Original Article

DOI: 10.4274/haseki.galenos.2026.15238
Med Bull Haseki 2026;64(4):333-344

Semantic Architecture and Knowledge Evolution of Endoscopic Submucosal Dissection Research: A Computational Bibliometric Analysis

Huseyin Demir[1], Tuncay Akar[2], Suleyman Mertoglu[3], Merve Tekinarslan[4]

[1]*Izmir Katip Celebi University Faculty of Economics and Administrative Sciences, Department of Health Management, Izmir, Türkiye*
[2]*Mugla Sıtkı Kocman University Institute of Health Sciences, Department of Health Management, Mugla, Türkiye*
[3]*Izmir Provincial Directorate of Health, Izmir, Türkiye*
[4]*Gumushane University Faculty of Health Sciences, Department of Health Management, Gumushane, Türkiye*

Abstract

Aim: The overarching intellectual structure, dominant thematic clusters, and longitudinal evolution of the endoscopic submucosal dissection (ESD) literature have yet to be systematically mapped. This study aims to characterize the semantic and thematic architecture of ESD research through computational text analytics and to reveal latent structural patterns and temporal dynamics within this domain.

Methods: A corpus of 1,000 peer-reviewed publications was processed through an eight-stage computational pipeline that incorporated word embedding-based semantic analysis, unsupervised clustering, Latent Dirichlet Allocation (LDA) topic modeling, inter-topic network analysis, and temporal trend modeling. The primary outcome was characterization of the semantic architecture of ESD research, whereas secondary outcomes included topic relationships, temporal evolution, and model performance metrics.

Results: Semantic analysis revealed two dominant clusters: a technically oriented domain encompassing device innovation and procedural optimization and a clinically oriented domain addressing therapeutic outcomes, organ-specific protocols, and adverse event management. Topic modeling identified nine coherent thematic categories, with limited integration between technical and clinical research trajectories. Temporal analysis revealed a pronounced post-2024 shift toward upper-gastrointestinal applications, particularly esophageal and gastric ESD, with relative stagnation in colorectal investigations. The optimal semantic clustering solution achieved a silhouette score of 0.215, while the nine-topic LDA model demonstrated a coherence score of 0.39.

Conclusion: This study provides the first data-driven macro-structural characterization of the ESD literature, revealing a persistent bifurcation between technical innovation and clinical application streams. These findings offer a strategic framework to help clinicians, investigators, and curriculum developers identify priority areas for advancement in therapeutic endoscopy.

Keywords: Endoscopic submucosal dissection, treatment outcome, bibliometrics, semantics, cluster analysis, policy

Introduction

Endoscopic submucosal dissection (ESD), first introduced in Japan during the late 1990s (1), represents a major advancement in therapeutic endoscopy. Developed to overcome the limitations of endoscopic mucosal resection (EMR) (2,3), ESD enables en bloc resection of large and morphologically heterogeneous neoplastic lesions without lymphovascular invasion (4,5), conferring significant histopathological advantages in margin assessment and invasion depth determination. The inability of conventional EMR to achieve en bloc resection in lesions larger than 20 mm, combined with elevated recurrence

Corresponding Author: Huseyin Demir, MD, Izmir Katip Celebi University Faculty of Economics and Administrative Sciences, Department of Health Management, Izmir, Türkiye
E-mail: huseyin.demir@ikc.edu.tr **ORCID:** orcid.org/0000-0002-8990-7228
Received: 21.01.2026 **Accepted:** 17.07.2026 **Publication Date:** 24.09.2026

Cite this article as: Demir H, Akar T, Mertoglu S, Tekinarslan M. Semantic architecture and knowledge evolution of endoscopic submucosal dissection research: a computational bibliometric analysis. Med Bull Haseki. 2026;64(4):333-344

rates following piecemeal resection (6,7), underscores ESD's clinical value. Recent studies indicate that ESD is the best way to treat superficial gastric neoplasias and early gastric cancer (2). It is also being used more and more to treat esophageal squamous cell carcinoma and some colorectal cases, especially lesions that are larger than 20 mm or that are not suitable candidates for conventional polypectomy (8-10). Despite its clinical impact, ESD remains technically demanding, with complication rates of 3.5%, 3.3%, and 4.6% for gastric, esophageal, and colorectal applications, respectively (11-13). Submucosal fibrosis and limited operator expertise in Western settings have further constrained broader implementation, though structured training initiatives may mitigate these barriers (14,15).

The ESD literature spans diverse domains, including clinical efficacy, complication management, technological innovation, and organ-specific protocols (16,17). However, no study has examined this corpus at the macrostructural level to characterize thematic clustering, latent semantic architecture, or temporal evolution—the gap that limits the systematic understanding of the field's research organization and strategic priorities. Natural language processing (NLP) and unsupervised machine learning methodologies, particularly word embedding-based semantic analysis and Latent Dirichlet Allocation (LDA), offer robust frameworks for revealing latent thematic structures within large medical corpora (18). The thematic trajectory of ESD literature may further be interpreted through Rogers' diffusion of innovations framework (19), wherein early feasibility and instrumentation studies reflect initial adoption stages, while contemporary focus on outcomes, complication protocols, and organ-specific applications signals progression toward implementation and confirmation phases. We hypothesized that computational analysis of the ESD literature would reveal a structurally fragmented research landscape, organized along two parallel but largely disconnected trajectories: one centered on technical innovation and procedural optimization, and the other focused on clinical outcomes and organ-specific applications, with a discernible temporal shift toward upper gastrointestinal domains over the 2015-2025 period.

