two different levels of salt, low level (blue), high level (red) to dysfunction rat (SS; n=9) and compare with healthy rat (SSBN13; n=6)

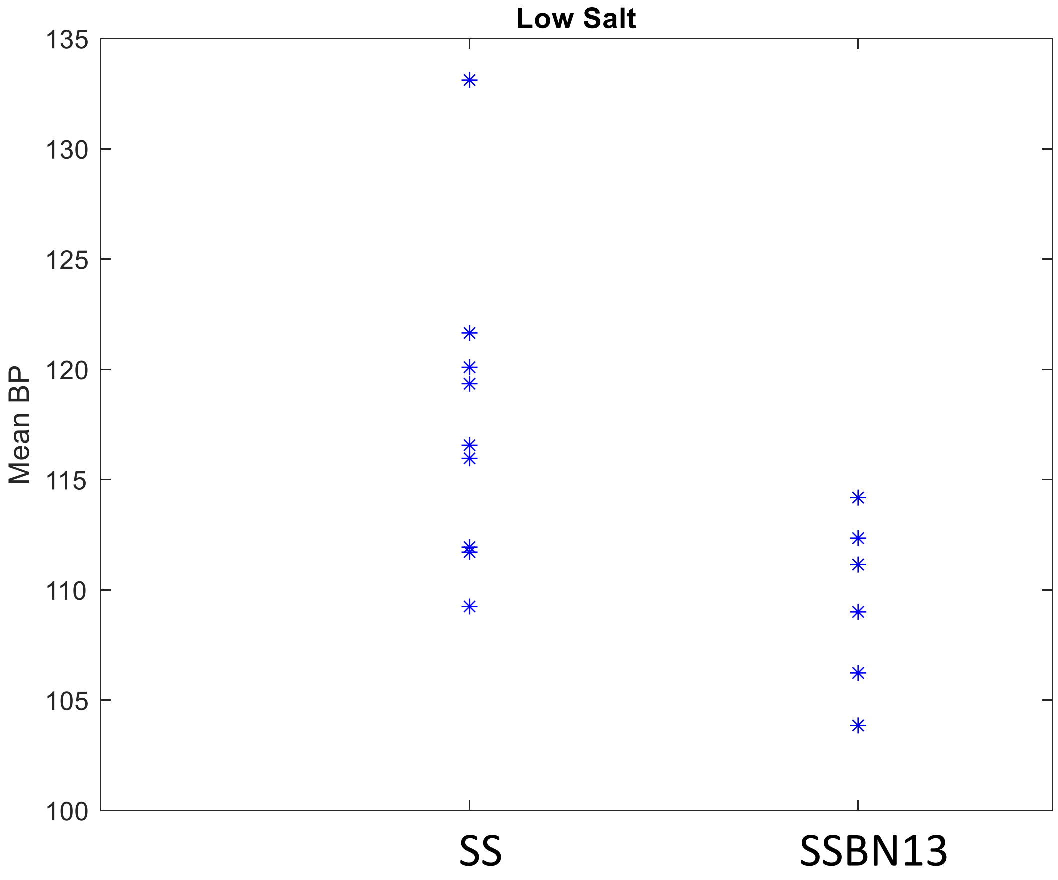


Project 1: Hypertension

Features:

Mean Blood Pressure (BP)

