Laboratory of Bioelectric and Bioenergetic Systems

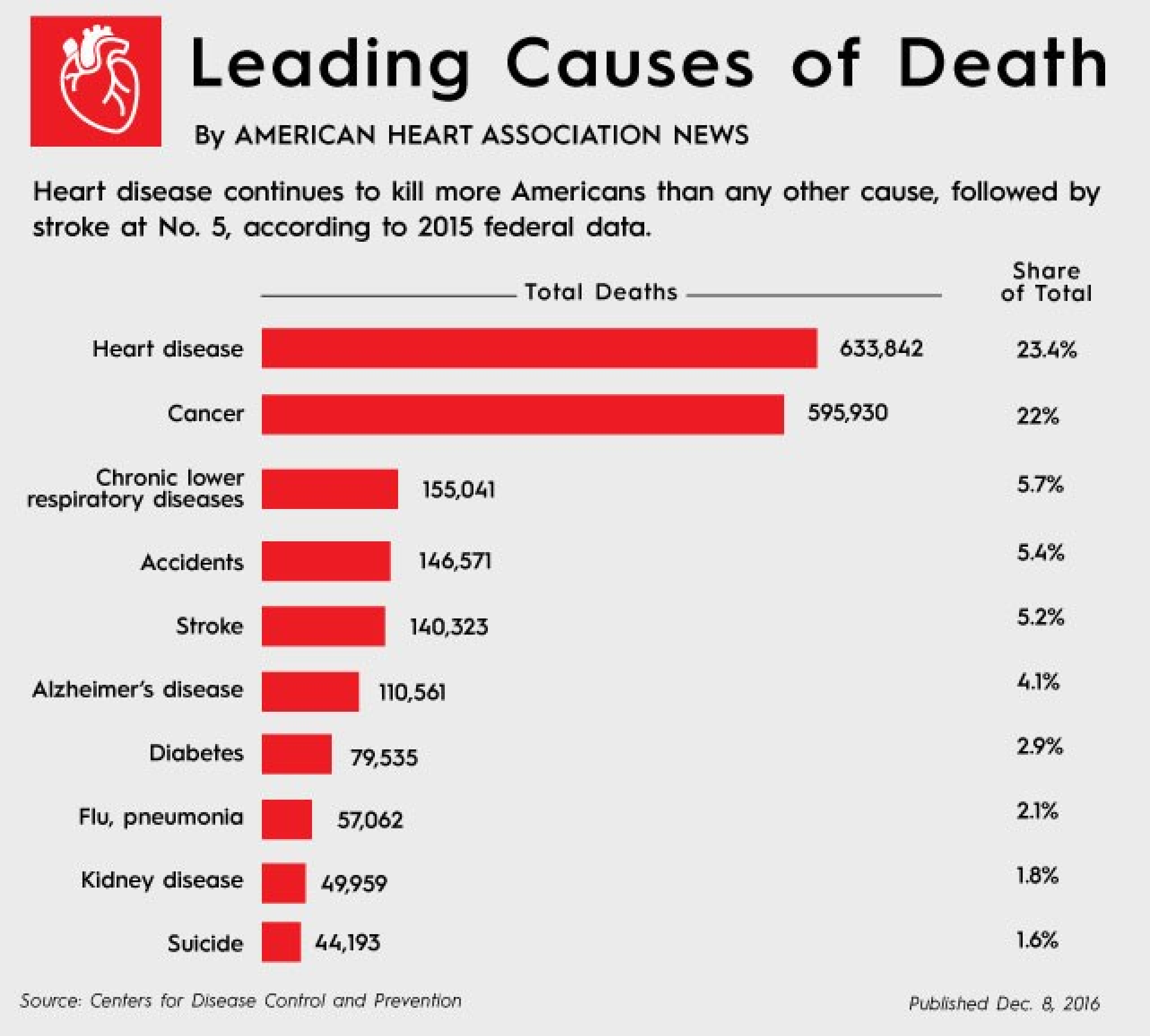
TECHNION
Israel Institute of Technology

How to gain biomedical big data?

Associate Professor Yael Yaniv, PhD
Biomedical Engineering Faculty
Technion, IIT
Haifa, Israel

MOTIVATION: Heart disease and unsatisfied treatment

- Heart diseases are the number 1 killer in the western world.
- Up to date, there is not satisfactory drug treatments.
- “No change in the number of deaths attributed to heart failure has been observed between 1995 and 2011.”

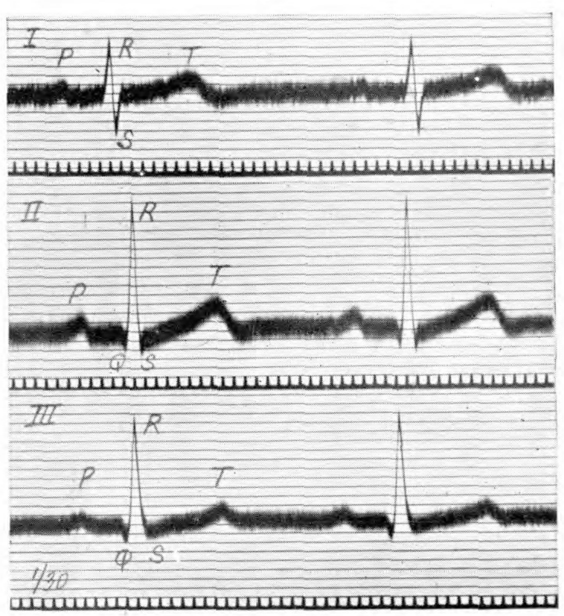


MOTIVATION: Heart disease and unsatisfied diagnostics

"No change in the number of deaths attributed to heart failure has been observed between 1995 and 2011."

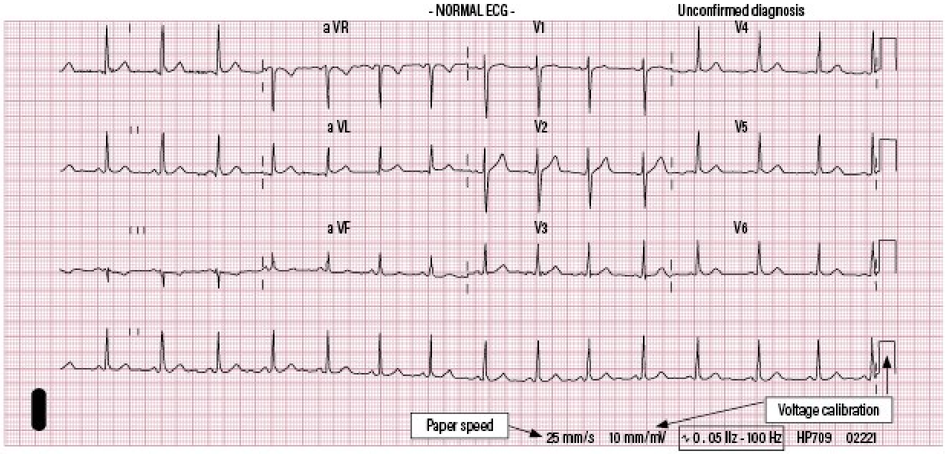
1995

2011



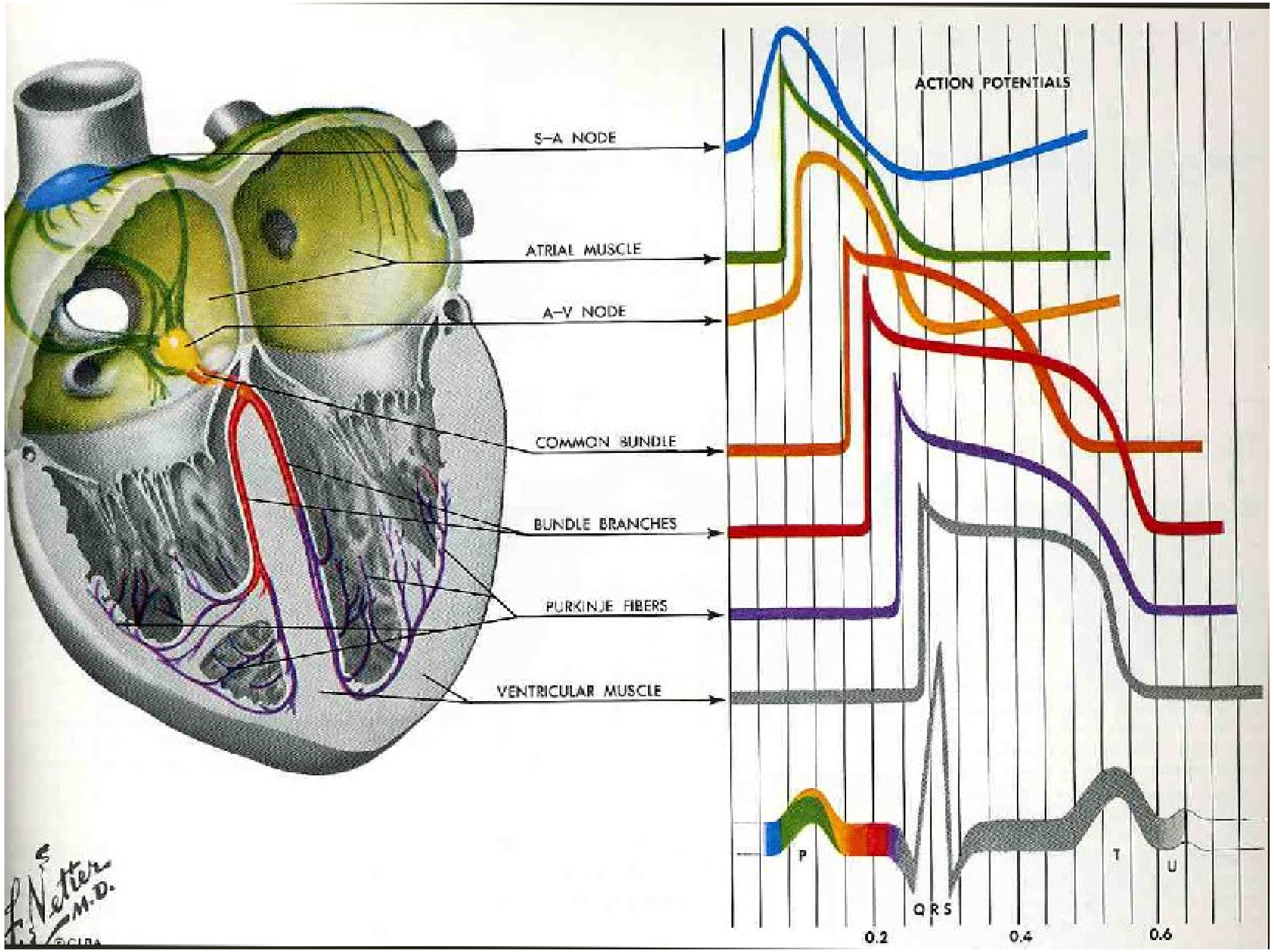
Thomas Lewis,
The Mechanism and Graphic Registration of the Heart Beat, 3rd ed. Shaw & Sons Ltd., London, **1925**

2019



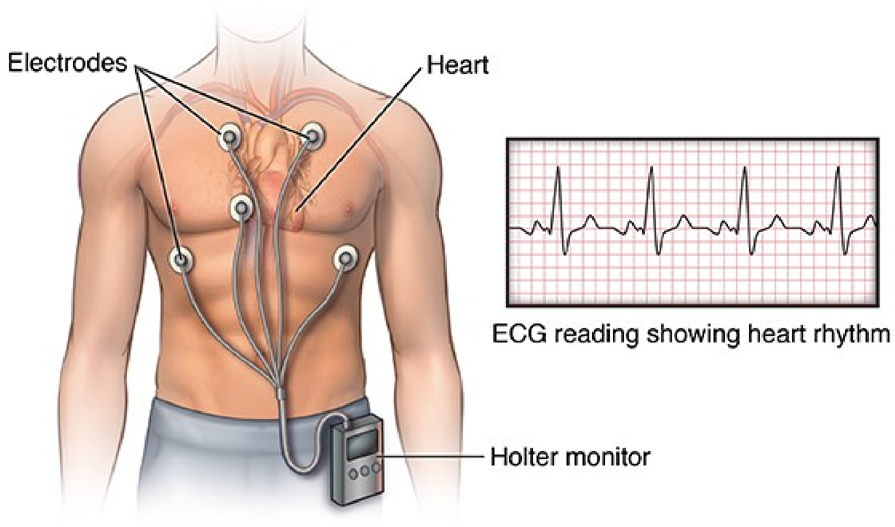
2019

INTRODUCTION: Conduction system of the heart

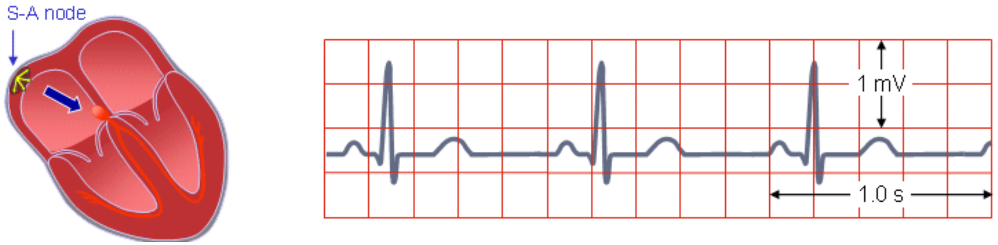


One lead electrocardiogram

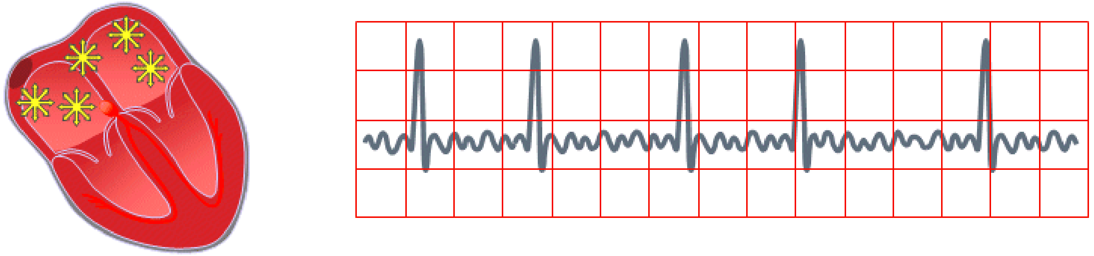
Holter monitor with ECG reading



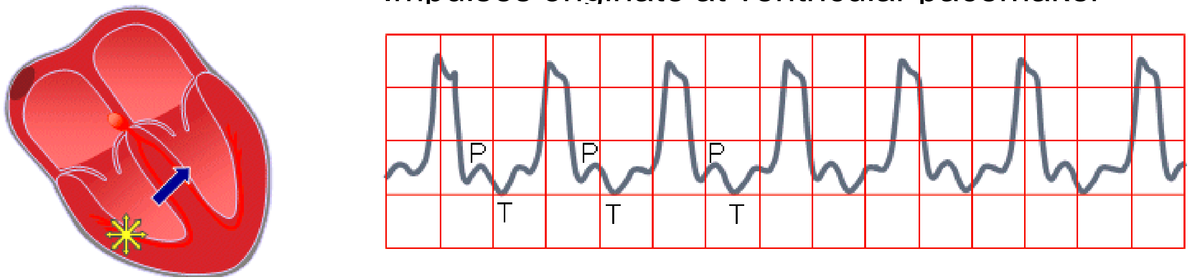
Normal sinus rhythm
Impulses originate at SA node at normal rate



