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

Kyoko Sawamura*, Miwa Sugitani*, Ryosuke Matsuo*, Tomoyoshi Yamazaki*,
Kenji Araki*, Masato Oguchi*, Haruo Yokota†, Hieu Hanh Le*
* Ochanomizu University, Tokyo, Japan
† Josai University, Tokyo, Japan
E-mail: g2120518@is.ocha.ac.jp

***Abstract*—Healthcare institutions analyze electronic medical records (EMRs) to identify treatment variations and optimize practices. While comparing data across multiple institutions reveals diverse medical treatments, existing clustering methods often fail to capture this complexity because they rely on a single representative pattern. Furthermore, the stability of these clustering results is rarely evaluated. To address these gaps, this study proposes a clustering method that reflects institutional characteristics by considering all frequent medical order patterns. We apply sequential pattern mining to EMRs and utilize dynamic time warping (DTW) to calculate inter-institutional distances, accounting for the temporal structure of medical sequences. These distances then inform a hierarchical clustering process. Experimental evaluations using real-world EMR data demonstrate the effectiveness of the proposed method and the improved stability of its results. By capturing the diverse characteristics of medical orders, this approach facilitates a more accurate and reliable comparison of institutional treatment practices.**

***Index Terms*—Clustering, Clinical Pathway, Electronic Medical Records**

I. Introduction

In recent years, individual health-care institutions have actively analyzed electronic medical records (EMRs) to support best practices. By analyzing EMRs, it becomes possible to construct typical time series for medical orders frequently administered to patients with specific diseases [1], evaluate the commonalities and differences between treatment processes [2], and estimate the underlying factors causing variations in medical procedures [3]. Furthermore, following the increasing availability of EMRs from multiple health-care institutions, comparing patterns in the frequent medical orders extracted from different institutions enables a more intuitive understanding of the institutional differences in medical treatments for specific diseases. However, as the number of target institutions increases, common treatments tend to differ, which makes it necessary to efficiently cluster multiple health-care institutions so that similar institutions can be compared.

To compare multiple health-care institutions [4], this study applies sequential pattern mining (SPM) to the EMRs from multiple health-care institutions to extract sequence variants (SVs), which represent the frequent medical orders in each institution. Merged sequence variants (MSVs) were then generated from these SVs and clustering was performed on the resulting MSVs to compare the patterns in frequent medical orders across institutions.

However, this method calculates distances between institutions using only a single representative pattern from the frequent medical orders for each institution. As a result, this method does not sufficiently capture the diversity of medical order characteristics in each institution. Consequently, it is difficult to accurately measure the similarity of medical orders between institutions, which may lead to suboptimal clustering performance. In addition, the clustering results have not been evaluated adequately, especially stability.

To address these issues, this study proposes a clustering method that more accurately reflects institutional characteristics by considering the patterns in frequent medical orders extracted from each health-care institution's EMRs and appropriately computing the interinstitutional distances. Specifically, SPM is applied to the EMRs from each institution to extract patterns from frequent medical orders. Then, to account for the temporal structure of medical order sequences, the distances between patterns and between institutions are calculated based on dynamic time warping (DTW). Finally, hierarchical clustering is performed using the computed interinstitutional distances and the results evaluated. In the experimental evaluation, real-world data for medical orders are used to verify the effectiveness of the proposed distance computation method and the stability of the clustering results.

The contributions of this study are summarized as follows:

- We proposed a method for calculating interinstitutional distances using DTW to cluster health-care institutions.
- Using real-world data from 27 health-care institutions, our method demonstrated improved stability compared with existing clustering methods using quantitative evaluation based on the silhouette score and the Dunn index.

The remainder of this paper is organized as follows. Section II summarizes the literature. Section III presents the proposed clustering algorithm and evaluation methods. Section IV reports the clustering results obtained by applying

the proposed method to real-world data from multiple health-care institutions. Finally, Section V concludes the paper and discusses future work.

