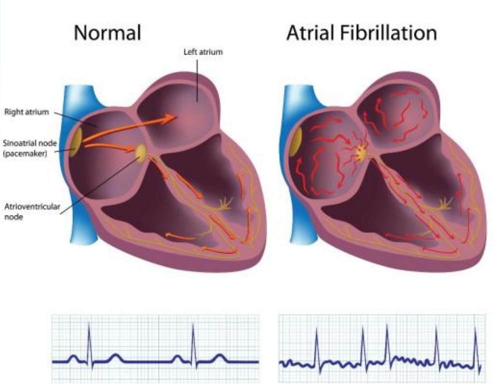
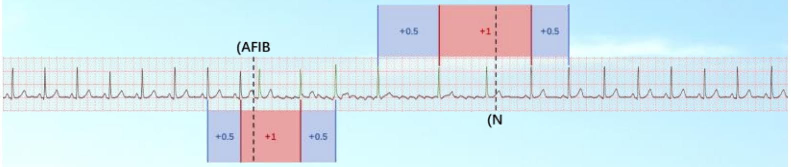
# Course Project Introduction

Xuegong Zhang
Oct. 7, 2021

## Course Project 1

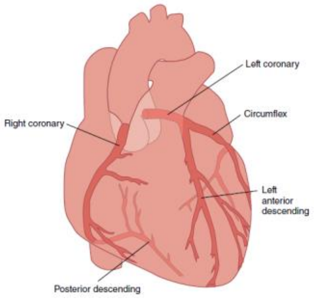
- Disease Classification/Detection in ECG signals
  - Atrial Fibrillation 房颤

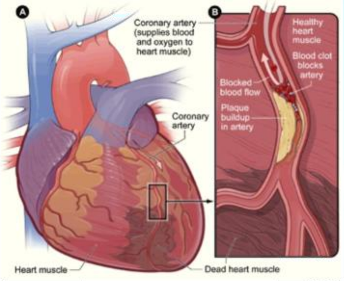


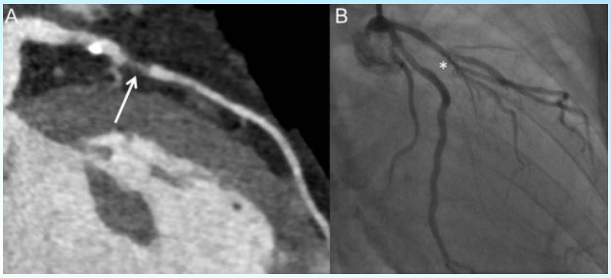


# Course Project 2

- Coronary CT Angiography (CCTA) Image Segmentation
  - Coronary Heart Disease (CHD) 冠心病
  - Coronary CT angiography  冠脉CTA、冠脉CT增强扫描、冠状动脉成像
  - Need to find data first





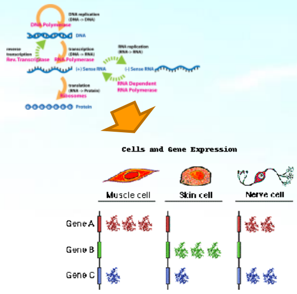


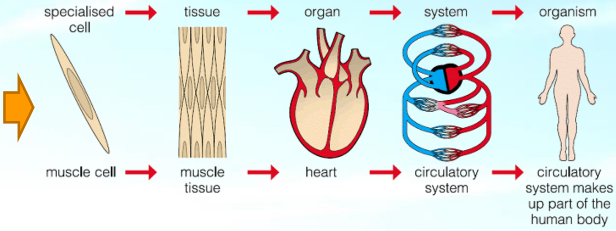
Coronary CT angiography (CCTA) in the detection of coronary artery disease. CCTA was performed in a 59-year-old man with atypical chest pain, positive family history for coronary artery disease, hypercholesterolaemia and inconclusive stress test. Curved multiplanar reconstruction of the left anterior descending coronary artery (LAD) (A) shows a severe coronary lesion due to a non-calcified atherosclerotic plaque in the proximal part of the vessel (arrow). Invasive coronary angiography (B) confirms the obstructive coronary lesion in the LAD (star).

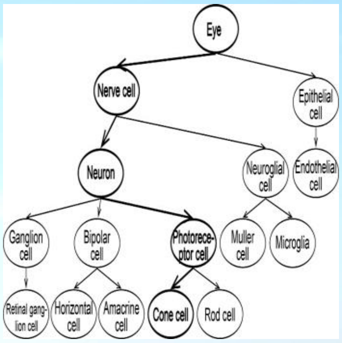
A. Rossi et al, Cornoray CT angiography for patients with suspected coronary artery disease, *Heart*, 100: 976-984, 2014

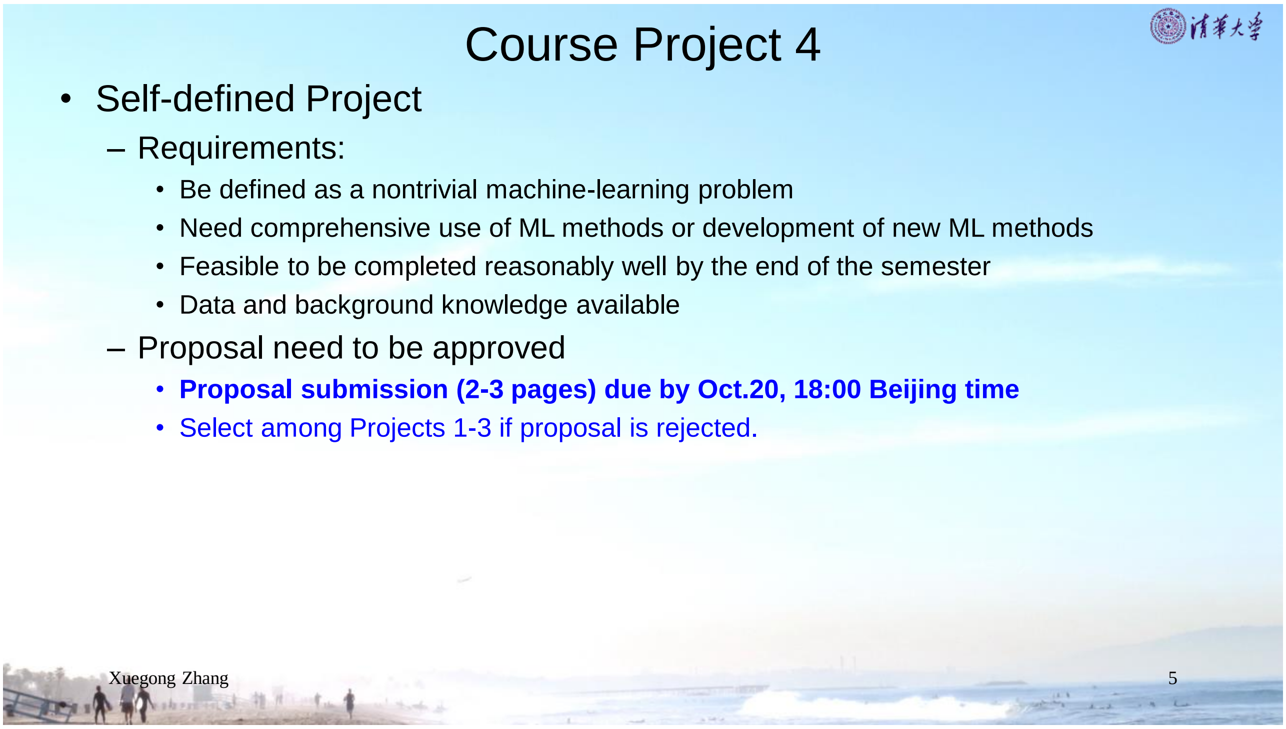
# Course Project 3

- Hierarchical Cell-type Classification
  - Same genes, different expressions → different cell types, tissues, organs
  - Hierarchical classification of cell types based on single-cell RNA-seq data