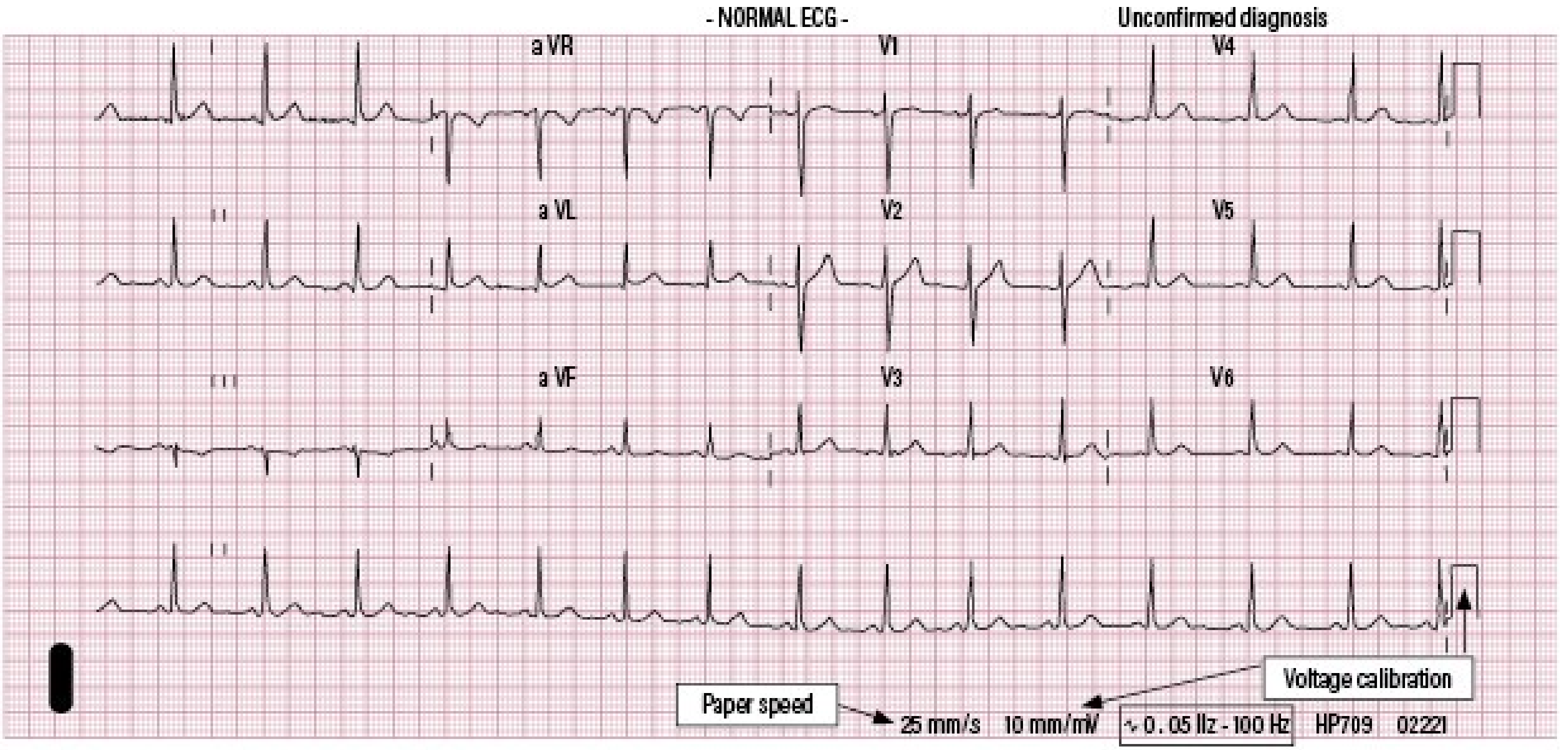
Atrial Fibrillation
Impulses have chaotic, random pathways in atria



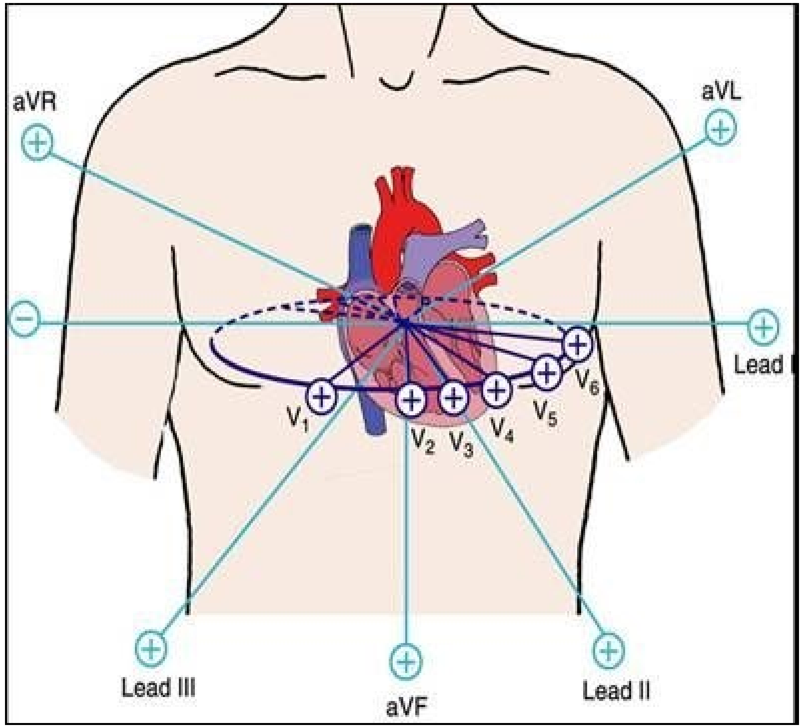
Ventricular tachycardia
Impulses originate at ventricular pacemaker



12 lead ECG

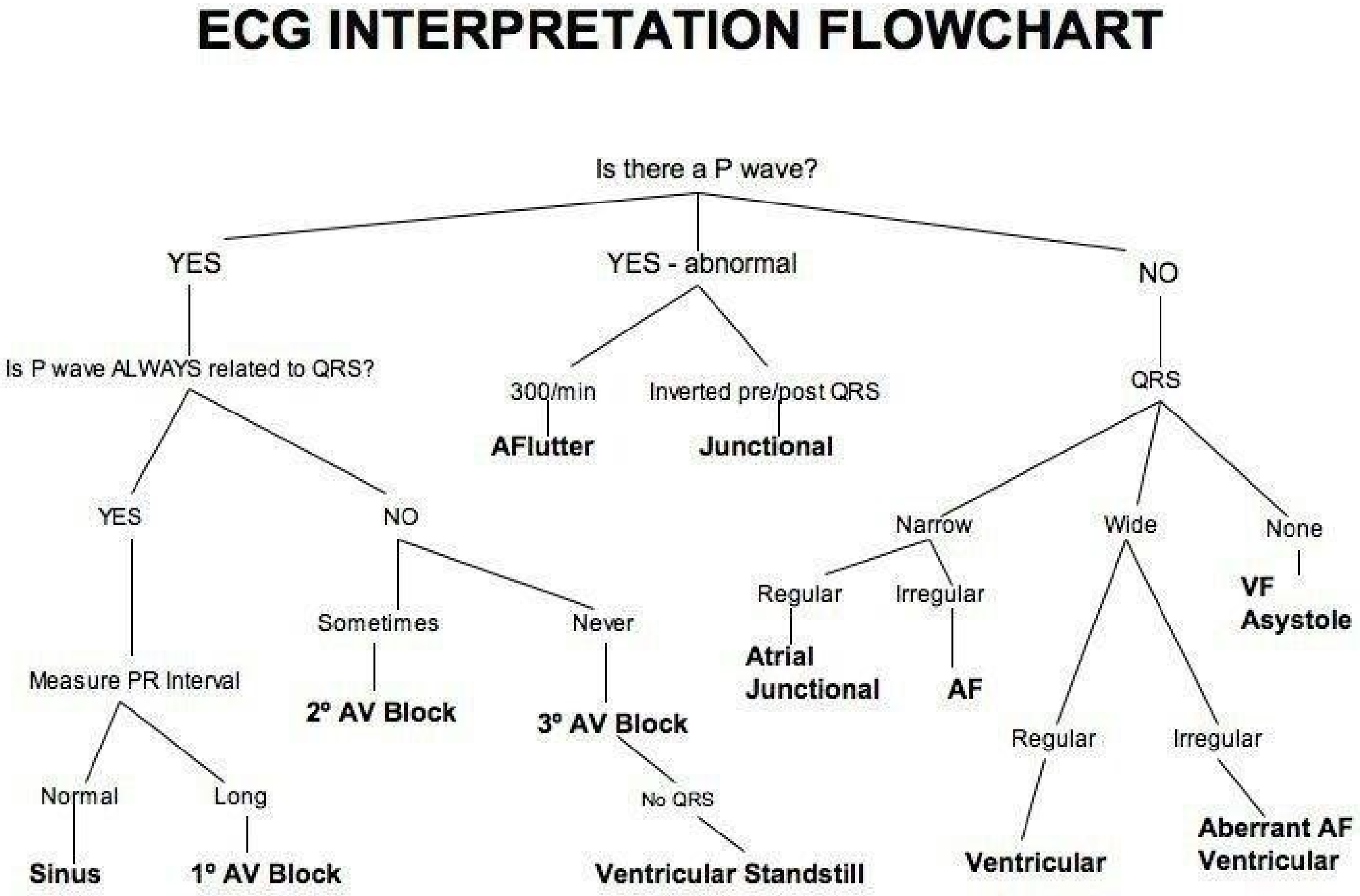


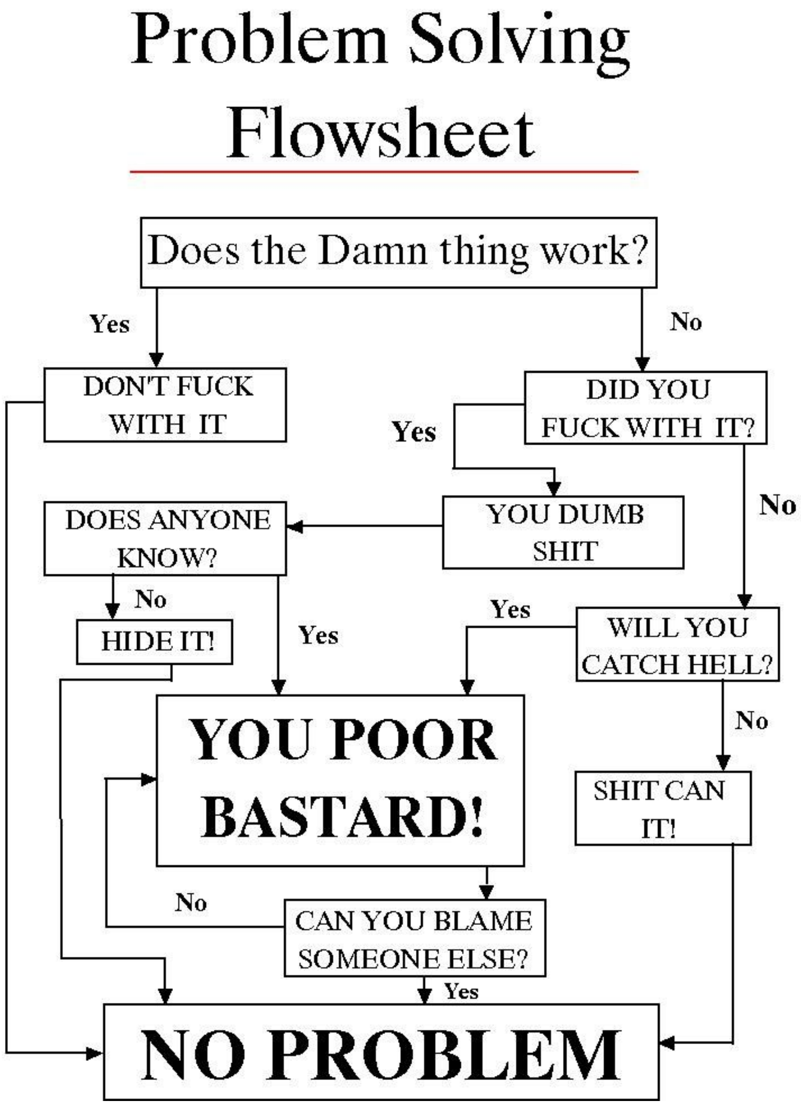
- NORMAL ECG -
Unconfirmed diagnosis
a VR
V1
V4
a VL
V2
V5
a VF
V3
V6
Voltage calibration
Paper speed
25 mm/s
10 mm/mV
0.05 Hz - 100 Hz
HP709
02221



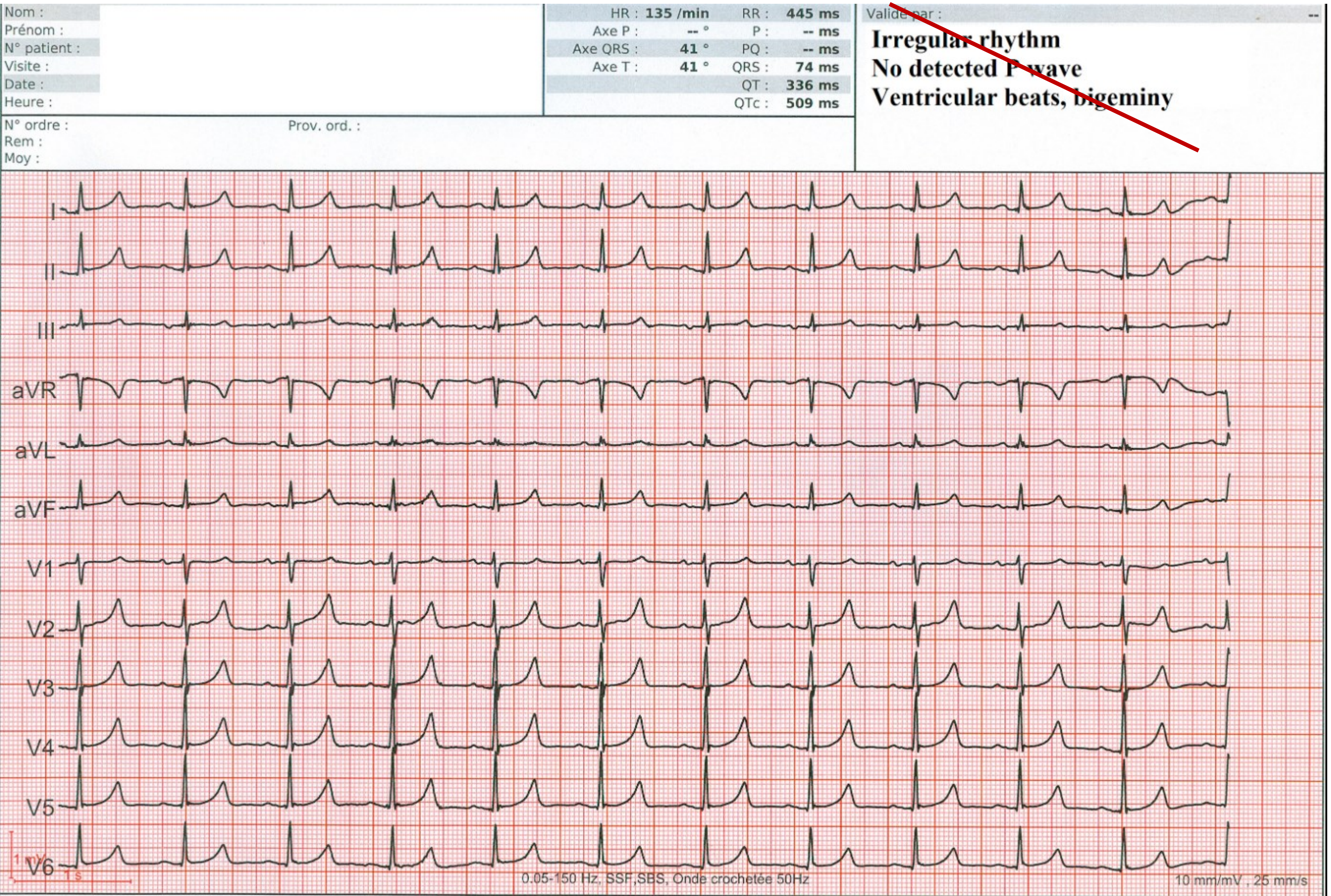
aVR
aVL
Lead I
V_1
V_2
V_3
V_4
V_5
V_6
Lead III
aVF
Lead II

How to diagnose 12 lead ECG?