II. Background and Related Work

A. Clinical Pathways

According to the definition of the Japanese Society for Clinical Pathways [5], a clinical pathway is a standardized care plan that includes patient conditions, medical treatment goals, and evaluation and recording processes. Clinical pathways also provide a method for improving the quality of health care because deviations from care plans can be analyzed. In addition, electronic clinical pathways are a health-care management approach that uses information and communication technologies to create standardized care plans, supports the delivery of care based on these plans, records patient-specific treatment processes and evaluations, and facilitates the aggregation and analysis of individual cases that deviate from normal treatment plans.

B. Extraction Method for Identifying Patterns in Frequent Medical Orders

T-PrefixSpan [1] extracts medical order sequences for each disease from institutions' EMRs and identifies frequent closed patterns in medical order sequences in addition to considering time intervals based on date information. In this study, we adopt the method proposed by Sugitani et al. [6] which is inspired by T-PrefixSpan to extract patterns from the frequent medical orders for each disease. This method uses the following definitions.

Definition 1 (Reordering Rule for Medical Orders on the Same Day)**.** *Assuming that the execution times of medical orders are not available, when multiple medical orders are recorded on the same day, they are reordered according to the following priority:*

$$\begin{bmatrix} 1. & \textit{Surgery} \\ 2. & \textit{Medication} \\ 3. & \textit{Examination} \\ 4. & \textit{Clinical Procedure} \end{bmatrix}$$

Furthermore, when multiple medical orders exist within the same category, they are reordered lexicographically based on a predefined dictionary, which standardizes the medical order sequence and is expected to improve the accuracy of extracting patterns from frequent medical orders.

Definition 2 (Medical Order Sequence)**.** *The medical order sequence for a patient p_i is defined as follows:*

$$S_{p_i} = \langle s_1, s_2, \ldots, s_n \rangle$$

where s_j represents the j-th medical order and is defined as $s_j = (t_j, a_j)$. Here, t_j denotes the elapsed number of days until the medical order is performed and a_j represents the type of medical order (e.g., surgery, medication, examination).

Definition 3 (Medical Sequence Database)**.** *A medical sequence database (MSDB) consists of medical order sequences from multiple patients. The dataset D in the MSDB is defined as follows:*

$$D = \{(SID_1, S_{p_1}), (SID_2, S_{p_2}), \ldots, (SID_m, S_{p_m})\}$$

Here, SID_i denotes the identifier of the i-th patient, and S_{p_i} represents the medical order sequence for that patient.

Definition 4 (Frequent Sequences)**.** *For a health-care institution, H, the set of frequent sequences considering time intervals, denoted as $fs(H)$, is extracted from the MSDB using T-PrefixSpan. A frequent sequence $fs(h) \in fs(H)$ is represented as: $fs(h) = \langle (a_1, x_1), (a_2, x_2), \ldots, (a_n, x_n) \rangle$ Here, $x_j = t_{j+1} - t_j$ represents the time interval between medical orders a_j and a_{j+1}.*

C. Clustering Evaluation Metrics

1) Silhouette Score: The silhouette score is a metric used to evaluate clustering performance [7]. Assume that a dataset D is partitioned into k clusters $C_1, C_2, \ldots, C_k$. The following two measures are computed:

$$a(o) = \frac{\sum_{o' \in C_i, o' \neq o} dist(o, o')}{|C_i| - 1}$$

$$b(o) = \min_{C_j : 1 \leq j \leq k, j \neq i} \left\{ \frac{\sum_{o' \in C_j} dist(o, o')}{|C_j|} \right\}$$

Here, $a(o)$ represents the average distance between a data point o and all other points within the same cluster C_i. However, $b(o)$ represents the minimum average distance between o and all points in any other cluster, indicating how well o is separated from the neighboring clusters. The silhouette score $s(o)$ is calculated as follows:

$$s(o) = \frac{b(o) - a(o)}{\max\{a(o),\, b(o)\}}$$

The value of $s(o)$ ranges from -1 to 1. A value close to 1 indicates that the cluster containing o is compact and well separated from other clusters, which is a desirable clustering result. Conversely, a negative value indicates that o is closer to points in other clusters than to those in its own cluster, implying an undesirable clustering result.

