# FEATURE EXTRACTION USING MATLAB IN < 30 MINUTES

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

# OUTLINE

- INTRODUCTION
- DIFFERENT TYPES OF FEATURES
- DISCUSSION

# OUTLINE

- INTRODUCTION
- DIFFERENT TYPES OF FEATURES
- DISCUSSION

# INTRODUCTION

- Why do we need to derive features from data ?

- To test an hypothesis
- To classify different groups
- To predict outcomes

# INTRODUCTION


**Source**: https://en.wikipedia.org/wiki/30_Minute_Meals

**30**
Minute
Feature Extraction

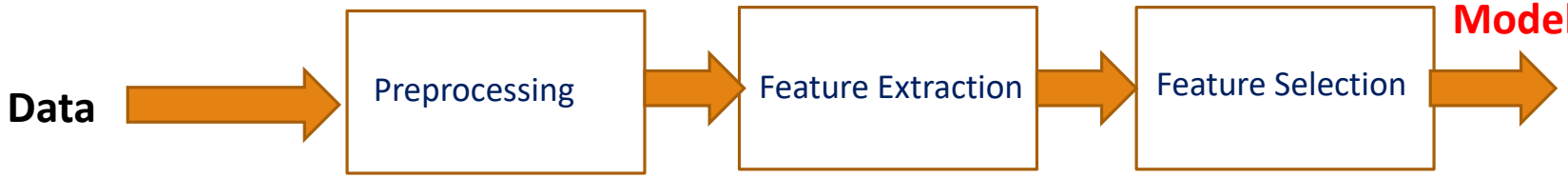
**Statistical or Machine Learning Models**

**Data** → Preprocessing → Feature Extraction → Feature Selection →

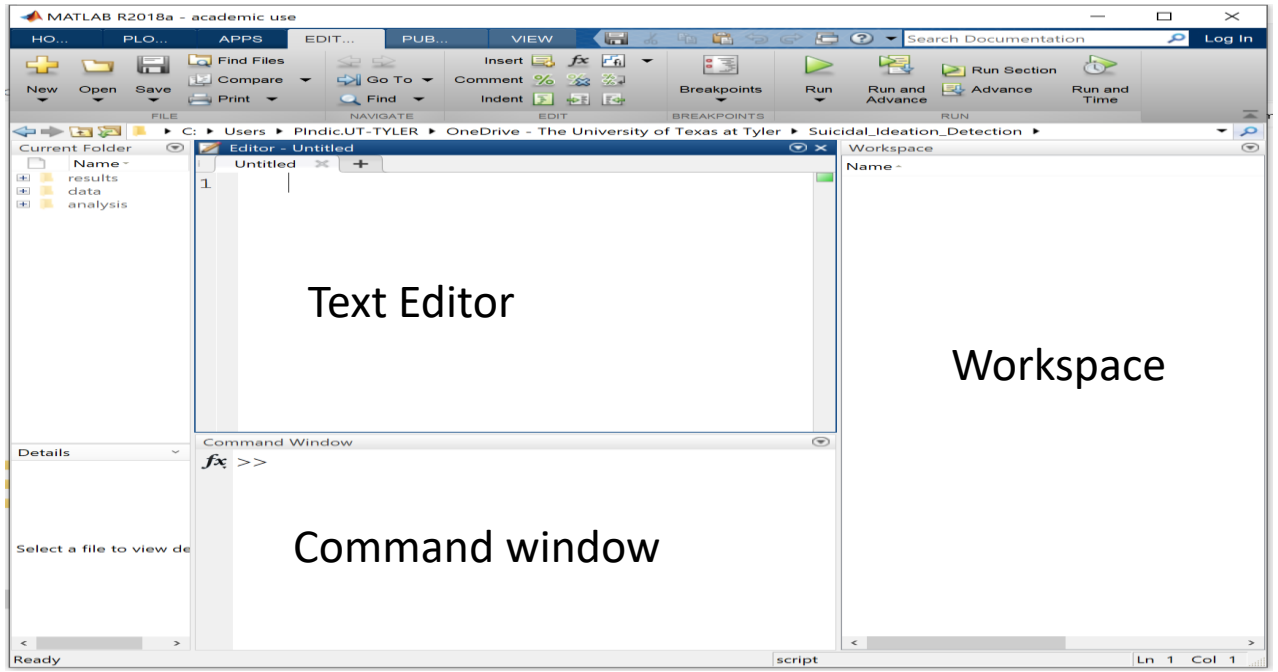
# INTRODUCTION


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

**30**
Minute
Feature Extraction



# INTRODUCTION

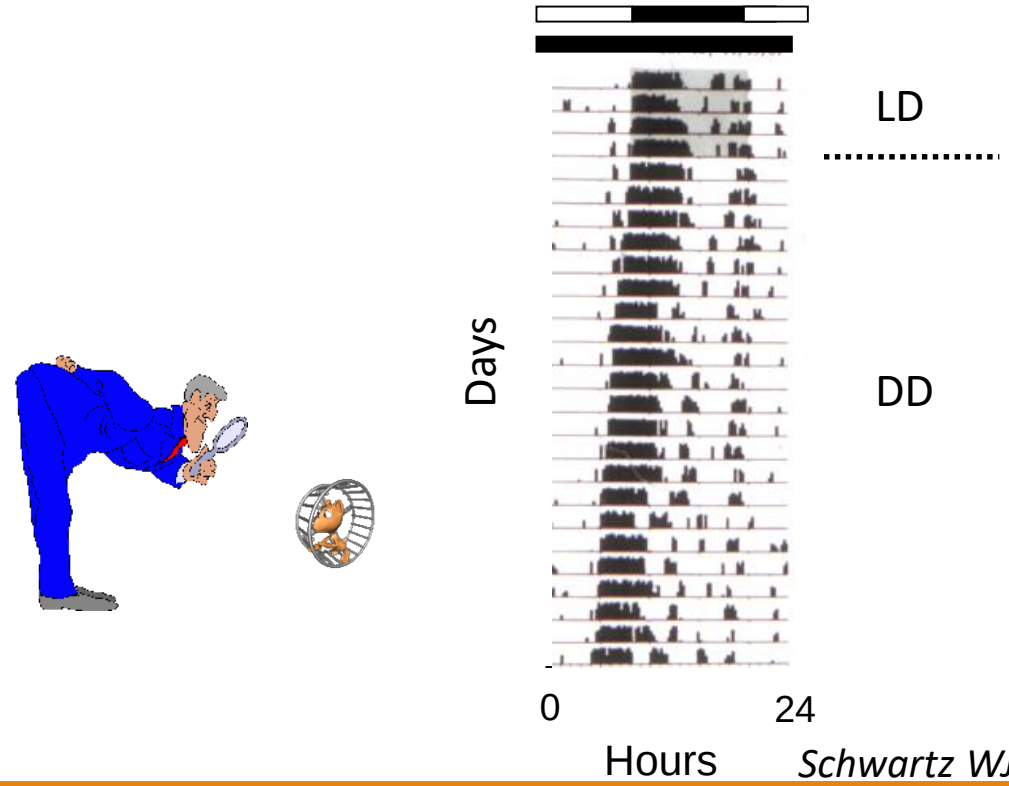



**30**
Minute
Feature Extraction

**Data** → Preprocessing → Feature Extraction → Feature Selection → **Statistical or Machine Learning Models**

# INTRODUCTION

- Feature extraction is the process of transforming raw data into numerical features while preserving the information of the original data set.