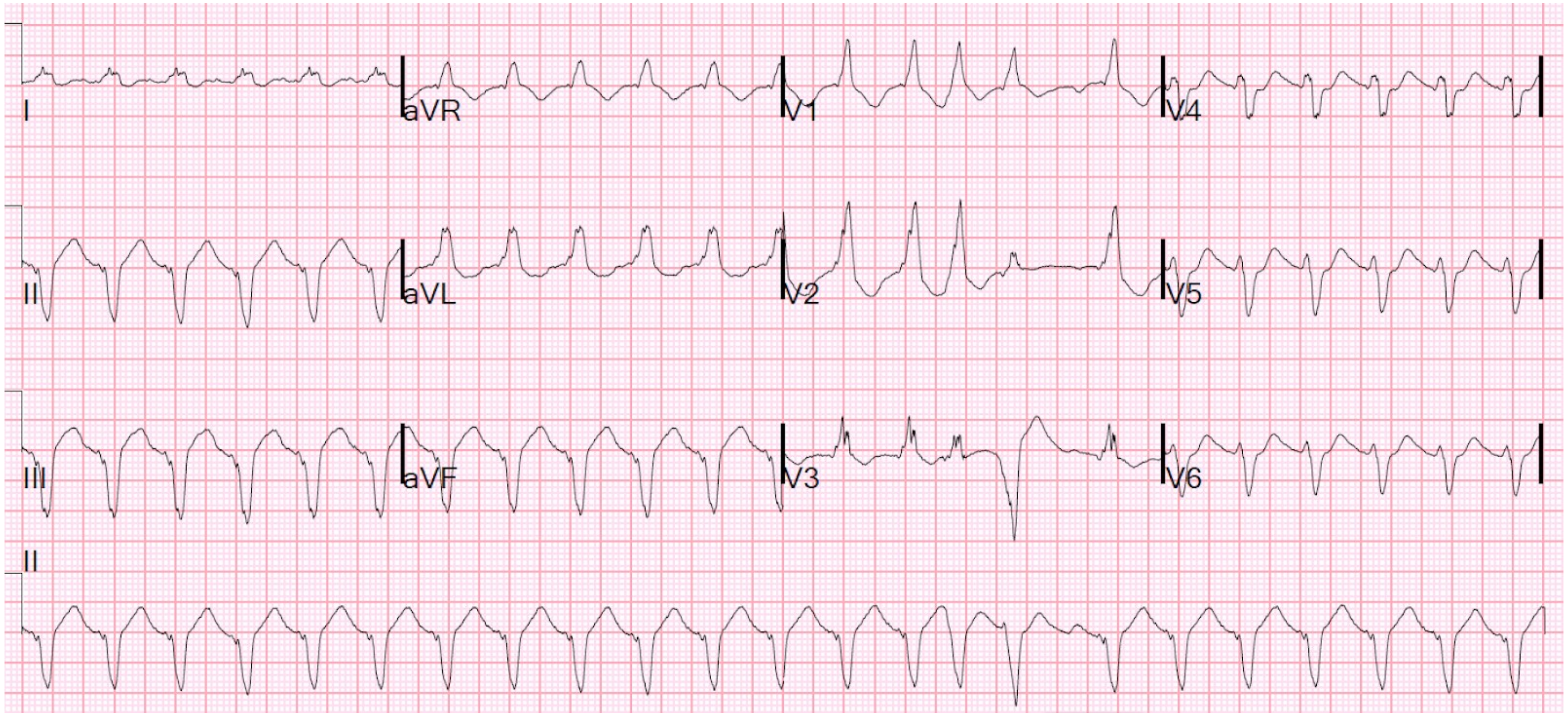
Current challenges to automatically diagnose 12 lead ECG



Unrecognized regular sinus rhythm with double counting of the heart rate (135 beats/min) due to T-wave oversensing.

Jürg Schläpfer, Hein J. Wellens,
Computer-Interpreted Electrocardiograms: Benefits and Limitations,
Journal of the American College of Cardiology,
Volume 70, Issue 9,
2017,
Pages 1183-1192,
ISSN 0735-1097,
https://doi.org/10.1016/j.jacc.2017.07.723.

Current challenges to diagnose 12 leads ECG by cardiologist- Dr. Smith's ECG Blog



What is the rhythm? Is it ventricular tachycardia (VT)?

Eventually: RCA acute occlusion

One might think this is VT, but if you look at beats 15 and 16 across the bottom, they are PVCs, and the 2nd PVC is followed by a P-wave and then a conducted beat, proving that the rhythm is sinus rhythm.
If you assess the PR interval, then go to other complexes, you can see that the P-wave in the first 14 complexes, and in 18-23, is superimposed upon the T-wave.

Current challenges to automatically diagnose 12 lead ECG

Data

Correct labeling

Many examples

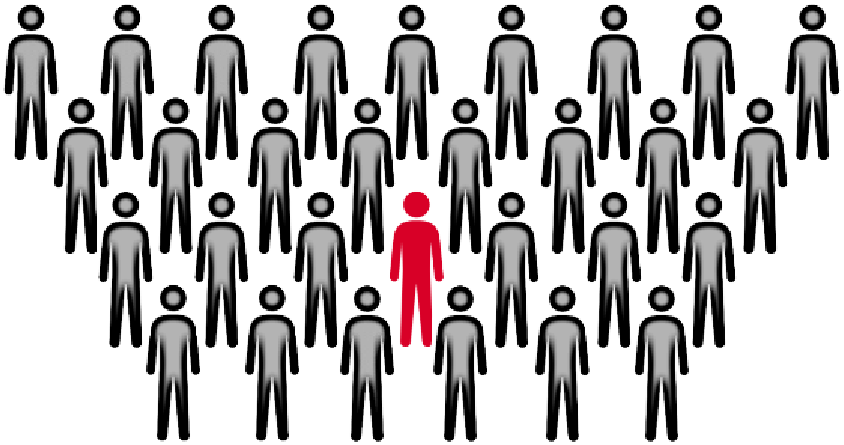
Main Topics

- Challenges related to big data in biomedical engineering field
- Example of automatic classification of cardiac diseases

BigData@Heart

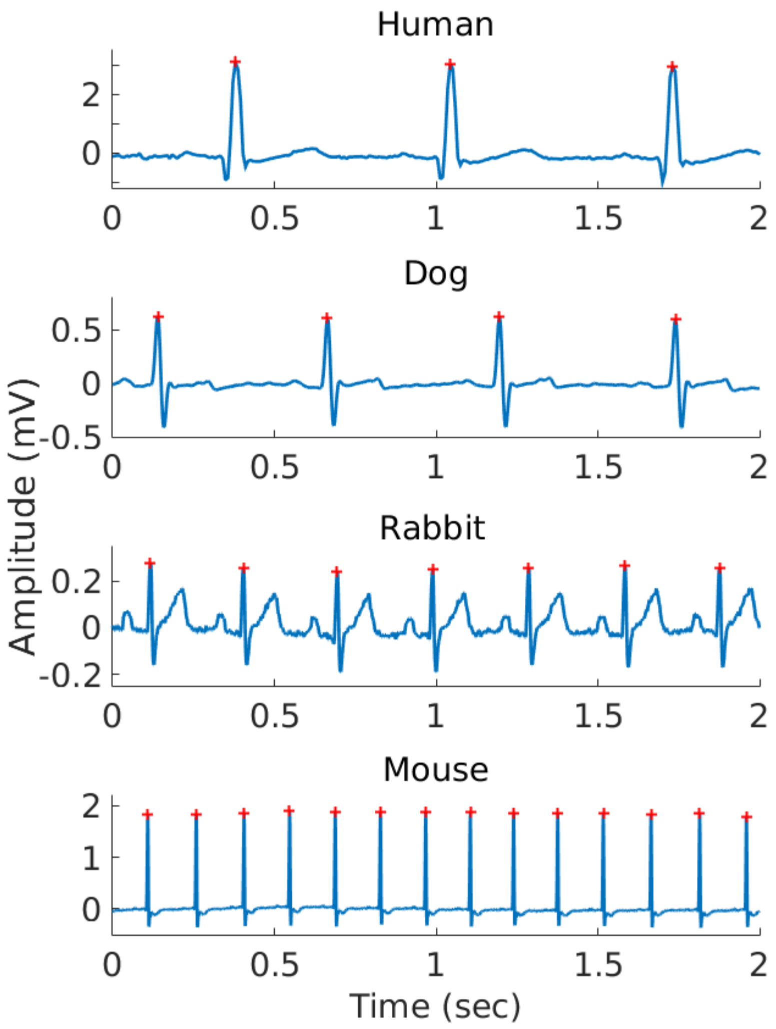
Big data in biomedical engineering field challenge:

(1) Small data set

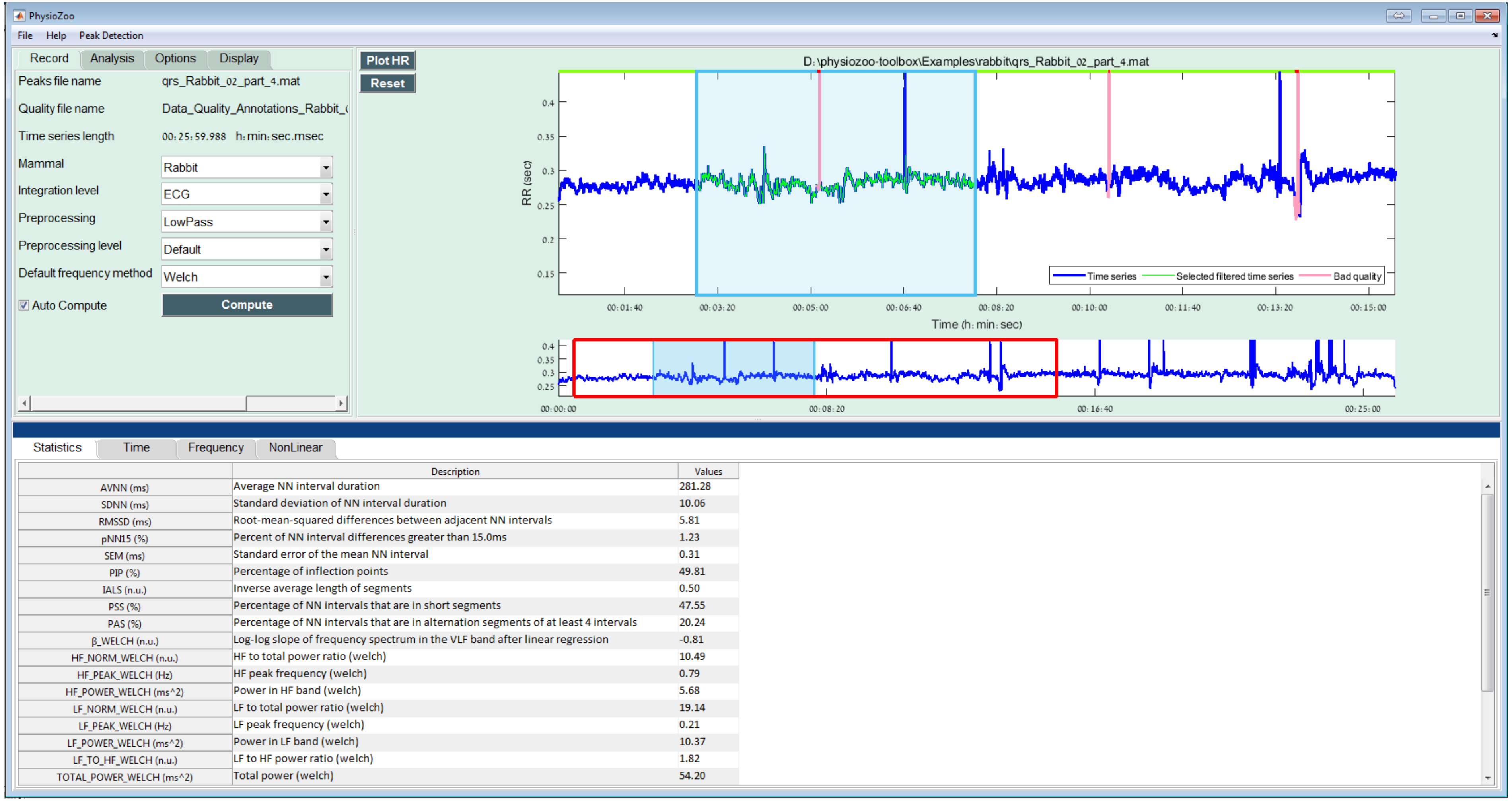


Animal data: Solution for small dataset

Animal data

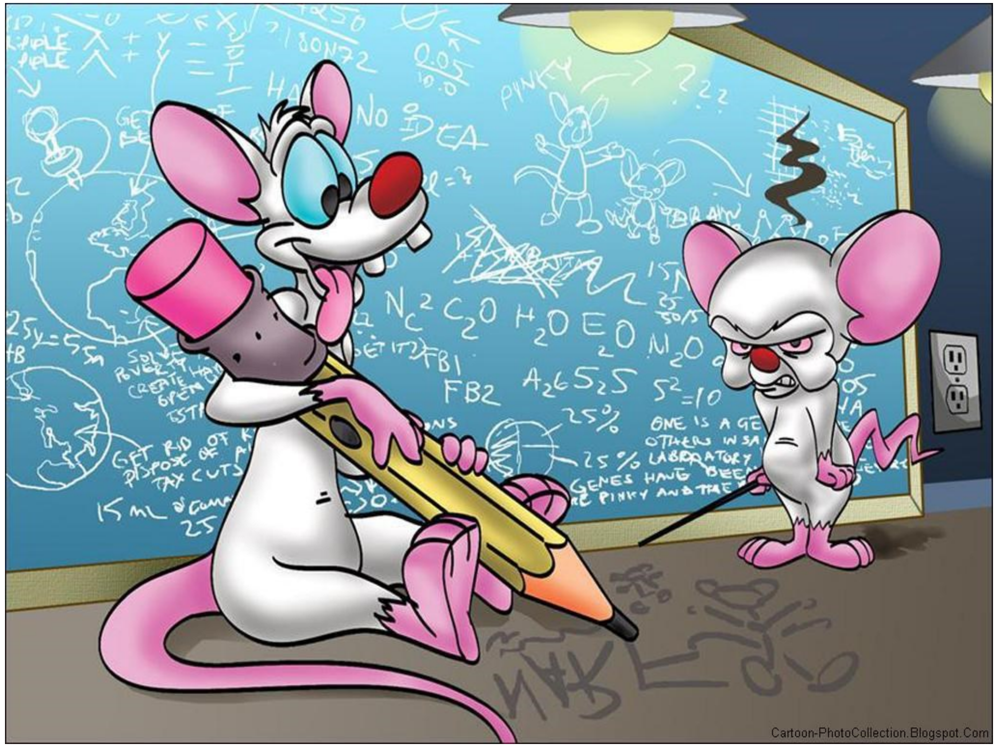


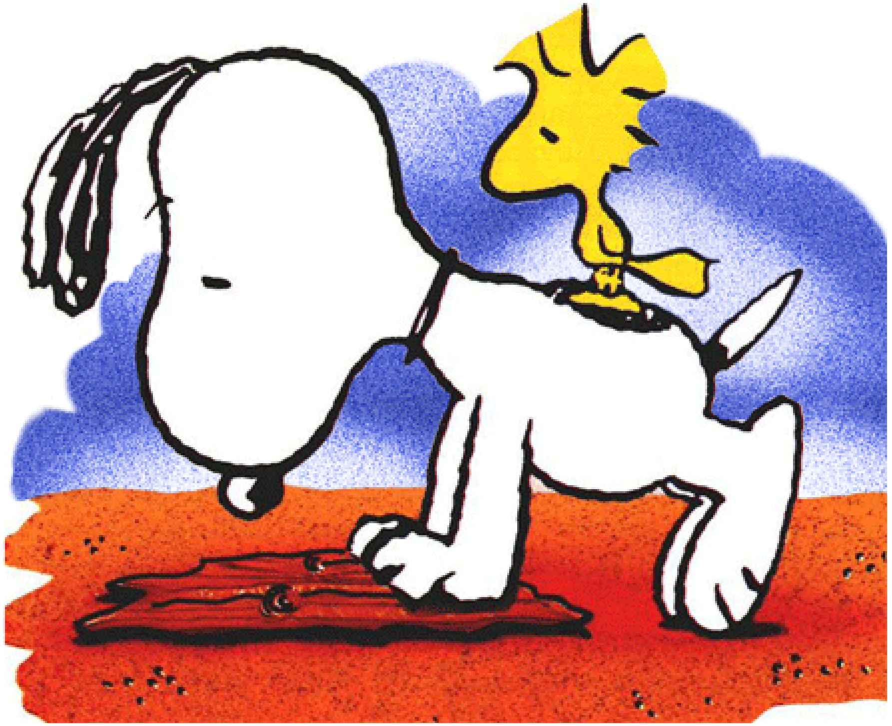
PhysioZoo: Animal analysis program



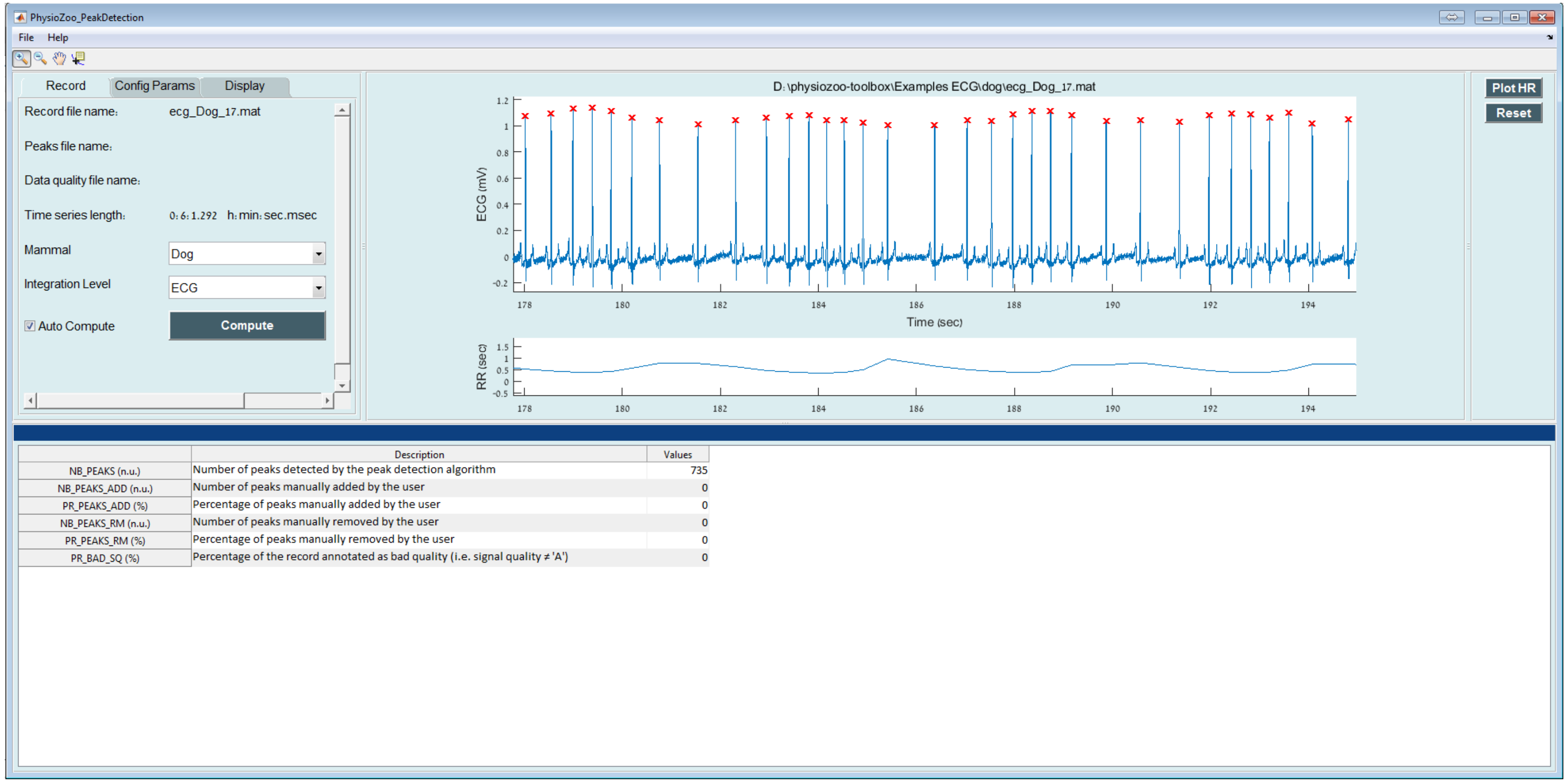
Behar and Rosenberg et al et al. *Frontiers in Physiology 2018*

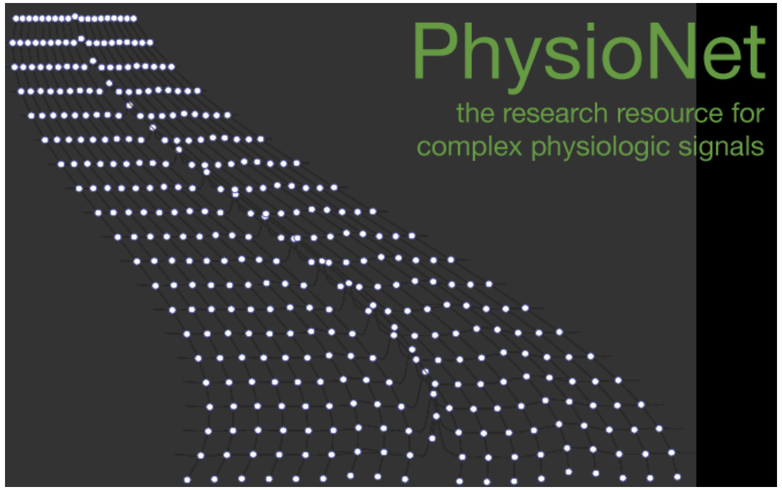
Big data in biomedical engineering field challenge:
(2) Data Sharing





PhysioZoo: Solution for data sharing

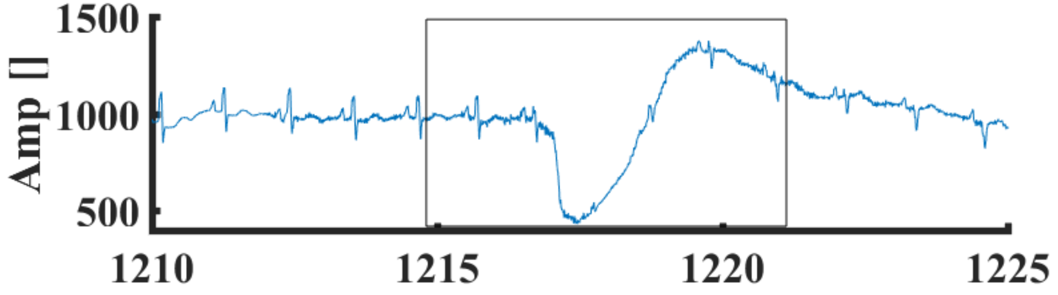




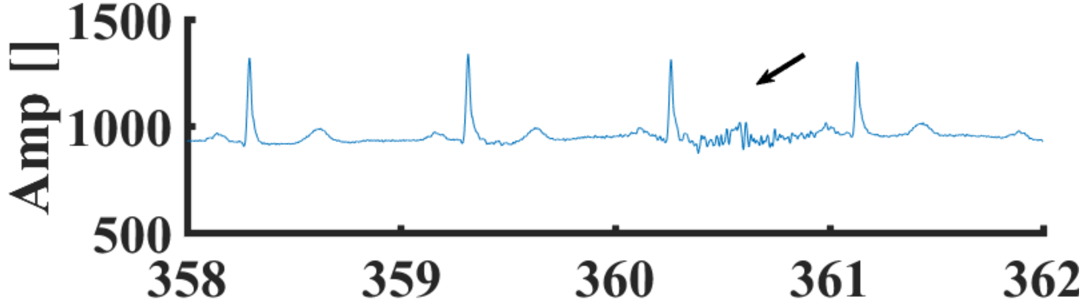
Behar and Rosenberg et al et al. *Frontiers in Physiology 2018*