Standard Deviation of BP

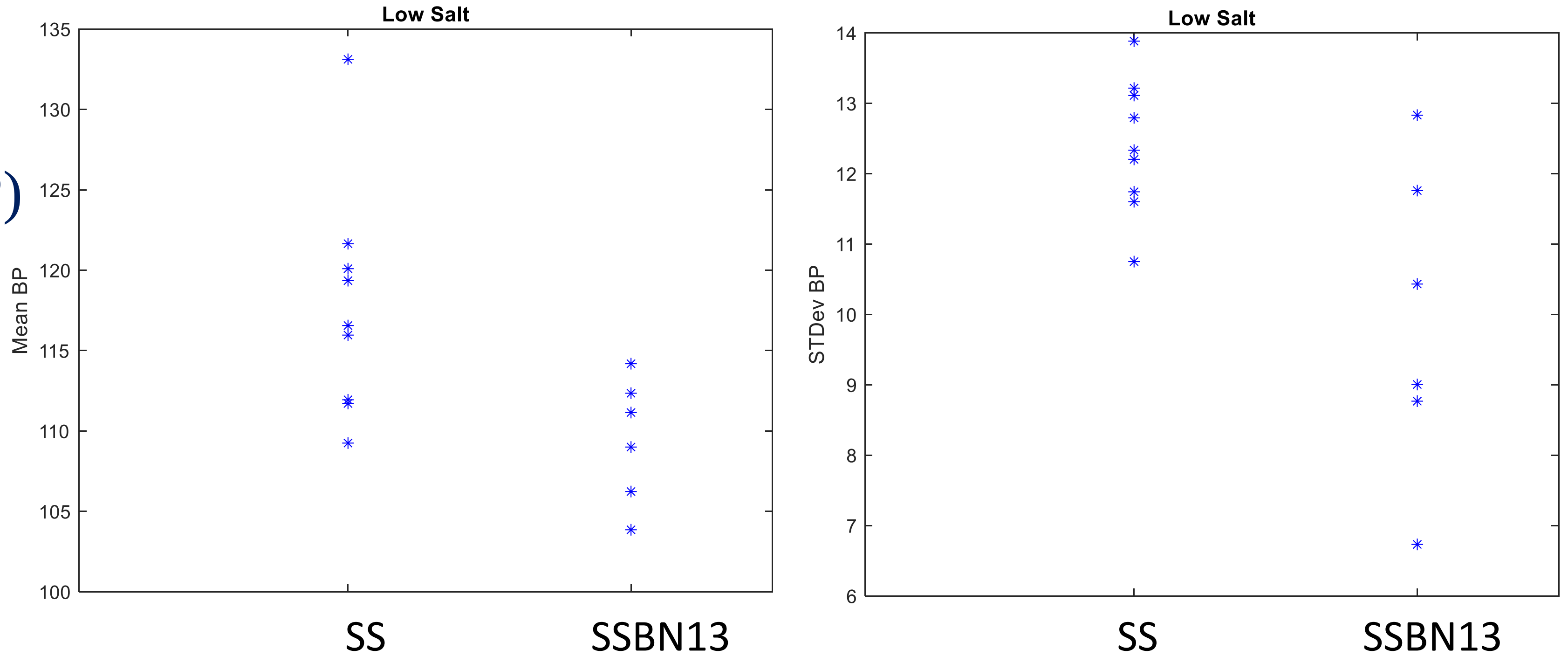


Project 1: Hypertension

Features:

Mean Blood Pressure (BP)