DNA Polymerase
DNA replication (DNA -> DNA)
DNA
reverse transcription
Rev. Transcriptase
transcription (DNA -> RNA)
RNA Polymerase
RNA replication (RNA -> RNA)
(+) Sense RNA
(-) Sense RNA
translation (RNA -> Protein)
Ribosomes
RNA Dependent RNA Polymerase
Protein
Cells and Gene Expression
Muscle cell
Skin cell
Nerve cell
Gene A
Gene B
Gene C



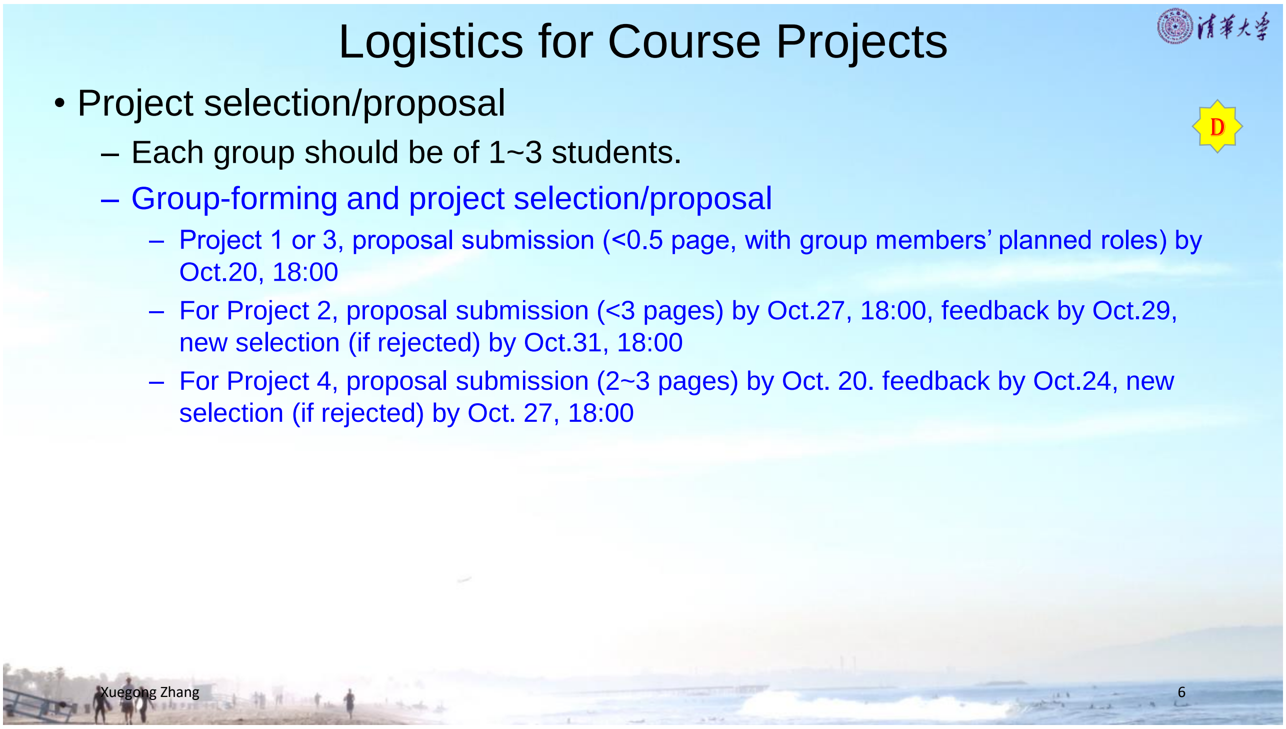




# Course Project 4

- Self-defined Project
  - Requirements:
    - Be defined as a nontrivial machine-learning problem
    - Need comprehensive use of ML methods or development of new ML methods
    - Feasible to be completed reasonably well by the end of the semester
    - Data and background knowledge available
  - Proposal need to be approved
    - **Proposal submission (2-3 pages) due by Oct.20, 18:00 Beijing time**
    - Select among Projects 1-3 if proposal is rejected.

