Clustering Health-care Institutions Based on the Dynamic Time Warping Similarity of Frequent Medical Orders

Kyoko Sawamura[1)], Miwa Sugitani[1)], Ryosuke Matsuo[1)], Tomoyoshi Yamazaki[1)]

Kenji Araki[1)], Masato Oguchi[1)], Haruo Yokota[2)], Le Hieu Hanh[1)]

1) Ochanomizu University, 2) Josai University

Table of Contents

Introduction

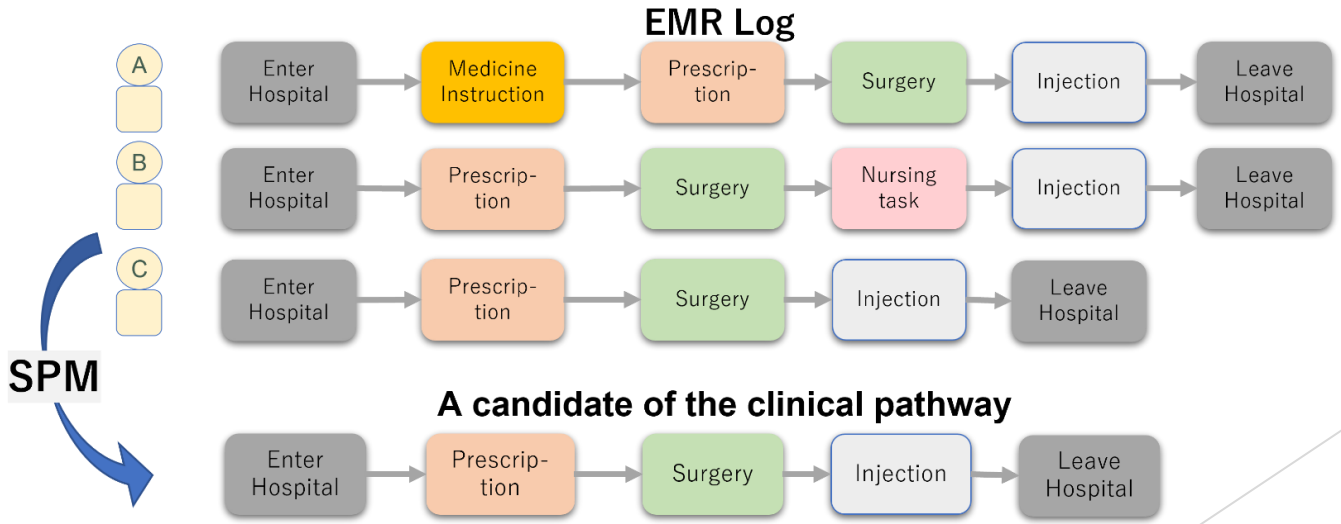
Background

- In recent years, electronic medical record data analysis by medical institutions has been advancing to improve the efficiency and safety of medical procedures.
- To make it easier to understand comparisons of medical procedures for specific diseases, clustering analyses of multiple medical institutions are also conducted.
- However, currently, clustering evaluation is insufficient, and features are calculated for one disease per hospital.