2) Dunn Index: The Dunn index is another metric used to evaluate clustering performance [8]. Assume that a dataset D is partitioned into k clusters $C_1, C_2, \ldots, C_k$, and let o denote a data point. The cluster compactness Δ is defined as follows:

$$\Delta = \max_{C(o_i) = C(o_j)} \{d(o_i, o_j)\}$$

This represents the maximum distance between any two points within the same cluster. The separation between different clusters δ is defined as follows:

$$\delta = \min_{C(o_i) \neq C(o_j)} \{d(o_i, o_j)\}$$

This represents the minimum distance between any two points belonging to different clusters. The Dunn index is defined as

the ratio of these two quantities, $DI = \delta/\Delta$, where a larger value of the Dunn index indicates that the separation between clusters is large relative to the cluster compactness, which corresponds to a desirable clustering result. However, since the Dunn index relies on the maximum and minimum distances within and between clusters, it is sensitive to outliers.

D. Related Work

In this highly active research field, multiple extensions for SPM have been designed to meet specific experimental needs. Major approaches related to this work include frequency-based SPM and time-interval SPM. A well-known SPM algorithm is the Apriori-based frequent pattern mining algorithm [9]. When using large datasets, however, this time-consuming method generates many irrelevant patterns. To exclude these irrelevant patterns, PrefixSpan [10] was proposed to mine the complete set of patterns while reducing the effort of candidate pattern generation by exploring prefix projection. To improve efficiency further, CSpan [11] was proposed for mining closed sequential patterns. For each sequence, if there is no supersequence with the same support that also includes the sequence, it is a closed sequence. Uragaki et al. evaluated the constructed clinical pathways and proposed a method for recommending new branches and variants [1]. Le et al. proposed a method for detecting differences in clinical pathways [2]. In addition, Le et al. analyzed the factors underlying these differences [3]. To compare multiple health-care institutions in this study, the distances between patterns observed in frequent medical orders from three or more institutions were defined. Clustering was performed based on these distances to classify the target hospitals into several clusters, followed by analyses within each cluster [4]. Although the final MSVs were visualized as graphs, the representation based solely on MSVs was insufficient to fully capture the characteristics of each medical institution.

Moreover, Valerie et al. [12] conducted clustering on ordered item sets; however, since only a single item set was assigned to each element, their method cannot be applied to the data in this study, where multiple item sets may correspond to a single element.

III. Proposed Method

This section describes the proposed method. First, we explain the method for extracting frequent medical orders using the approach proposed by Sugitani et al. [6], followed by the method for determining the appropriate minimum support (minsup) used in the extraction process. Next, we describe the method for calculating distances between multiple health-care institutions using DTW. Finally, we present the clustering method and evaluation approach.

A. Extraction Method for Frequent SVs

Patterns in frequent medical orders for each disease at each medical institution are extracted. In this study, we adopt the method proposed by Sugitani et al. [6], which was described in Section II-B, to extract patterns from the frequent medical orders for each disease. Since the set of extracted patterns varies depending on the minsup, it is necessary to determine which patterns should be used for clustering. Multiple minsup values are tested to select an appropriate minsup that yields a set of patterns most consistent with actual clinical practice.

