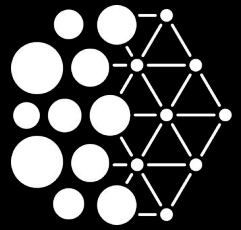
Institut
québécois
d'intelligence
artificielle

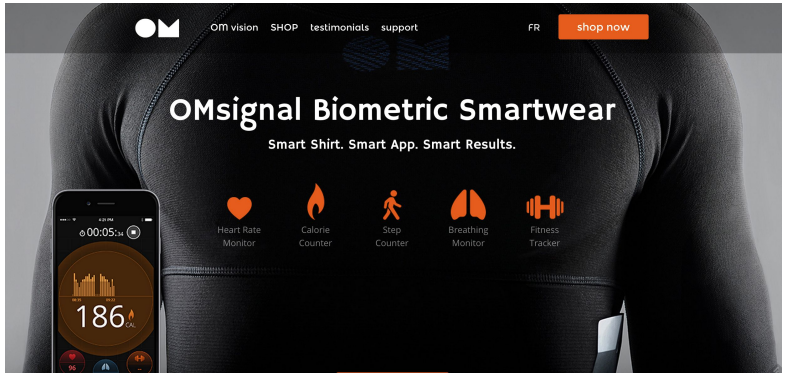


Mila

OMsignal Project: ECG Processing

Arsene Fansi-Tchango, PhD

Company



1,8 billion+
heartbeats recorded

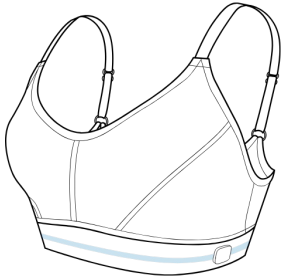
500 million+
breaths recorded

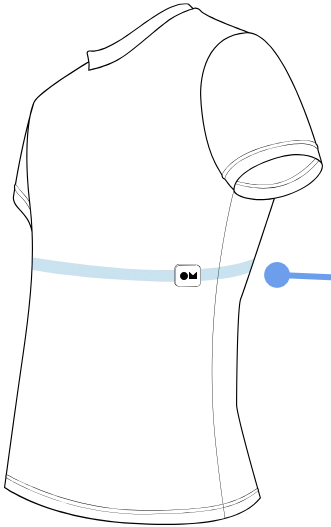
330 thousand+
hours of data collected

Make personal health and wellness central to our daily lives, through the world's most advanced biosensing apparel platform.

Technology

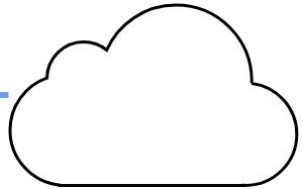
Garments
The apparel picks up the body's signals using strategically placed ECG, Respiration, and Physical Activity sensors.





Recording Module
The recording module transmits the signal to your smartphone, live.

Cloud + AI
The data is sent to the Cloud to be further analysed using advanced algorithms and AI.