Related Research

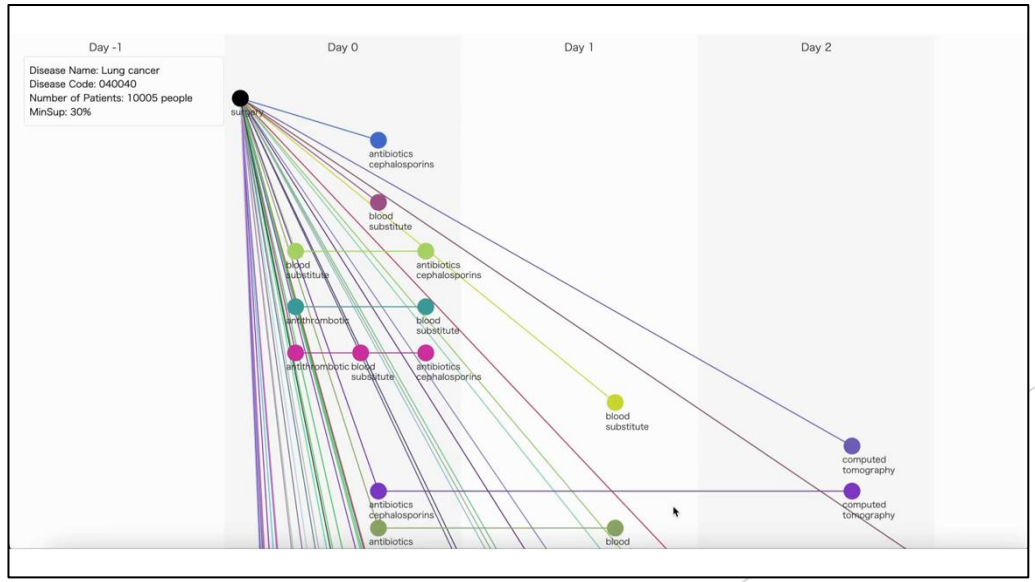
Electronic Medical Record (EMR) Data

- Contains information such as diseases, medical order, and patient information.
- Includes temporal information (e.g., dates and times).
- Enables tracking of the sequence and timing of medical procedures, treatments, and other healthcare events conducted for a patient.



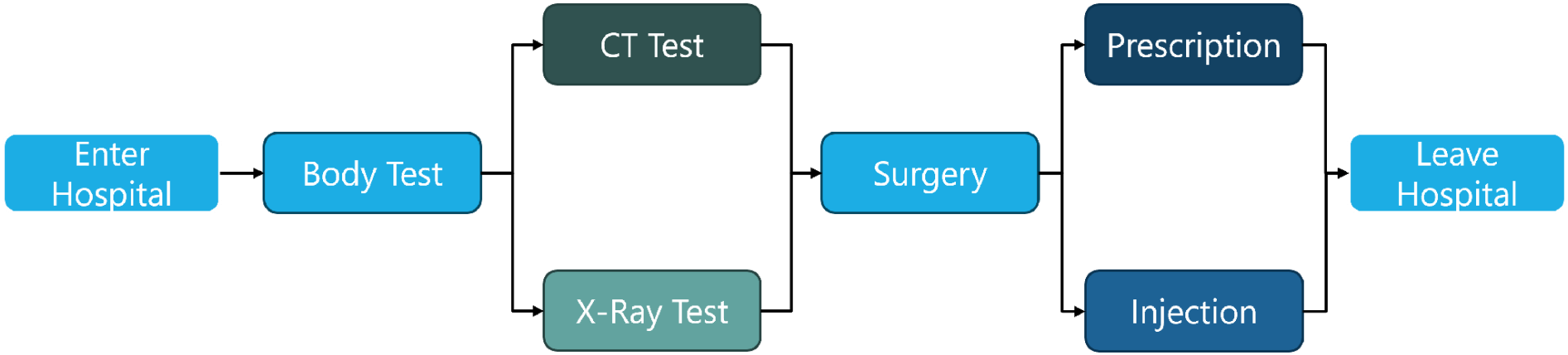
Extraction of Frequently Occurring Medical Instruction Patterns Using Time Interval SPM

- Time interval SPM retains temporal elements in addition to the order of medical orders, enabling analysis of the timings of the clinical process and extracting frequently occurring medical instruction patterns with time-dependent characteristics.
- Medical order sequences for each disease are extracted from electronic medical record data, and frequent closed patterns are extracted as frequently occurring medical order sequences considering time intervals based on date information [Sugitani et al., CBMS2025].



Sequence Variant (SV)

- Frequently occurring medical instruction patterns are extracted as sequencing variants (SV) [Honda et al., DMAH2017].
  - Frequently Extracted Medical Instruction Patterns from SPM Represent SVs considering branching.