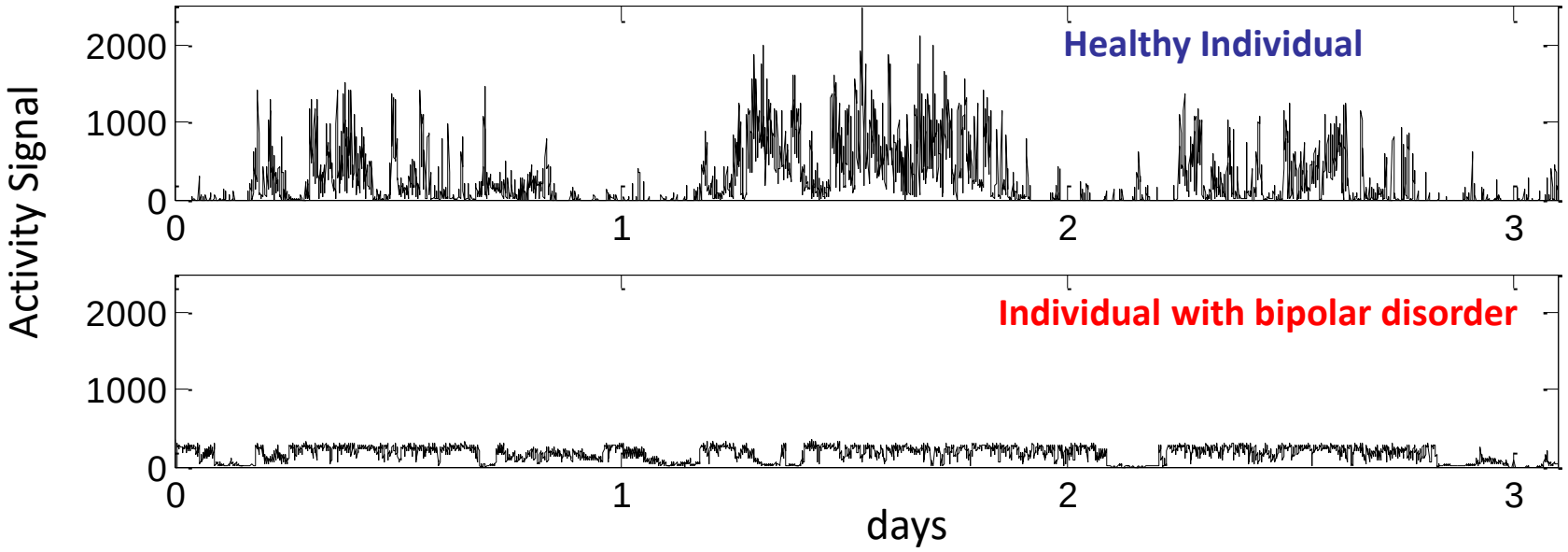



*Schwartz WJ & Zimmerman P, J. Neurosci, 1990, 10, 3685-3694*

# INTRODUCTION

- Feature extraction identifies most discriminating characteristics in signals.

**Philips Actiwatch 2**



# OUTLINE

- INTRODUCTION
- DIFFERENT TYPES OF FEATURES
- DISCUSSION

# DIFFERENT TYPES OF FEATURES

- VARIABLE
- STATISTICAL FEATURES
- SPECTRAL FEATURES
- NONLINEAR FEATURES

# DIFFERENT TYPES OF FEATURES

- VARIABLE

fitlm

| Weight | Acceleration | MPG |
|---|---|---|
| 3504 | 12 | 18 |
| 3693 | 11.5 | 15 |
| 3436 | 11 | 18 |
| 3433 | 12 | 16 |
| 3449 | 10.5 | 17 |

```
lm = fitlm(tbl,'MPG~Weight+Acceleration')
```

```
lm =


Linear regression model:
    MPG ~ 1 + Weight + Acceleration

Estimated Coefficients:
                    Estimate        SE         tStat       pValue
                   _________    _________    _______    _________

    (Intercept)       45.155       3.4659     13.028    1.6266e-22
    Weight         -0.0082475    0.00059836   -13.783    5.3165e-24
    Acceleration      0.19694      0.14743     1.3359      0.18493
```

**MPG = a + b Weight + c Acceleration**

```
Number of observations: 94, Error degrees of freedom: 91
Root Mean Squared Error: 4.12
R-squared: 0.743,  Adjusted R-Squared 0.738
```

# DIFFERENT TYPES OF FEATURES

- VARIABLE

Real Estate Data

Command Window

```
>> lm=fitlm(housing)

lm =


Linear regression model:
    median_house_value ~ [Linear formula with 9 terms in 8 predictors]

Estimated Coefficients:
```

| | Estimate | SE | tStat | pValue |
|---|---|---|---|---|
| **(Intercept)** | -3.5854e+06 | 62901 | -57.001 | 0 |
| **longitude** | -42730 | 717.09 | -59.588 | 0 |
| **latitude** | -42510 | 676.95 | -62.796 | 0 |
| **housing_median_age** | 1157.9 | 43.389 | 26.687 | 2.9463e-154 |
| **total_rooms** | -8.2497 | 0.79426 | -10.387 | 3.2948e-25 |
| **total_bedrooms** | 113.82 | 6.9306 | 16.423 | 3.1889e-60 |
| **population** | -38.386 | 1.0841 | -35.407 | 1.4597e-266 |
| **households** | 47.701 | 7.5466 | 6.3209 | 2.6535e-10 |
| **median_income** | 40298 | 337.21 | 119.5 | 0 |