Xuegong Zhang 5

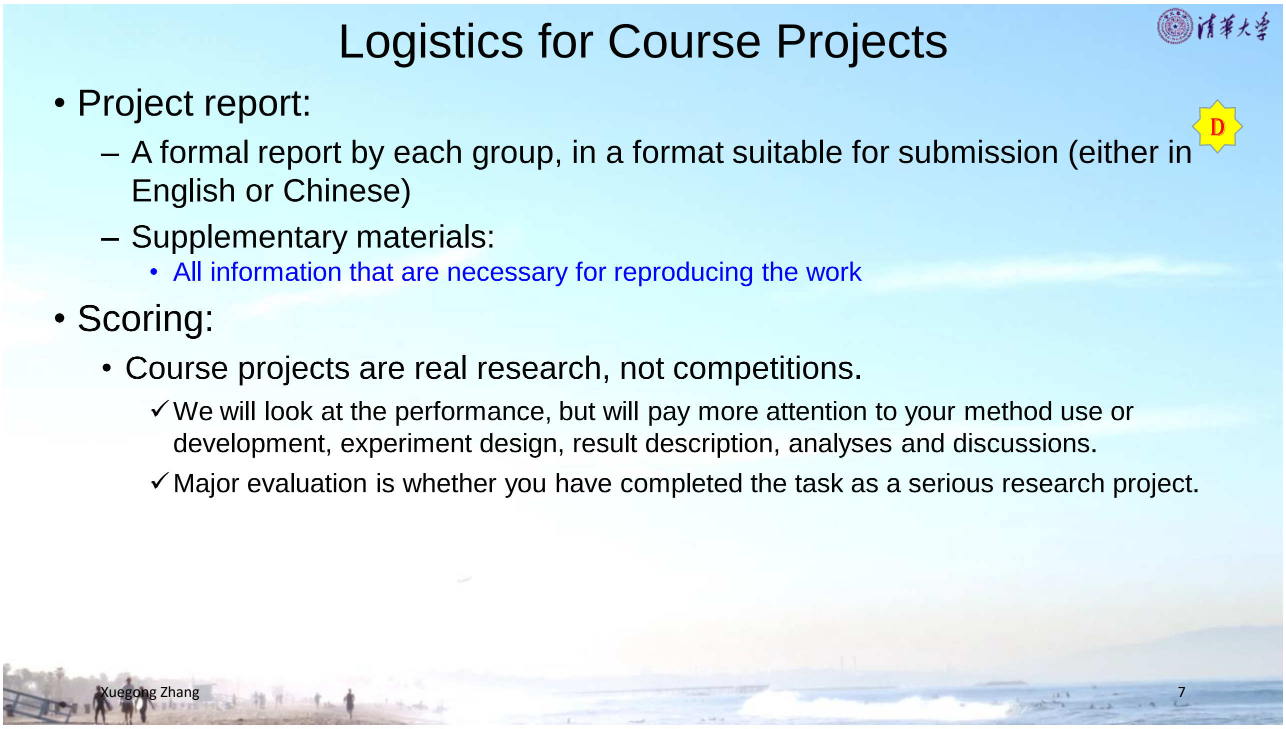


# Logistics for Course Projects

D

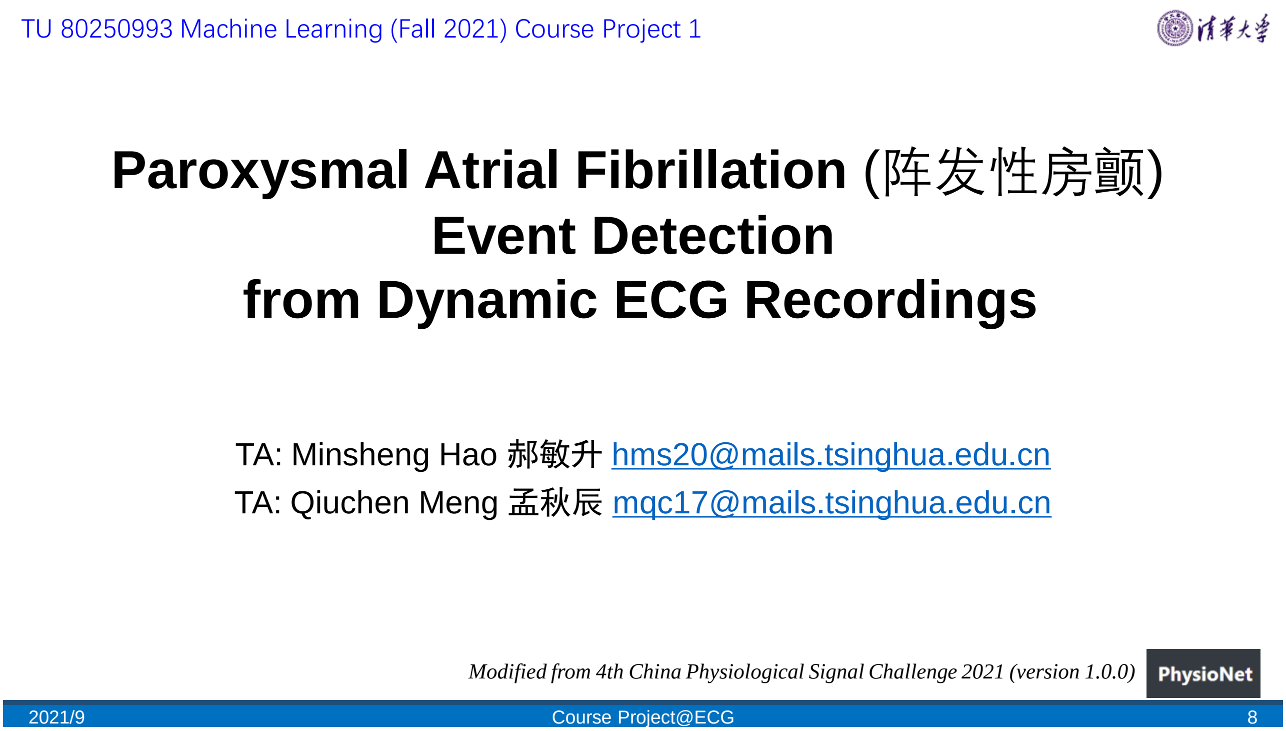
- Project selection/proposal
  - Each group should be of 1~3 students.
  - Group-forming and project selection/proposal
    - Project 1 or 3, proposal submission (<0.5 page, with group members' planned roles) by Oct.20, 18:00
    - For Project 2, proposal submission (<3 pages) by Oct.27, 18:00, feedback by Oct.29, new selection (if rejected) by Oct.31, 18:00
    - For Project 4, proposal submission (2~3 pages) by Oct. 20. feedback by Oct.24, new selection (if rejected) by Oct. 27, 18:00

Xuegong Zhang 6



## Logistics for Course Projects

- Project report:
  - A formal report by each group, in a format suitable for submission (either in English or Chinese)
  - Supplementary materials:
    - All information that are necessary for reproducing the work
- Scoring:
  - Course projects are real research, not competitions.
    - ✓ We will look at the performance, but will pay more attention to your method use or development, experiment design, result description, analyses and discussions.
    - ✓ Major evaluation is whether you have completed the task as a serious research project.



TU 80250993 Machine Learning (Fall 2021) Course Project 1

# Paroxysmal Atrial Fibrillation (阵发性房颤) Event Detection from Dynamic ECG Recordings

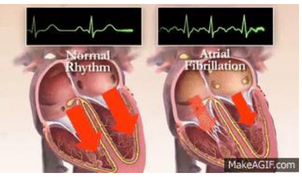
TA: Minsheng Hao 郝敏升 hms20@mails.tsinghua.edu.cn

TA: Qiuchen Meng 孟秋辰 mqc17@mails.tsinghua.edu.cn

*Modified from 4th China Physiological Signal Challenge 2021 (version 1.0.0)* PhysioNet

# Overview

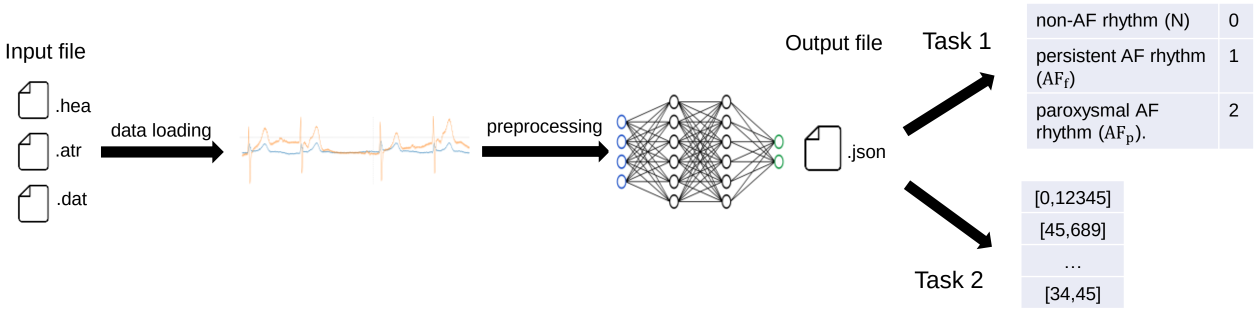
- Electrocardiogram (ECG) is a signal of voltage versus time and can capture cardiac abnormalities.
- We have 1,425 two channels time-series ECG records sampled at 200 Hz, which are obtained from the normal, persistent AF ($\mathrm{AF_f}$) and paroxysmal AF ($\mathrm{AF_p}$) patients.
- You need to
  - ***classify these records into three types***
  - ***detect the onset and end locations of the abnormal heartbeat episode***



*Atrial fibrillation (AF) is the most frequent arrhythmia. Early detection and classification of PAF is of great value for surgery options, drug intervention and diagnosis and treatment of various clinical complications.*

