MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

@Sage_ZJ

Temporal Phenotyping for End-stage Renal Disease Using Longitudinal Electronic Health Records

Shengqiang CHI

Associate researcher

Zhejiang Lab

AIDH IMIA

Introduction

- Disease Subtyping
  - Use electronic health record (EHR) data to distinguish different subtypes of patients, uncover disease characteristics, and understand the underlying disease pathology
  - Traditional methods: clustering analysis of initial static phenotypic data of patients
    - Fail to fully utilize the clinical process information in long-term longitudinal EHR data
  - Existing studies on the subtyping of end-stage renal disease (ESRD)
    - Use the blood pressure fluctuation patterns during dialysis
    - Use immune expression data

Introduction

- Process Mining
  - Extract clinical events: including diagnoses, laboratory test results, microbiology test results, procedures and medications
  - Create a clinical event log: sort these clinical events in chronological order
  - Mine patterns: use the clinical event log to mine different patterns in the sequence of clinical events
  - Subtype classification: derive the subtype classification of patient disease phenotypes from these patterns
  - Limitations
    - Fail to handle the event association and sequential relationship across multiple visits
    - The resulting clinical process is complex and lacks interpretability

Methods

- Data
  - Data source: The First Affiliated Hospital of Zhejiang University
  - Cohort: Patients with ESRD who received hemodialysis treatment between 2000 and 2016, with a minimum follow-up period of 90 days
  - Observation window
    - Start: the initial hemodialysis
    - End: endpoint events(kidney transplantation and death) OR the last follow-up or visit
  - Clinical events: diagnoses, abnormal laboratory tests, procedures and medications

Methods

- Topic Model
  - Preprocessed the clinical process data by treating the clinical events in a single visit as documents and labeling visits
  - Latent Dirichlet allocation algorithm: obtained topic probability distributions for these documents. The topic with the highest probability was assigned as the topic of each visit.
  - Determine the optimal number of topics (K)
    - Consistency
    - Redundancy
    - Importance
    - Perplexity

Methods

- Clinical process mining
  - Algorithm
    - Each visit was represented as an event log
    - Consecutive visits with the same event names were merged
    - Calculated the support of events
      - support definition: the number of patients who experienced a specific sequence of events divided by the total number of patients: $S = N_S/N$
    - Added events with a support greater than a threshold to the clinical process until no more frequent events could be extracted or the maximum length was reached

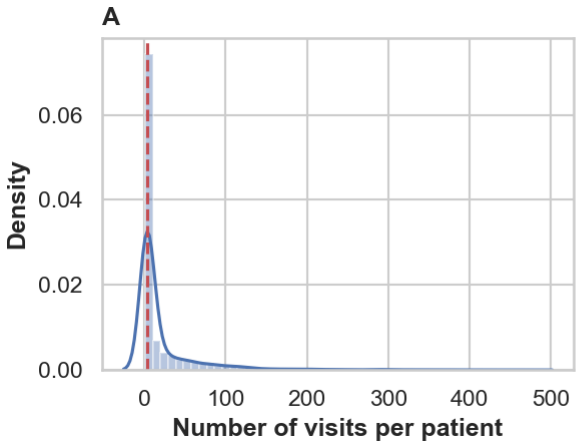
Results

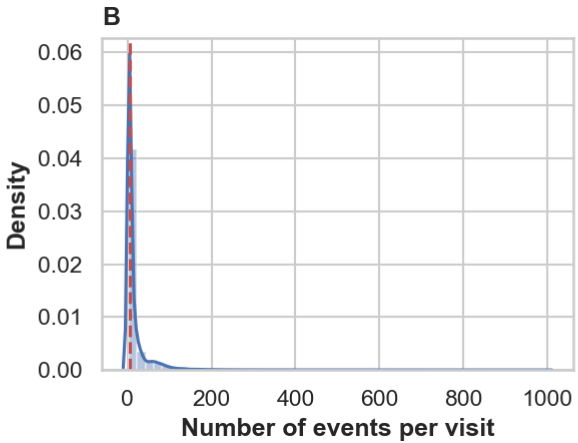
- Data preprocessing
  - Combined medications whose have similar treatment efficacy
  - Excluded visits with only one clinical event
  - Excluded patients with less than 3 visits