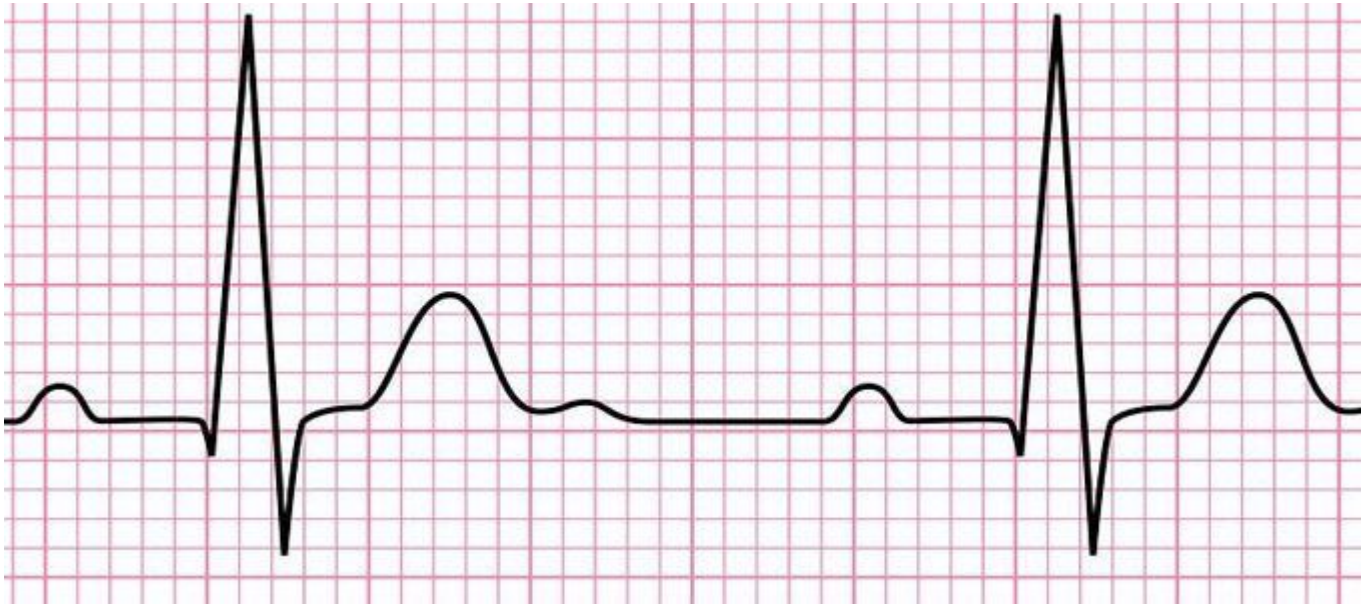


Operational Challenges

- **Easy** to collect unlabeled data
  - Huge amount of data captured under different conditions
    - running / walking / sitting / sleeping, etc...
    - different levels of signal to noise ratio

- **Hard** to label this data for supervised learning
  - Experts (e.g., medical doctors) are expensive
  - Time demanding
    - E.g., walk through all the samples of a signal

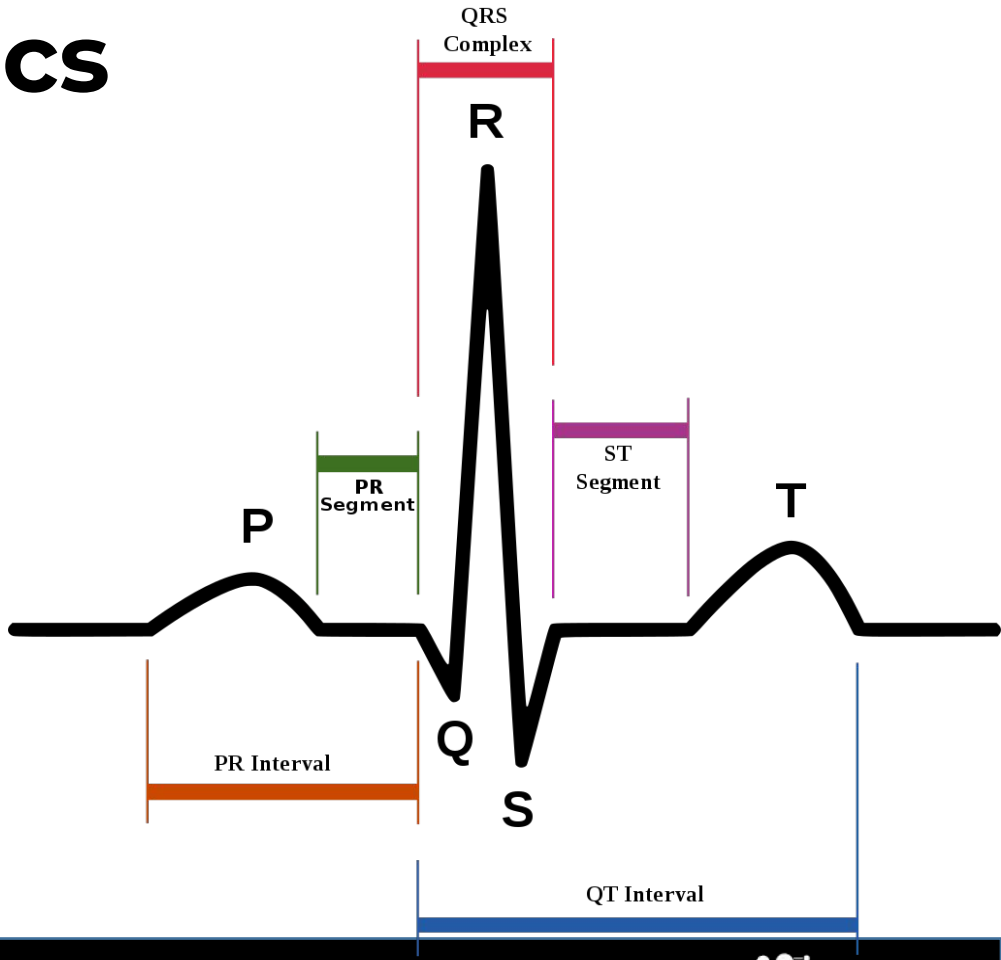
ECG Example (1 lead)