# Pipeline

- Tasks
  - classify the heart rhythm
  - detect the location of onsets and ends of $\mathrm{AF_p}$ episodes in dynamic ECG



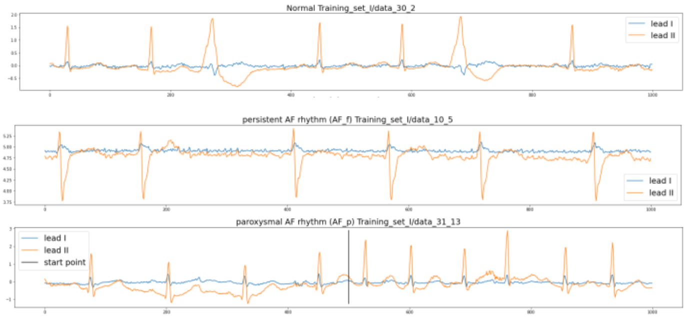
| | |
|---|---|
| non-AF rhythm (N) | 0 |
| persistent AF rhythm ($\mathrm{AF_f}$) | 1 |
| paroxysmal AF rhythm ($\mathrm{AF_p}$). | 2 |

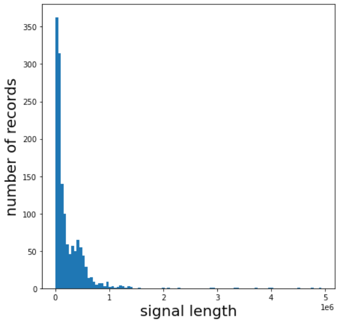
| |
|---|
| [0,12345] |
| [45,689] |
| … |
| [34,45] |

*Wang, X., Ma, C., Zhang, X., Gao, H., Clifford, G., & Liu, C. (2021). Paroxysmal Atrial Fibrillation Events Detection from Dynamic ECG Recordings: The 4th China Physiological Signal Challenge 2021 (version 1.0.0). PhysioNet. https://doi.org/10.13026/ksya-qw89.*
*Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220.*

# Data

- **1,425** (1,138 for training, 286 for testing) ECG records which are obtained from 12-lead Holter or 3-lead wearable ECG monitoring devices.
- Each ECG record extracted from **lead I** and **lead II** of the long-term dynamic ECGs and sampled at **200 Hz**. Most of the data have less than 1,000,000 time points.
- To avoid ambiguity in annotation, an AF episode is limited to contain **no less than 5** heartbeats.



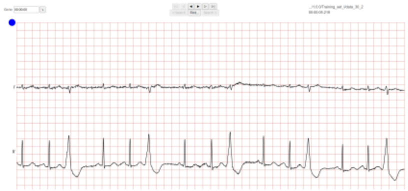


# Data

- Data format: WFDB MIT format
  - Signal files (.dat) : digitized signals
  - Header files (.hea) : classification label, signal length, signal frequency etc.
  - Annotation files (.atr) : the onset and end locations of the atrial fibrillation interval; the R peak location.
- You can use the ***wfdb.rdsamp*** and ***wfdb.rdann*** from python ***wfdb*** package to read these files: *https://wfdb.readthedocs.io/en/latest/wfdb.html*
- The training and test sample lists, sample code and all data are on the Tsinghua cloud:

  https://cloud.tsinghua.edu.cn/d/5a65c16c3ac74741a747/.
- Online visualization:

  https://physionet.org/lightwave/?db=cpsc2021/1.0.0



# Tasks

**Task1**

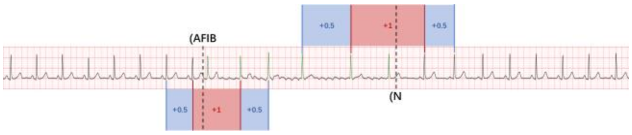
- To classify the rhythm types: non-AF rhythm (N), persistent AF rhythm ($\mathrm{AF_f}$) and paroxysmal AF rhythm ($\mathrm{AF_p}$).
- *Evaluation:* The *ur_calculate* function in the *score_2021.py* can calculate the classification score using the scoring matrix below.

| | | Prediction | | |
|---|---|---|---|---|
| | | N | $\mathrm{AF_f}$ | $\mathrm{AF_p}$ |
| Answer | N | +1 | -1 | -0.5 |
| | $\mathrm{AF_f}$ | -2 | +1 | 0 |
| | $\mathrm{AF_p}$ | -1 | 0 | +1 |

# Tasks

**Task2**

- To locate the onset and end points for all AF episode predictions.
- *Evaluation:* There are only reward scores in this task. The *ue_calculate* function in the *score_2021.py* can calculate the task2 score ($U_e$). And the final score $U$ is the sum of the task1 and task2 score.



$$U = \frac{1}{N}\sum_{i=1}^{N}\left(U_{r_i} + \frac{M_{a_i}}{\max\{M_{r_i}, M_{a_i}\}} \times U_{e_i}\right)$$

$U_r$- Task1 score
$U_e$- Task2 score

N-number of records

$M_a$ the amount of annotated AF episodes
$M_r$ the amount of predicted AF episodes

**All these evaluation functions *ur_calculate* and *ue_calculate* can be found in the *score_2021.py***

# Input & Output

- **Input:** Note that you could only use the signal data as the input data, all the other information like R peak location etc. in the annotation files can only be used as the auxiliary label.

- **Output:** A JSON file for each record as the output of both two tasks. It contains the onset and end locations:

*{'predict_endpoints': [[2169.0, 39395.0], [41344.0, 52822.0]]}*

- If the ECG record is classified as **N**, the predict_endpoints field should be an **empty list**. If the current ECG record is classified as $\boldsymbol{AF_f}$, the provided onset and end locations should be **the first and last** record points. Otherwise ($\boldsymbol{AF_p}$), the file will contain at least one pair of locations.