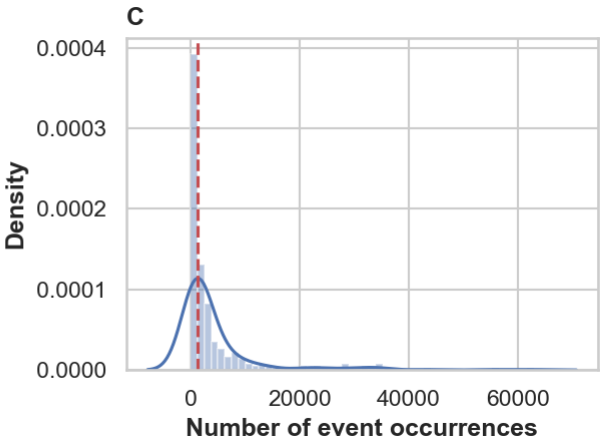
- Data statistics
  - 73,090 visits from 2,010 patients
  - 67 clinical events: 15 diagnoses
    21 abnormal laboratory tests
    31 medications

Results

- Data statistics





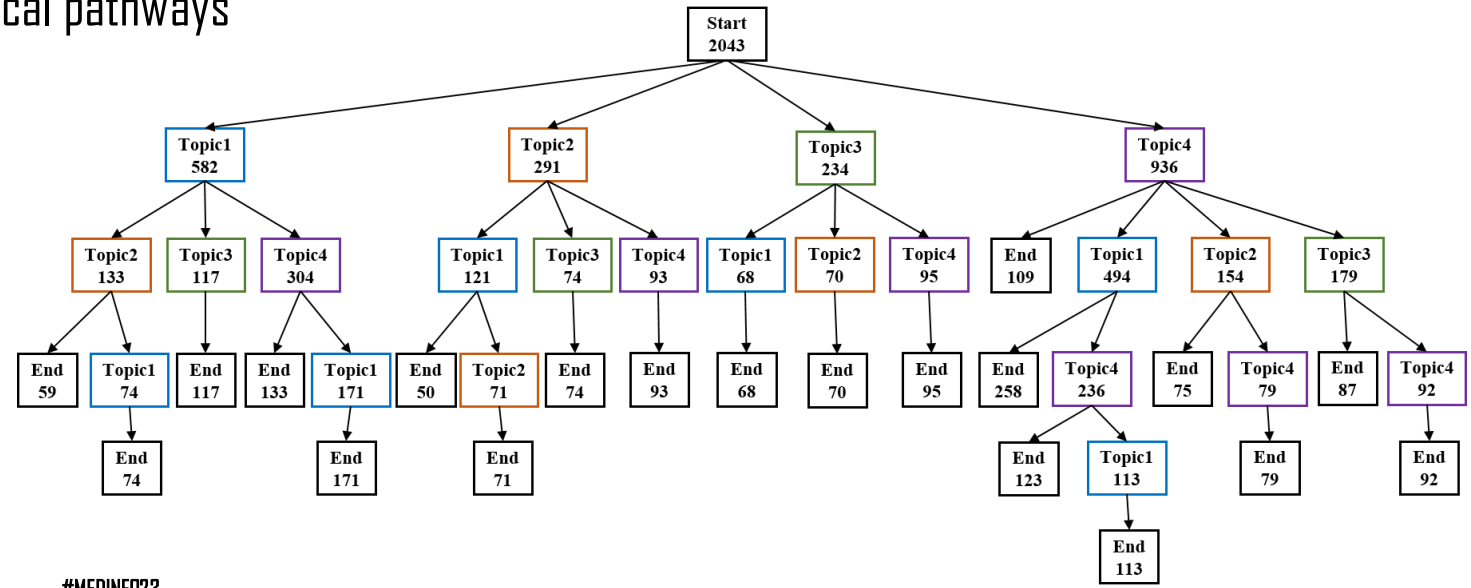


Results

Topic	Topic keywords and their probabilities
1	***Multiple complications and their treatments (17 keywords)*** ***Keywords:*** Hypertension (0.095), anti-peptic ulcer (0.086), antimicrobials (0.077), anticoagulants (0.053), anemia (0.050), amino acids (0.045), antacidosis (0.038), sodium channel blockers (0.036), anti-platelet drugs (0.036), hyperparathyroidism (0.035), calcium supplements (0.033), mucopolysaccharide polysulfate cream (0.033), sedative-hypnotics (0.030), laxatives (0.030), anti-anginal (0.025), nitrates (0.024), diuretics (0.024)
2	***Hemodialysis treatment, abnormal bone metabolism, anemia and their treatments (7 keywords)*** ***Keywords:*** Abnormal bone mineral metabolism (0.180), calcium supplements (0.172), anti-anemia drugs (0.113), levocarnitine (0.109), anti-platelet drugs (0.064), diabetes (0.06), amino acids (0.052)
3	***Hemodialysis treatment, anemia, cardiovascular diseases and their treatments (8 keywords)*** ***Keywords:*** anti-anaemic agents (0.148), anti-anginal (0.147), levocarnitine (0.107), calcium channel blockers (0.100), anti-abnormal bone mineral metabolism (0.086), vitamins (0.063), beta-blockers (0.061), angiotensin II receptor blockers (0.049)
4	***Abnormal laboratory test results (11 keywords)*** ***Keywords:*** routine blood test (0.120), routine kidney function test (0.110), routine urine test (0.102), routine potassium, sodium, chloride, calcium, magnesium and phosphorus measurement (0.098), routine blood lipid test (0.072), routine liver function test (0.066), bone metabolism marker measurement (0.053), glucose measurement (0.048), routine myocardial enzyme profile test (0.038), routine coagulation test (0.035), ultrasensitive C-reactive protein measurement (0.026)