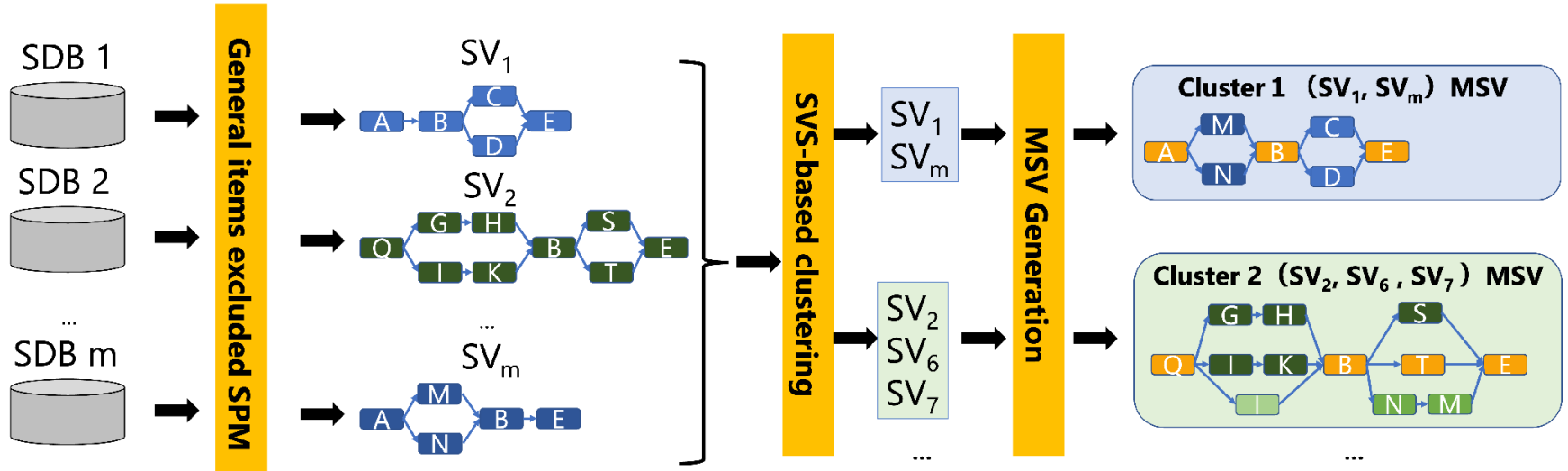


Comparison of Frequent Medical Order Patterns among Three or More Medical Institutions [Le et al., BDH2024]

- Clustering can be performed using representative SVs from each healthcare institution to identify and analyze the characteristics of the institutions.
- Clustering is essential to support improvements in medical practices by visualizing the commonalities and differences in frequent medical order patterns.
  - Calculate the distance between medical institutions using their representative frequent medical order patterns.
  - Merged SV is extracted from all frequently occurring medical orders from medical institutions belong in to the same group.

Comparison of Frequent Medical Order Patterns among Three or More Medical Institutions [Le et al., BDH2024]

- A method to efficiently compare SVs from more than 3 medical institutions.



Comparison of Frequent Medical Order Patterns among Three or More Medical Institutions [Le et al., IEEE2024]

[Issues]

- Insufficient clustering accuracy due to inaccurate representation of medical institution characteristics.

- Clustering results are not sufficiently evaluated, making the accuracy of clustering unclear.

Research Objectives

Objectives of This Study and Approaches to Achieve It

[Purpose]

- Solving two issues of clustering accuracy and insufficient evaluation in existing methods [Le et al., MIPR2024].

[Approaches]

- Calculate the distance between all frequently occurring medical instruction patterns and between medical institutions using the dynamic time warping.
- Evaluating clustering results.

Proposal Methods

Flow of the Proposal Method

① Extraction Method for Frequent SVs

Sequential pattern mining is applied from the electronic medical records of medical institutions to extract frequent medical orders from each institution.

② Distance Calculation Using Dynamic Time Warping

To account for the temporal order of medical procedures, dynamic time warping was used to calculate distances between frequent medical orders and healthcare institutions.

③ Hierarchical Clustering

Based on the distances between medical institutions calculated in the previous step, we cluster medical institutions using hierarchical clustering.