Workspace

Name

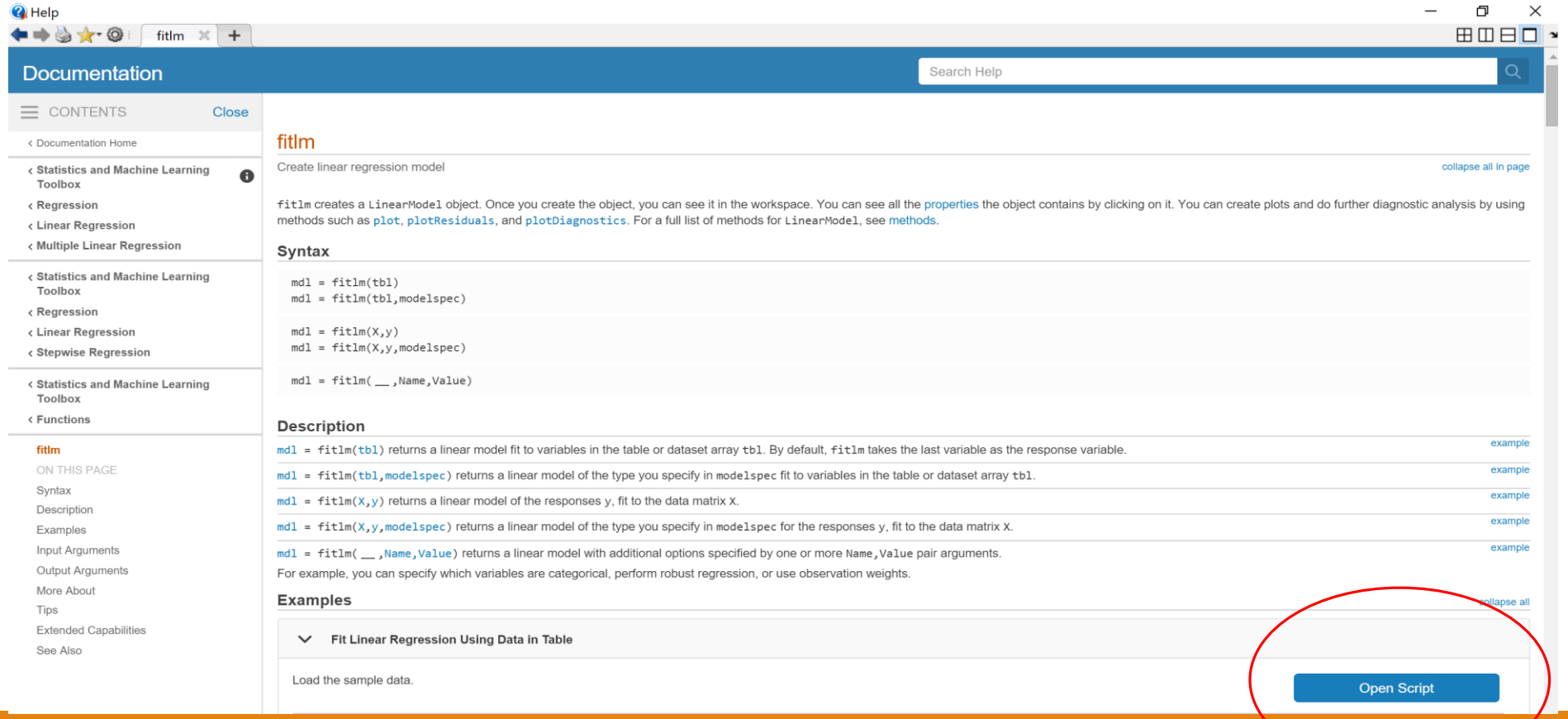
housing

lm

# DIFFERENT TYPES OF FEATURES

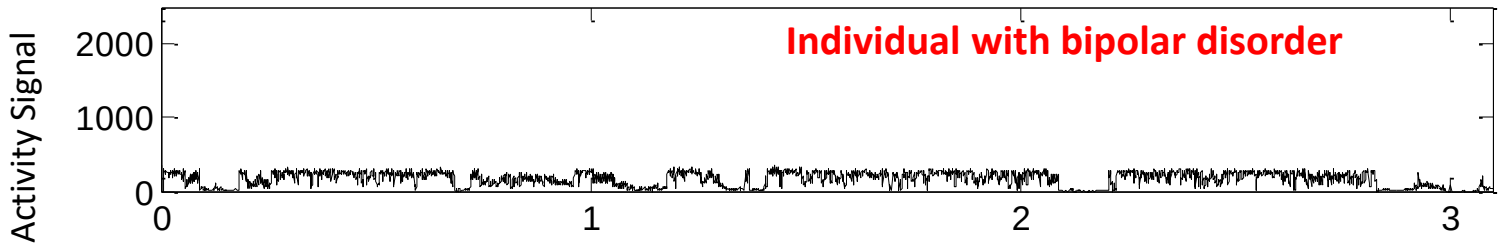
- VARIABLE

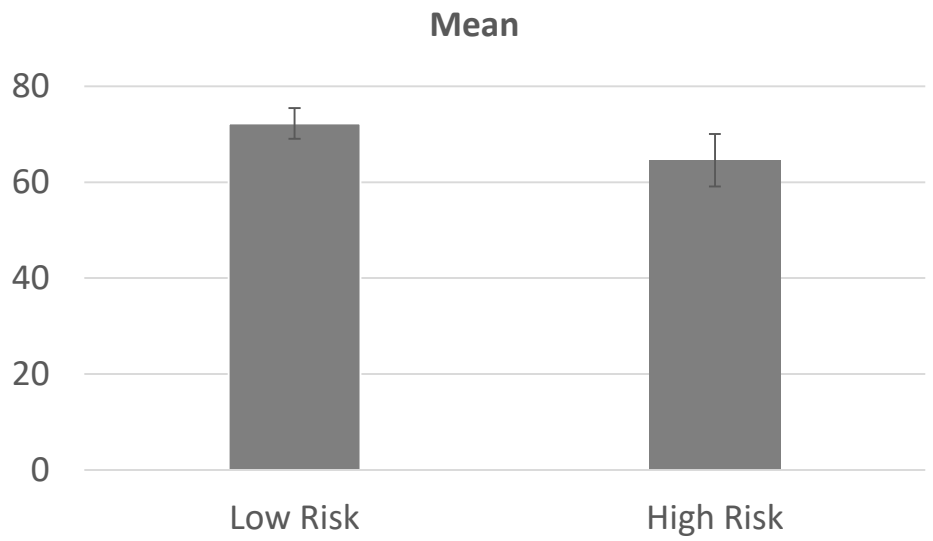
**Don’t want to write the code?**



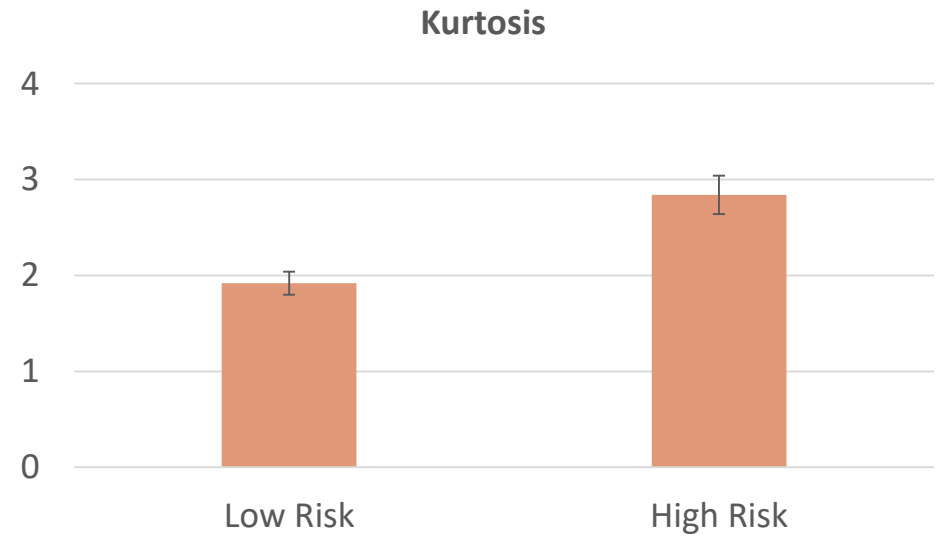
# DIFFERENT TYPES OF FEATURES

- STATISTICAL FEATURES



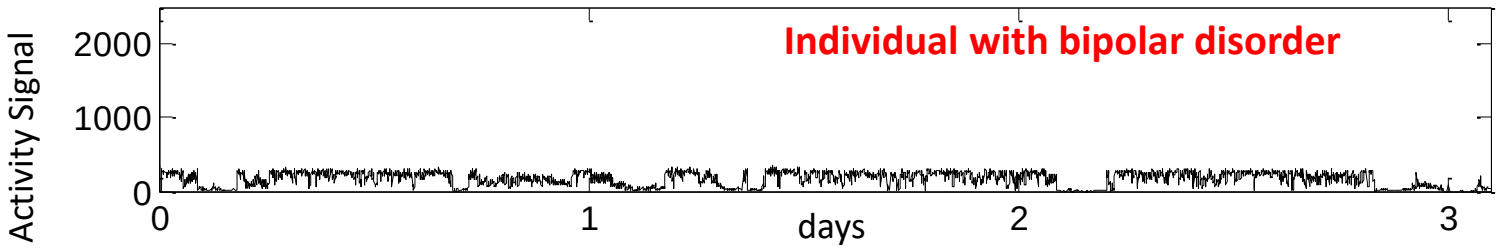




N =128



# DIFFERENT TYPES OF FEATURES

- STATISTICAL FEATURES




**Correlation with Self Reported Suicidal Ideation**

Mean : ***r*** = -0.17 ***p*** = 0.05

Variance : ***r*** = -0.05 ***p*** = 0.53

Skewness : ***r*** = 0.23 ***p*** = 0.007

Kurtosis : ***r*** = 0.18 ***p*** = 0.03

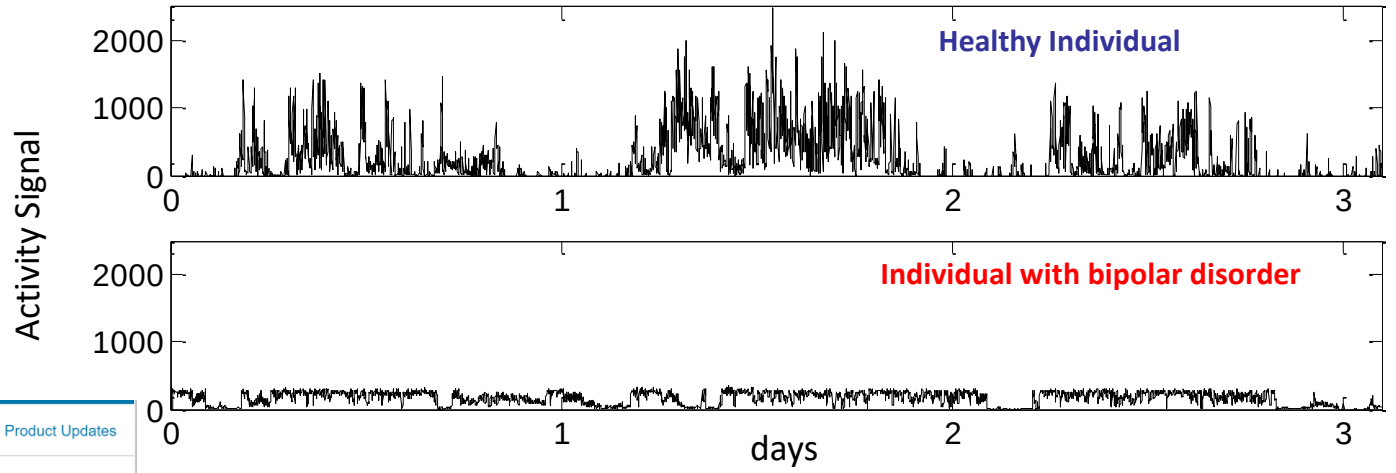
**MATLAB functions**

```
mean(filename)
variance(filename)
skewness (filename)
kurtosis(filename)
[r,p]=corr(resultsfilename(:,1),resultsfilename(:,5));
```

*Atluri, Salvatore and Indic (unpublished)*

# DIFFERENT TYPES OF FEATURES

- SPECTRAL FEATURES




Documentation | Examples | Functions | Apps | Videos | Answers | Trial Software | Product Updates

## Spectral Features

R2020b

Spectral features provide frequency-domain metrics on your data. To compute spectral features, you must already have a power spectrum or an order spectrum variable.

### Spectrum

- **Spectrum** — Choose from the available spectrum variables. The software brings up the plot of that variable for reference, and converts the plot from log scale to linear scale.