Unlike conventional bibliometric studies that primarily rely on citation counts, co-authorship networks, or keyword co-occurrence analyses, the present study integrates NLP with word embedding–based semantic analysis and LDA topic modeling to uncover the latent conceptual organization of the ESD literature. This integrated computational framework enables a deeper exploration of semantic relationships, thematic structures, and temporal knowledge evolution beyond the capabilities of traditional bibliometric approaches. In this context, the primary aim of this study is to systematically characterize the semantic, thematic, and temporal architecture of ESD research through computational text analytics, thereby revealing latent structural patterns and dynamics within this specialized domain. A secondary aim is to identify dominant thematic clusters, inter-topic relationships, and longitudinal research trends within the field. If these objectives are achieved, this analysis is expected to provide clinicians, investigators, and medical educators with an empirically grounded roadmap for prioritizing future research agendas, optimizing procedural training curricula, and informing health policy decisions related to ESD implementation.

Materials and Methods

Compliance with Ethical Standards

This study is a retrospective analysis of publicly available publication data retrieved from bibliometric databases. The analysis did not involve any human subjects, primary data collection, or interventions. Since all data used in this research were already in the public domain and freely accessible, and no identifiable personal information or sensitive data were processed, ethical approval from an institutional review board was not required.

Study Design

This study employed a retrospective computational bibliometric design based on scientific publications, integrating bibliometric methods with NLP and computational text analytics, including latent semantic analysis and topic modeling, to systematically map the intellectual structure and temporal evolution of the ESD literature. To achieve this objective, the analysis was conducted using an eight-step computational pipeline (Figure 1), enabling a systematic and multidimensional evaluation of the ESD research landscape.

The pipeline shown in Figure 1 consists of data collection, preprocessing, word embedding-based semantic vector generation, semantic clustering, topic modeling, inter-topic correlation analysis, temporal dynamics evaluation, model performance measurement, and visualization of results.

Step 1: Data Source, Search Strategy, and Research Unit

The data used in this study were obtained from the Web of Science (WoS) Core Collection (20), a widely recognized and trusted source for examining scientific literature in bibliometric and text analytics research. The study was conducted under the assumption that publications retrieved from the WoS database adequately reflect developments in the relevant scientific field. The unit of analysis for the study consists of the 1,000 most-cited articles identified

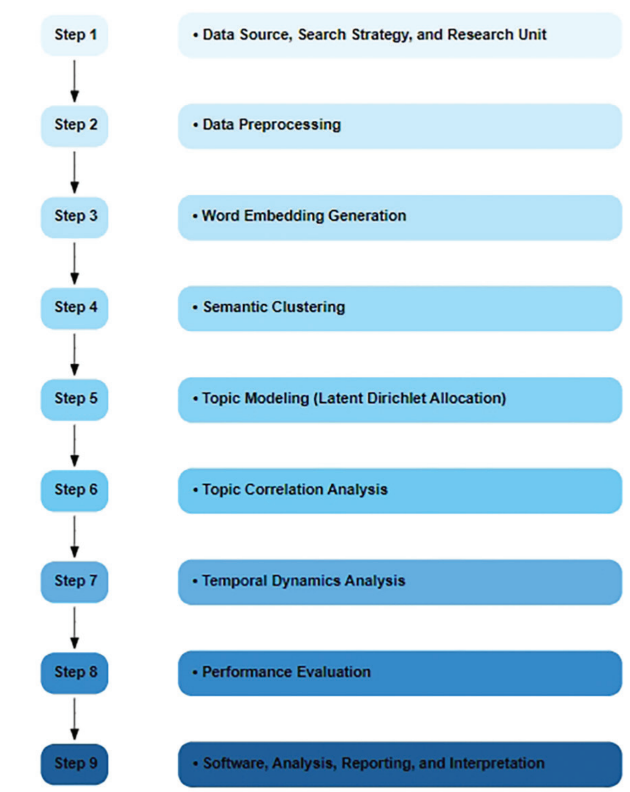

Figure 1. Eight-step computational workflow used for semantic and thematic analysis of the ESD literature. The workflow includes data collection, text preprocessing, Word2Vec-based document embedding, semantic clustering, LDA topic modeling, topic correlation analysis, temporal trend analysis, model performance evaluation, and data visualization

ESD: Endoscopic submucosal dissection, LDA: Latent Dirichlet Allocation

in the database and published in the last decade (2015-2025). The selection of the 1,000 most-cited publications within the 2015-2025 period was guided by both methodological and substantive considerations. Citation frequency serves as a widely validated proxy for scientific influence; highly cited publications disproportionately shape the conceptual vocabulary, methodological norms, and thematic priorities of subsequent research, collectively constituting the intellectual core of a discipline. The 2015-2025 window ensures analytical coherence by capturing the contemporary phase of ESD development, characterized by procedural maturation and expanding organ-specific applications. The threshold of 1,000 articles was chosen to balance corpus richness—sufficient for statistically robust topic modeling and word-embedding analysis—against analytical tractability, as excessively large, heterogeneous corpora risk introducing noise that may obscure latent structural patterns. This citation-based delimitation strategy is consistent with established practice in computational literature analyses. Inclusion criteria were defined as follows: the study title must contain "endoscopic submucosal dissection," publications must be in English, must be research articles, must be published in citation-indexed journals (SCI, SSCI, SCI-Expanded, ESCI, or SCOPUS), and must be classified under health and/or medicine-related categories within WoS.