From http://www.onlinebiologynotes.com/electrocardiogram-ecg-working-principle-normal-ecg-wave-application-of-ecg/

ECG Characteristics

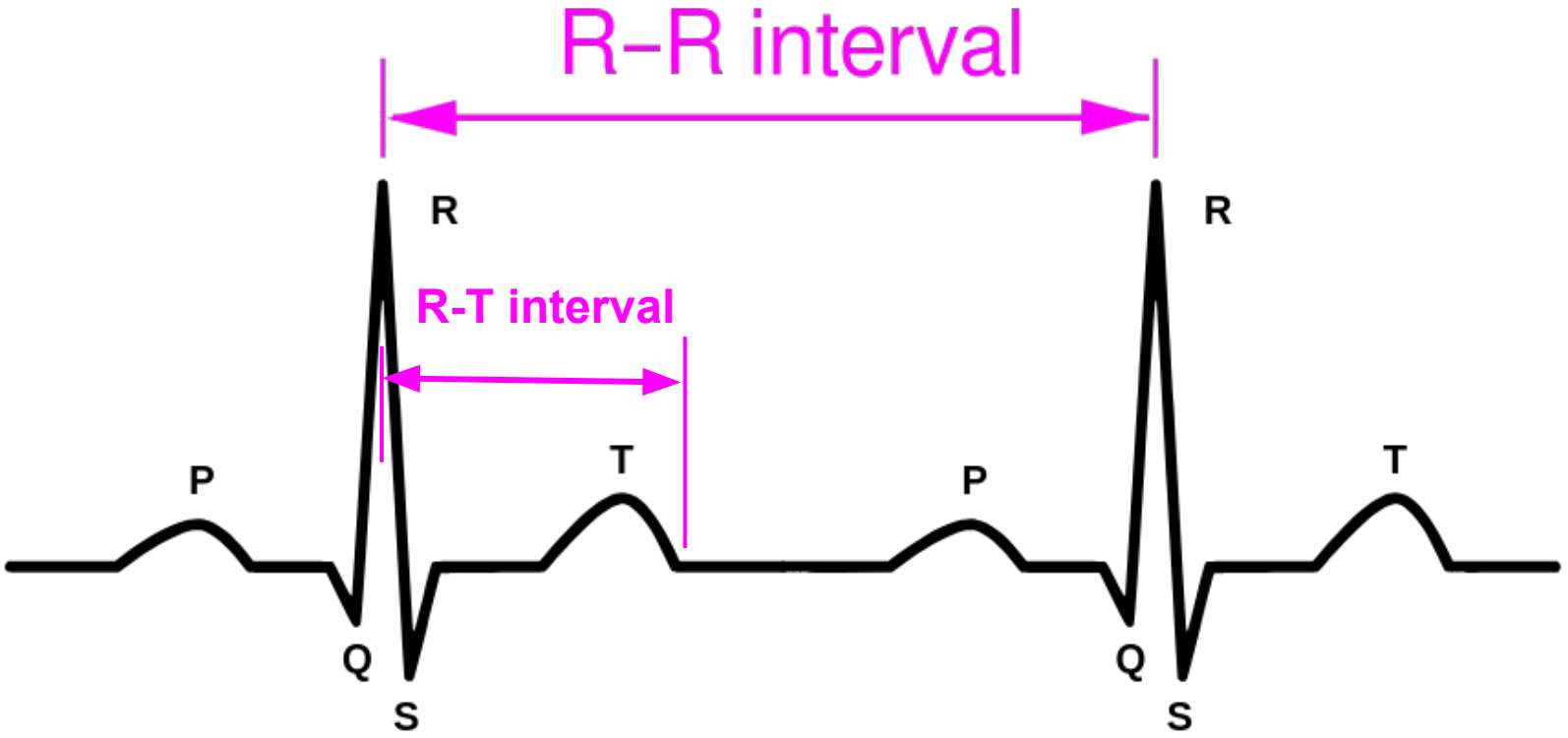
- **Fiducial points**: P, Q, R, S, T
- **P-Wave**:
  - Indicates atrial depolarization (systole)
- **QRS wave:**
  - Represents the ventricular depolarization (systole)
- **T- wave:**
  - Indicates ventricular repolarization (diastole)
- **P-R interval:**
  - Represents the time required for an impulse to travel through the atria
- **S-T segment:**
  - Represents the time when ventricular fibres are fully depolarized