### Spectral Peaks

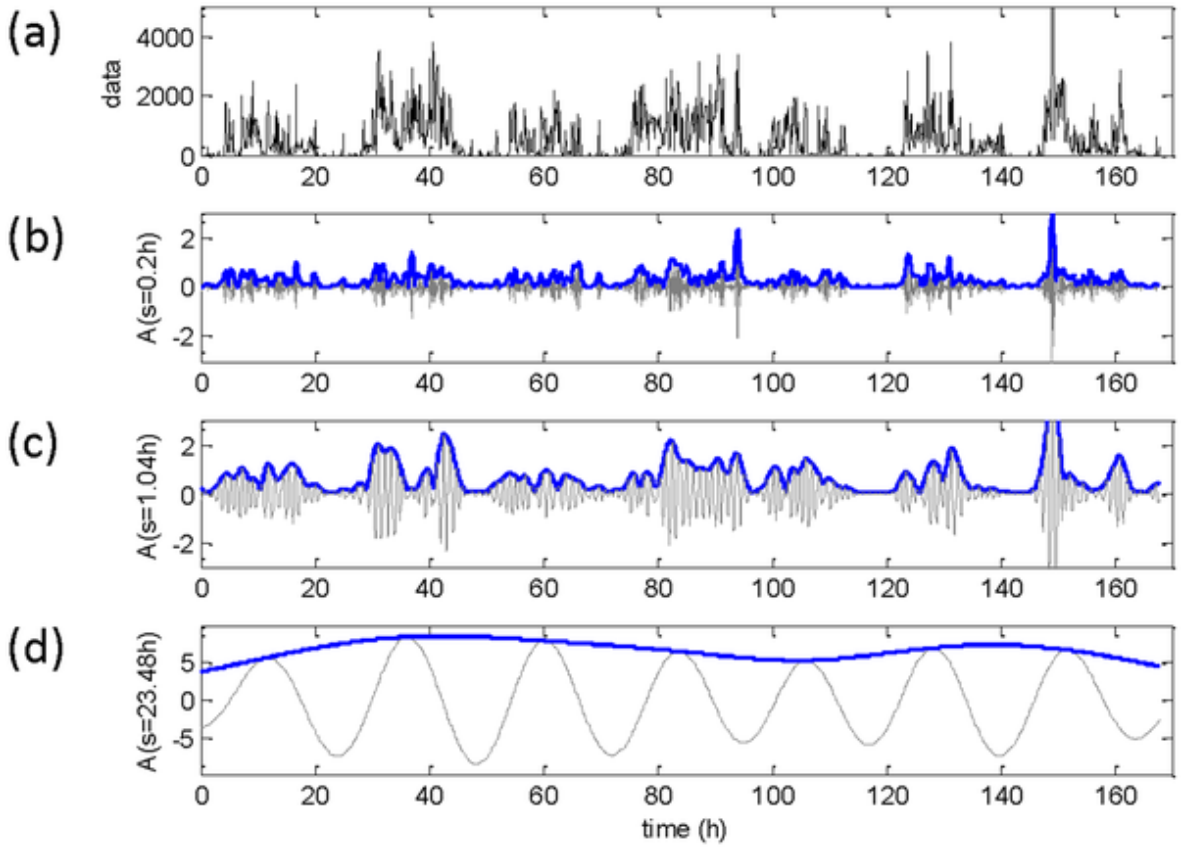
- **Peak amplitude** — Generate a feature based on the amplitude of the peaks.
- **Peak frequency** — Generate a feature based on the frequency of the peaks.
- **Peak value lower threshold** — Constrain peak size to exclude low-amplitude peaks. For more information, in `findpeaks`, see the `MinPeakHeight` name-value pair argument.
- **Number of peaks** — Number of peaks to generate features for. The software selects N most prominent peaks in the chosen frequency band, going in the descending amplitude order. For more information, in `findpeaks`, see the `NPeaks` name-value pair argument.
- **Minimum frequency gap** — Specify a minimum frequency gap. If the gap between two peaks is less than this specification, the software ignores the smaller peak of the pair. For more information, in `findpeaks`, see the `MinPeakDistance` name-value pair argument.
- **Peak excursion tolerance** — Specify the minimum prominence of a peak. The prominence of a peak measures how much the peak stands out due to its intrinsic height and its location relative to other peaks. For more information, in `findpeaks`, see the `MinPeakProminence` name-value pair argument.

# DIFFERENT TYPES OF FEATURES

- SPECTRAL FEATURES

**Wavelet transform**

wavelets(filename)




*Indic P, Salvatore P, Maggini C, et al. (2011) Scaling Behavior of Human Locomotor Activity Amplitude: Association with Bipolar Disorder. PLOS ONE 2011 6(5): e20650.*

