Track patient recovery in real-time by processing streaming data

BIOMEDICAL DATA DESIGN

TA: Haoyin Xu

Group:Zhenyu Xiao
Haobin Zhou
Yimeng Xu
Emma Cardenas

Content

01
Proposal

02

Data Preprocessing(Heart rate)

02 Data Preprocessing(Heart rate)

1. Extract sub-categories patient id from cardiovascular

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import data_toolbox

# import diagnosis.csv
df_diagnosis = pd.read_csv('diagnosis.csv')
df_diagnosis.sort_values(by=['patientunitstayid', 'diagnosisoffset'], inplace=True)

# select cardiovascular patients
df_cardiovascular = df_diagnosis[df_diagnosis['diagnosisstring'].str.contains('cardiovascular')]
# print(df_cardiovascular)


# get shock patient
shock_patient = df_cardiovascular[df_cardiovascular['diagnosisstring'].str.contains('shock')]
# print(shock_patient)

# get ventricular patient
ventricular_patient = df_cardiovascular[df_cardiovascular['diagnosisstring'].str.contains('ventricular')]
# print(ventricular_patient)

# get chest pain patient
chest_pain_patient = df_cardiovascular[df_cardiovascular['diagnosisstring'].str.contains('chest pain')]

# get arrhythmias patient
arrhythmias_patient = df_cardiovascular[df_cardiovascular['diagnosisstring'].str.contains('arrhythmias')]

# put id together
df_wanted = pd.concat([shock_patient, ventricular_patient, chest_pain_patient, arrhythmias_patient])
# print(df_wanted)

# Get the patient ids from df_wanted & sort the patient id
# patient_id_all multiple entry patient's stayid
patient_id_all = df_wanted['patientunitstayid'].unique()
patient_id_all.sort()
print(patient_id_all)
```

```
[ 143870  151179  151900 ... 3351297 3352230 3352231]
```

2. Exclude patient whose unitvisitnumbe>1

```
# import patient.csv
df_patient = pd.read_csv('patient.csv')
df_patient.sort_values(by=['patientunitstayid'], inplace=True)
df_patient_buf = df_patient[df_patient['patientunitstayid'].isin(patient_id_all)]
df_1time_patient = df_patient_buf[df_patient_buf['unitvisitnumber']==1]
# print(df_1time_patient)

# select the patient id from df_1time_patient
patient_id = df_1time_patient['patientunitstayid'].unique()
print(f'Total number of patients: {len(patient_id)}')
```

```
Total number of patients: 915
```

915 Patients valid

02 Data Preprocessing(Heart rate)

```python
# define heartrate preprocessing function
def normal_heartrate(num):
    """
    Function to normalize heart rate values.

    Parameters:
        num: the originial input value
    Return:
        num: the normalized output value
    """
    # Return null values direcly
    if pd.isna(num):
        return num
    # Remove values out of range
    elif num > 300 or num < 0:
        return np.nan
    # Return normal values directly
    else:
        return num
```

Filter the abnormal ones

```python
# extract heart rate from df_vitalPeriodic
HR = df_vitalPeriodic[['patientunitstayid', 'observationoffset', 'heartrate']]
print(f'First 5 rows of HR: \n{HR.head()}')

# exclude abnormal heart rate values
HR.loc[:, 'heartrate'] = HR['heartrate'].apply(normal_heartrate)

# save HR to csv file (uncomment the code to save)
# HR.to_csv('HR.csv', index=False)

value_position_dict = {}
first_occurrences = []
for idx, value in enumerate(HR['patientunitstayid']):
    # if the value is not in the dictionary, add it and create index
    if value not in value_position_dict:
        value_position_dict[value] = idx
        first_occurrences.append(idx)

first_occurrences.append(len(HR))
# create first occurrence index for every patient
HR_index = pd.Series(first_occurrences)
print(f'First 5 rows of HR_index: \n{HR_index.head()}')

# double check the index is correct
# print(HR.iloc[HR_index].head())
# print(HR.iloc[[156, 157, 158, 159]])
# print(HR.iloc[[1015, 1016, 1017, 1018]])
```

```
First 5 rows of HR:
     patientunitstayid  observationoffset  heartrate
628             143870                  7       44.0
574             143870                 12       42.0
543             143870                 17       41.0
580             143870                 22       41.0
519             143870                 27       41.0
First 5 rows of HR_index:
0       0
1     158
2    1017
3    1708
4    2501
dtype: int64
```

Sorted by a 2×1 vector

Example: how to use HR & HR_index