From https://en.wikipedia.org/wiki/Electrocardiography

Mila

ECG Characteristics



Mila

OMsignal Project

- **Goal**: develop an **unsupervised/semi-supervised** representation learning approach that produces representations useful for tasks that have little labeled data:

    - Identification of the user
    - Fiducial point distributional information
        - Mean of the PR-Interval (real value)
        - Mean of the RT-Interval (real value)
        - Standard deviation of the RR-Interval (real value)

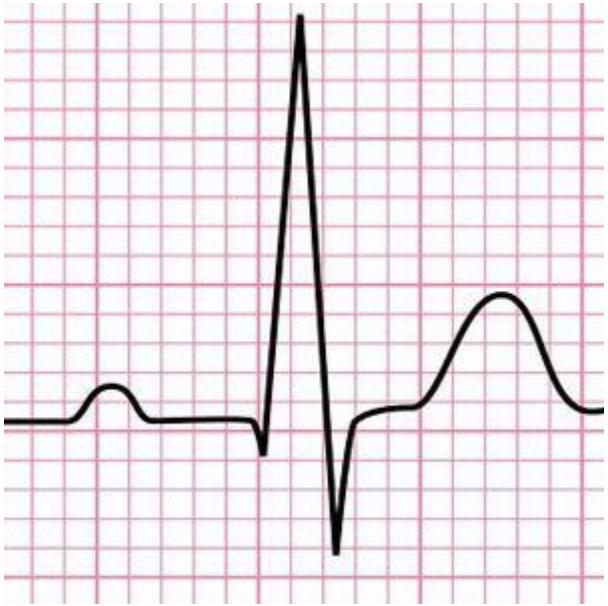
Data

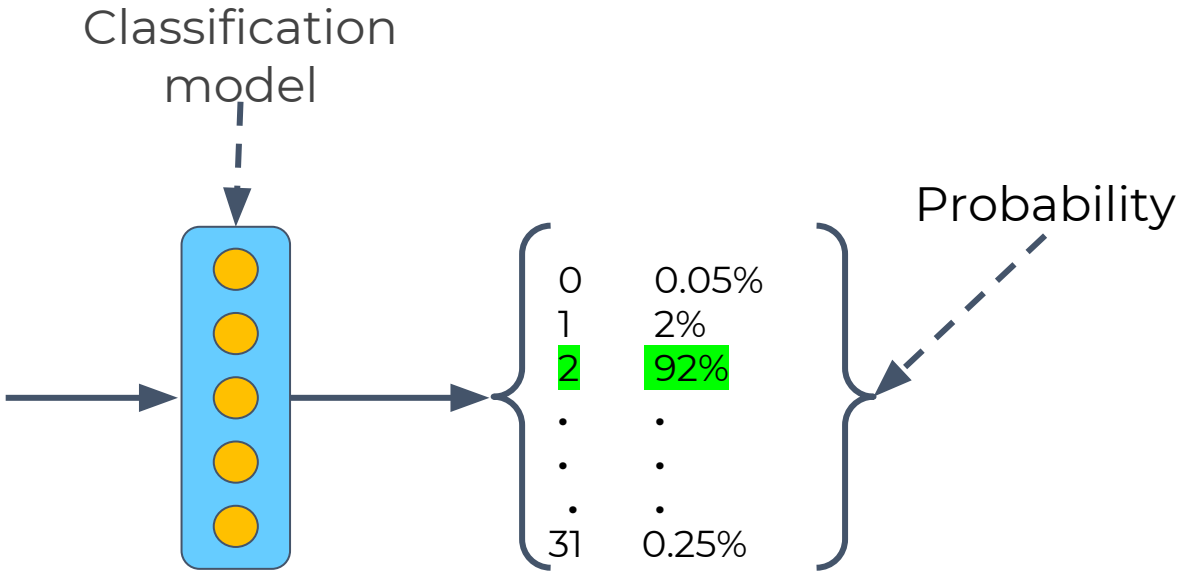
OMsignal MyHeart project:

- **Private** data
- **32** Participants
- ECG signals are divided into windows of 30 seconds each at 125 Hz (**3750** samples per window)
- Labeled data:
  - **15** windows for each participant are labeled
  - Among them, **5** windows are used as test data
  - The remaining **10** are provided as train/validation data

- Unlabeled data:
  - **657233** windows

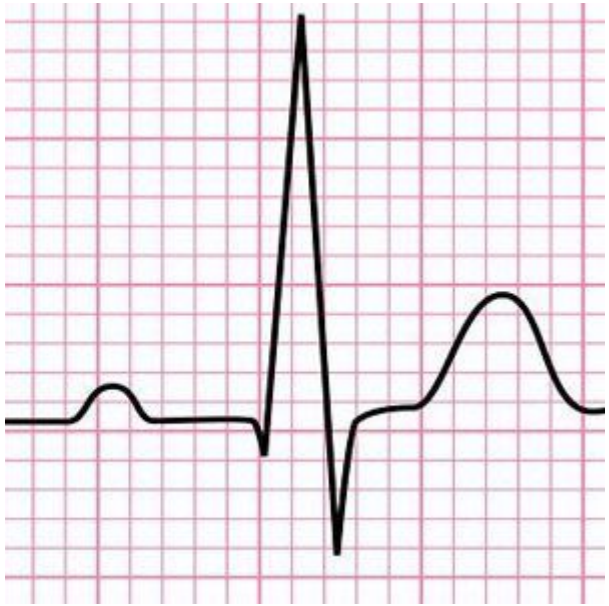
User Identification Task

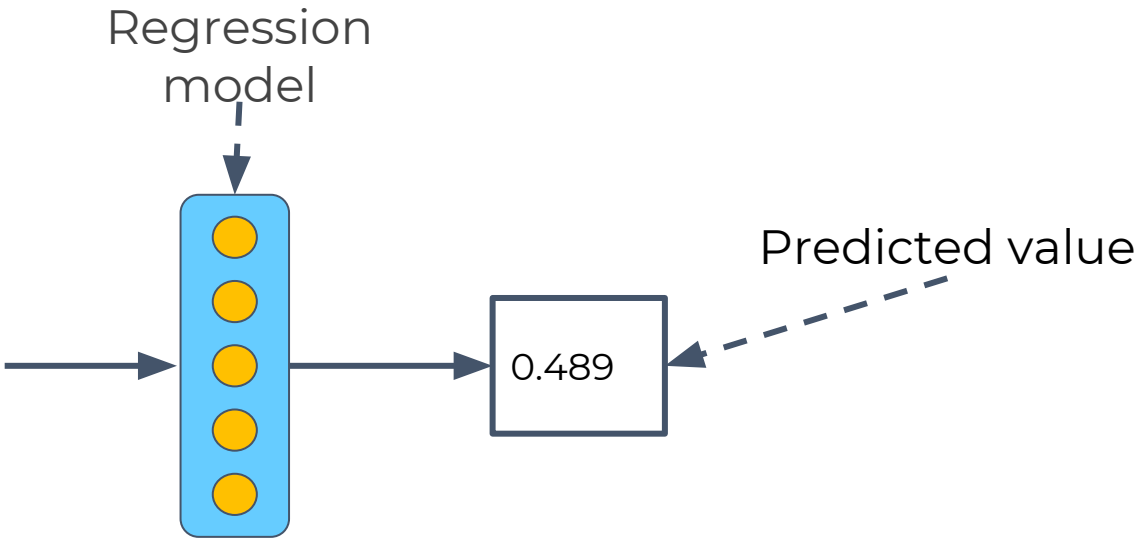




Regression Tasks

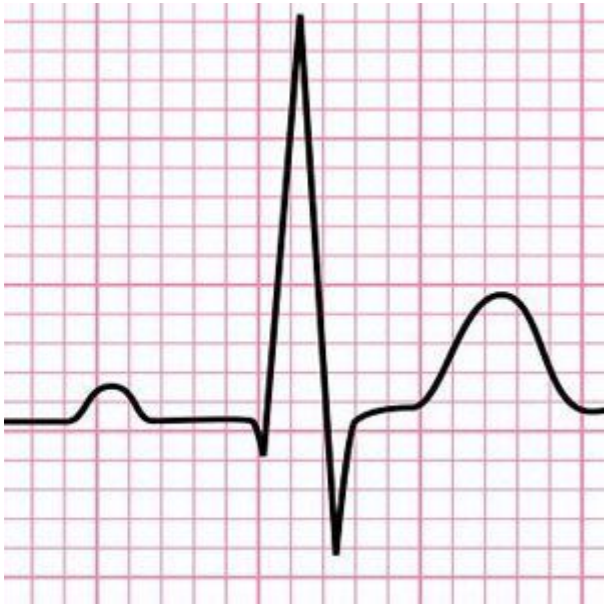
- Applicable for the prediction of the fiducial point statistics: **PR Mean, RT Mean, RR StdDev**