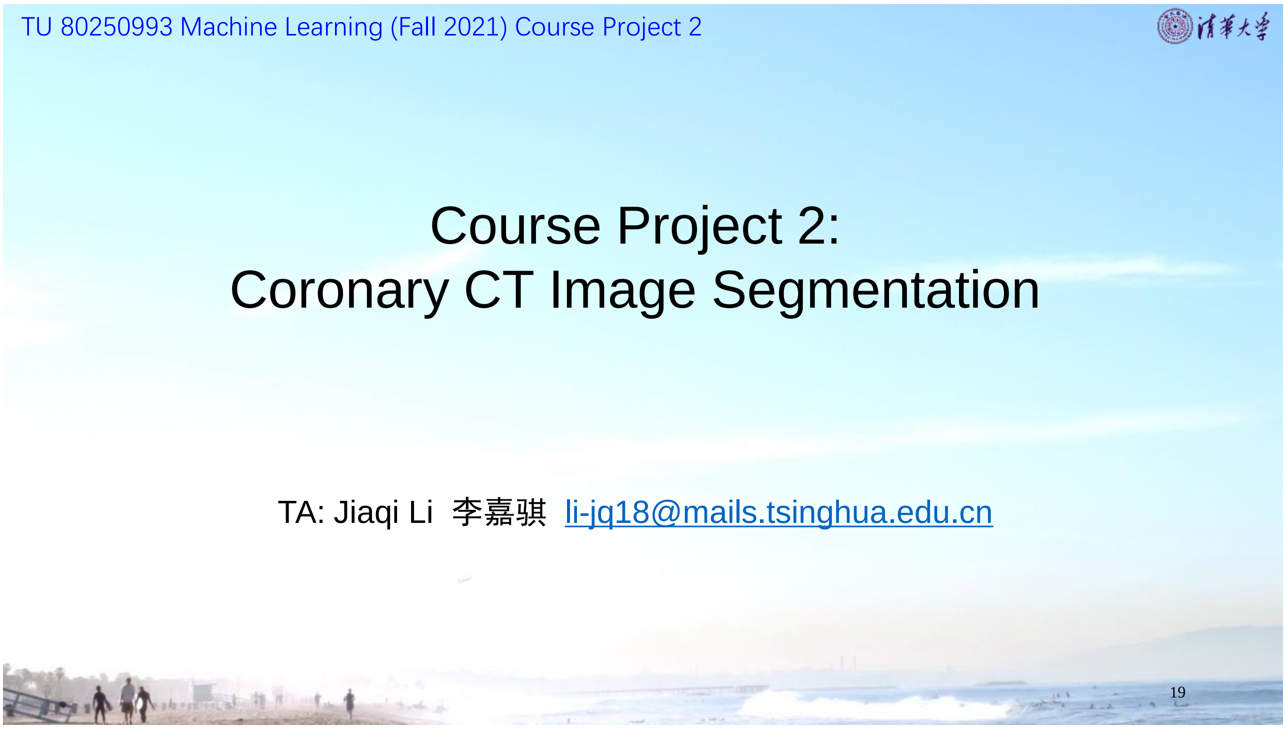
# Requirement

- You need to submit your code (python, Matlab or R) and report (PDF).

- Your final report should cover the details of your model and its performance on training and test datasets. Please report both your average classification score $U_r$ and final score $U$.

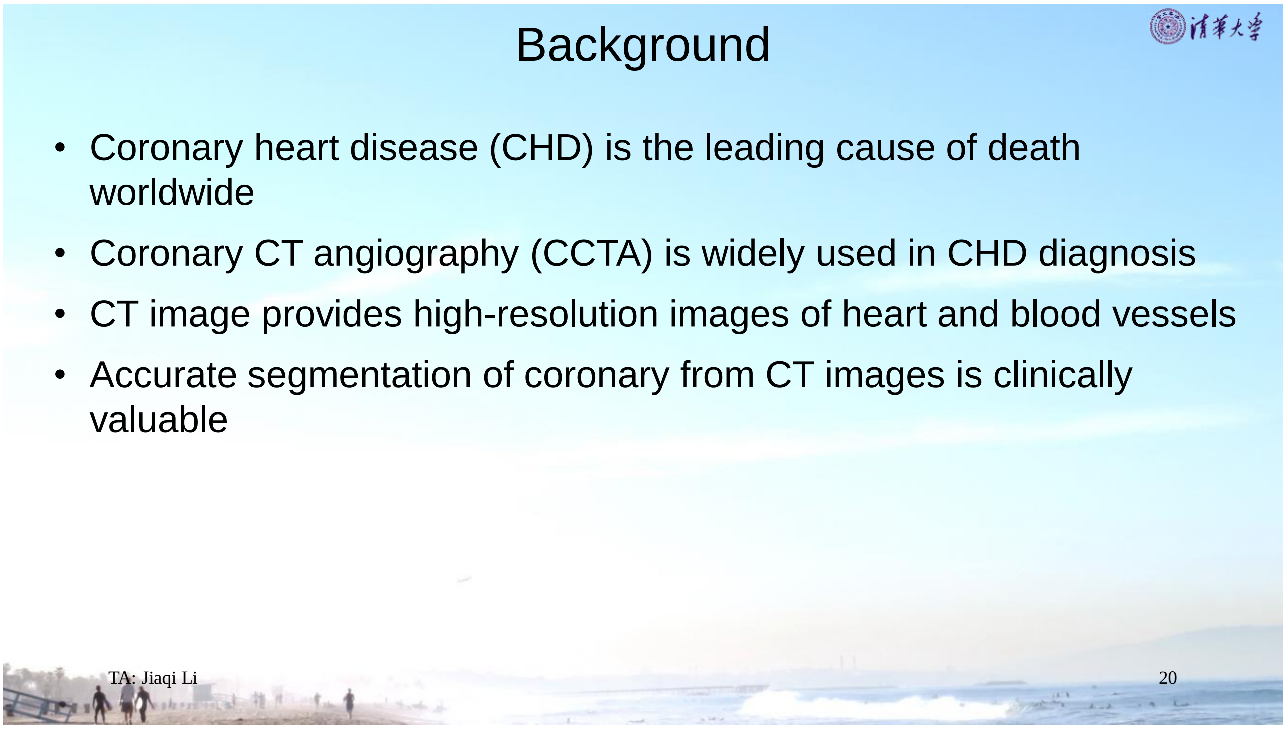
- We suggest that you build your own model. You can use pre-trained models to improve its performance.

- Your code must be executable and contain a readme file. We may run your model and check the outputs and test accuracy.



# Course Project 2:
# Coronary CT Image Segmentation

TA: Jiaqi Li 李嘉骐 li-jq18@mails.tsinghua.edu.cn

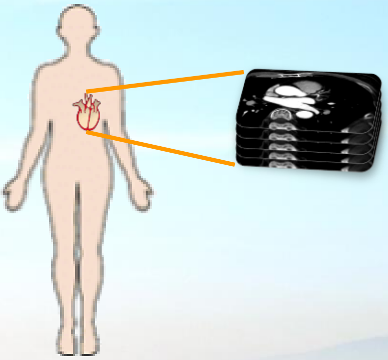


## Background

- Coronary heart disease (CHD) is the leading cause of death worldwide
- Coronary CT angiography (CCTA) is widely used in CHD diagnosis
- CT image provides high-resolution images of heart and blood vessels
- Accurate segmentation of coronary from CT images is clinically valuable

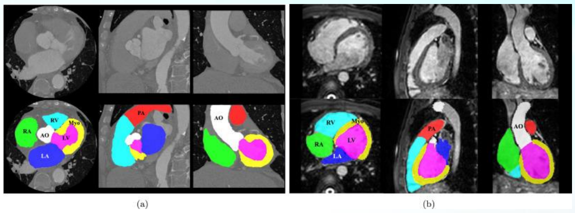
# Introduction to CCTA

- Coronary CT angiography (CCTA)
  - CT visualizes certain part of human body with image slices
  - Blood becomes brighter than nearby tissues in CT image after injecting contrast agent
  - More detailed inspection of vessels using CCTA
  - Common data formats: .dcm, .nii, .mhd, etc.



# Tasks

- Train a segmentation model to predict the pixel-level location of coronary in CT images.
- Detect and calculate the proportion of coronary occlusion.
- (Optional) Design some other applications using the segmented coronary from task 1.



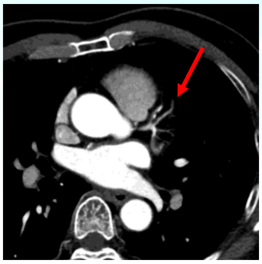
Example of image segmentation task

# Data

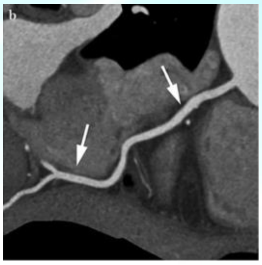
- You need to propose your own dataset
  - The dataset should contain at least 25 samples.
  - Each sample should have paired CT images (3D) of heart and pixel-level masks of coronary.
  - Data can be collected from public datasets (if available) or other datasets you can get usage permission.