```python
# if we want the i th patient's data (i starts from 0)
# use HR.iloc[HR_index[i]:HR_index[i+1]]
i = 0
print(f'HeartRate data for patient {i+1}: \n{HR.iloc[HR_index[i]:HR_index[i+1]]}')
```

```
HeartRate data for patient 1:
     patientunitstayid  observationoffset  heartrate
628             143870                  7       44.0
574             143870                 12       42.0
543             143870                 17       41.0
580             143870                 22       41.0
519             143870                 27       41.0
..                 ...                ...        ...
614             143870                772       50.0
584             143870                777       51.0
578             143870                782       48.0
572             143870                787       48.0
566             143870                792       49.0

[158 rows x 3 columns]
```

Index to search certain patients

Thank you

01 Goal & Literature Review

Treatment Method

- Beside making favorable lifestyle modifications, primary regimes for the prevention and treatment of CVDs include **lipid-lowering drugs**, **antihypertensives**, **antiplatelet** and **anticoagulation therapies**.

- **Interventional treatment** is the minimally invasive diagnosis and treatment of diseases under the guidance of medical imaging equipment (angiography, fluoroscopy, CT, MR, B ultrasound, etc.), percutaneous puncture, introduction of puncture needles, special catheters, guide wires and other precision instruments into the body's blood vessels.

- Cardiac Procedures and Surgeries:

Coronary Artery Bypass Grafting (CABG): Used to treat coronary artery disease by bypassing narrowed arteries with new blood vessels.