Standard Deviation of BP

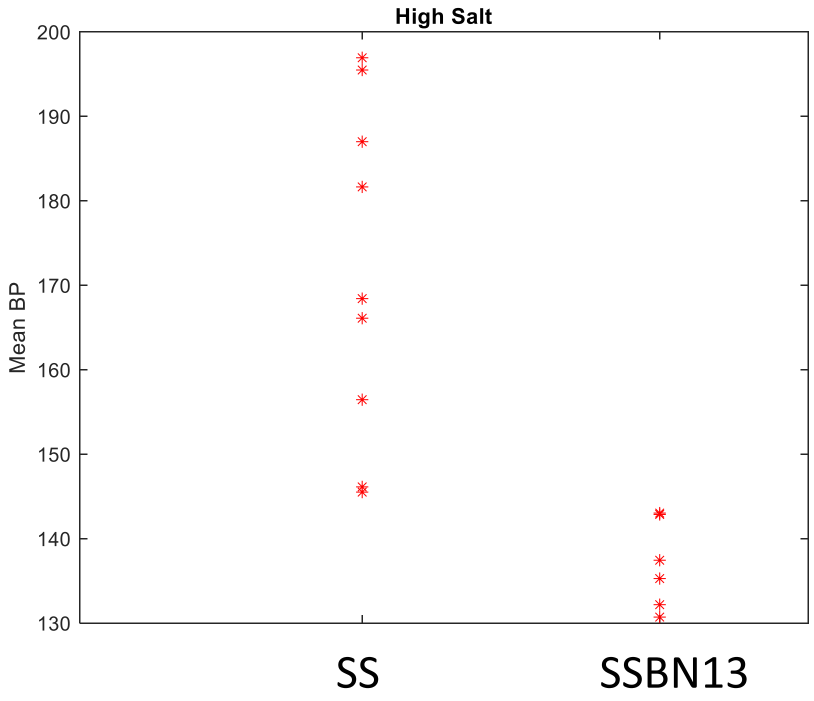


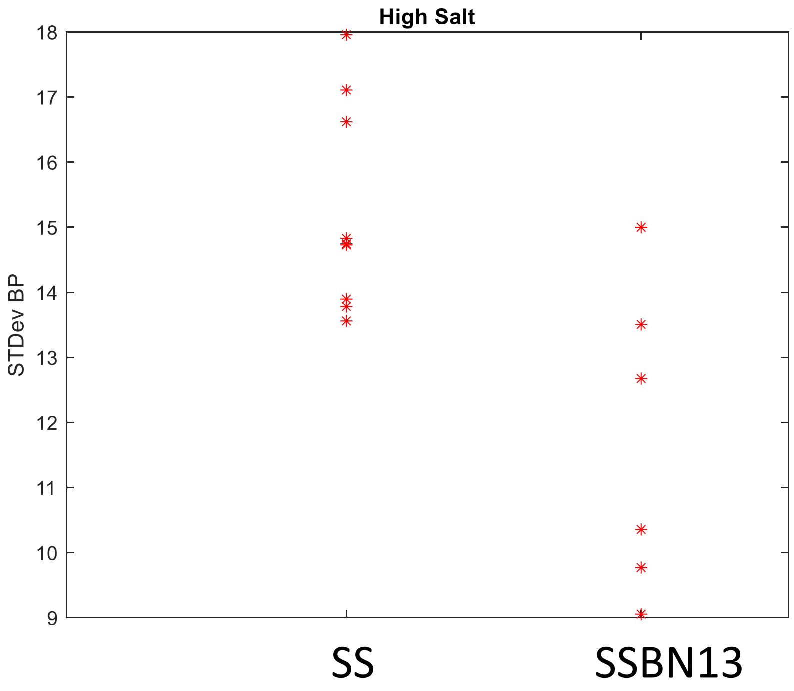
Project 1: Hypertension

Features:

Mean Blood Pressure (BP)