Step 2: Data Preprocessing

Data preprocessing was performed to improve the performance of subsequent NLP operations. In this context, the preprocessing procedure began with loading the data into the Google Colab (21) environment. Subsequently, all content in the dataset was converted to lowercase, thereby standardizing letter-case usage. Punctuation marks present in the data that do not provide any semantic contribution were systematically removed. All numerical characters found in the data were cleaned, assuming that numbers were unnecessary for data or semantic analysis. Additionally, common English stop words (e.g., "the," "is," and "a"), which are frequently used in text data but contribute little semantic content, were removed. In the final stage, the data were tokenized into a discrete word list that serves as the fundamental building block for subsequent analyses.

Step 3: Word Embedding Generation

Text data was transformed into dense numerical vector representations to more successfully reveal semantic relationships during the word embedding process (22). For this purpose, the Word2Vec model (23), commonly used in the literature for word embedding, was employed, and training was performed on the token lists generated for each abstract. The vector dimensionality in the model was set to 100, ensuring that each word was represented by a vector of that dimensionality. The choice of vector dimensionality reflects a balanced approach between semantic analysis and computational efficiency. Thus, for each abstract, a single document vector was generated by averaging the Word2Vec vectors of all the words in the abstract. This approach facilitated subsequent ML tasks by consolidating the semantic information of individual words into a fixed-size document vector.

Step 4: Semantic Clustering

The document embedding array was subjected to semantic clustering using the k-means algorithm (24). The elbow method (25) was used to determine the optimal number of clusters (k). This algorithm was run for k values from 2 to 10; the inertia for each k was calculated and plotted. Thus, the point at which the decrease in the inertia value sharply slowed was identified as the elbow point, and the optimal k value was determined to be 2. With this approach, each abstract was labeled with a semantic cluster label indicating membership in cluster 0 or 1, based on its proximity to the semantic cluster centers. In the next stage, document embeddings in very high dimensions were reduced to two dimensions using the t-Distributed Stochastic Neighbor Embedding (t-SNE) method (26). Thus, the t-SNE model was initialized using different parameters, including the number of components [2]; reproducibility [42]; perplexity [30], which balances local and global structure; and optimization [100]. The obtained two-dimensional coordinates were visualized.

Step 5: Topic Modeling (LDA)

The purpose of topic modeling is to uncover the latent thematic structure within the dataset (27). For this purpose, a corpus was created from tokenized texts, ensuring that each word was matched with a unique integer identifier. With this dictionary, words appearing in fewer than 5 documents or in more than 50% of documents were filtered and refined, thereby eliminating words that are either excessively sparse or excessively common and had low information value. After this process, a Bag-of-Words corpus (28) was created containing word ID and word frequency pairs for each document. Thus, the prepared corpus was subjected to training with the LDA model (29), and this process was carried out within the framework of values regarding the model's reproducibility [42] and the number of training iterations [10]. Prominent words for each topic were identified. The LDA model calculated the probability distribution over all topics for each abstract and assigned the topic with the highest probability as that abstract's dominant topic. Subsequently, word clouds for each topic were created from words and their weights using the WordCloud method (30) of the WordCloud library.

Step 6: Topic Correlation Analysis

The purpose of topic correlation analysis is to analyze the relationships and co-occurrence patterns between identified latent topics (31). In this context, topic distributions for each document were obtained from the LDA model. These data were organized as a topic matrix, in which rows represent documents and columns represent topic probabilities. Then, Pearson correlation (32) coefficients were calculated for all topic column pairs in the matrix, thus obtaining a topic correlation matrix showing the extent to which different topics co-occur across documents. This correlation matrix was visualized in the form of a heatmap using the heatmap function (33).

Step 7: Temporal Dynamics Analysis

The purpose of analysis of temporal dynamics is to examine how the prevalence of semantic clusters and topics has evolved. Data obtained from topic-matrix dataframe indices were consolidated with the original dataset, and averages of topic probabilities were calculated for each year, producing a dataframe that shows the annual average prevalence of each topic. The obtained time series data were visualized through line graphs using pyplot (34), thereby revealing the time-dependent development of each topic. Similarly, another data frame was created by calculating the proportion of documents in each semantic cluster per year. These data were also visualized using line graphs, which present in detail how the overall semantic structure changed over time and highlight prominent shifts between clusters.

Step 8: Performance Evaluation

The performance of the k-means clustering algorithm was analyzed using two fundamental metrics: the Silhouette score (35) and the Davies-Bouldin index (36). In this regard, the Silhouette score is a relative measure that quantifies an object's similarity to its own cluster relative to its similarity to other clusters. The Davies-Bouldin index evaluates the average similarity ratio of each cluster to its most similar cluster. In the topic modeling process, the interpretability of LDA topics was evaluated using the coherence score. Topic coherence is widely considered a sufficient and reliable metric for evaluating the quality of topics generated by LDA models, as it reflects the semantic interpretability of topics in a manner consistent with human judgment. Coherence measures have been shown to correlate strongly with human assessments of topic quality, supporting their use as a primary evaluation criterion in topic modeling (37).

Statistical Analysis

All computational analyses were performed using Google Colaboratory (21). Python (38) was used as the primary programming language for the analysis. The computational workflow was supported by various libraries commonly used in scientific research. These include NumPy (39) for numerical operations; Pandas (40) for data processing and tabular analyses; Matplotlib (34) and Seaborn (33) for visualization; Scikit-learn (41) for ML operations; Statsmodels (42) for time series modeling; and WordCloud (43) for word-cloud visualizations.

Results

Semantic Clustering Analysis Findings

Findings related to semantic clustering analysis are shown in Figure 2.