# Examples

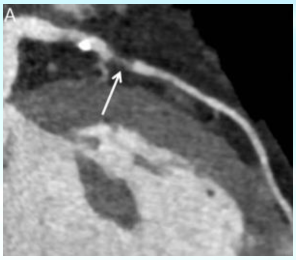
2D CT image of heart and coronary



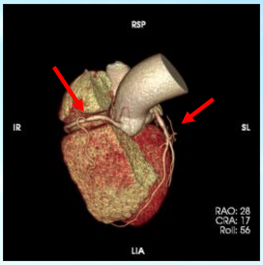
Normal coronary



Coronary with occlusion

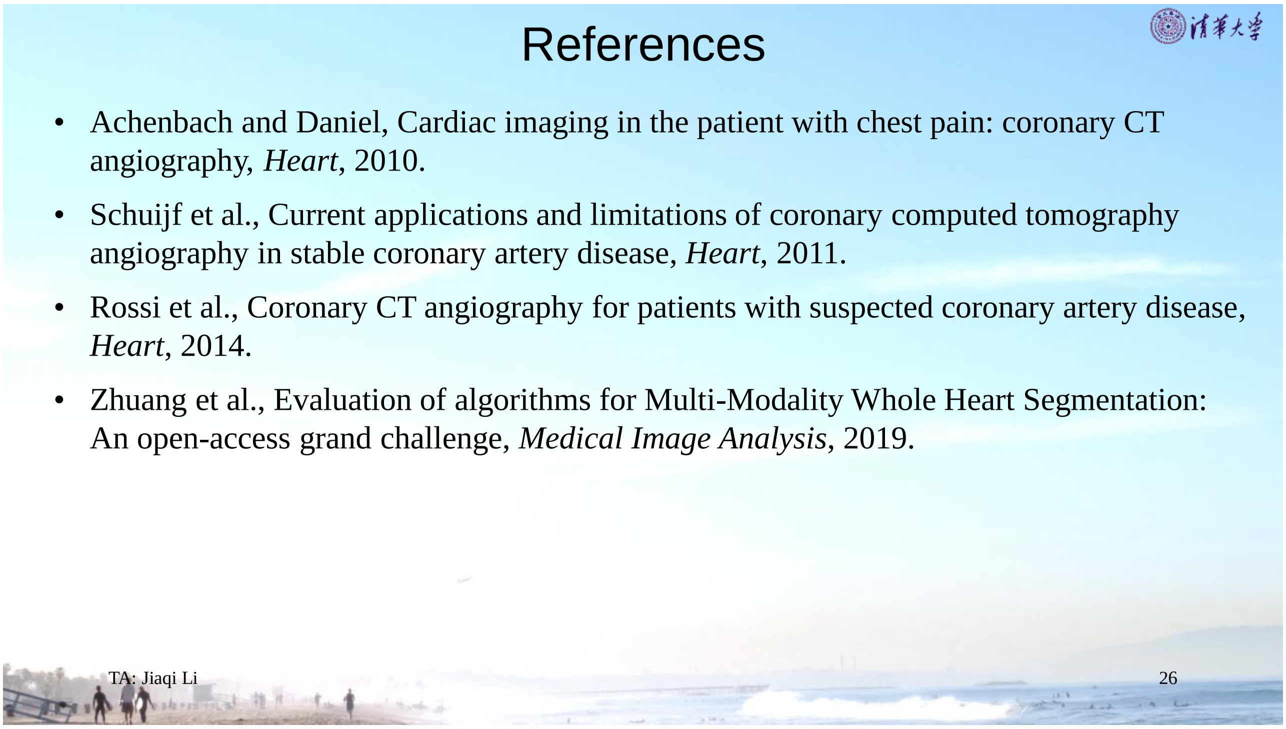


3D reconstructed heart model



# Requirements

- Write a proposal describing the collected dataset, experimental design and expected results.
- Proposal need to be approved
  - Proposal submission (< 3 pages) due by Oct.27, 18:00 Beijing time
  - Select among Projects 1-2 if proposal is rejected (announced by Oct.29, new selection due by Oct.31).
  - Discuss with TA Jiaqi Li (li-jq18@mails.tsinghua.edu.cn)



# References

- Achenbach and Daniel, Cardiac imaging in the patient with chest pain: coronary CT angiography, *Heart*, 2010.
- Schuijf et al., Current applications and limitations of coronary computed tomography angiography in stable coronary artery disease, *Heart*, 2011.
- Rossi et al., Coronary CT angiography for patients with suspected coronary artery disease, *Heart*, 2014.
- Zhuang et al., Evaluation of algorithms for Multi-Modality Whole Heart Segmentation: An open-access grand challenge, *Medical Image Analysis*, 2019.

TU 80250993 Machine Learning (Fall 2021) Course Project 3

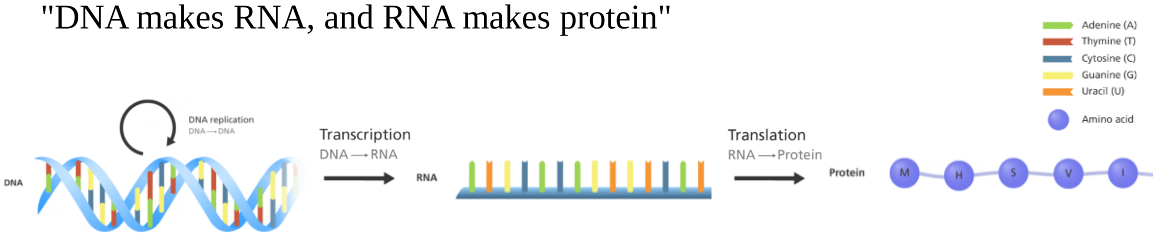
# Hierarchical Cell-type Classification

TA: Xi Xi 席曦 xix19@mails.tsinghua.edu.cn
TA: Yixin Chen 陈奕鑫 chenyx19@mails.tsinghua.edu.cn

## Background

- **Central Dogma**

"DNA makes RNA, and RNA makes protein"



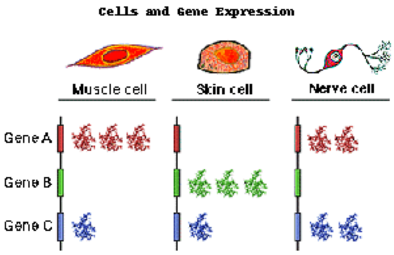
- **Gene**

A basic unit of heredity. A sequence of nucleotides in DNA or RNA that encodes the synthesis of a gene product, either RNA or protein.

# Background

- **Gene expression**

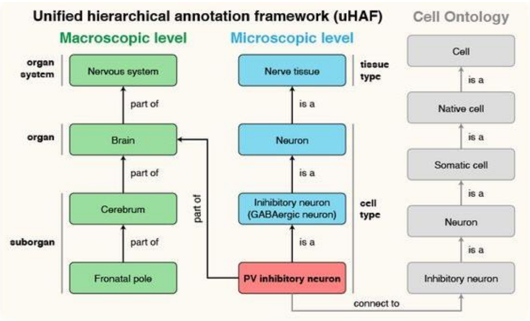
The process by which the instructions in DNA are converted into a functional product, such as a protein, **through transcription and translation**.