# DIFFERENT TYPES OF FEATURES

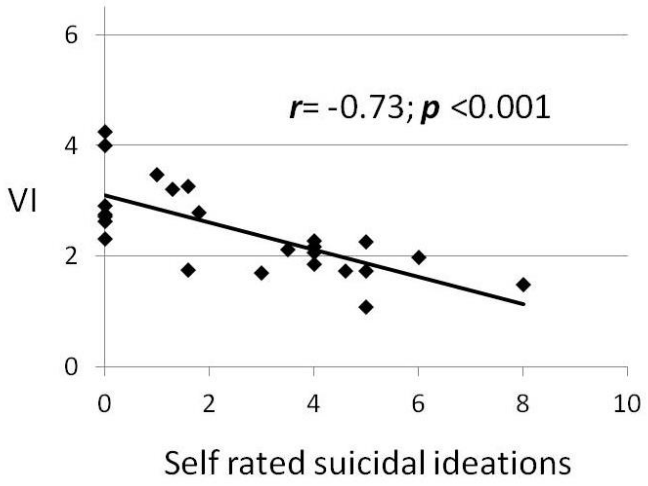
- SPECTRAL FEATURES

**MATLAB functions:**

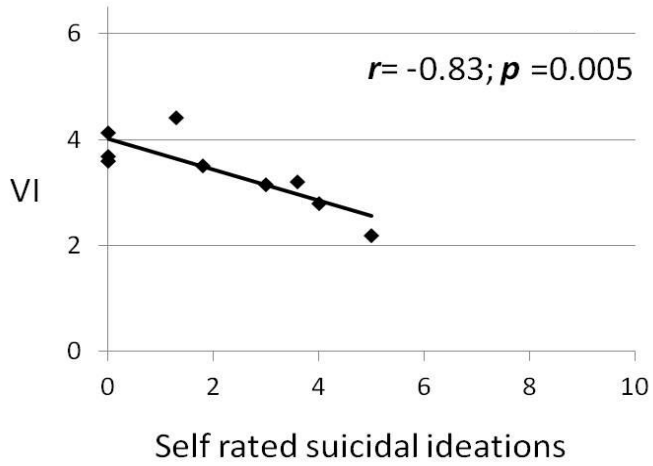
Wavelets
corr

**Individuals during major depression phase**

*N=24*



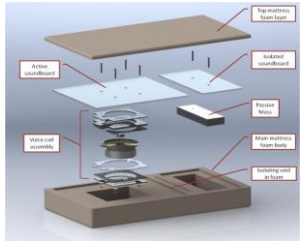

*N=1*

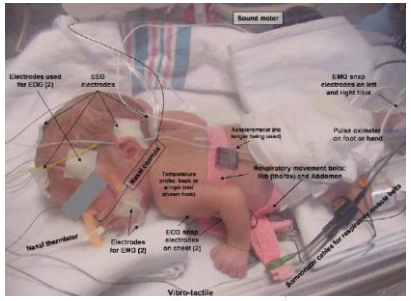


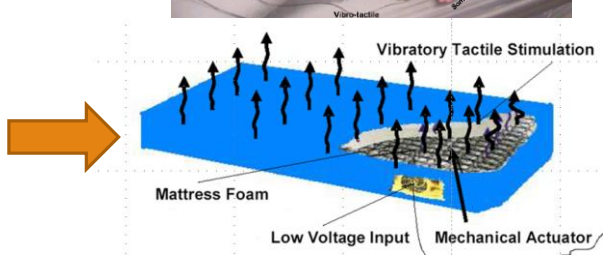

*Indic P, Murray G, Maggini C, et al. (2012) Multiscale Motility Amplitude Associated with Suicidal Thoughts in Major Depression. PLOS ONE 2012, 7: e38761*