B. Determination of Appropriate Minimum Support

In this study, the extracted patterns are compared with a core clinical pathway, which represents a basic treatment plan, and recall, precision, and F-measure are calculated. Among the tested minsup values, the one that achieves the highest average F-measure is selected. To determine an appropriate minsup, the extracted frequent SVs are compared with the standard clinical pathways used in actual medical settings, and recall, precision, and F-measure are calculated. Recall is defined as the proportion of correctly identified positive instances among all actual positive instances, and is calculated as follows:

$$\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Precision is defined as the proportion of correctly identified positive instances among all predicted positive instances, and is calculated as follows:

$$\frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

The F-measure is the harmonic mean of precision and recall, and is calculated as follows:

$$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

C. Distance Calculation Using Dynamic Time Warping

This study adopts DTW because it can manage time-series data and allows distance computation even when the lengths of medical order sequences differ.

1) Calculating Distances Between Sequences: The DTW distance between two sequences is defined as the sum of distances after aligning their lengths. Let $A = (a_1, a_2, \ldots, a_n)$ and $B = (b_1, b_2, \ldots, b_m)$ be sequences of length n and m, respectively. The DTW distance $DTW(A, B)$ is defined as follows:

$$f(t,i) = ||a_t - b_i|| + \min \begin{cases} f(t, i-1) \\ f(t-1, i) \\ f(t-1, i-1) \end{cases}$$

where $f(0,0) = 0$, $f(t,0) = f(0,i) = \infty$ for $(t \in \{1..n\}, i \in \{1..m\})$. Here, $||a_t - b_i||$ represents the distance between two elements. In this study, when comparing sequences, we assign a distance of 0 if the elements are identical and 1 if they are different and then apply the above computation.

Table I illustrates an example of calculating the DTW distance between two sequences: $\langle(0, \text{Surgery}), (1, \text{Medication}), (2, \text{Examination})\rangle$ and $\langle(0, \text{Surgery}), (1, \text{Examination}), (2, \text{Medication})\rangle$. First, the element $(0, \text{Surgery})$ in both sequences is identical, resulting in a distance of 0. Next, the elements $(1, \text{Medication})$ and $(1, \text{Examination})$ differ, giving a distance of 1. Similarly, the

TABLE I
EXAMPLE OF DTW DISTANCE CALCULATION BETWEEN TWO SEQUENCES

B/A	(0, Surgery)	(1, Medication)	(2, Examination)
(0, Surgery)	0	1	2
(1, Examination)	1	1	2
(2, Medication)	2	2	2

elements $(2, \text{Examination})$ and $(2, \text{Medication})$ also differ, resulting in a distance of 1. Therefore, the final DTW distance is 2.

2) Calculating Distances Between Health-care Institutions: Let $fs(H)$ and $fs(K)$ denote the sets of frequent sequences for health-care institutions H and K, respectively. Each element in these sets is defined according to Definition 4. Using the DTW method, the distance between institutions H and K is calculated as follows:

$$distance(H, K) = \frac{\sum d(fs(h_p), fs(k_q))}{\|fs(H)\| \times \|fs(K)\|}$$

Here, $fs(h_p) \in fs(H)$ and $fs(k_q) \in fs(K)$, and $\|fs(H)\|$ and $\|fs(K)\|$ denote the numbers of frequent sequences in $fs(H)$ and $fs(K)$, respectively. The term $d(fs(h_p), fs(k_q))$ represents the DTW distance between $fs(h_p)$ and $fs(k_q)$ where $fs(h_p)$ and $fs(k_q)$ can be represented as follows:

$$fs(h_p) = \langle (a_{h_1}, x_{h_1}), (a_{h_2}, x_{h_2}), \ldots, (a_{h_p}, x_{h_p}) \rangle$$

$$fs(k_q) = \langle (a_{k_1}, x_{k_1}), (a_{k_2}, x_{k_2}), \ldots, (a_{k_q}, x_{k_q}) \rangle$$

The DTW distance $d(fs(h_p), fs(k_q))$ is computed by assigning a cost of 0 when elements (i.e., pairs of medical orders and time intervals) match, and 1 otherwise.