④ Clustering Evaluation Method

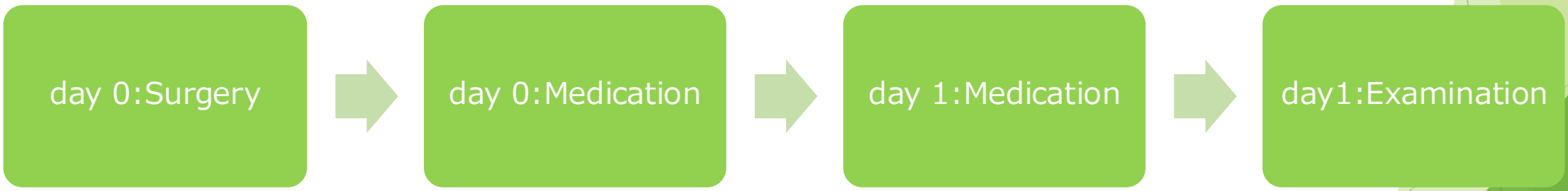
Evaluating clustering results using Silhouette scores and Dunn index.

①Extraction Method for Frequent SVs

- Extract frequently occurring medical instruction patterns by medical institution [Sugitani et al., CBMS2025].
- Since the pattern group extracted differs for each minimum support minsup, it is necessary to determine the patterns to target clustering.
- ✓ We tried multiple minsups to extract the pattern group closest to actual treatment to determine the appropriate minsup.

① Extraction Method for Frequent SVs
Appropriate minsup Decisions

- Patterns within the extracted pattern group were compared with the core clinical path, and recall, precision, and F-measure were calculated.
  - The core clinical pass is the minimum medical planning required.
  - The minsup with the highest average F-measure is considered appropriate minsup.



ex: Frequent Medical Order

②Distance Calculation Using DTW

Two Frequent Medical Order Patterns

- Adopts dynamic time warping (DTW) **because it considers time series and allows calculation independent on number differences of elements in each medical instruction column.**
- Suppose A = (day 0: surgery, day 1: medication, day 2: examination)
  B = (day 0: surgery, day 1: examination, day 2: medication).
- To calculate DTW distance, set the match to "1" and the mismatch to "0".

B/A	day 0: Surgery	day 1: Medication	day 2: Examination
day 0: Surgery	0	1	2
day 1: Examination	1	1	2
day 2: Medication	2	2	2

d(A, B) = 2

②Distance Calculation Using DTW

Two Institutions

※FMOP: Frequent Medical Order Patterns

- Two pairs were formed at two medical institutions, and the distance was calculated for each pair in the medical instruction line.
- The average distance between all pairs is taken as the distance between medical institutions.

Medical Institution X
・FMOP X1 ・FMOP X2

Medical Institution Y
・FMOP Y1 ・FMOP Y2 ・FMOP Y3

↓

Calculate the DTW distance "d(a, b)" for each pair
d(X1, Y1)、d(X1, Y2)、d(X1, Y3)
d(X2, Y1)、d(X2, Y2)、 d(X2, Y3)

↓

Calculate the distance between medical institutions X and Y as the average of the six values above