Valve Repair or Replacement: Repair or replace heart valves, such as mitral valve repair, aortic valve replacement, etc.

Cardiac Pacemaker or Defibrillator Implantation: Used to treat arrhythmias and regulate the heart's rhythm.

01

Goal & Literature Review

Risk Factors

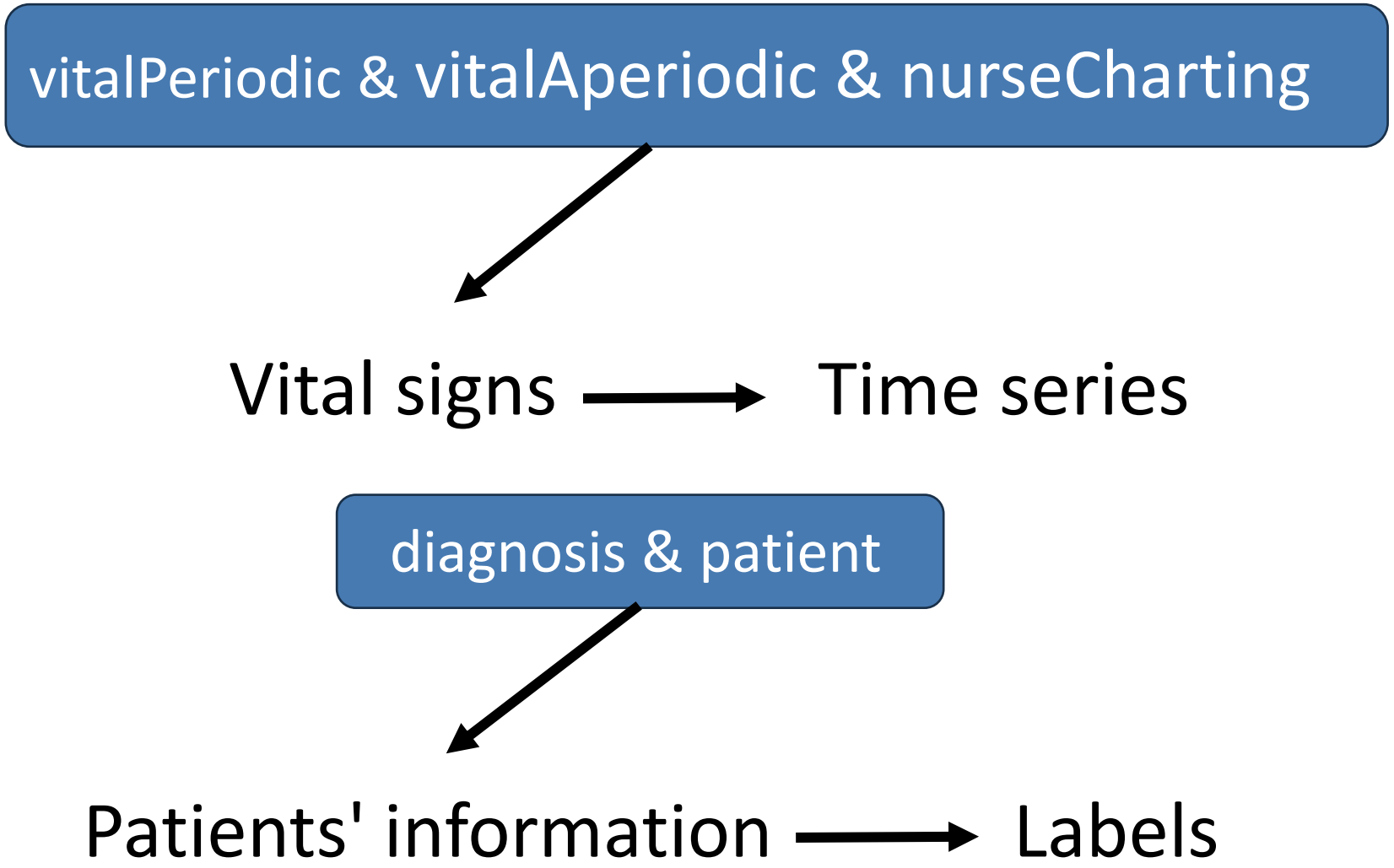
hypertension, hyperlipidemia, and diabetes

smoking, physical inactivity, alcohol abuse, unhealthy diet, obesity

genetic predisposition and family history of cardiovascular disease

high-sensitivity C-reactive protein (hs-CRP), ankle brachial pressure index, lipoprotein subclasses and particle concentration, lipoprotein (a), apolipoproteins AI and B, fibrinogen, leukocyte count, homocysteine, N-terminal pro-B-type natriuretic natriuretic peptide (NT-proBNP), and renal function markers. High blood phosphorus was also associated with risk factor. （as we mentioned in our last PPT）

03 Dataset Analysis



03 Dataset Analysis

diagnosisid	patientunitstayid	activeupondischarge	diagnosisoffset	diagnosisstring
4035907	143870	TRUE	10	cardiovascular|chest pain / ASHD|coronary artery disease
3843251	143870	TRUE	10	cardiovascular|post vascular surgery|s/p cartoid endarterectomy
3460672	143870	TRUE	10	cardiovascular|arrhythmias|bradycardia
3717065	151179	FALSE	29	cardiovascular|shock / hypotension|septic shock
4102418	151179	FALSE	120	cardiovascular|shock / hypotension|septic shock
3885168	151179	TRUE	3929	cardiovascular|shock / hypotension|septic shock
4053934	151179	TRUE	3929	cardiovascular|shock / hypotension|hypotension
3850876	151900	FALSE	148	cardiovascular|shock / hypotension|septic shock