The blue points in the graph are predominantly concentrated on the right side, particularly in the lower-right region, while the green points are predominantly located on the left side, especially in the upper-left region. Although the clusters are distinct, minimal overlap between points of the two clusters is observable in the middle section.

Topic Modeling Findings

The LDA model identified 9 main topics within the corpus, with a coherence score of 0.39. The results are summarized in Table 1.

Findings related to word clouds are presented in Figure 3.

Concepts such as "patient," "ESD," "endoscopic," "resection," "bleeding," "dissection," "gastric," and "colorectal" are encountered in nearly all topics. While the words "patient," "ESD," and "group" stand out in Topic 0, the words "bleeding" and "patients" are prominent in Topic 2. For Topics 3 and 5, the concepts of "submucosal"

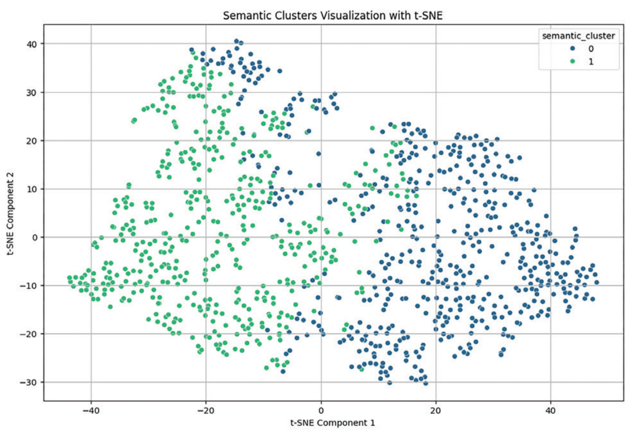

Figure 2. Two-dimensional visualization of semantic clusters generated using t-distributed stochastic neighbor embedding (t-SNE). Each point represents one publication abstract, while colors indicate cluster assignment derived from K-means clustering. Spatial proximity between points reflects semantic similarity among publications

and "resection" are prominent. In Topic 4, the word "perforation" stands out.

Findings on Topic Correlations

The heatmap showing correlations between topics is presented in Figure 4a.

Correlation analysis revealed that all inter-topic associations were negative, ranging from mild to moderate (-0.22 to -0.04). In this regard, the strongest negative correlations were observed between Topic 0 and Topic 2 (-0.21), Topic 3 (-0.22), and Topic 7 (-0.22), suggesting that documents heavily weighted toward Topic 0, centered on technical innovation and procedural optimization, are least likely to concurrently emphasize the thematic content represented by these three topics.

Findings on Temporal Dynamics Analysis

Time-dependent changes in clusters and topics are presented in Figure 4b.

Topics exhibited a relatively stable distribution over the years examined. While most topics fluctuate, ranging from 5% to 30% in annual document proportions, no dramatic change in topics was observed. However, Topic 7 showed a marked increase during the examined period, reaching approximately 75%.

The right graph (Cluster 0-blue; Cluster 1-orange) showed that, between 2015 and 2023, both clusters exhibited fluctuations at similar rates, with values generally ranging between 40% and 60%. Cluster 0 was more dominant, exceeding 60% in 2017 and 2022, while Cluster 1 remained at relatively lower levels during these periods. In the subsequent period (2024-2025), Cluster 0 showed a sharp increase exceeding 75%, while Cluster 1 experienced a dramatic decline to below 25%.

Table 1. Topics and prominent keywords

Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
esd	group	patients	esd	esd	esd	patients	esd	esd
patients	esd	bleeding	submucosal	patients	resection	esd	patients	bleeding
group	patients	esd	endoscopic	pecs	endoscopic	risk	resection	patients
resection	p	delayed	resection	perforation	dissection	factors	lesions	esophageal
p	esophageal	closure	dissection	p	submucosal	submucosal	endoscopic	stricture
endoscopic	gastric	group	cases	endoscopic	p	lesions	p	endoscopic
survival	endoscopic	endoscopic	procedure	risk	time	gastric	group	submucosal
recurrence	stricture	p	lesions	submucosal	speed	invasion	vs	lesions
outcomes	vs	risk	patients	factors	emr	cancer	colorectal	dissection
submucosal	ulcer	rate	study	study	procedure	ci	submucosal	gastric
rate	dissection	mucosal	colorectal	dissection	traction	endoscopic	rate	group
treatment	groups	submucosal	time	colorectal	cesd	lnm	study	p
gastric	study	colorectal	gastric	syndrome	rate	egc	treatment	postesd
dissection	significantly	study	performed	analysis	using	tumor	dissection	delayed
cancer	submucosal	gastric	p	group	study	fibrosis	bloc	risk

Figure 3. Word clouds representing the dominant keywords within each LDA-derived topic. Word size is proportional to the relative importance (weight) of each term within the corresponding topic
LDA: Latent Dirichlet Allocation