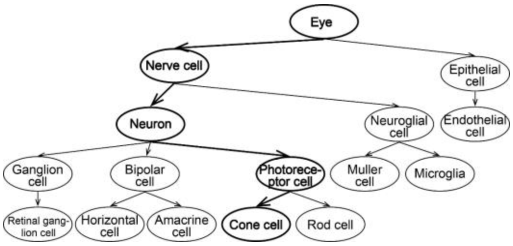


Same genes, different expressions
→ different cell types, tissues, organs

# Background

- A unified hierarchical cell annotation framework (**uHAF**) can be used to organize cell types in a **hierarchical tree structure**





Chen, S. *et al. A human ensemble cell atlas (hECA) enables* in data *cell sorting*. http://biorxiv.org/lookup/doi/10.1101/2021.07.21.453289 (2021)

# Background

- **Hierarchical classification**

A special classification task, where the classes involved in the problem are hierarchically structured.

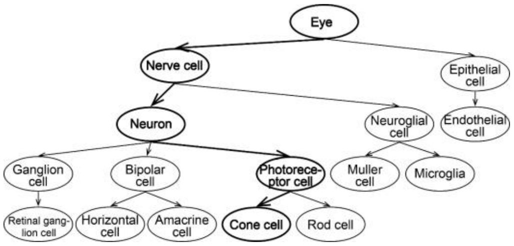
✓**Hierarchical single-label classification**

A hierarchical classifier can assign at most one class at any given level of the hierarchy.

- **Hierarchical multi-label classification**

A hierarchical classifier can assign more than one class at any given level of the hierarchy.

An example of hierarchical classification of cell types on uHAF tree:

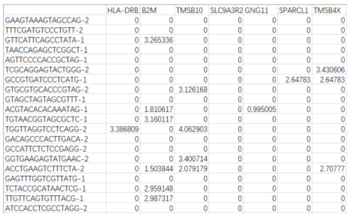


# Goal

- Hierarchical single-label classification of cell types on uHAF tree according to gene expressions

Gene

Cell

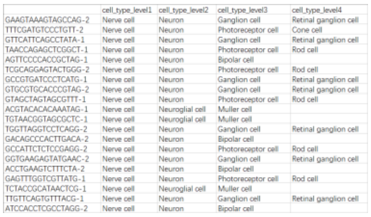


| | HLA-DRB: | B2M | TMSB10 | SLC9A3R2 | GNG11 | SPARCL1 | TMSB4X |
|---|---|---|---|---|---|---|---|
| GAAGTAAAGTAGCCAG-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TTTCGATGTCCCTGTT-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GTTCATTCAGCCTATA-1 | 0 | 3.265336 | 0 | 0 | 0 | 0 | 0 |
| TAACCAGAGCTCGGCT-1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AGTTCCCCACCGCTAG-1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TCGCAGGAGTACTGGG-2 | 0 | 0 | 0 | 0 | 0 | 0 | 3.430606 |
| GCCGTGATCCCTCATG-1 | 0 | 0 | 0 | 0 | 0 | 2.64783 | 2.64783 |
| GTGCGTGCACCCGTAG-2 | 0 | 0 | 3.126168 | 0 | 0 | 0 | 0 |
| GTAGCTAGTAGCGTTT-1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ACGTACACACAAATAG-1 | 0 | 1.810617 | 0 | 0 | 0.995005 | 0 | 0 |
| TGTAACGGTAGCGCTC-1 | 0 | 3.160117 | 0 | 0 | 0 | 0 | 0 |
| TGGTTAGGTCCTCAGG-2 | 3.386809 | 0 | 4.062903 | 0 | 0 | 0 | 0 |
| GACAGCCCACTTGACA-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GCCATTCTCTCCGAGG-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GGTGAAGAGTATGAAC-2 | 0 | 0 | 3.400714 | 0 | 0 | 0 | 0 |
| ACCTGAAGTCTTTCTA-2 | 0 | 1.503844 | 2.079179 | 0 | 0 | 0 | 2.70777 |
| GAGTTTGGTCGTTATG-1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TCTACCGCATAACTCG-1 | 0 | 2.959148 | 0 | 0 | 0 | 0 | 0 |
| TTGTTCAGTGTTTACG-1 | 0 | 2.987317 | 0 | 0 | 0 | 0 | 0 |
| ATCCACCTCGCCTAGG-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Gene expressions of a cell**

**A machine learning model**

Cell types in different levels in the hierarchy



| | cell_type_level1 | cell_type_level2 | cell_type_level3 | cell_type_level4 |
|---|---|---|---|---|
| GAAGTAAAGTAGCCAG-2 | Nerve cell | Neuron | Ganglion cell | Retinal ganglion cell |
| TTTCGATGTCCCTGTT-2 | Nerve cell | Neuron | Photoreceptor cell | Cone cell |
| GTTCATTCAGCCTATA-1 | Nerve cell | Neuron | Ganglion cell | Retinal ganglion cell |
| TAACCAGAGCTCGGCT-1 | Nerve cell | Neuron | Photoreceptor cell | Rod cell |
| AGTTCCCCACCGCTAG-1 | Nerve cell | Neuron | Bipolar cell | |
| TCGCAGGAGTACTGGG-2 | Nerve cell | Neuron | Photoreceptor cell | Rod cell |
| GCCGTGATCCCTCATG-1 | Nerve cell | Neuron | Ganglion cell | Retinal ganglion cell |
| GTGCGTGCACCCGTAG-2 | Nerve cell | Neuron | Ganglion cell | Retinal ganglion cell |
| GTAGCTAGTAGCGTTT-1 | Nerve cell | Neuron | Photoreceptor cell | Rod cell |
| ACGTACACACAAATAG-1 | Nerve cell | Neuroglial cell | Muller cell | |
| TGTAACGGTAGCGCTC-1 | Nerve cell | Neuroglial cell | Muller cell | |
| TGGTTAGGTCCTCAGG-2 | Nerve cell | Neuron | Ganglion cell | Retinal ganglion cell |
| GACAGCCCACTTGACA-2 | Nerve cell | Neuron | Bipolar cell | |
| GCCATTCTCTCCGAGG-2 | Nerve cell | Neuron | Photoreceptor cell | Rod cell |
| GGTGAAGAGTATGAAC-2 | Nerve cell | Neuron | Ganglion cell | Retinal ganglion cell |
| ACCTGAAGTCTTTCTA-2 | Nerve cell | Neuron | Bipolar cell | |
| GAGTTTGGTCGTTATG-1 | Nerve cell | Neuron | Photoreceptor cell | Rod cell |
| TCTACCGCATAACTCG-1 | Nerve cell | Neuroglial cell | Muller cell | |
| TTGTTCAGTGTTTACG-1 | Nerve cell | Neuron | Ganglion cell | Retinal ganglion cell |
| ATCCACCTCGCCTAGG-2 | Nerve cell | Neuron | Bipolar cell | |

**Hierarchical cell types of the cell**

# Data

- **Data overview**

| Data | | | |
|---|---|---|---|
| **Dataset** | **Sample size** | **Feature data file** | **Label file** |
| **TrainingSet** | 14346 | menon_train_data.csv | menon_train_label.csv |
| **ValidationSet** | 4782 | menon_validation_data.csv | menon_validation_label.csv |
| **TestSet** | 21979 | lukowski_test_data.csv | - |
| **Organization of hierarchical-structured classes** | | | |
| **uHAF tree** | classes_tree.csv | | |

* TestSet may have labels that were not present in TrainingSet or uHAF tree.