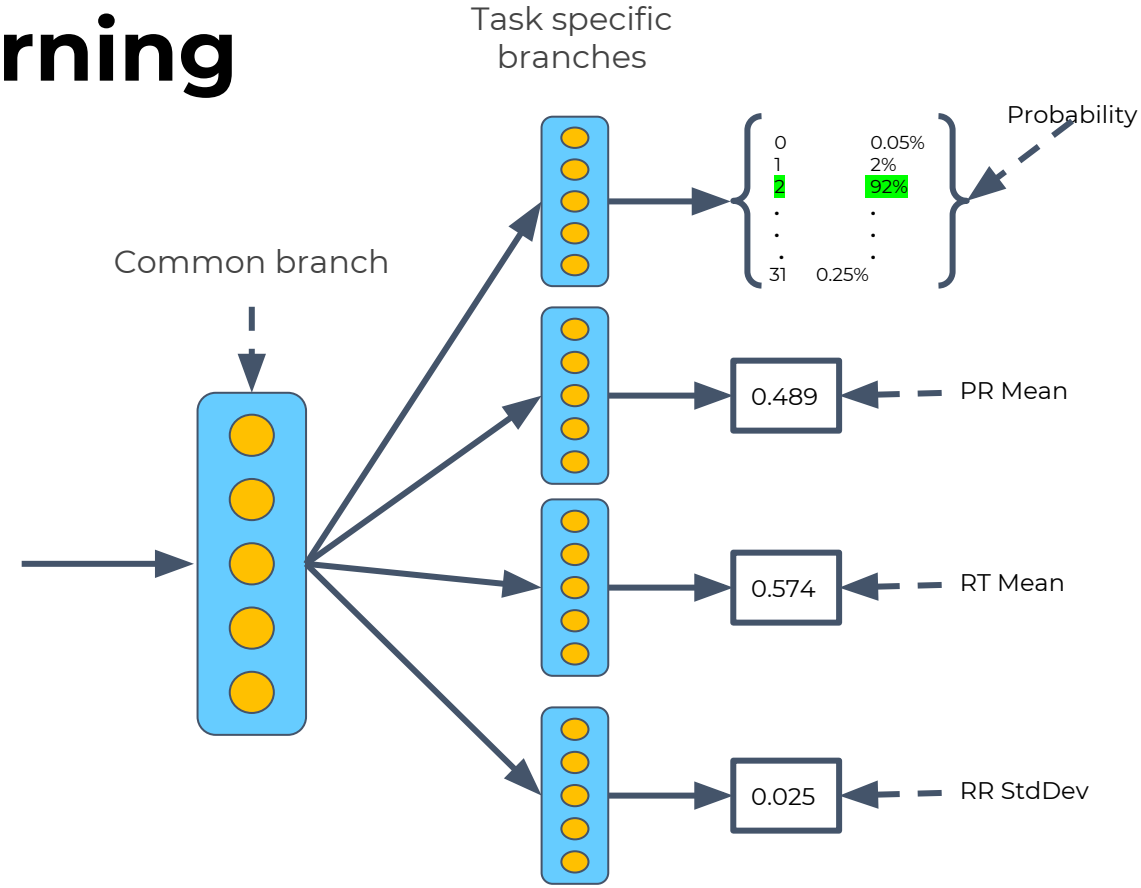




Mila

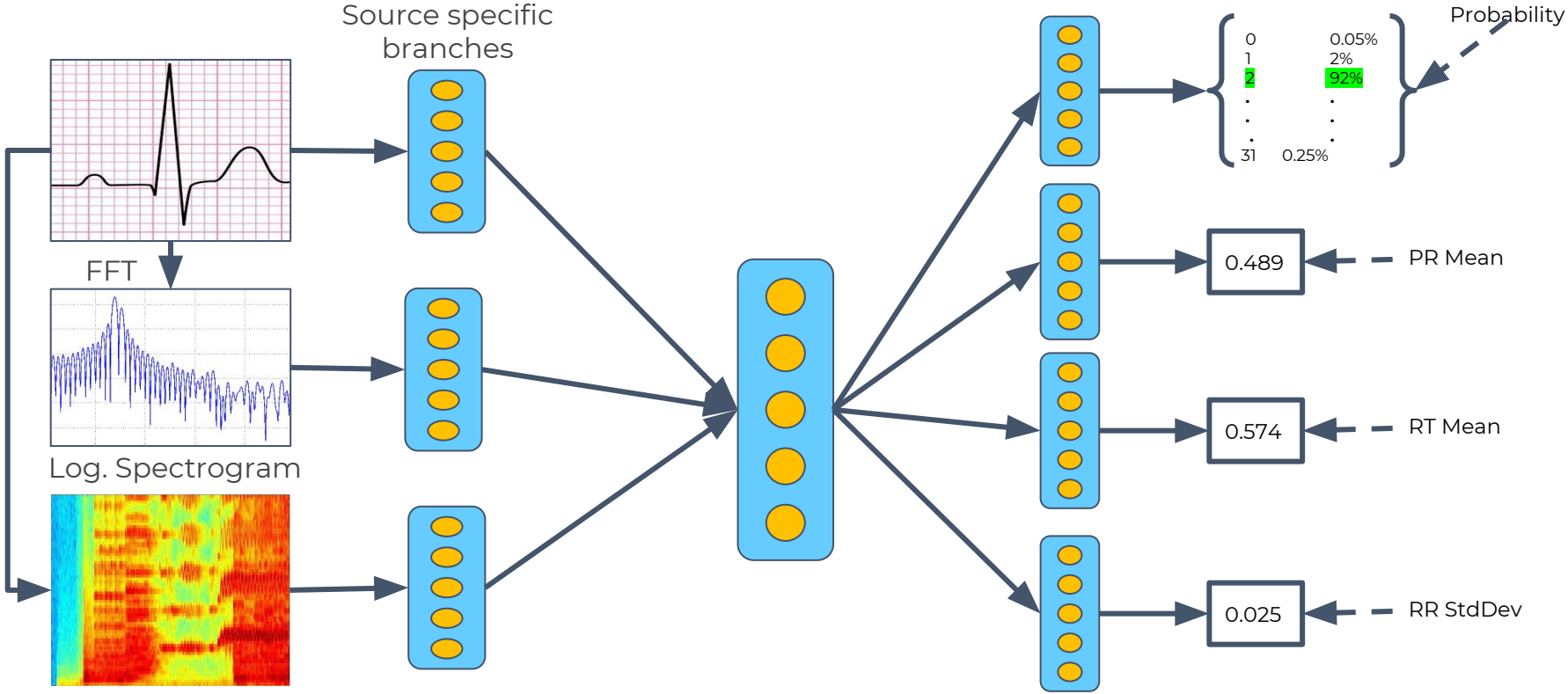
Multi-task Learning





Mila

Multi-source Multi-task Learning



Mila

Dealing with Unlabeled data

Goal: Efficient way to integrate knowledge from the unlabeled data

- **Unsupervised + Supervised Learning**
  - **Step 1**: Auto-Encoder to learn representations
  - **Step 2**: Supervised training based on representations extracted from the trained encoder
- **Semi-supervised Learning**
  - One step process.
  - Possible approaches (combined with the supervised loss):
    - Reconstruction loss (unlabeled data) - auto encoder
    - Regularization loss (unlabeled data) based on some assumptions (e.g. invariance of the output to small amounts of noise added to the input signal)

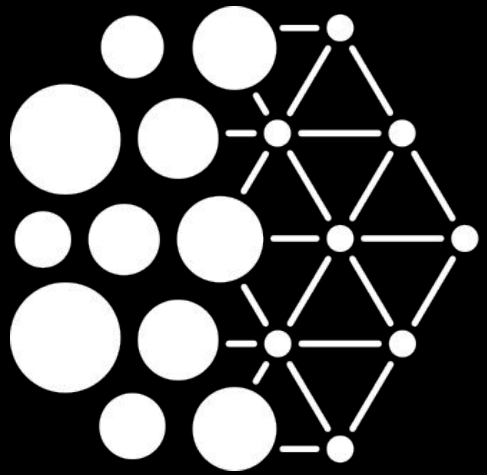
Official evaluation metrics

- Classification task
  - Macro Average Recall Score (sklearn.metrics.recall_score)
- Regression tasks
  - Kendall Correlation Score for each task (scipy.stats.kendalltau)
- Overall Score:
  - All individual scores are clipped at zero
  - Geometric mean of the scores of the 4 tasks

Informative evaluation metrics

- Cross Entropy for the classification task
- PR_Mean MSE (Mean Squared Error)
- RT_Mean MSE
- RR_StdDev MSE
- etc...

Quebec
Artificial