Standard Deviation of BP



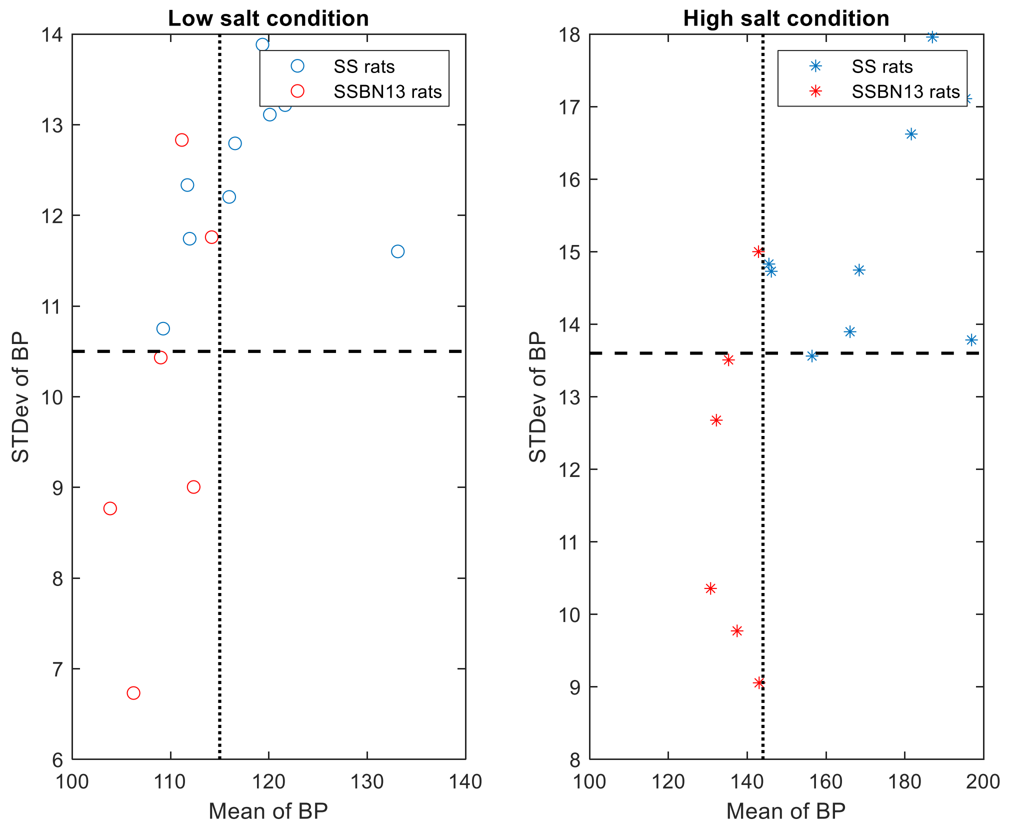


Project 1: Hypertension

Is there any predictability ?

Mean Blood Pressure (BP)

Standard Deviation of BP



Project 2: Dehydration Detection

Salivary Markers for Quantitative Dehydration Estimation During Physical Exercise

Matthias Ring, *Student Member, IEEE*, Clemens Lohmueller, Manfred Rauh, Joachim Mester, and Bjoern M. Eskofier, *Member, IEEE*

***Abstract*—Salivary markers have been proposed as noninvasive and easy-to-collect indicators of dehydrations during physical exercise. It has been demonstrated that threshold-based classifications can distinguish dehydrated from euhydrated subjects. However, considerable challenges were reported simultaneously, for example, high intersubject variabilities in these markers. Therefore, we propose a machine-learning approach to handle the**

osmolality have been shown to track total body water (TBW) loss during physical exercise [2]. The determination of plasma osmolality, however, involves invasive withdrawing of a blood sample and separation of the plasma compartment [3, Ch. 19].

Therefore, salivary osmolality and other salivary markers have been proposed as noninvasive and easy-to-collect alterna-

Paper 2

Project 2: Dehydration Detection

Hypothesis: To test the hypothesis that markers of saliva can detect dehydration

Paper 2

Project 2: Dehydration Detection

Features:

Amylase
Chloride
Cortisol
Cortisone
Osmolality
Potassium
Proteins

Paper 2

Project 3: Oxygen desaturation

Smokers vs Non Smokers

frontiers in Physiology

ORIGINAL RESEARCH
published: 02 August 2017
doi: 10.3389/fphys.2017.00555

Check for updates

Pattern Analysis of Oxygen Saturation Variability in Healthy Individuals: Entropy of Pulse Oximetry Signals Carries Information about Mean Oxygen Saturation

Amar S. Bhogal and Ali R. Mani *

UCL Division of Medicine, University College London, London, United Kingdom

Pulse oximetry is routinely used for monitoring patients' oxygen saturation levels with little regard to the variability of this physiological variable. There are few published studies on oxygen saturation variability (OSV), with none describing the variability and its pattern in a healthy adult population. The aim of this study was to characterize the pattern of OSV using several parameters; the regularity (sample entropy analysis), the self-similarity [detrended fluctuation analysis (DFA)] and the complexity [multiscale entropy (MSE) analysis]. Secondly, to determine if there were any changes that occur with age. The study population consisted of 36 individuals. The "young" population consisted of 20 individuals [Mean (±1 SD) age = 21.0 (±1.36 years)] and the "old" population consisted of 16 individuals [Mean (±1 SD) age = 50.0 (±10.4 years)]. Through DFA analysis, OSV was shown to exhibit fractal-like patterns. The sample entropy revealed

OPEN ACCESS

Edited by:
Radhakrishnan Nagarajan,
University of Kentucky, United States

Reviewed by:
Damian Kelty-Stephen,

Paper 3

Project 3: Oxygen desaturation

Smokers vs Non Smokers

Hypothesis: To test the hypothesis features of oxygen desaturation can detect the smokers from non smokers

Paper 3

Project 3: Oxygen desaturation

Smokers vs Non Smokers

Features:

Mean
Variance
Sample Entropy
Multiscale entropy

Paper 3

THANK YOU

Mohammed Alenazi, Graduate

Pravitha Ramanand, PhD, Postdoc

ORS Research Design & Data Analysis Lab

Office of Research and Scholarship