Results

- Clinical pathways



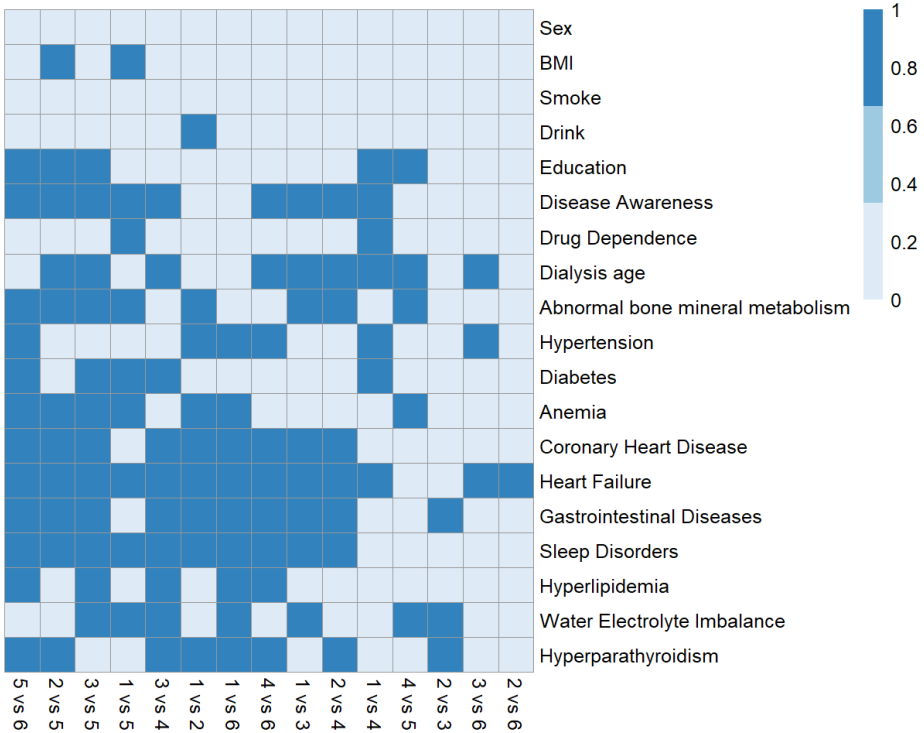
Results

- Subtypes
  - Fine-grained subtyping

Subtype	Treatment process	Subtype description
Subtype 1	Topic 2 alternates with Topic 4	Abnormal bone metabolism and its treatments, ALTRs
Subtype 2	Topic 2 alternates with Topic 3	Abnormal bone metabolism and its treatments, cardiovascular disease and its treatments
Subtype 3	Topic 1 alternates with Topic 2	Abnormal bone metabolism and its treatments, diagnosis and treatment of multiple complications
Subtype 4	Topic 3 alternates with Topic 4	Cardiovascular diseases and their treatments, ALTRs
Subtype 5	Topic 1 alternates with Topic 4, Topic 4	ALTRs, and/or diagnosis and treatments of multiple complications
Subtype 6	Topic 1 alternates with Topic 3	Cardiovascular diseases and their treatments, diagnosis and treatment of multiple complications

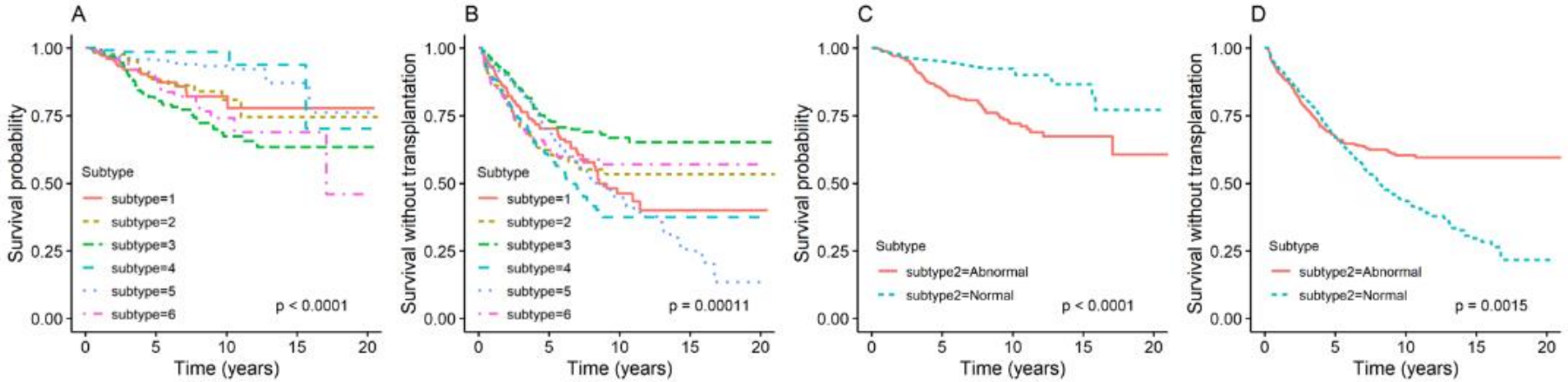
Results

- Subtypes
  - Coarse-grained subtyping
    - Normal group (2,3,6) vs Abnormal group (1,4,5)
    - The abnormal group had abnormal laboratory test results (ALTRs) earlier than the normal group
  - Both subtyping methods demonstrated statistically significant differences
    - demographics, lifestyles, disease characteristics and comorbidities



Results

- Survival Analysis



Conclusions

- Utilized topic model and process mining techniques for ESRD patients subtyping based on longitudinal EHR data.
- Captured differences in disease progression and clinical outcomes with interpretable and clinically relevant subtype classifications.
- Highlights:
  - Comprehensive Disease Understanding
    - Mining disease subtypes from process data provides a dynamic assessment of disease progression.
    - Reveals unique characteristics and progression of each patient subtype.
  - Data-Driven Clinical Event Analysis
    - Topic modeling summarizes clinical event information into visit labels.
    - Offers a streamlined and comprehensive analysis of complex clinical processes.

Thank you!