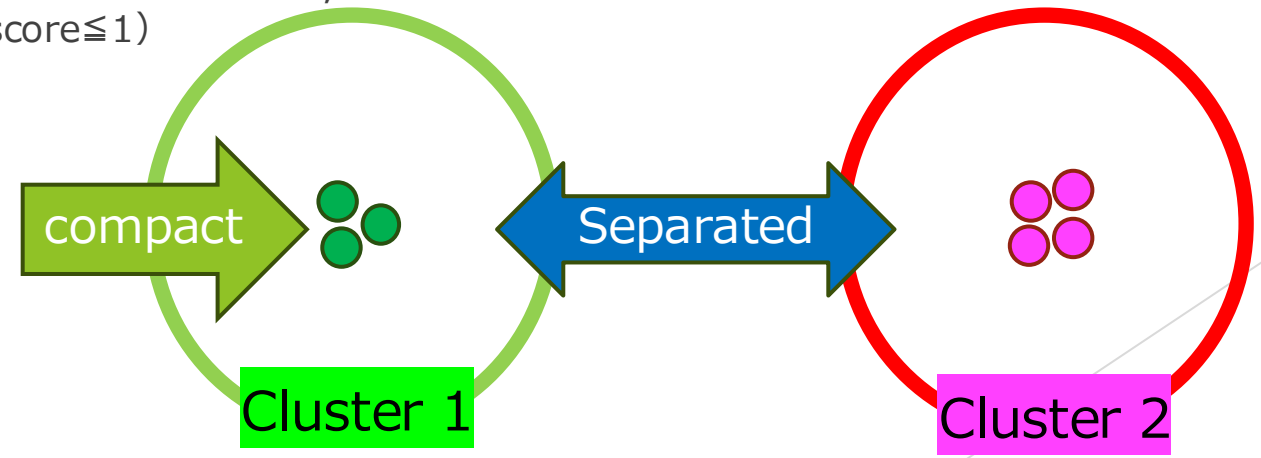
③Hierarchical Clustering

- To compare with existing methods, hierarchical clustering is used, similar to existing method.

- Other clustering methods using distance can also be applied.

④Clustering Evaluation Method

- Internal evaluation metrics are adopted because it is currently difficult to evaluate clustering results using ground truth clusters defined in real-world clinical settings.
- Specifically, this study uses the Silhouette score and the Dunn index.
  - A larger **Dunn index** indicates greater separation between clusters relative to cluster cohesion, indicating desirable clustering results. (0≦Dunn index)
  - The closer the **Silhouette score** is to 1, the more compact it is and sufficiently separated from other clusters, which can be considered a desirable clustering result. (-1≦Silhoutte score≦1)



Experiments

Experimental Subjects and Evaluation Indicators

- Compare existing method [Le et al., MIPR2024] and proposed method to verify improvements and stability in the accuracy.
- Clustering adopts and compares five methods:
  Ward's method, single linkage, complete linkage, average linkage, and centroid linkage.
- Evaluated by Dunn index and Silhouette score.

Experimental Environment and Datasets

- Using anonymized processed electronic medical record datasets collected from 27 medical institutions between 2015 and 2024.

※This study was reviewed and approved by the Review Committee for Research Use of Data of the Life Data Initiative and was conducted in accordance with the approved guidelines.