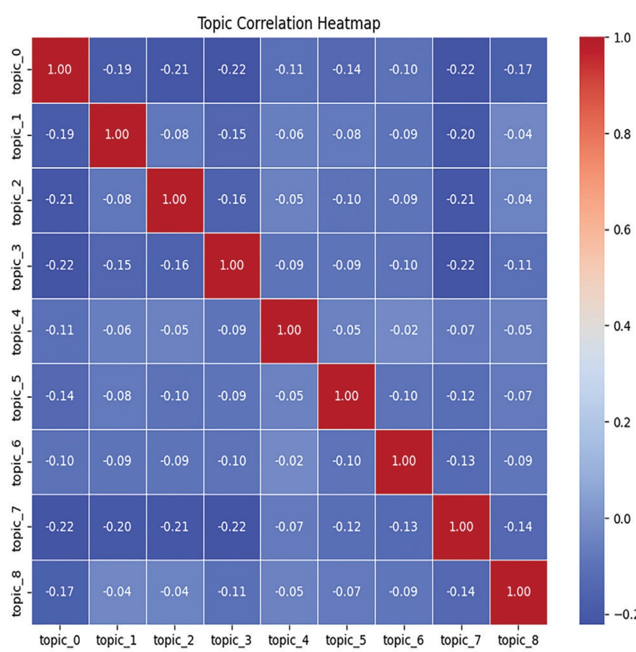

Figure 4a. Heatmap illustrating Pearson correlation coefficients between LDA-derived topics. Warmer and cooler colors represent the strength and direction of correlations, indicating the degree of thematic overlap or separation between topics
LDA: Latent Dirichlet Allocation

Findings on Cluster-Topic Distribution

Findings on cluster and topic distribution are presented in Figure 5a.

Accordingly, a balanced distribution between the two clusters was observed during the 2015-2023 period. Although both clusters had similar values of approximately 60 in 2017, Cluster 0 stood out in 2018. However, Cluster 1 lagged behind Cluster 0 in this regard. Subsequently, similar fluctuations were observed between 2019 and 2023, and, generally, no major differences were observed between the two clusters.

The right-hand graph shows values corresponding to the average probability distribution of topics (0-8) within two semantic clusters. Topic 0 (0.29) and Topic 7 (0.31) stood out in Cluster 0. Other topics had low probabilities in Cluster 0 (between 0.01 and 0.11).

Findings on Model Performance

Performance graphs for the semantic clustering and topic modeling processes are presented in Figure 5b.

The highest Silhouette score was obtained with two semantic clusters and was approximately 0.215. The right-hand graph presents the coherence score for LDA topic modeling. Findings from the LDA model show that nine topics have the highest coherence scores, suggesting that the model achieves a satisfactory level of semantic coherence and that the resulting topics are meaningful and interpretable for further analysis.

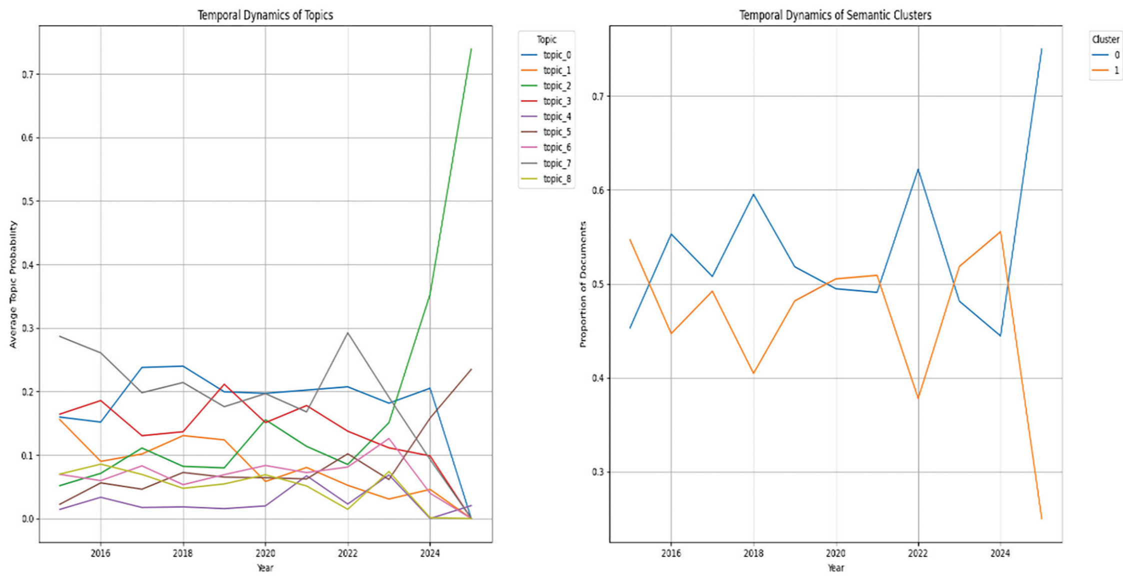

Figure 4b. Annual temporal evolution of semantic clusters and LDA-derived topics between 2015 and 2025. The graphs illustrate changes in the relative prevalence of semantic clusters and thematic topics over time, highlighting shifts in research focus within the ESD literature