Big data in biomedical engineering field challenge:
(3) Noise

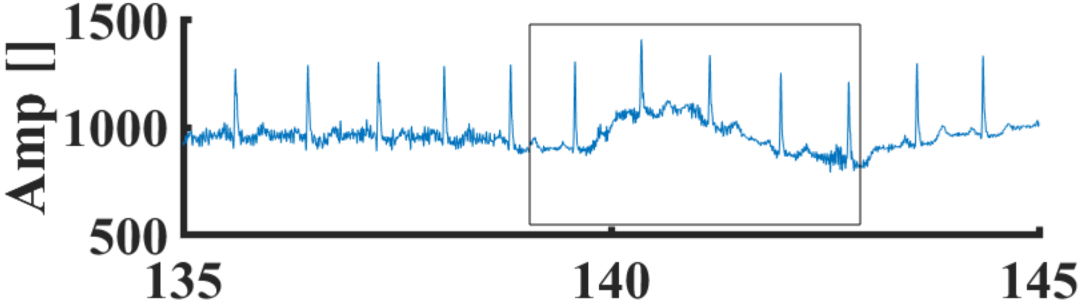
Sudden movement



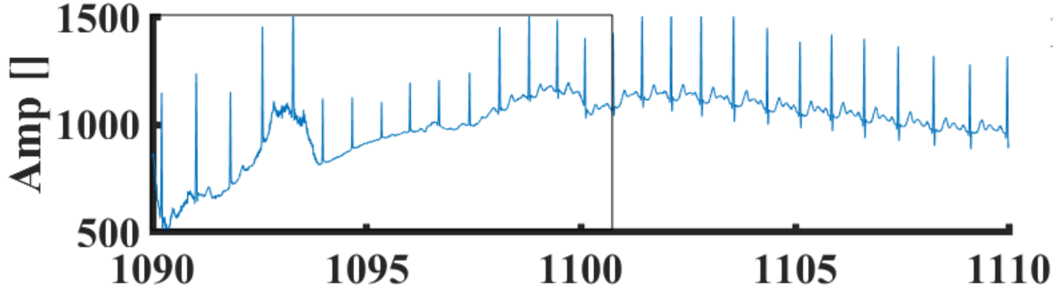
Environmental noise



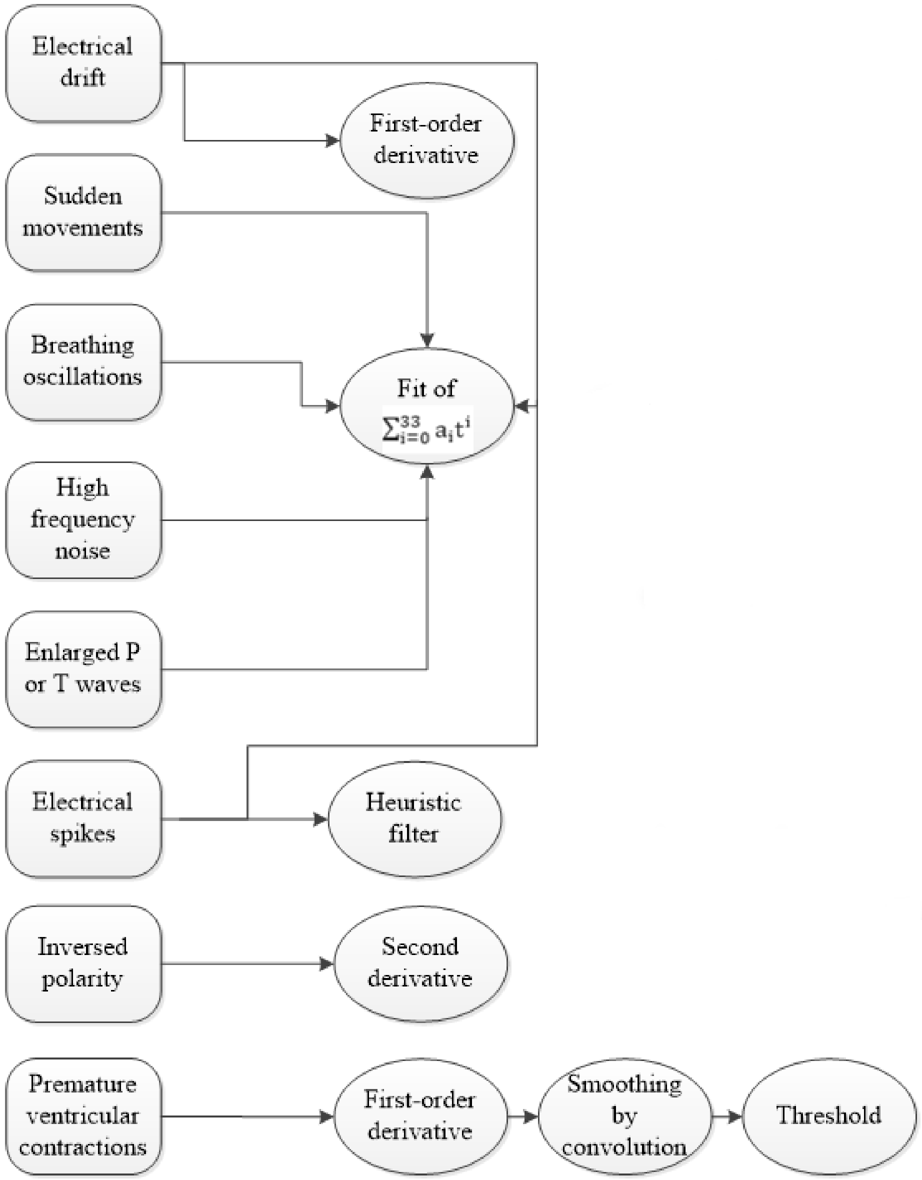
Breathing oscillations



Electrical drift

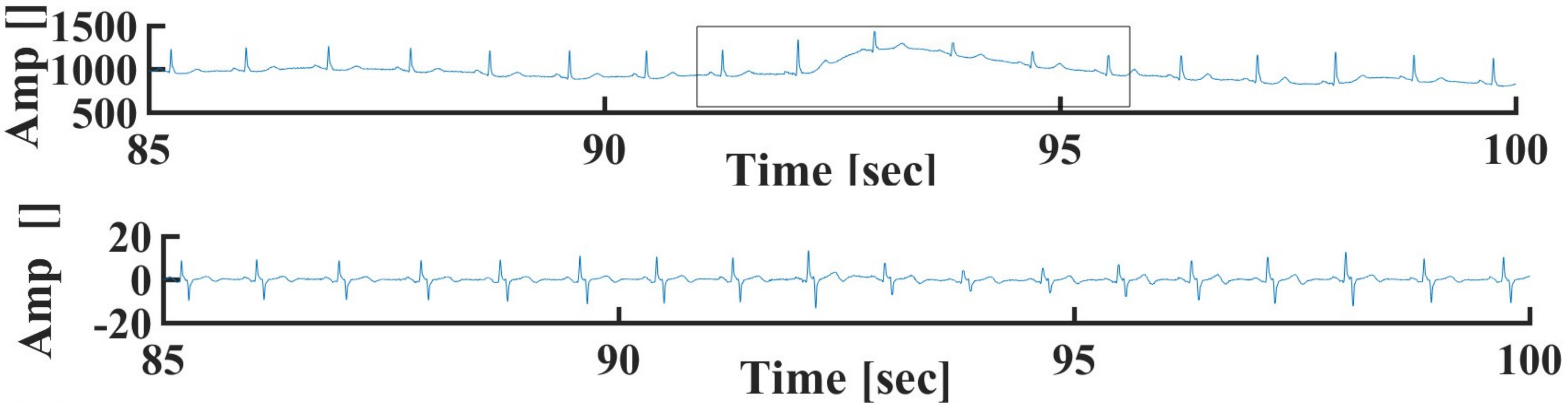


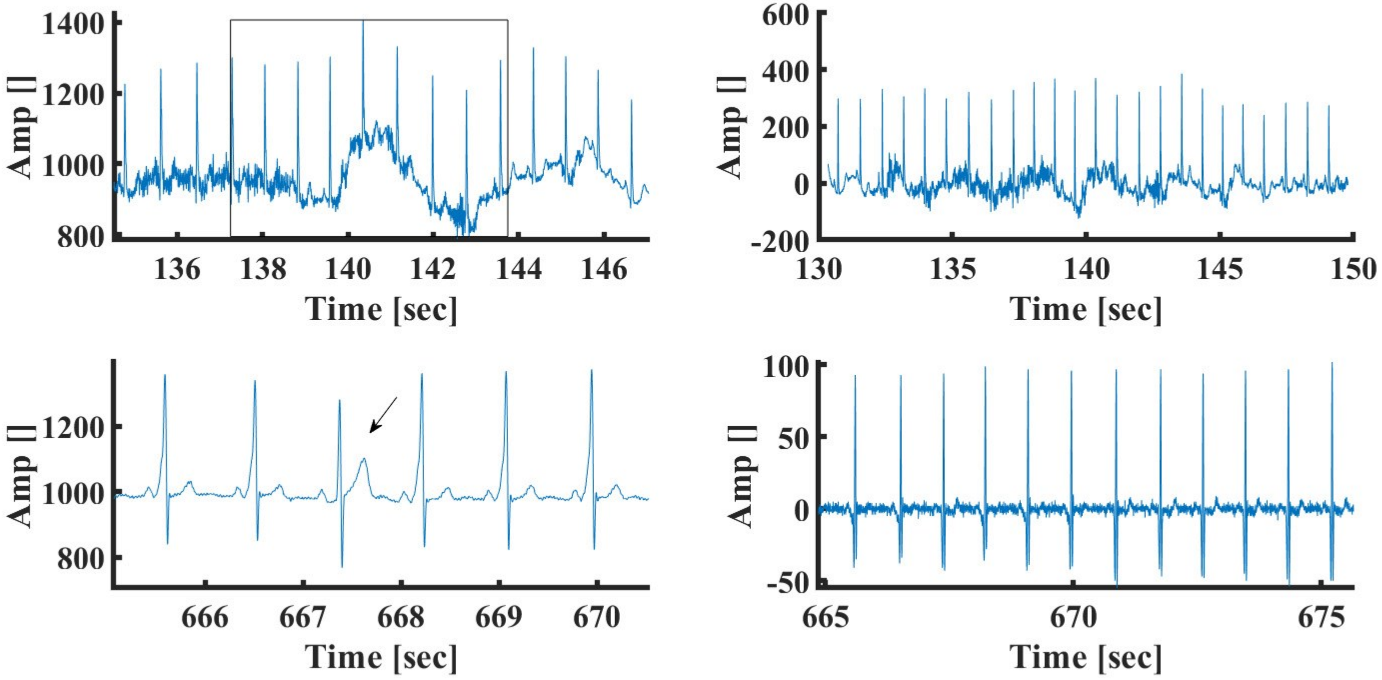
Solution for noise problem



Gliner et al. *JMIR mHealth and uHealth* 2018

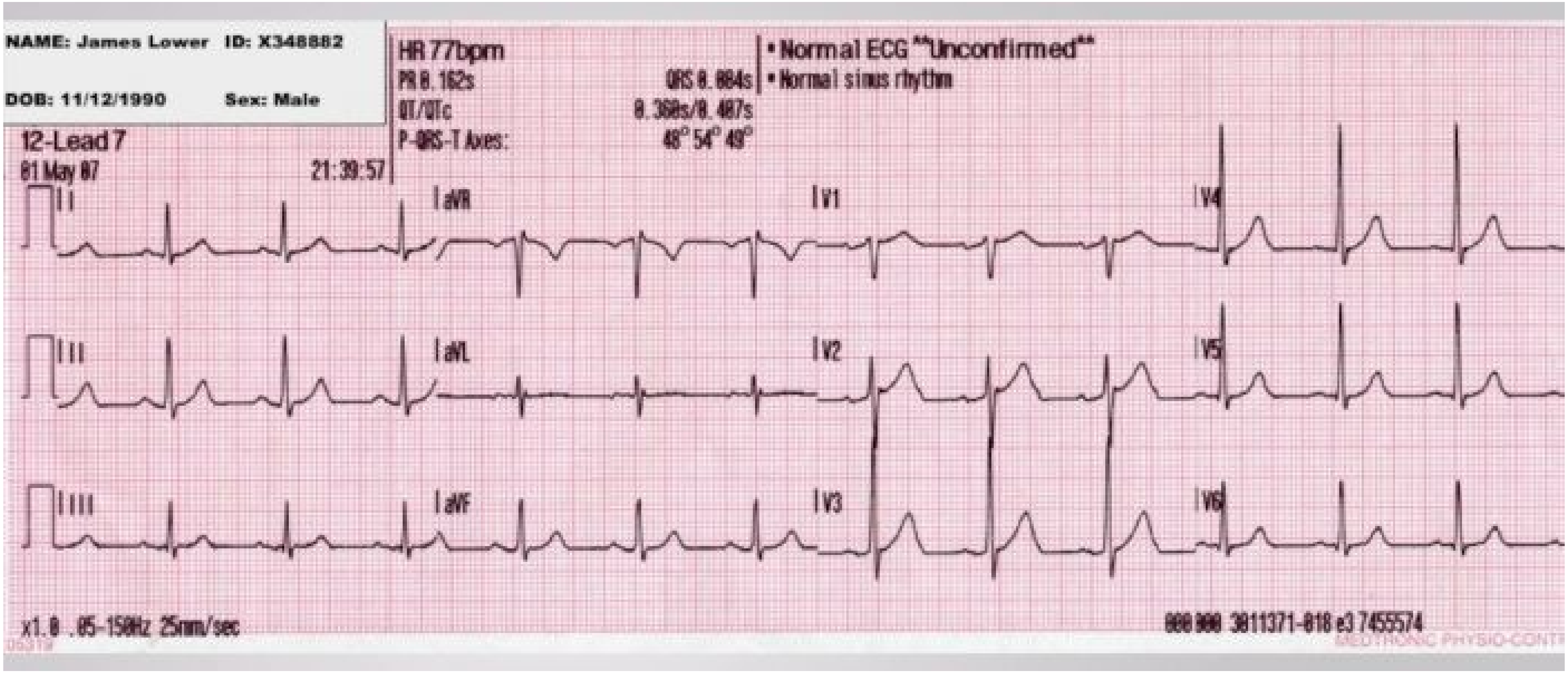
Solution for noise problem



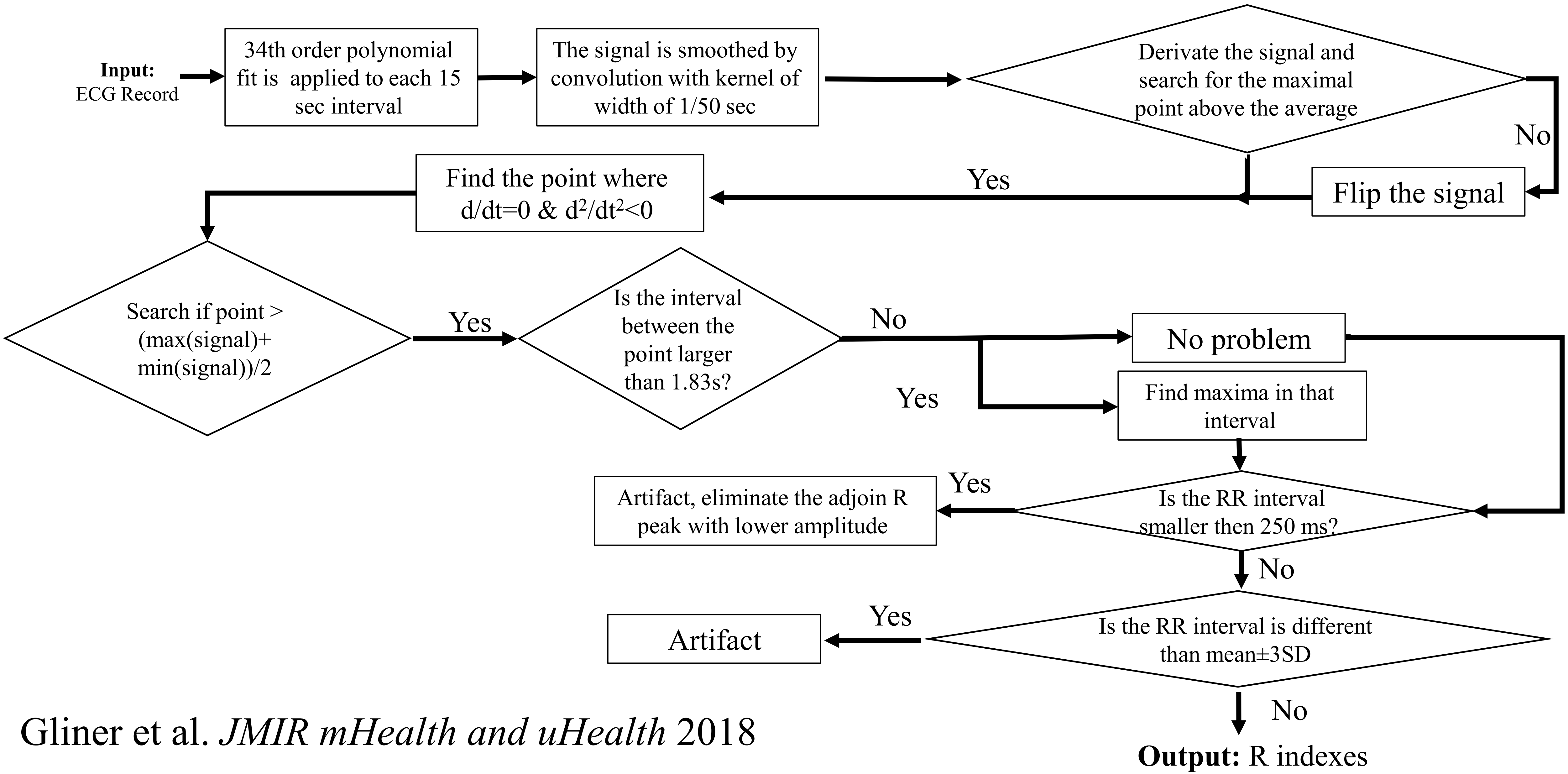


Gliner et al.
JMIR mHealth and uHealth
2018

Big data in biomedical engineering field challenge:
(4) labeling



Automatic R peak detector algorithm



Gliner et al. *JMIR mHealth and uHealth* 2018

R-Peak detector performance

Non-AF (%)	gqrs algorithm	Pan et al. algorithm	Behar et al. algorithm	Current algorithm
False negative	0.15	0.49	0.47	**0.14**
False positive	0.39	0.33	1.19	**0.28**
Positive prediction	99.6	99.7	98.8	**99.7**

AF (%)	gqrs algorithm	Pan et al. algorithm	Behar et al. algorithm	Current algorithm
False negative	0.42	0.72	2.09	**0.34**
False positive	0.79	0.67	5.45	**0.25**
Positive prediction	99.2	99.3	95.7	**99.7**

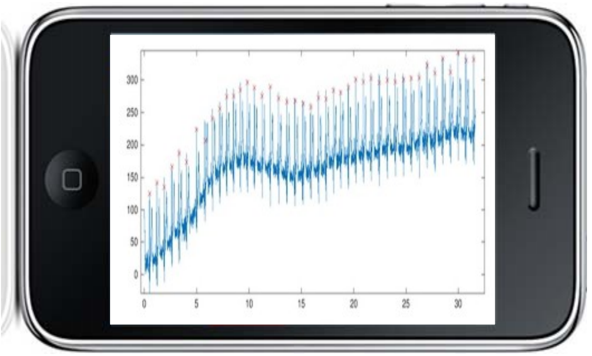
Total (%)	gqrs algorithm	Pan et al. algorithm	Behar et al. algorithm	Current algorithm
False negative	0.30	0.30	1.21	**0.24**
False positive	0.62	0.62	3.91	**0.27**
Positive prediction	99.4	99.4	97.2	**99.7**

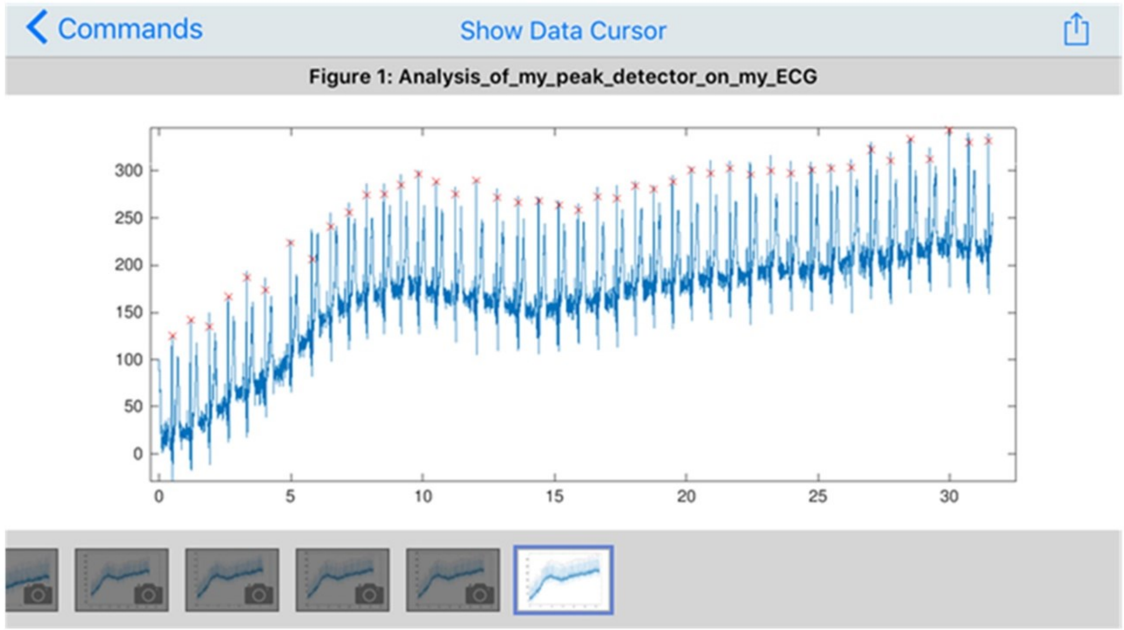
Big data in biomedical engineering field challenge :