Intelligence
Institute



Mila

Université de Montréal

Block 1 instructions / expected timeline

	2019/01/14 week	2019/01/21 week	2019/01/28 week	2019/02/04 week
Tasks / Homework	• Data visualization • Data augmentation	• Code the data loader for the provided dataset • (optional) Implement a supervised single-task model for the identification task.	• Implement a supervised multi-task model	• Write a short report summarizing the work, and results • (Peer-) Review of other teams' code
Objectives / Deliverables	• Have a clear understanding of the data	• Data loader • (optional) Single task model	• Multi-task model	• Produce documented code and report summarizing the experimental work • Provide model for blind test set evaluation • Complete the peer code review

Block 2 instructions / expected timeline

	2019/02/11 week	2019/02/18 week	2019/02/25 week	2019/03/11 week
Tasks / Homework	• Review code and reports from previous block • TensorboardX • Code Data loader for unlabeled data	• Implement multi-task solution which leverage unlabeled data	• Continue implementing multi-task solution which leverage unlabeled data • Hyper parameter tuning	• Write a short report summarizing the work, and results • (Peer-) Review of other teams' code
Objectives / Deliverables	• Have a clear understanding of the data they will manipulate • Data loader for unlabeled data	• Choose the design pattern for incorporating unlabeled data into the training process	• Multi-task model with unlabeled data	• Produce documented code and report summarizing the experimental work • Provide model for blind test set evaluation • Complete the peer code review

Block 3 instructions / expected timeline

	2019/03/18 week	2019/03/25 week	2019/04/01 week	2019/04/08 week
Tasks / Homework	• Review code and reports from previous block • New direction of improvement/ architecture of the multi-task model with unlabeled data	• Improve or Implement a new multi-task solution which leverage unlabeled data	• Continue implementing multi-task solution which leverage unlabeled data • Hyper parameter tuning	• Write a short report summarizing the work, and results • (Peer-) Review of other teams' code
Objectives / Deliverables	• Have a clear understanding of the data they will manipulate • Have a clear understanding of the different approaches	• Choose the design pattern for incorporating unlabeled data into the training process	• Multi-task model with unlabeled data	• Produce documented code and report summarizing the experimental work • Provide model for blind test set evaluation • Complete the peer code review