Number of records	53,699
Number of diseases	8

Clustering Rules

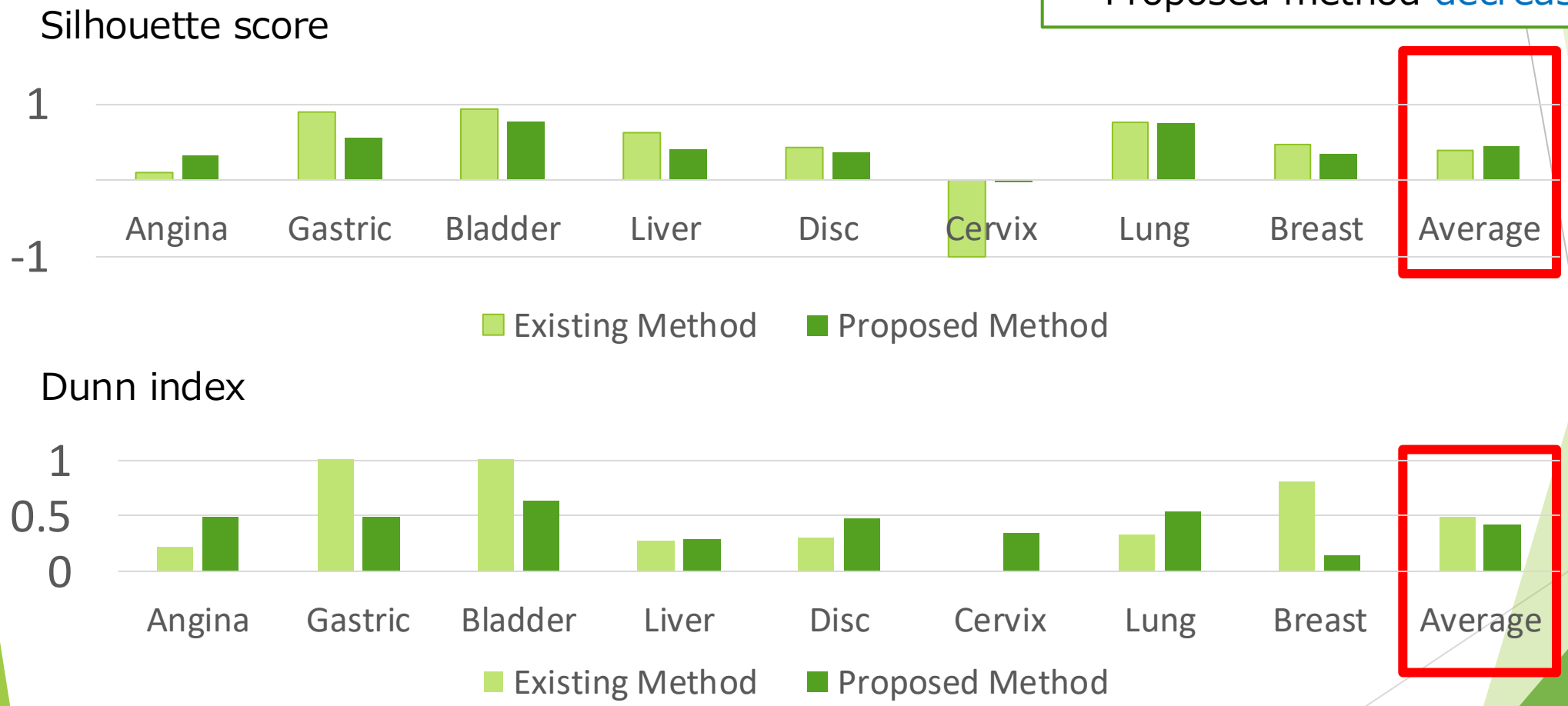
- Avoid generating excessively small clusters.

The clustering conditions are summarized as follows:

1. The minimum cluster size is 3.
2. The Dunn index is not infinite.
3. If these conditions are not satisfied, the Dunn index is set to 0 and the Silhouette score is set to -1.