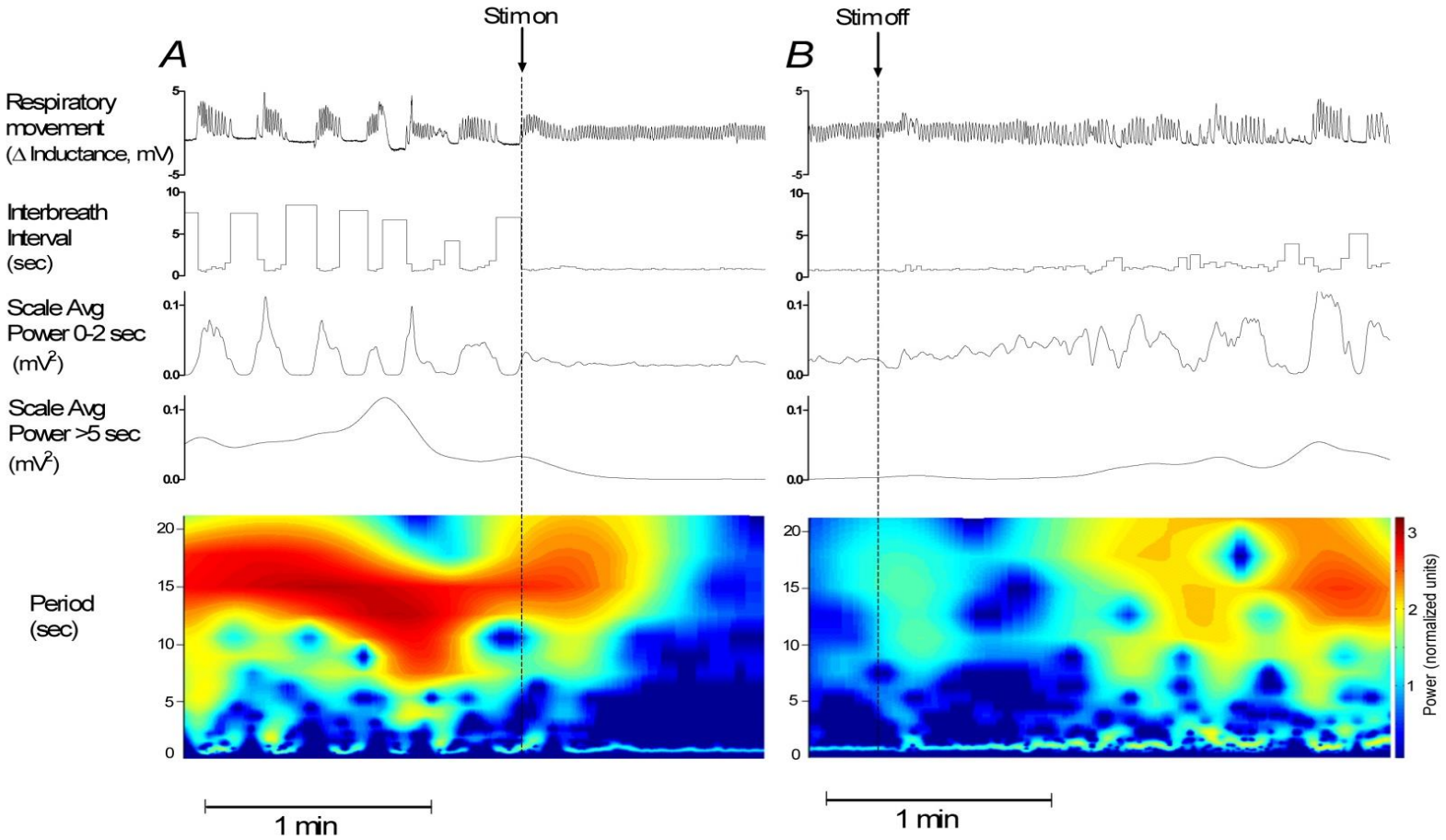
# DIFFERENT TYPES OF FEATURES

- SPECTRAL FEATURES





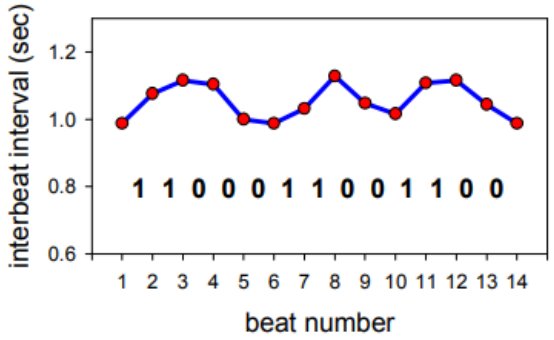


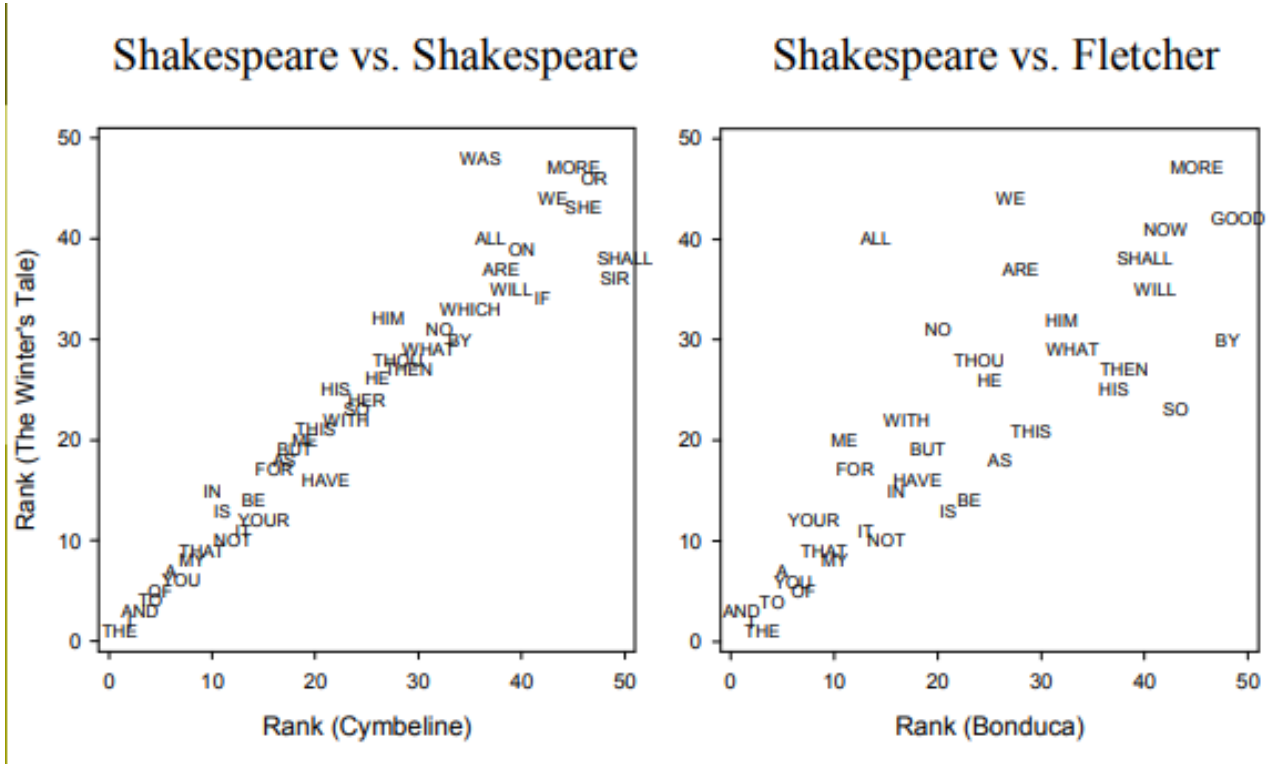

*Bloch-Salisbury E, Indic P, Bednarek F, and Paydarfar D, J Appl Physiol., 2009, 107: 1017-1027*

# DIFFERENT TYPES OF FEATURES

## ➢NONLINEAR FEATURES



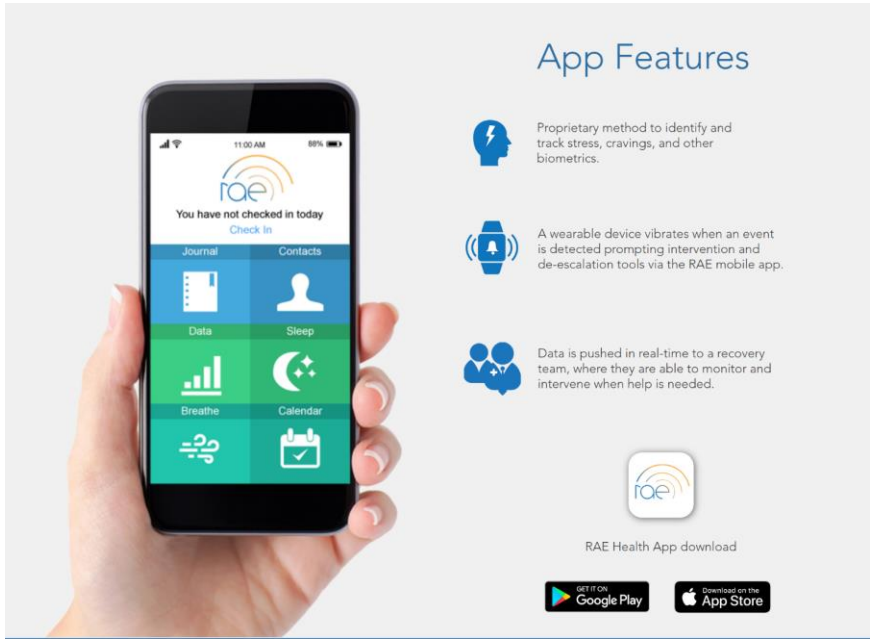

- Fluctuation Analysis
- Pattern Analysis
- Fractal Analysis
- Information Categorization Approach
- Power Law
- Entropy
- Dimension