LDA: Latent Dirichlet Allocation, ESD: Endoscopic submucosal dissection

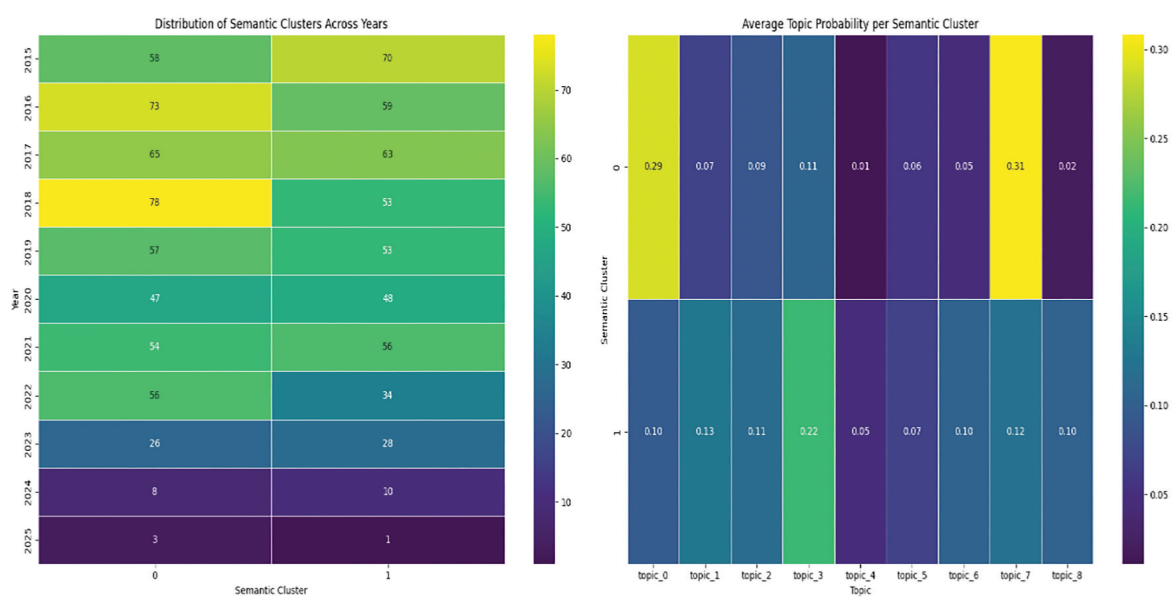

Figure 5a. Distribution of semantic clusters over time and average topic probabilities within each semantic cluster. The left panel illustrates annual changes in semantic cluster distribution, whereas the right panel presents the mean probability of each LDA-derived topic within Cluster 0 and Cluster 1

LDA: Latent Dirichlet Allocation

Discussion

This study represents one of the first comprehensive analyses using NLP and ML methods to reveal the semantic and thematic structure of the ESD literature. Since its development in Japan in the late 1990s (1) to overcome the limitations of EMR (2,3), ESD has emerged as a significant paradigm shift in therapeutic endoscopy. Our findings demonstrate that the literature surrounding the procedure is organized around two main axes: technical innovation and procedural development, and clinical outcomes and organ-specific applications.

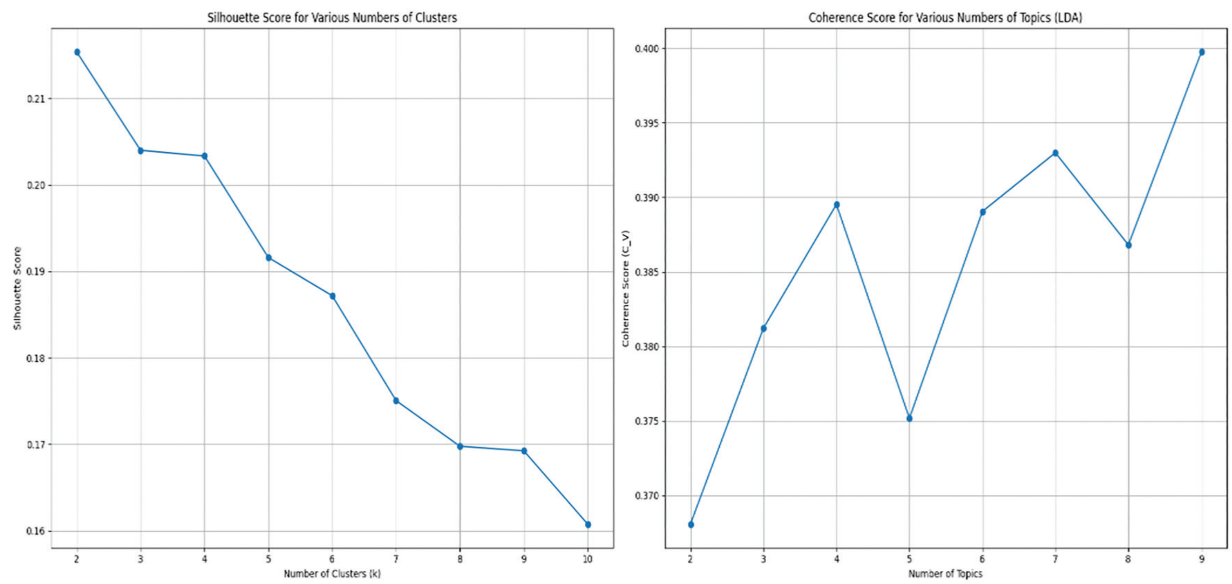

Figure 5b. Performance evaluation of semantic clustering and topic modeling. The left panel presents silhouette scores across candidate cluster numbers, whereas the right panel illustrates coherence scores for different LDA topic models
LDA: Latent Dirichlet Allocation

The semantic clustering analysis revealed that the ESD literature exhibits a fundamental structural divergence. This bifurcation was consistently observed in both t-SNE visualizations and LDA-based topic modeling, revealing two parallel research streams: clinical and technical. Studies related to clinical outcomes and organ-specific applications are positioned together in semantic space, whereas themes of dissection methods, device development, and procedural efficiency constitute a separate, technical research area. This divergence indicates that the causal relationships between technical innovation and clinical performance are not addressed holistically in the literature. K-means clustering was performed with k=2, resulting in a low silhouette score of 0.21. This relatively low value indicates limited separation between clusters, suggesting that the chosen number of clusters may oversimplify the underlying structure of the data. While the two-cluster solution provides a basic overview of the dataset, it may fail to capture more nuanced patterns that could be revealed with alternative clustering strategies or a higher number of clusters. Future analyses should consider testing additional k values and validating cluster stability using complementary metrics to ensure that the identified groupings accurately reflect the complexity of the dataset. These limitations highlight the need for cautious interpretation of the clustering results and underscore that the findings should be considered exploratory rather than definitive.