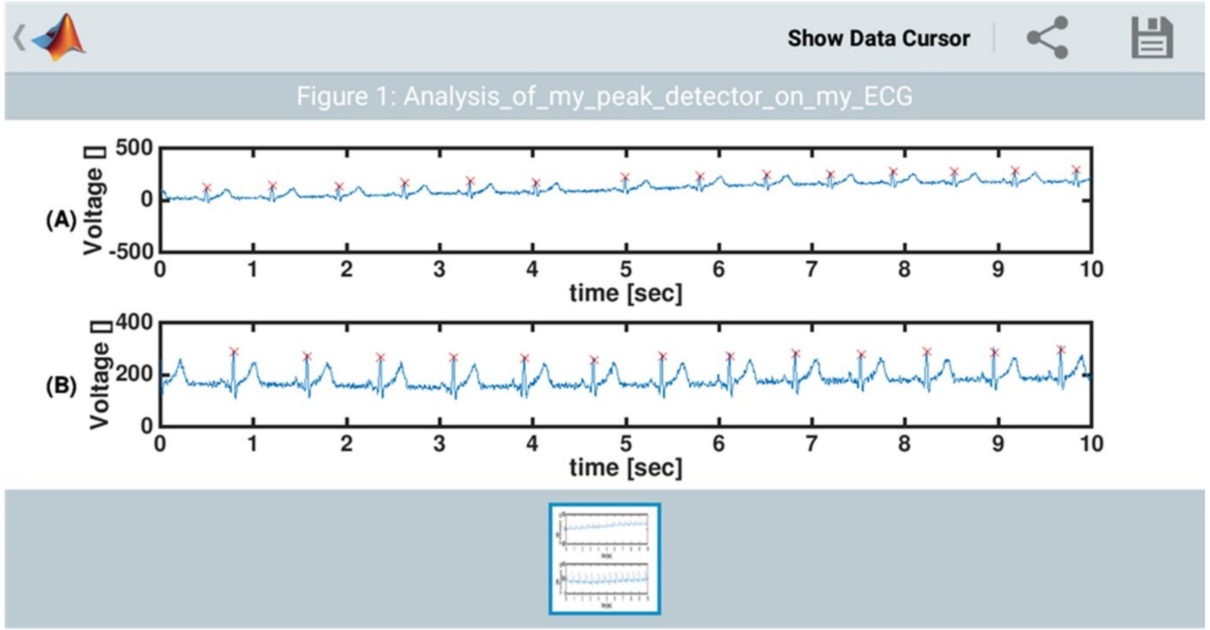
(5) Live analysis

1. Download the Matlab Mobile application and install it on the cell phone or tablet
2. Connect to your Matlab Cloud account and download the source code
3. Acquire data using the ECG device and upload to the cloud
4. Download the ECG data from the cloud using Matlab Mobile
5. Process the data and annotate R- peaks

Peak detection on mobile phones







Gliner et al. *JMIR mHealth and uHealth* 2018

Big data in biomedical engineering field challenges
SUMMARY 1

- Small data set
- Data Sharing
- Noise
- Labeling
- Live analysis
- Novel analytic tools for big data
- Phenomenological description

General Aim, Hypothesis and Challenges

Aim:

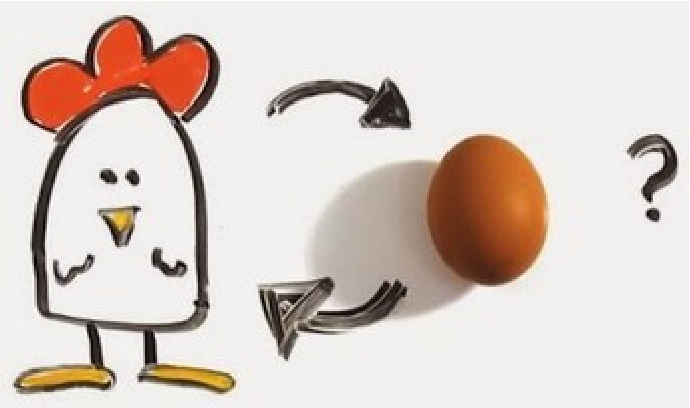
We aim to create a tool for accurate 12- lead ECG classification.

Hypothesis:

Given big data, we would be able to come with a tool for 12- lead ECG classification.

Challenges:

- No big data of 12- lead ECG exist:
  - No sharing
  - No standards
- No accurate tools for 12-lead ECG classification exist



How to overcome the challenges of 12 leads automatic diagnostics

1. Develop a first prototype of a 12-lead ECG classifier based on existing supervised machine learning methods.

- PhysioNet Challenge 2017 - Single-lead classification of atrial fibrillation and normal vs. abnormal rhythm.
- The China Physiological Signal Challenge (CPSC) 2018- classifier for 12-lead ECG for the 9 most common morphological and rhythm disorders.

2. Design a system to gather big data:

- General platform for accumulation and classification of 12-lead ECG data

Single lead classification-Introduction

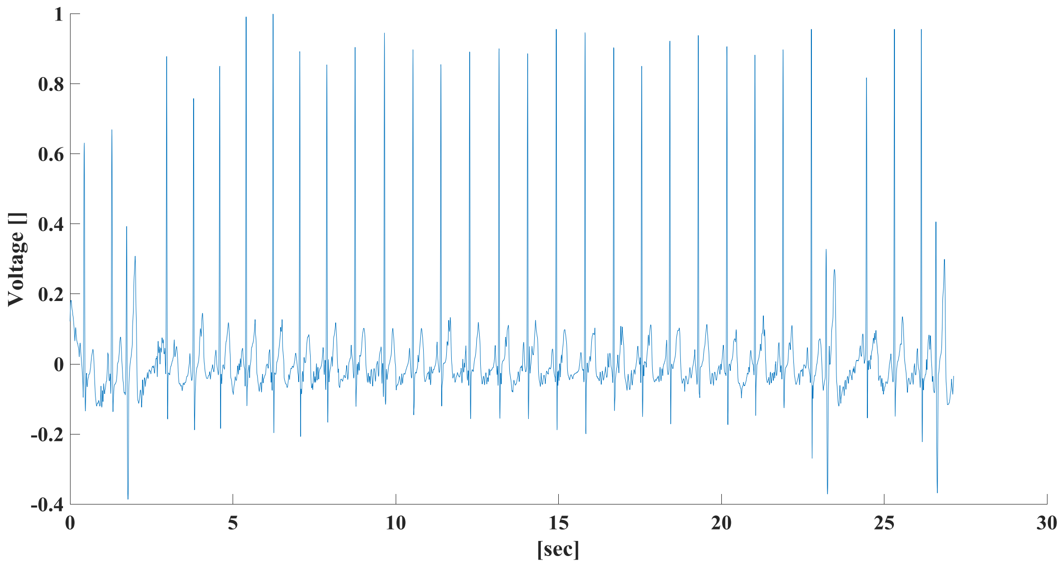
- 2 datasets: a training set of 8,528 single lead ECG recordings from 9s to just over 60s and a test set containing 3,658 ECG recordings of similar length.
- All recordings consisted of one bipolar channel recorded by an AliveCor device.
- The data was sampled at 300 Hz and filtered by a band pass filter in the device itself.
- The training set data was annotated by a cardiologist to one of the four types: normal (~59.5% of the recordings), AF (~10% of the recordings), other rhythm (~30% of the recordings), and noisy (~0.5% of the recordings).
- Classified database of short single ECG lead recordings divided into 4 categories:

- Normal rhythm

- Atrial fibrillation

- Noisy segment

- Other rhythm disturbances



Single lead classification-Methods

- To test the binary classification (AF/Non-AF, Healthy/Other), we used the training set for both training and learning.
- We divided the training set into 7 parts, 6 of which were used for learning and 1 to test the algorithm's performance. Each part of the divided database consisted of 60.43% normal sinus rhythm, 9.05% AF, 30% other rhythm disturbances, and 0.54% too noisy to be classified.
- After learning on the training set and testing on a test set, we again divided that data into 7 parts and repeated this procedure 7 times.

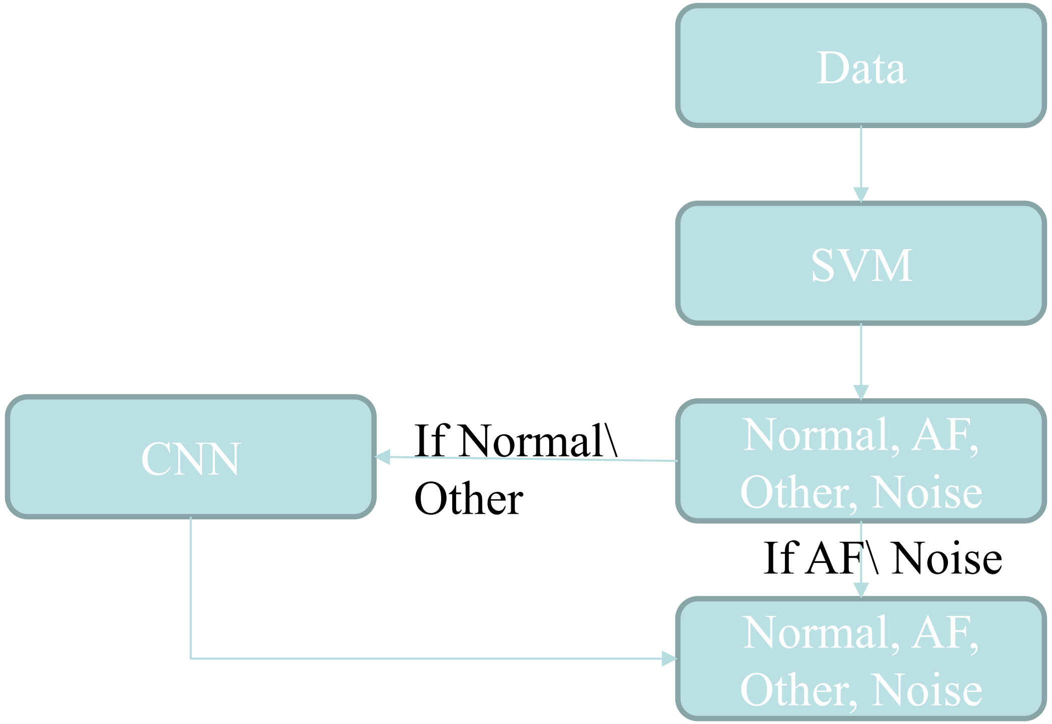
$$F_1 = \left(\frac{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}}{2}\right)^{-1} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}.$$

$$\text{Normal rhythm: } F_{1n} = \frac{2 \times Nn}{\sum N + \sum n}$$

Single lead classification-Results

(1) Classification of the data into 4 categories with SVM and NN:

	Predicted classification			
	Normal	AF	Other	Noise
Normal	4621	22	374	33
AF	31	574	127	6
Other	601	88	1734	33
Noisy	51	5	55	173



Single lead classification-Results

Prediction of Challenge results by 7 cross-folds of the training set:

	F_{1N}	F_{1A}	F_{1O}	F_1
Average (per run)	0.89	0.80	0.73	0.81
Challenge	0.90	0.81	0.70	0.80

Average results obtained on the validation subset for 7 cross-folds of the training set:

	F_{1AF}	F_{1NAF}	F_{1F}	F_{1A}	F_{1NAF}	F_{1F}	F_{1AF}	F_{1NAF}	F_{1F}
	Quadratic SVM			Linear Discriminant			RusBoosted Tree		
Average	0.80	0.98	0.89	0.75	0.97	0.86	0.69	0.96	0.82

$$F_1 = \left(\frac{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}}{2} \right)^{-1} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}.$$

Single lead classification-Results

(2) Classification of the data into AF and non-AF:
Three methods of classification were tested on the same features:

- SVM
- Linear Discriminant
- RusBoosted Trees

	F_{1AF}	F_{1NAF}	F_{1F}	F_{1A}	F_{1NAF}	F_{1F}	F_{1AF}	F_{1NAF}	F_{1F}
	Quadratic SVM			Linear Discriminant			RusBoosted Tree		
Average	0.80	0.98	0.89	0.75	0.97	0.86	0.69	0.96	0.82

Single lead classification-Results

(3) Classification of the data into normal sinus vs. abnormal rhythm:

Three methods of classification were tested on the same features:

- SVM
- Linear Discriminant
- RusBoosted Trees

	F_{1NS}	F_{1NNS}	F_{1S}	F_{1NS}	F_{1NNS}	F_{1S}	F_{1NS}	F_{1NNS}	F_{1S}
	Quadratic SVM			Linear Discriminant			RusBoosted Tree		
Average	0.88	0.81	0.85	0.85	0.74	0.80	0.86	0.79	0.82

Single lead classification-Conclusions

Our algorithm obtained a total score (F1) of 0.80 on the hidden dataset (placing 18th–24th out of all the algorithms participating in the challenge; places 18–24 received the same score). Our algorithm was also able to classify AF vs. non-AF and normal vs. abnormal (arrhythmia or noise) records.

12 lead classification-Introduction

- 12 lead ECG
- 9 most common morphological and rhythm disorders

Normal: $F_{11} = \frac{2N_{11}}{N_{1x}+N_{x1}}$

AF: $F_{12} = \frac{2N_{22}}{N_{2x}+N_{x2}}$

I-AVB: $F_{13} = \frac{2N_{33}}{N_{3x}+N_{x3}}$

LBBB: $F_{14} = \frac{2N_{44}}{N_{4x}+N_{x4}}$

RBBB: $F_{15} = \frac{2N_{55}}{N_{5x}+N_{x5}}$

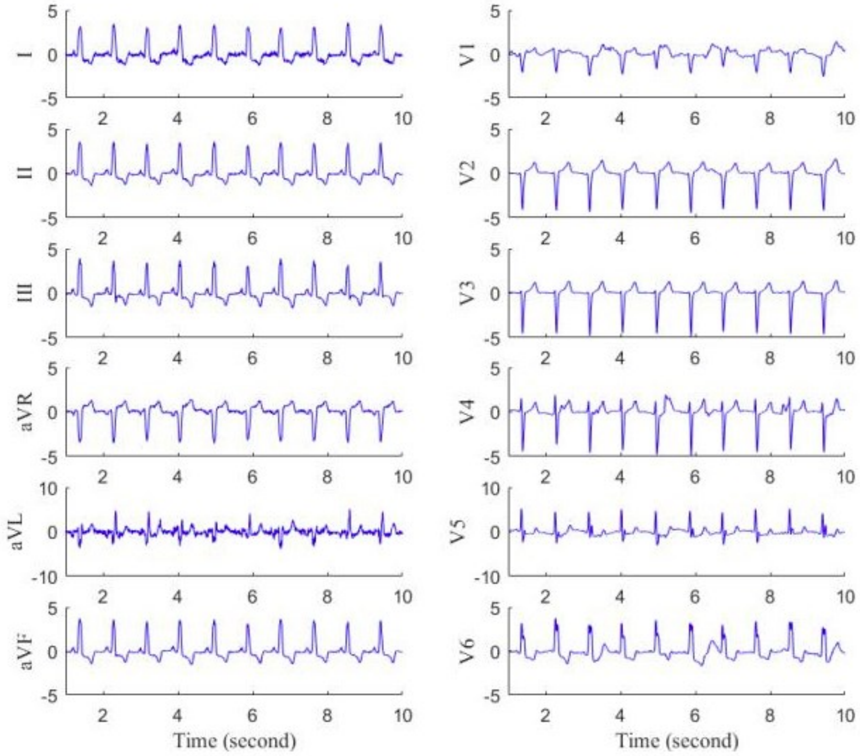
PAC: $F_{16} = \frac{2N_{66}}{N_{6x}+N_{x6}}$

PVC: $F_{17} = \frac{2N_{11}}{N_{7x}+N_{x7}}$

STD: $F_{18} = \frac{2N_{11}}{N_{8x}+N_{x8}}$

STE: $F_{19} = \frac{2N_{99}}{N_{9x}+N_{x9}}$

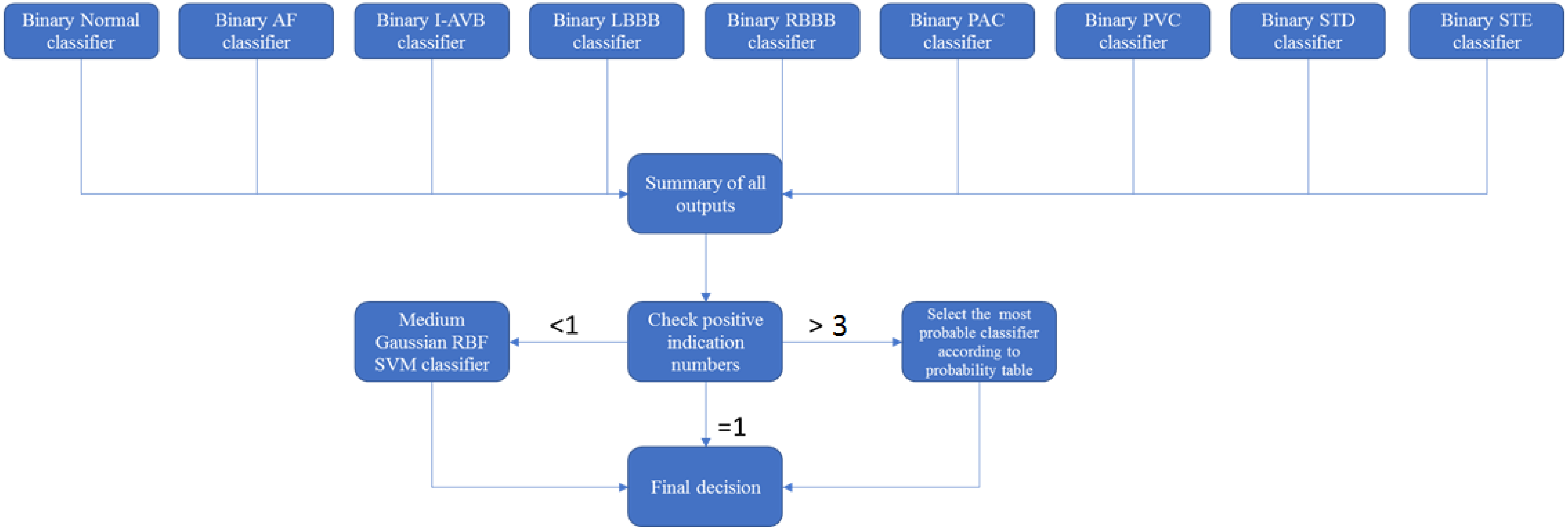
Type	#recording
Normal	918
Atrial fibrillation (AF)	1098
First-degree atrioventricular block (I-AVB)	704
Left bundle branch block (LBBB)	207
Right bundle branch block (RBBB)	1695
Premature atrial contraction (PAC)	556
Premature ventricular contraction (PVC)	672
ST-segment depression (STD)	825
ST-segment elevated (STE)	202
Total	**6877**



$$F_1 = \frac{F_{11} + F_{12} + F_{13} + F_{14} + F_{15} + F_{16} + F_{17} + F_{18} + F_{19}}{9}$$