**Key point 1**

# Data

- **Data description**

**uHAF tree file:**

The hierarchical organization (4 levels) of cell types including all labels of samples in TrainingSet and ValidationSet.

| **Root** | **Level 1** | **Level 2** | **Level 3** | **Level 4** |
|---|---|---|---|---|
| Eye | | | | |
| | Nerve cell | | | |
| | | Neuron | | |
| | | | Photoreceptor cell | |
| | | | | Rod cell |
| | | | | Cone cell |
| | | | Bipolar cell | |
| | | | | Amacrine cell |
| | | | | Horizontal cell |
| | | | Ganglion cell | |
| | | | | Retinal ganglion cell |
| | | Neuroglial cell | | |
| | | | Muller cell | |
| | | | Microglia | |
| | Epithelial cell | | | |
| | | Endothelial cell | | |

# Data

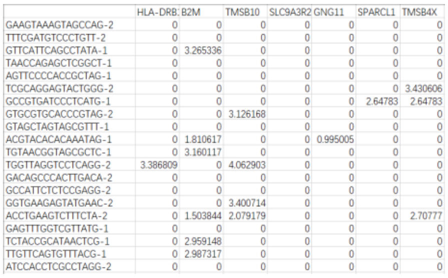
- **Data description**

### Feature data file:

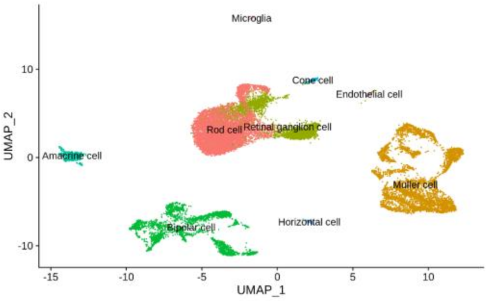
A cell (sample)-by-gene (feature) matrix, or gene expression matrix.

Gene

Cell



| | HLA-DRB: | B2M | TMSB10 | SLC9A3R2 | GNG11 | SPARCL1 | TMSB4X |
|---|---|---|---|---|---|---|---|
| GAAGTAAAGTAGCCAG-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TTTCGATGTCCCTGTT-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GTTCATTCAGCCTATA-1 | 0 | 3.265336 | 0 | 0 | 0 | 0 | 0 |
| TAACCAGAGCTCGGCT-1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AGTTCCCCACCGCTAG-1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TCGCAGGAGTACTGGG-2 | 0 | 0 | 0 | 0 | 0 | 0 | 3.430606 |
| GCCGTGATCCCTCATG-1 | 0 | 0 | 0 | 0 | 0 | 2.64783 | 2.64783 |
| GTGCGTGCACCCGTAG-2 | 0 | 0 | 3.126168 | 0 | 0 | 0 | 0 |
| GTAGCTAGTAGCGTTT-1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ACGTACACACAAATAG-1 | 0 | 1.810617 | 0 | 0 | 0.995005 | 0 | 0 |
| TGTAACGGTAGCGCTC-1 | 0 | 3.160117 | 0 | 0 | 0 | 0 | 0 |
| TGGTTAGGTCCTCAGG-2 | 3.386809 | 0 | 4.062903 | 0 | 0 | 0 | 0 |
| GACAGCCCACTTGACA-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GCCATTCTCTCCGAGG-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GGTGAAGAGTATGAAC-2 | 0 | 0 | 3.400714 | 0 | 0 | 0 | 0 |
| ACCTGAAGTCTTTCTA-2 | 0 | 1.503844 | 2.079179 | 0 | 0 | 0 | 2.70777 |
| GAGTTTGGTCGTTATG-1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TCTACCGCATAACTCG-1 | 0 | 2.959148 | 0 | 0 | 0 | 0 | 0 |
| TTGTTCAGTGTTTACG-1 | 0 | 2.987317 | 0 | 0 | 0 | 0 | 0 |
| ATCCACCTCGCCTAGG-2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |



# Data

- **Data description**

### Label file:

- Only TrainingSet and ValidationSet were provided with the label file;
- Each cell has at most one identity in a hierarchy;
- Not all cells can be assigned to a leaf-node cell type. Some can at most be assigned to an internal-node cell type.

Key point 2

| cell_id | cell_type_level1 | cell_type_level2 | cell_type_level3 | cell_type_level4 |
|---|---|---|---|---|
| GAAGTAAAGTAGCCAG-2 | Nerve cell | Neuron | Ganglion cell | Retinal ganglion cell |
| TTTCGATGTCCCTGTT-2 | Nerve cell | Neuron | Photoreceptor cell | Cone cell |
| CTTTCGGGTAGCCAGA-2 | Epithelial cell | Endothelial cell | | |
| AGTTCCCCACCGCTAG-1 | Nerve cell | Neuron | Bipolar cell | |
| CGCGTGACAAAGACGC-1 | Nerve cell | Neuron | Bipolar cell | Amacrine cell |
| GAAGGGTGTACTCCGG-1 | Nerve cell | Neuron | Bipolar cell | Horizontal cell |

# Tasks

Predict the hierarchical classes of ValidationSet and TestSet samples based on the model trained on TrainingSet. We require that:

**Key point 1**: the model should be able to decide at which level in the hierarchy the classification should stop, since not all samples can be assigned to a leaf-node class. However, the classification should be as fine as possible.

**Key point 2**: the model should consider the situation when query samples belong to classes that were not present in any TrainingSet sample, and should label them as “Unclassified” in the corresponding level in the hierarchy.

# Requirements

- Submit code and report (PDF). Submit supplementary tables containing the cell type prediction results on **ValidationSet** and **TestSet** (as “menon_train_label.csv”).
- Your final report should cover the details of your model and its performance on **ValidationSet**. Report $F_1$ scores or AUROC scores in each level in the hierarchy.
- Please make sure to give clear and detailed descriptions on your algorithms in the report, especially addressing the two key points described in “Tasks”. We will mostly judge according to your algorithm designs instead of the scores.
- Your code must be executable and contain a readme file. We may run your model and reproduce the outputs.

# Reference

- Silla, C. N. & Freitas, A. A. A survey of hierarchical classification across different application domains. *Data Min Knowl Disc* **22**, 31–72 (2011).
- Menon, M. *et al.* Single-cell transcriptomic atlas of the human retina identifies cell types associated with age-related macular degeneration. *Nat Commun* **10**, 4902 (2019).
- Lukowski, S. W. *et al.* A single-cell transcriptome atlas of the adult human retina. *EMBO J* **38**, (2019).
- Chen, S. *et al. A human ensemble cell atlas (hECA) enables* in data *cell sorting*. http://biorxiv.org/lookup/doi/10.1101/2021.07.21.453289 (2021) doi:10.1101/2021.07.21.453289.

**See you next week**

**for**

**Support Vector Machines**