D. Hierarchical Clustering

Using the distances between frequent sequences and between health-care institutions calculated in Section III-C, hierarchical clustering is performed as described previously. Other clustering methods based on distance measures can also be applied.

E. Clustering 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, which are described in detail in Sections II-C1 and II-C2, respectively.

IV. EXPERIMENTS

This section describes the experimental evaluation of the proposed method. First, the experimental methodology and environment are presented. Finally, the experimental results are reported.

TABLE II
EXPERIMENTAL ENVIRONMENT

Item	Description
Programming Language	Python
Library	PrefixSpan implementation
Database	PostgreSQL ver. 16.6
Database Driver	Psycopg

A. Experimental Method

In this study, we compare the clustering results obtained using all frequent sequences extracted by the proposed method with those obtained using an existing method [4], which employs a single sequence generated by merging the frequent medical order sequences extracted from each health-care institution. For example, two sequences $\langle (1, \text{Examination}) \rangle$ and $\langle (0, \text{Medication}), (2, \text{Examination}) \rangle$, are integrated into a single sequence, such as $\langle (0, \text{Medication}), (1, \text{Examination}), (2, \text{Examination}) \rangle$ while preserving temporal order. Clustering performance is evaluated using the silhouette score and the Dunn index.

When extracting frequent SVs, the minsup varies (range, 0.3–0.8). The recall, precision, and F-measure processes described in Section III-B are calculated, and the minsup value with the highest average F-measure is selected.

Clustering performance is evaluated using the silhouette score and Dunn index, which are internal validation metrics. A silhouette score closer to 1 indicates higher intracluster cohesion and intercluster separation.

To avoid generating excessively small clusters, the minimum cluster size is set to 3, and clusters smaller than this threshold are merged into the nearest cluster. In addition, the optimal number of clusters is determined as the one that maximizes the silhouette score while ensuring that the Dunn index is not infinite. The clustering conditions are summarized as follows:

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

As representative clustering methods, five hierarchical clustering approaches are compared: Ward's method, single linkage, complete linkage, average linkage, and centroid linkage.

B. Experimental Environment and Dataset

In this study, the implementation was conducted in the environment shown in Table II. The Python library `linkage` was used to perform hierarchical clustering and the results were visualized using dendrograms. The silhouette score was computed using the `sklearn.metrics` library in Python, while the Dunn index was calculated using the `validclust` library.

For the evaluation using real-world data, we used an anonymized dataset of EMRs collected from multiple health-care 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 (Approval

TABLE III
SCHEMA OF THE REAL-WORLD DATASET

Data	Data Schema	Relation to Medical Order Category
Patient	Facility ID, Patient ID, Admission Date, Age, Sex, Body Mass Index	Attribute Information
Date	Facility ID, Patient ID, Date, Admission Date, Discharge Date	Sequence Length
Diagnosis	Facility ID, Patient ID, Admission Date, ICD-10 Code	Disease Name
Surgery	Facility ID, Patient ID, Admission Date, Surgery Date, K Code	Surgery
Medication	Facility ID, Patient ID, Admission Date, Administration Date, Drug Classification Code	Medication
Examination	Facility ID, Patient ID, Admission Date, Measurement Date, Test Name, Test Value, Unit	Examination
Clinical Procedure	Facility ID, Patient ID, Admission Date, Procedure Date, Receipt Code	Clinical Procedure
Vital Signs	Facility ID, Patient ID, Admission Date, Measurement Date, Item Name, Value, Unit	—
Disease	Facility ID, Patient ID, Admission Date, Diagnosis Procedure Combination Code	Attribute Information

TABLE IV
REAL DATASET: NUMBER OF PATIENTS BY DISEASE

Disease	Num. Patients
Malignant neoplasm of breast (Breast)	12,841
Angina pectoris/chronic ischemic heart disease (Angina)	11,044
Malignant neoplasm of lung (Lung)	10,005
Bladder tumor (Bladder)	9,720
Malignant neoplasm of stomach (Gastric)	5,961
Malignant neoplasm of liver and intrahepatic bile ducts (Liver)	2,016
Intervertebral disc degeneration/herniation (Disc)	1,194
Malignant neoplasm of cervix and uterus (Cervix)	888
Total	**53,669**

No. 2024_MIL_0004_A001) and was conducted in accordance with the approved guidelines. This dataset has undergone strict anonymization to protect patient privacy and the medical order codes, and related representations were standardized to enable comparative analyses across different health-care institutions.