Example of the 12-lead ECG waveforms
(Left bundle brunch block, from A0011)

12 lead classification-Methods



Flowchart of classification using the tournament method

12 lead classification-Methods

Feature types:

(1) Time domain features

(2) Average heart rate and heart rate variability

(3) Signal entropy

(4) Intra-beat temporal interval variability

(5) Single beat morphology

(6) Average normalized cross-correlation with saved (left and right) bundle branch block QRS complexes

Methods which were considered:

Linear kernel SVM, quadratic kernel SVM, radial basis kernel SVM, cubic kernel SVM, regression methods, bagged trees, boosted trees, RUSboosted trees and k-nearest neighbor

12 lead classification-Results

Our F_1 scores were as following:

Total: $F_1 = 0.704$

AF: $F_{AF} = 0.885$

Block: F_{Block} =0.796

Premature contraction: $F_{PC} = 0.698$

ST-segment change: $F_{ST} = 0.636$

The highest score in the challenge was that of Tsai-Min Chen et al.:

Total: $F_1 = 0.837$

AF: $F_{AF} = 0.933$

Block: F_{Block} =0.899

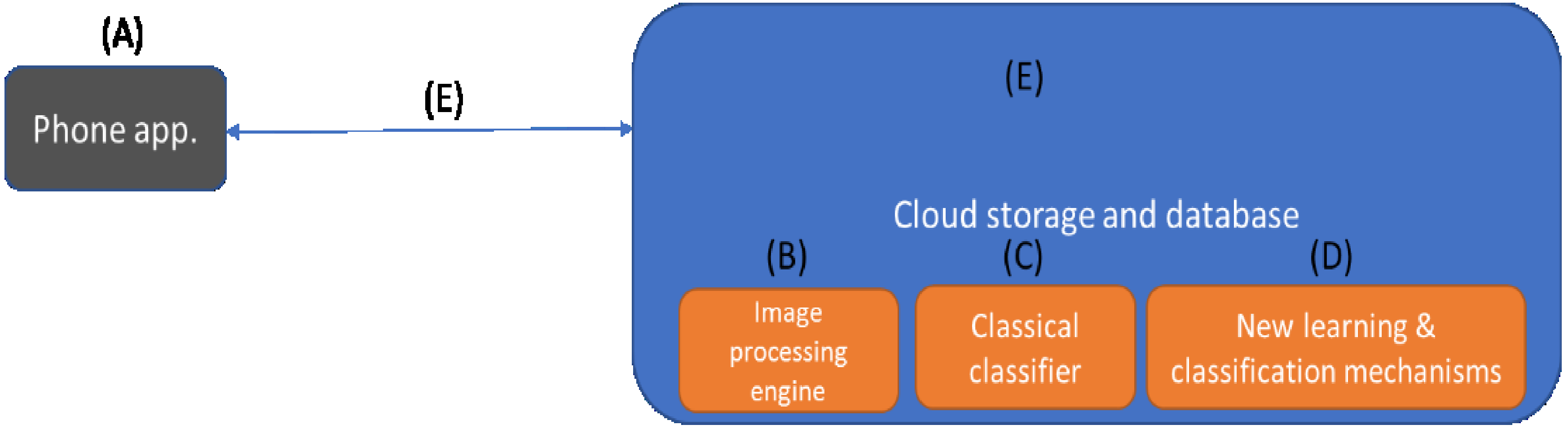
Premature contraction: $F_{PC} = 0.847$

ST-segment change: $F_{ST} = 0.779$

12 lead classification-Conclusions

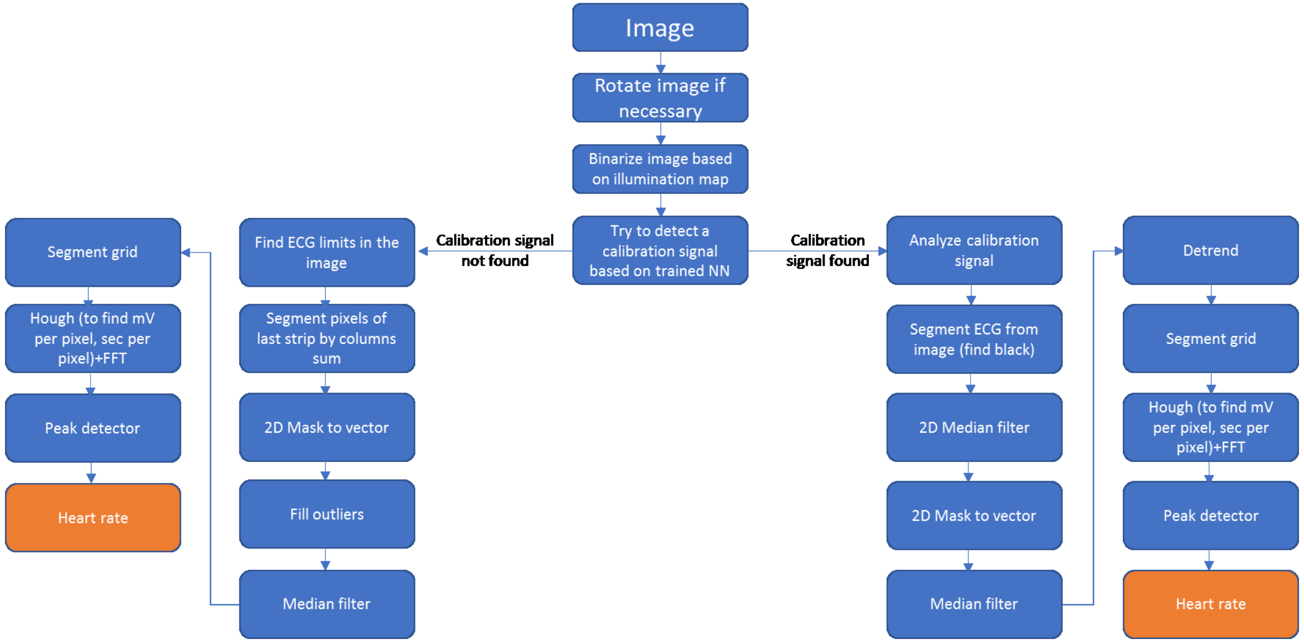
In terms of ranking, we reached 2nd place among Matlab users and 20th place overall. In terms of score, the scoring for CPSC2018 uses a F1 measure, which is an average of the nine F1 values from each classification type.

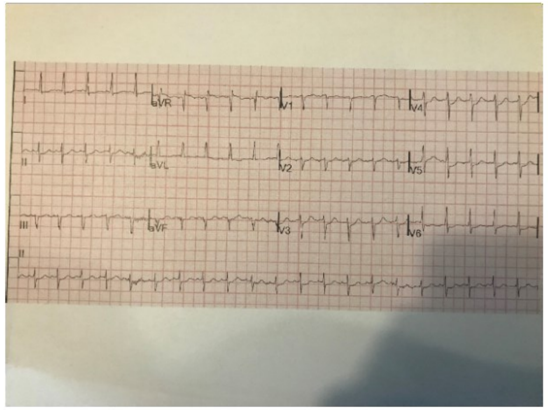
General platform for accumulation and classification of 12-lead ECG data-Introduction



General system topology. (A) Front-end mobile application, (B) Image processing engine for extracting the ECG vectors from the 12-lead ECG image, (C) SVM classifier of 12-lead digitized data, (D) Deep learning classification engine, (E) Database and storage of both the processing engines and the incoming data.

General platform for accumulation and classification of 12-lead ECG data-Methods

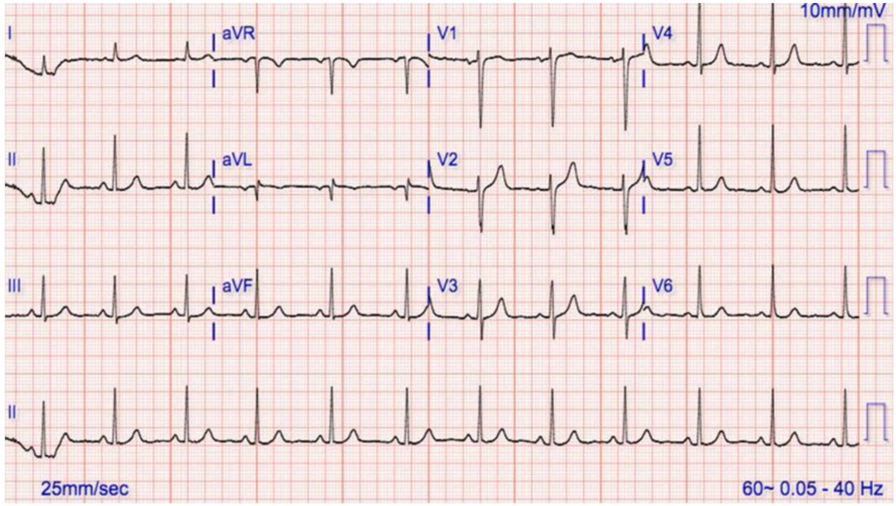




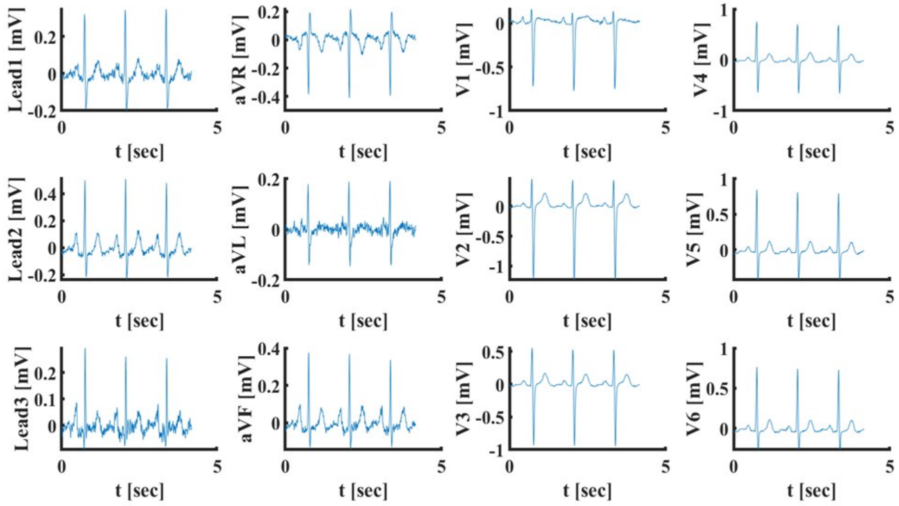
(A)

(B)

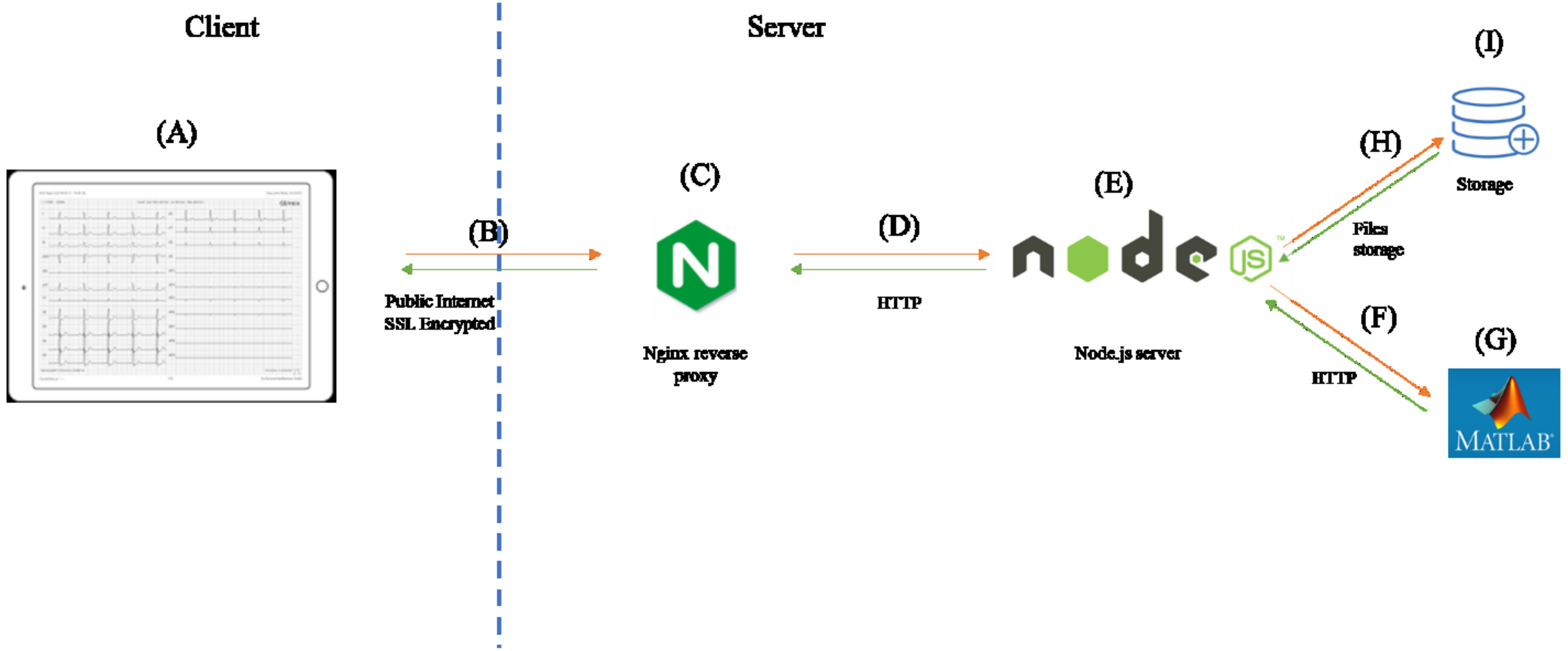
(A)