3707280	151900	FALSE	939	cardiovascular|shock / hypotension|septic shock
4192192	151900	FALSE	939	cardiovascular|chest pain / ASHD|acute coronary syndrome
3379776	151900	TRUE	2895	cardiovascular|chest pain / ASHD|acute coronary syndrome
3892141	151900	TRUE	2895	cardiovascular|shock / hypotension|septic shock
3678632	152954	FALSE	39	cardiovascular|shock / hypotension|signs and symptoms of sepsis (SIRS)
3977729	152954	FALSE	39	cardiovascular|ventricular disorders|congestive heart failure
4144394	152954	FALSE	219	cardiovascular|shock / hypotension|signs and symptoms of sepsis (SIRS)
3757248	152954	FALSE	219	cardiovascular|ventricular disorders|congestive heart failure

04

Next Step

04 Next Step

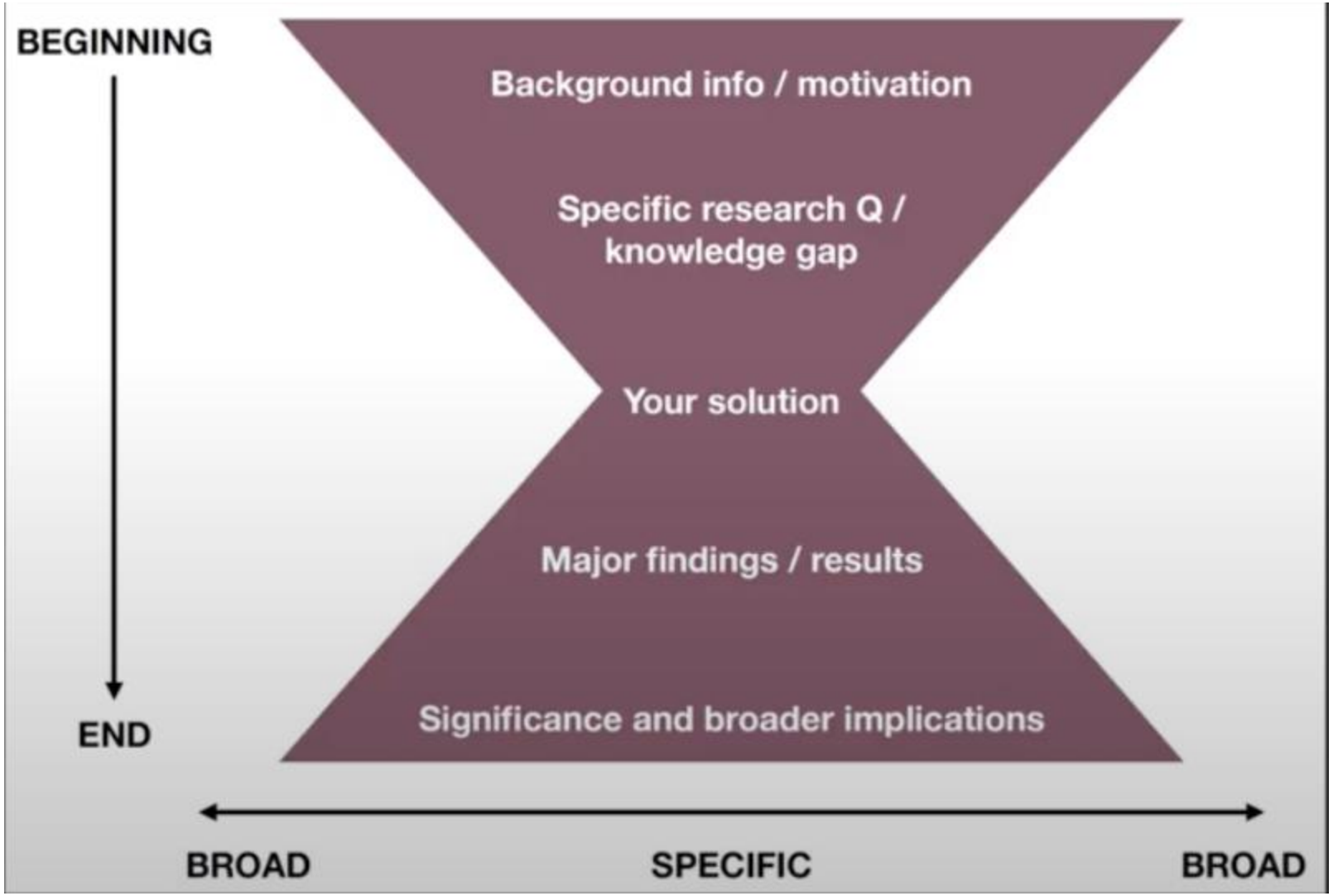
1.Extract more meaningful data
1.1 Extraction
1.2 Interpolation, Correction

2.Replicate the deep learning model as baseline and try new models.

References

Flora G D, Nayak M K. A brief review of cardiovascular diseases, associated risk factors and current treatment regimes[J]. Current pharmaceutical design, 2019, 25(38): 4063-4084.

Content


BEGINNING
Background info / motivation
Specific research Q / knowledge gap
Your solution
Major findings / results
Significance and broader implications
END
BROAD
SPECIFIC
BROAD

01

Cardiovascular Diseases

1.Overview of cardiovascular disease

2.Why is it important to predict symptoms, how can it help doctors?

3.Why it needs to be real-time?

4.Why we want to make a prediction of best treatment?

01 Goal & Literature Review

Goal(now)

- Death rate
- Risk of cardiovascular disease
- Date of discharge from hospital

Goal(in the future)

- Predict symptoms that will develop
- How to give treatments