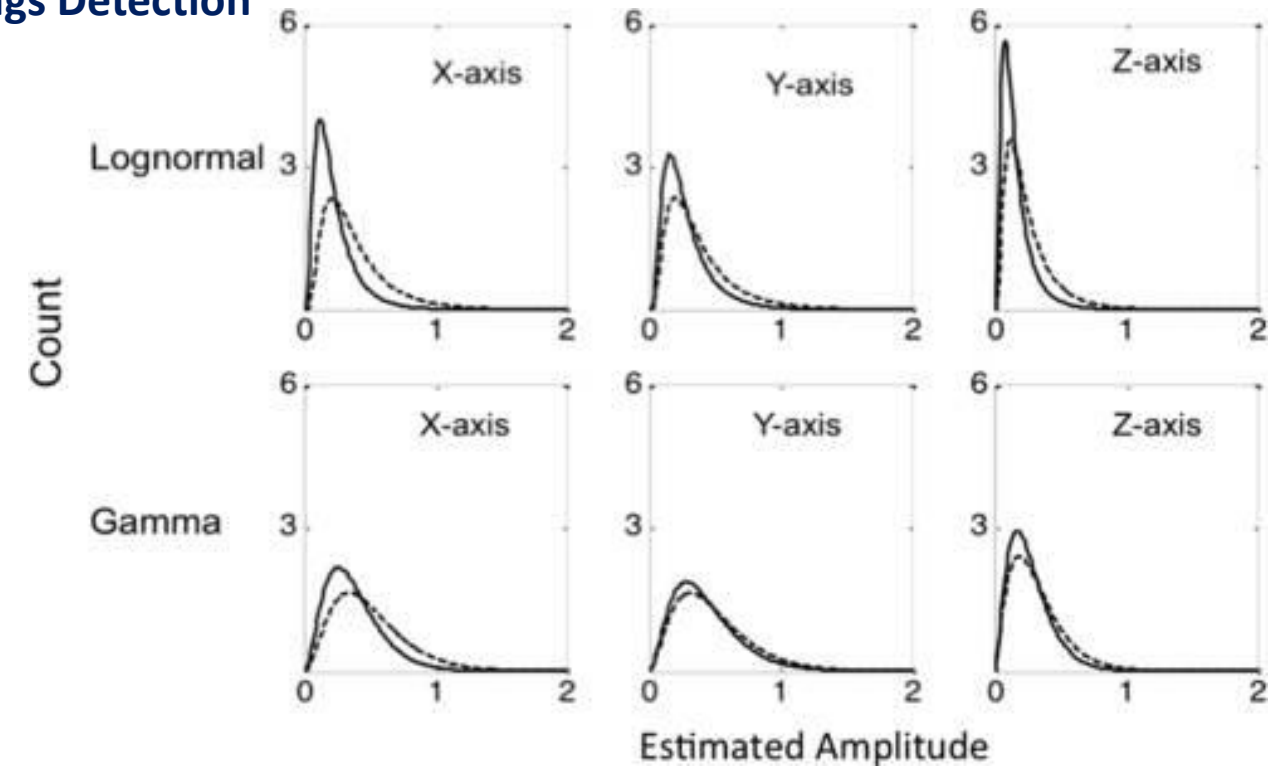

*Yang CC, Peng CK, Yien HW, Goldberger AL. Information categorization approach to literary authorship disputes. Physica A. 2003; 329:473-483.*

# DIFFERENT TYPES OF FEATURES

- NONLINEAR FEATURES




**Cravings Detection**




Carreiro, S, Chintha KK, Shrestha S*, Chapman B, Smelson D, Indic P. Wearable sensor based detection of stress and craving in patients during treatment for substance use disorder: A mixed methods pilot study. Drug and Alcohol Dependence. 2020, 107929

# DIFFERENT TYPES OF FEATURES

- Additional MATLAB Code

**Physionet:** https://physionet.org/

**TISEAN:** https://www.pks.mpg.de/~tisean/

# SUMMARY

**ORS Research Design & Data Analysis Lab**

**Office of Research and Scholarship**

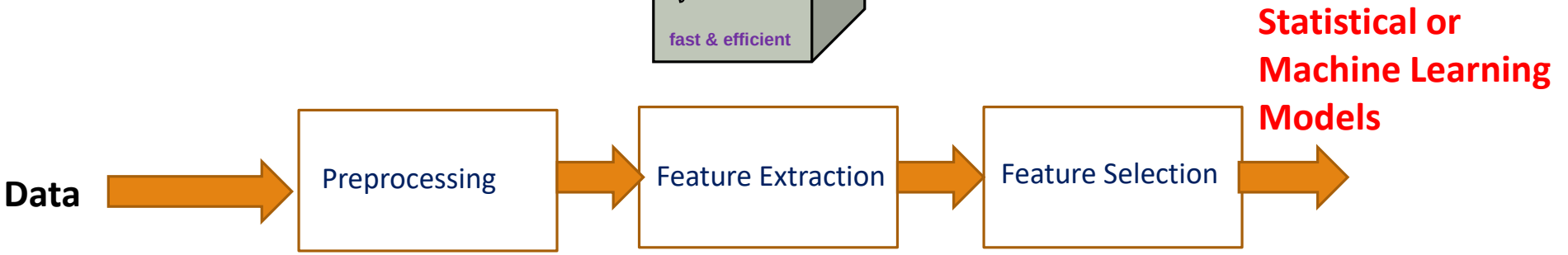
- Identification of Features
- Develop MATLAB code for feature extraction
- Set up database
- Assist with experiment protocol and data analysis
- Machine Learning

**30**
Minute
Feature Extraction

# SUMMARY

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

- **Ready to go features**

# THANK YOU

**Current Students:**

**Sloke Shrestha, UG**

**Mohammed Alenazi, Graduate**

**Pravitha Ramanand, PhD, Postdoc**

**Joshua Stapp, Graduate**

**Former Students:**

Apurupa Amperayani (PhD Student, Arizona State University)
Jonathan Wells (PhD Student, UT Austin)
Pallavi Atluri
Keerthi Chintha (Data Scientist, Wabtec Corporation)
Selorm Darkey (Business Intelligent Analyst, Taylor Solutions)

# THANK YOU

## ORS Research Design & Data Analysis Lab

### Office of Research and Scholarship

SBIR: RAE (Realize, Analyze, Engage) - A digital biomarker based detection and intervention system for stress and carvings during recovery from substance abuse disorders.
***PIs: M. Reinhardt, S. Carreiro, P. Indic***

STARs Award

The University of Texas System
***P. Indic (PI, UT Tyler)***

Department of Veterans Affairs

*Design of a wearable sensor system and associated algorithm to track suicidal ideation from movement variability and develop a novel objective marker of suicidal ideation and behavior risk in veterans.*
Clinical Science Research and Development Grant (approved for funding),
***P. Indic (site PI, UT-Tyler)***
***E.G. Smith (Project PI, VA)***
***P. Salvatore (Investigator, Harvard University)***

*Design of a wearable biosensor sensor system with wireless network for the remote detection of life threatening events in neonates*.

National Science Foundation Smart & Connected Health Grant
***P. Indic (Lead PI, UT-Tyler)***
***D. Paydarfar (Co PI, UT-Austin)***
***H. Wang (Co PI, UMass Dartmouth)***
***Y. Kim (Co PI, UMass Dartmouth)***

*Pre-Vent*

National Institute Of Health Grant
***P. Indic (Analytical Core PI, UT-Tyler)***
***N. Ambal (PI, Univ. of Alabama, Birmingham)***

*Wearable system for the detection of addiction*
***P. Indic (PI, UT-Tyler)***
***M. Reinhart (PI, ContinueYou, LLC***
***S. Carriero, (PI. Univ. of Mass. Med. School)***

# DISCUSSION