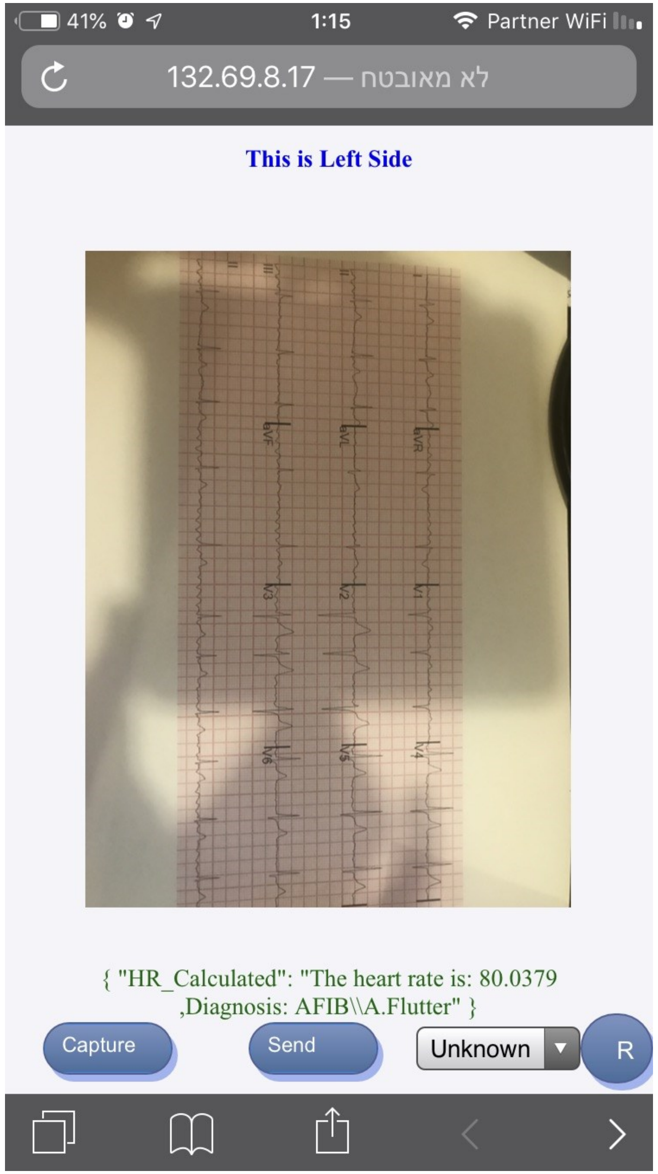
(B)



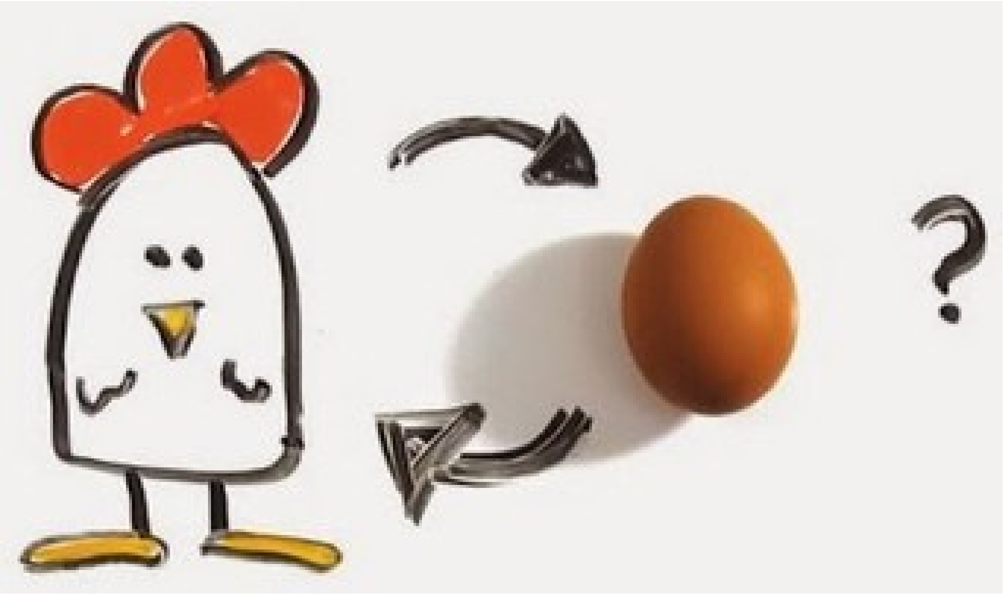
General platform for accumulation and classification of 12-lead ECG data-Results



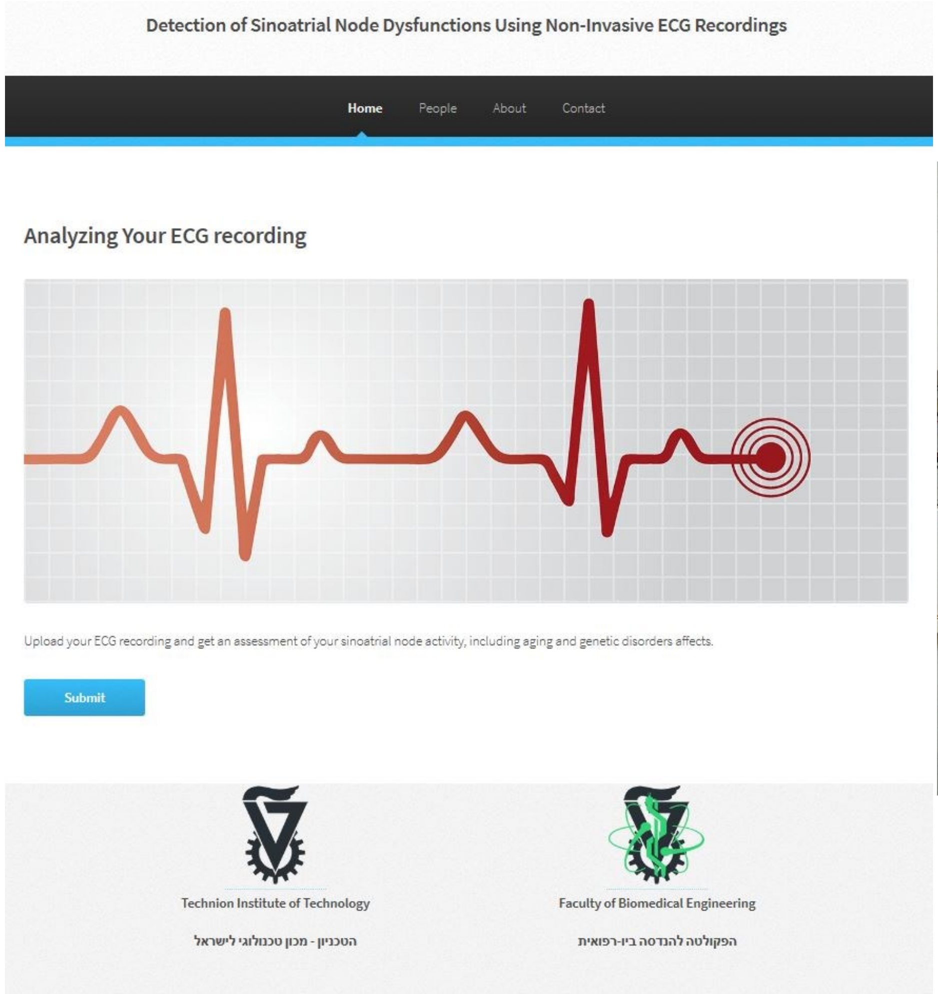
(A) Client side: any mobile device with a camera and connection to the internet, (B) Communication with the front-end part of the server using public internet (SSL Encrypted), (C) A powerful Nginx reverse proxy server responsible for receiving files and other data from the user, (D) HTTP protocol for communication between the reverse proxy server and the operation server, (E) Operation server responsible for routing the data either to the processing unit or to user, (F) HTTP protocol used for communication between the operation server and the processing unit, (G) Matlab environment running all the image processing, signal processing , machine learning and other algorithms, (H) network traffic data storage , and (I) storage location on the server, which contains all the data the server received in its lifetime .

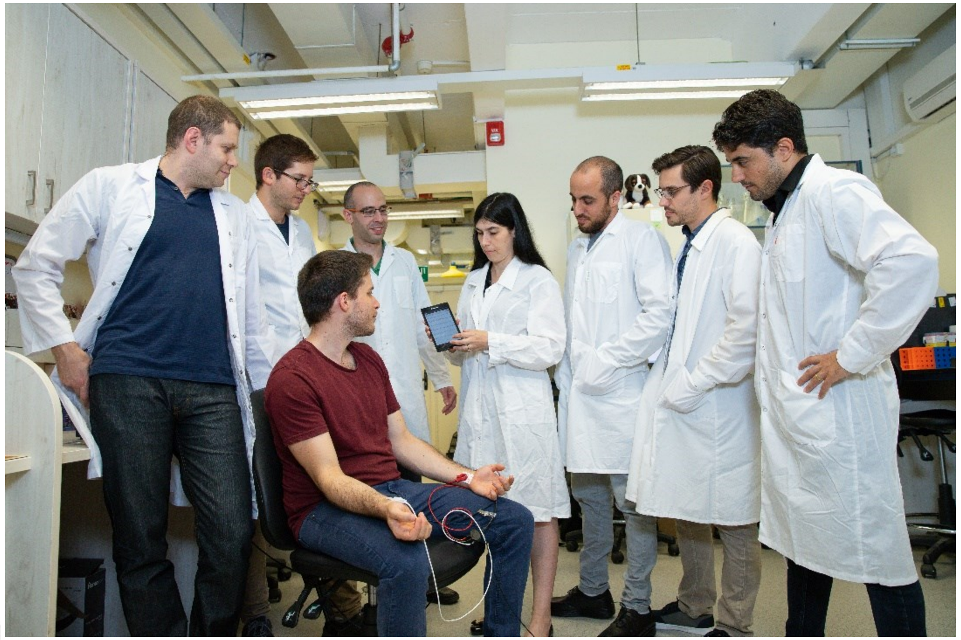


Take home message 1: **The best way to solve the chicken and the egg question is to create the chicken and the egg**

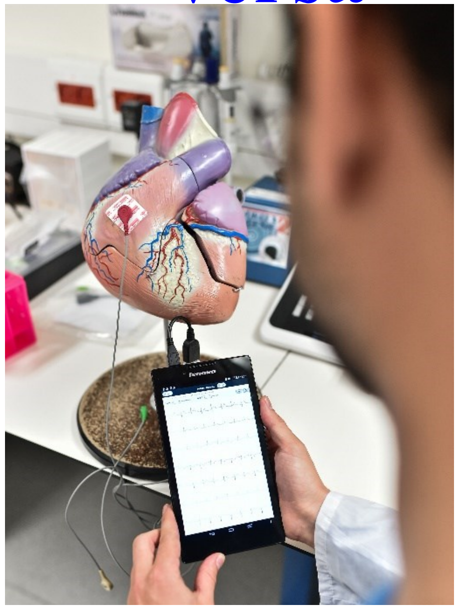


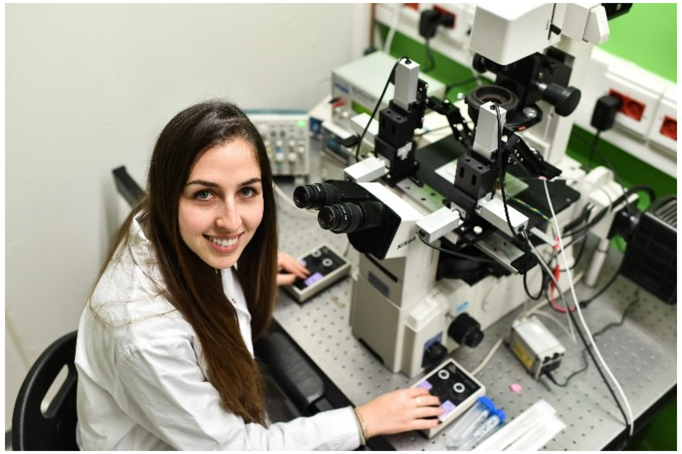
Take home message 2: AI Computer engineer is the new Jewish mam dream

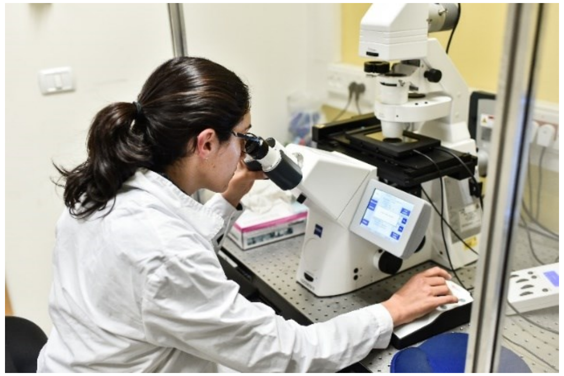


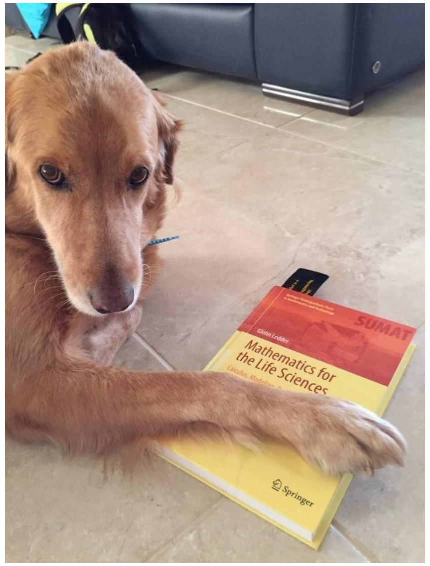


Future work: From bench to bedside and vice versa









The best way to predict the future is to create it.
Peter Drucker

Acknowledgments

Bioelectric and Bioenergetic System Lab

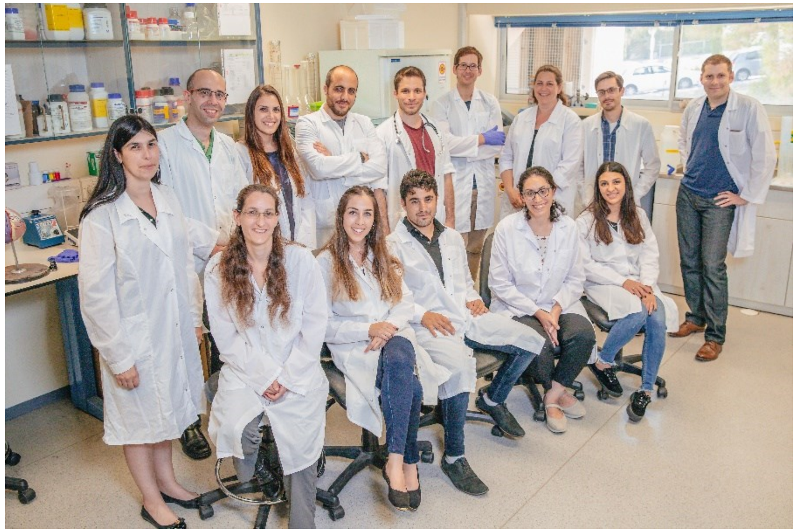
- **Ido Weiser Bitoun, MSc**
- Inbar Brosh, PhD
- Moran Davoodi, BSc
- Limor Arbel Ganon, MSc
- **Vadim Gliner, MSc**
- Daphna Marbach, MSc
- Savyon Mazgaoker, MSc
- Noa Kirschner Peretz, MSc
- **Aviv Rosenberg, MSc**
- Sofia Segal, MSc

Alumni

- David Kamoun, MSc
- Joachim Behar, PhD

Mobile Health Lab/ PhysioZoo

- **Alexandra Alexandrovich, PhD**
- Gal Eidelsztein, BSc
- Yonatan Elul, BSc
- Noam Keidar, BSc
- **Eugene Konyukhov, MD**
- **Ori Shemla, BSc**



Collaborators

- Prof. Ana M. Gomez, Inserm, France
- Prof. Edward G. Lakatta , NIA/NIH, US
- Prof. Shi-Qiang Wang, Beijing University, China
- Prof. Jin Zhang, UCSD, US
- Prof. Jose Jalife, UM, US
- Prof. Assaf Schuster, Technion, Israel
- Prof Alex Bronstein, Technion, Israel

Joy Ventures

Ministry of Science, Technology and Space

Yad Hanadiv ياد هَنديف יד הנדיב

ISRAEL SCIENCE FOUNDATION הקרן הלאומית למדע

רשות החדשנות Israel Innovation Authority

Questions?