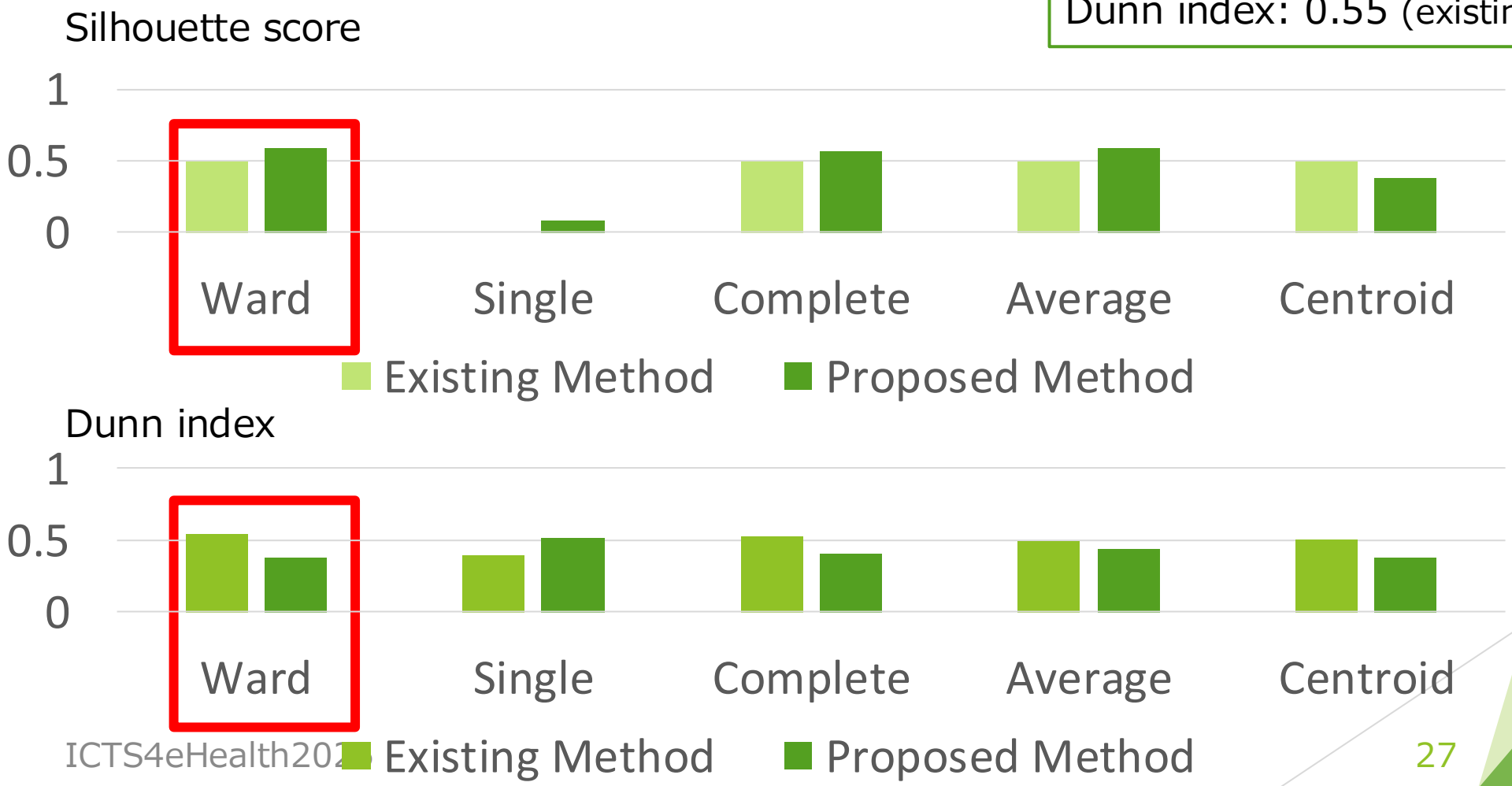
Evaluation Indicator Results By Disease

- Average Silhouette score:
  Proposed method improved by 0.039.
- Average Dunn index:
  Proposed method decreased by 0.066.



Evaluation Indicator Results By Method

From evaluation results by method, the Ward's method achieved the highest indicator results
Silhouette score: 0.59 (proposed method)
Dunn index: 0.55 (existing method)



Evaluation of Clustering Stability

	Proposed Method	Existing Method
Standard Deviation of the Silhouette Score (by Disease)	0.234	0.588
Standard Deviation of the Silhouette Score (by Method)	0.198	0.200
Standard Deviation of the Dunn Index (by Disease)	0.147	0.360
Standard Deviation of the Dunn Index (by Method)	0.048	0.050
Cases That Did Not Satisfy the Clustering Conditions	2	7

- For all outcomes, the proposed method had a lower standard deviation.
- The number of cases that did not meet clustering rules was lower than the proposed method.

➢ The proposed method achieved **more stable clustering by disease and method than existing method.**

Summary and Future Work

Summary and Future Work

[Summary]

- Existing method represented each medical institution with a single pattern of frequent medical order patterns, which posed challenges in not fully reflecting the characteristics of the medical institutions. Additionally, there was an issue of insufficient evaluation of clustering accuracy.
- In the proposed method, the distance between all frequently occurring SVs extracted by time interval SPM is calculated using DTW, and the average is used as the distance between medical institutions, enabling hierarchical clustering.
- Evaluation experiments using real data showed that the proposed method showed more stable clustering performance than existing method.

[Future Work]

- We plan to perform clustering using available information.
  - [example] Grouping by number of hospital beds and medical facility.

Thank you for listening!