Table III shows the schema of the real-world dataset used in this study while Table IV presents the major diseases included in the dataset and the corresponding number of patients. The tables show a considerable variation in the number of patients across diseases, reflecting the actual distribution of diseases and data-collection conditions in real-world clinical settings. This dataset contains records for a total of 53,669 patients, which enables the evaluation of the proposed method under realistic conditions without bias toward specific diseases.

TABLE V
CASES THAT DID NOT SATISFY THE CLUSTERING CONDITIONS

Disease	Method Type	Clustering Method
Angina pectoris/chronic ischemic heart disease	Existing method	Single
Malignant neoplasm of cervix and uterus	Proposed method	Single
Malignant neoplasm of cervix and uterus	Existing method	Ward
Malignant neoplasm of cervix and uterus	Existing method	Single
Malignant neoplasm of cervix and uterus	Existing method	Complete
Malignant neoplasm of cervix and uterus	Existing method	Average
Malignant neoplasm of cervix and uterus	Existing method	Centroid
Malignant neoplasm of breast	Proposed method	Single
Malignant neoplasm of breast	Existing method	Single

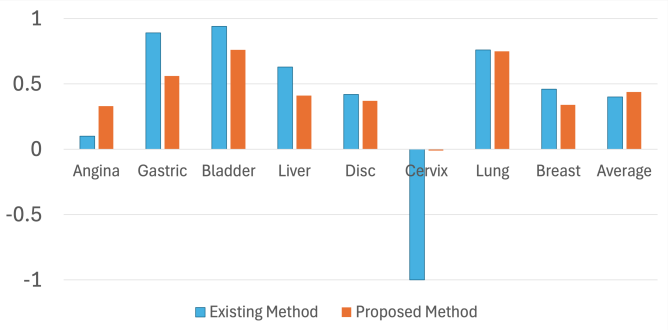

Fig. 1. Silhouette scores for each disease

C. Experimental Results

Figures 1 and 2 show the silhouette scores and Dunn index values for each disease. The average silhouette scores were 0.439 and 0.400 for the proposed and existing methods, respectively, indicating that the result from the proposed method was higher by 0.039. In contrast, the average Dunn index was 0.424 and 0.490 for the proposed and existing methods, indicating that the result from the proposed method was lower by 0.066.

Figures 3 and 4 show the silhouette scores and Dunn index values for each clustering method. As can be observed from the figures, there is no significant difference between the proposed and existing methods when results are compared across clustering methods. This finding suggests that differences in clustering performance are mainly attributable to variations in disease-specific sequences rather than the clustering methods themselves. In addition, among the clustering methods, Ward's method achieved the highest performance in terms of both silhouette score and Dunn index for both the proposed and existing methods.

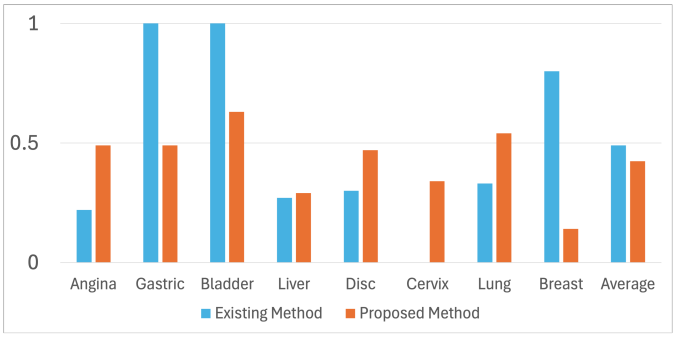

Fig. 2. Dunn index values for each disease

TABLE VI
STANDARD DEVIATION

	Proposed Method	**Existing Method**
Silhouette scores (by disease)	0.234	0.588
Silhouette scores (by method)	0.198	0.200
Dunn index (by disease)	0.147	0.360
Dunn index (by method)	0.048	0.050