This structural separation is particularly significant given that ESD's clinical success fundamentally depends on technical proficiency. The technique's ability to achieve en bloc resection of lesions exceeding 20 mm—where conventional EMR requires piecemeal resection, with associated high local recurrence rates (7)—depends entirely on the operator's technical skill and the available instrumentation. The procedure involves submucosal lifting with viscous agents such as sodium hyaluronate followed by precise dissection with specialized knives (11), and the reported complication incidences of 3.5%, 3.3%, and 4.6% for gastric, esophageal, and colorectal ESD, respectively (12), underscore the intimate connection between technical execution and clinical outcomes. Yet our analysis reveals that these two dimensions are largely explored in isolation in the literature. This situation highlights that clinical needs are not taken as the primary basis for health innovation; technological advancements remain largely disconnected from the problems encountered in healthcare; consequently, the convergence between clinical practice and technology is becoming increasingly difficult.

The lack of clinical-technical integration emerges as an important limiting factor in the development of ESD and suggests that research designs need to be structured with a more holistic approach. This gap is particularly critical given that factors such as submucosal fibrosis caused by prior interventions like biopsy or marking can increase dissection difficulty and reduce en bloc success rates (13), creating a direct pathway from technical challenges to clinical outcomes. The semantically weak connections among complication-focused themes further reveal that risk prediction models are still immature, despite the procedure's well-documented association with bleeding and perforation risks. Accordingly, there is a clear need in the literature for modeling studies aimed at predicting certain risks, such as delayed bleeding.

The temporal analysis revealed a marked increase in esophageal and gastric studies after 2024, whereas colorectal ESD research remained relatively stagnant. This thematic shift provides empirical support for the diffusion of innovations model (19), which explains the adoption of health technologies through distinct stages. The observed pattern reflects the natural progression from technical feasibility studies to widespread clinical implementation.

Early-stage ESD research focusing on technical feasibility and instrument innovation corresponds to the "knowledge acquisition" and "persuasion" stages of the innovation adoption process (19). The current concentration of publications on upper gastrointestinal applications, particularly the acceptance of ESD as the standard treatment for superficial gastric neoplasias and early-stage gastric cancer (2), as well as its increasing adoption as first-line treatment for superficial esophageal squamous cell carcinoma, indicates progression to the "implementation" and "confirmation" stages. Applications carried out in more limited centers during the technical feasibility and early-experience phases have expanded rapidly upon reaching a certain maturity, thereby creating thematic density in the literature. In this context, the evidence suggests that the next phase of ESD research should move beyond adoption-focused discourse toward risk stratification and outcome optimization, thereby reinforcing the transition from widespread implementation to a more data-driven and precision-oriented confirmation stage in clinical practice.

The concentration of publications in organ-specific areas is also consistent with learning-curve dynamics. The relative stagnation in colorectal ESD research, despite EMR remaining the dominant method in this region (8), may reflect the unique technical challenges posed by this anatomical site. The distal colon and rectum present specific difficulties related to wall thickness, peristalsis, and anatomical configuration, which may have slowed the adoption curve relative to upper gastrointestinal applications. This finding suggests that colorectal ESD may still be in an early phase of the diffusion process, potentially transitioning between the "persuasion" and "decision" stages. With the adoption of ESD in this field, the accumulation of experience and knowledge is expected to yield substantially greater benefits for clinical practice.

This study goes beyond classical bibliometric approaches and presents an innovative framework that objectively reveals the structural characteristics of the literature through word embedding-based semantic analysis. LDA-based topic modeling (18) made the thematic centers of the literature quantitatively visible, while t-SNE-based visualization enabled better understanding of high-dimensional semantic relationships. The combined use of these methods allowed the ESD literature to be evaluated from a structural perspective, revealing patterns that cannot be detected through traditional systematic reviews (18).

These findings are critical to guiding clinical practices and research priorities. The marked increase in publications on the upper gastrointestinal system demonstrates that clinical practice increasingly focuses on these areas and that training programs need to allocate more resources to esophageal and gastric dissection techniques. The scarcity of sufficiently experienced operators in Western countries, which has limited widespread adoption, can be addressed through structured training programs (14,15) informed by the learning-curve dynamics revealed in our analysis.

The structural divergence between technical and clinical research reveals that procedural innovation is not evaluated in an integrated manner alongside clinical outcomes. This situation demonstrates the need for more holistic research in critical areas such as management of complications, selection of dissection strategies, and effectiveness of device use. Given that ESD provides superior outcomes in specific clinical contexts (8)—particularly in lesions exceeding 20 mm, cases with suspected superficial submucosal invasion, or cases unsuitable for conventional snare polypectomy, especially in the distal colon and rectum (9,10)—the integration of technical and clinical research streams could optimize patient selection and procedural planning.

The technique's ability to obtain intact specimens provides histopathological advantages by enabling clear assessment of lateral and vertical surgical margins and accurate analysis of invasion depth (4,5). Such diagnostic precision both enhances the success of curative treatment and reduces the risk of recurrence; however, our analysis suggests that the relationship between the quality of technical execution and these histopathological outcomes remains underexplored in the literature.