However, the existing method frequently did not satisfy the clustering conditions. Table V lists the cases where the clustering conditions were not satisfied. Among these, seven out of nine cases correspond to the existing method, indicating that such failures occurred 7.5 times more often than in the proposed method. In particular, for malignant neoplasms of the cervix and uterus, clustering could not be performed using any method under the existing approach. Therefore, the relatively high silhouette scores and Dunn index values observed for the existing method are attributable only to a subset of successfully clustered diseases and do not reflect the method's overall performance.

Furthermore, as shown in Table VI, the proposed method exhibits lower standard deviation across all evaluation results, indicating more stable clustering performance than the existing method. These results demonstrate that the proposed method enables more stable clustering across a wider range of diseases.

V. Conclusion

A. Summary

In this study, we proposed a clustering method based on interinstitutional distances, which were computed using DTW to accurately capture the similarities and differences between the observed patterns in frequent medical orders across multiple health-care institutions. Conventional methods represent each institution using only a single representative pattern for frequent medical orders, which limits their ability to fully capture institutional characteristics. In addition, insufficient evaluation of clustering performance was a limitation of earlier approaches.

In the proposed method, all SVs extracted from each health-care institution were considered. The distances between SVs were computed using DTW, and interinstitutional distances defined by averaging these values. This distance computation accounts for temporal differences in medical order sequences. The experimental results using real-world data showed that Ward's method achieved the highest performance among clustering methods for both the proposed and existing approaches. Furthermore, in the evaluation experiment using real-world data, the results showed that Ward's method achieved high accuracy for both the proposed and existing clustering methods. In addition, the existing method exhibited 7.5 times more clustering rule violations compared with the proposed method. Furthermore, the proposed method showed a lower standard deviation by an average of 0.143 for both disease-wise and method-wise evaluations. Therefore, the proposed method has more stable clustering performance than the existing method. These findings indicate that the proposed method is a robust and practical approach for clustering health-care institutions.

B. Future Work

As future work, we plan to perform clustering using available information, such as the number of hospital beds and types of health-care institutions and compare those results with the outcomes of the proposed method. In addition, we aim to present the study results to medical professionals and evaluate its practical applicability in real-world clinical settings.

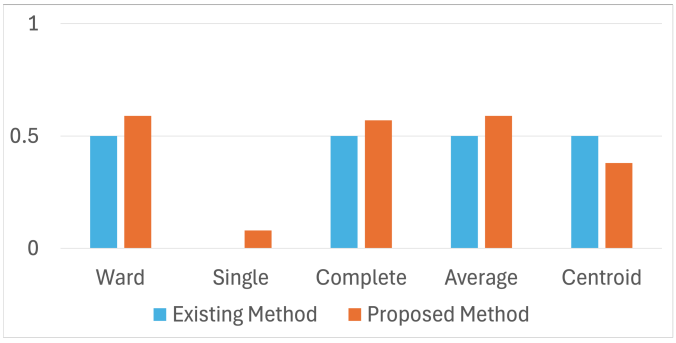

Fig. 3. Silhouette scores for each clustering method

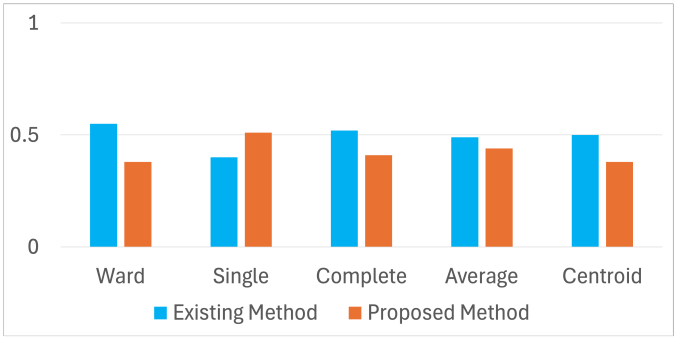

Fig. 4. Dunn index values for each clustering method

ACKNOWLEDGMENT

This work was partially supported by a Grant-in-Aid for Scientific Research (B) (#24K02943) from the Japan Society for the Promotion of Science.

REFERENCES

[1] K. Uragaki, T. Hosaka, Y. Arahori, M. Kushima, T. Yamazaki, K. Araki, and H. Yokota, "Sequential Pattern Mining on Electronic Medical Records with Handling Time Intervals and the Efficacy of Medicines," in *Proc. IEEE Workshop on ICT Solutions for Health*, pp. 20–25, 2016.

[2] Y. Honda, M. Kushima, T. Yamazaki, K. Araki, and H. Yokota, "Detection and Visualization of Variants in Typical Medical Treatment Sequences," in *Proc. the Int. Workshop on Data Management and Analytics for Medicine and Healthcare (DMAH)*, pp. 88–101, 2017.

[3] H. H. Le, T. Yamada, Y. Honda, T. Sakamoto, R. Matsuo, T. Yamazaki, K. Araki, and H. Yokota, "Methods for Analyzing Medical-Order Sequence Variants in Sequential Pattern Mining for Electronic Medical Record Systems," *ACM Trans. Comput. Healthcare*, vol. 4, no. 1, pp. 3:1–3:28, 2023.

[4] H. H. Le, Y. Yasumitsu, R. Matsuo, T. Yamazaki, and H. Yokota, "A Clustering-Based Sequence Variants Analysis Method for Electronic Medical Records of Multi-Medical Institutions," in *Proc. the 7th IEEE Int. Conf. on Multimedia Information Processing and Retrieval (MIPR)*, pp. 653–659, 2024.

[5] Japanese Society for Clinical Pathways, "Clinical Pathways," [Online]. Available: https://www.jscp.gr.jp/index_en.html

[6] M. Sugitani, R. Matsuo, T. Yamazaki, K. Araki, H. Yokota, M. Oguchi, and H. H. Le, "Extracting and Visualizing Frequent Medical Instruction Patterns with Statistical Insights from Multi-Institutional Electronic Medical Record Data," in *Proc. the 38th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 514–515, 2025.

[7] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *J. Comput. Appl. Math.*, pp. 53–65, 1987.

[8] J. C. Dunn, "Well-Separated Clusters and Optimal Fuzzy Partitions," *J. Cybernetics*, pp. 95–104, 1974.

[9] A. Rakesh and S. Ramakrishnan, "Mining Sequential Patterns," in *Proc. the 11th Int. Conf. on Data Engineering (ICDE)*, pp. 3–14, 1995.

[10] H. Jiawei, P. Jian, M.-A. Behzad, P. Helen, C. Qiming, D. Umeshwar, and H. Mei-Chun, "Prefixspan: Mining Sequential Patterns Efficiently by Prefix-projected Pattern Growth," in *Proc. the 17th Int. Conf. on Data Engineering (ICDE)*, pp. 215–224, 2001.

[11] R. V. Purushothama and V. G. Saradhi, "Mining Closed Sequential Patterns in Large Sequence Databases," in *International Journal of Database Management Systems*, vol. 7, no. 1, pp. 29–39, 2015.

[12] V. Guralnik and G. Karypis, "A Scalable Algorithm for Clustering Sequential Data," in *Proc. the 1st Int. Conf. on Data Mining (ICDM)*, pp.179–186, 2001.