A fixed corpus size of 1,000 publications, selected on a citation-priority basis, constitutes a notable limitation of this study. This threshold, while methodologically convenient, is analytically arbitrary: it imposes a rigid ceiling that automatically excludes any publication—however recent, innovative, or clinically significant—that falls outside the citation-ranked top tier at the time of data extraction. In a rapidly evolving procedural field such as ESD, where AI-based risk prediction, organ-specific protocol refinement, and novel technical variants are actively emerging, this cutoff systematically filters out the studies most likely to signal where the field is heading rather than where it has been. The resulting corpus is therefore best understood not as a representative sample of ESD research, but as a citation-weighted retrospective of its most consolidated subset. Future bibliometric and NLP-based investigations should consider recency-stratified or citation-agnostic sampling strategies to counterbalance this structural skew and more accurately capture the field's emerging frontier.

Study Limitations

This study has several limitations that should be acknowledged. Most critically, the analysis is restricted to abstract-level text, which constitutes a significant methodological constraint. Abstracts inherently omit procedural details, statistical nuances, subgroup-level findings, and contextual clinical information embedded in full-text publications, which may carry substantial thematic weight. This restriction likely introduces systematic bias into the semantic and topical structures identified, as abstracts selectively emphasize outcomes over methodology and context. Consequently, the thematic architecture revealed in this study should be interpreted as a representation of how ESD research is framed and communicated, rather than a complete reflection of its intellectual depth.

A second and equally consequential limitation concerns selection bias introduced by the citation-based corpus construction strategy. By restricting the analysis to the top 1,000 most-cited publications, this study systematically privileges well-established, highly cited work to the detriment of recent contributions that have not yet had sufficient time to accumulate citations. Given that citation accumulation is inherently a slow, cumulative process—often spanning several years post-publication—this sampling approach almost certainly underrepresents studies published toward the latter portion of the 2015–2025 window, including emerging technical innovations, novel AI-based applications, and early-stage organ-specific protocols, which may not yet be reflected in the citation record. As a direct consequence, the thematic and temporal structures identified in this study should be understood to characterize the established, citation-validated core of ESD research rather than its full, current state; nascent research trajectories, cutting-edge methodological developments, and recently emerging subfields are likely underrepresented or entirely absent from the derived topic structure. This limitation is particularly consequential for the temporal trend analysis, as citation-driven undersampling of recent publications may artificially compress or obscure both the acceleration in upper-gastrointestinal and AI-related research that more recent, lower-citation studies might otherwise reveal. Finally, the 2015-2025 analytical window, which is temporally defined, may be further influenced by publication delays, geographic disparities in research output, and field-specific innovation cycles—factors that interact with and potentially compound the citation-based selection bias described above.

Notwithstanding these limitations, this study's primary strength lies in its novel application of NLP and unsupervised machine learning to the ESD literature, offering an objective, data-driven alternative to traditional narrative or systematic reviews. Word embedding-based semantic analysis captures nuanced conceptual relationships between research constructs that conventional keyword-based approaches fail to detect, while LDA topic modeling systematically exposes latent thematic clusters and structural divergences across the corpus. The integration of temporal analysis further enabled the longitudinal mapping of research trajectories—a dimension largely absent from existing ESD literature reviews. Collectively, these methodological layers render the study's findings reproducible, scalable, and transferable to other domains of procedural medicine seeking macro-structural characterization of their research landscapes.

Future ESD research should prioritize four key areas that together define the field's methodological and clinical trajectory. First, the current literature inadequately bridges technical innovation and clinical outcomes; multidisciplinary studies that jointly evaluate both dimensions within unified analytical frameworks are urgently needed to close this gap. Second, organ-specific comparative research—spanning esophageal, gastric, and colonic ESD—remains scarce despite the markedly distinct difficulty levels and complication profiles that characterize each anatomical site. Third, AI-based clinical decision support models capable of real-time prediction of complication risks, such as perforation and bleeding, represent a critical yet underdeveloped frontier, reflecting the broader immaturity of current risk-prediction frameworks in this domain. Fourth, and perhaps most pressing for global access to the procedure, quantitative modeling of ESD learning curves and standardized, competency-based performance metrics are essential to guide training programs, particularly in Western countries, where the scarcity of operators continues to constrain the expansion of this technique. Taken together, these four priorities—integrated technical-clinical evaluation, organ-specific comparative modeling, AI-driven adverse event prediction, and standardized competency assessment—chart a coherent agenda for advancing both methodological rigor and clinical relevance of ESD research.

Conclusion

This study provides a systematic mapping of the semantic and thematic architecture of the ESD literature using computational text analytics, extending beyond conventional bibliometric approaches. The analysis identifies two structurally distinct thematic clusters within

ESD research: one centered on technical innovation and procedural optimization, and the other on clinical outcomes and organ-specific therapeutic protocols. Co-occurrence and topic-relationship analyses show limited overlap between these two clusters, indicating that technical and clinical themes are predominantly addressed in separate bodies of literature rather than within integrated studies. Temporal analysis demonstrates a measurable shift in publication focus toward upper gastrointestinal applications, particularly esophageal and gastric ESD, in recent years. This shift is consistent with a broader transition observed in the data, from early-stage feasibility studies toward organ-specific clinical implementation. Future studies integrating transformer-based language models and full-text semantic analysis may further improve the understanding of knowledge evolution in ESD research.

Ethics

Ethics Committee Approval: This study was based exclusively on publicly available publication data retrieved from bibliometric databases. The analysis did not involve any human subjects, primary data collection, or interventions. Since all data used in this research were already in the public domain and freely accessible, and no identifiable personal information or sensitive data were processed, ethical approval from an institutional review board was not required for this bibliometric analysis.

Informed Consent: Informed consent was not applicable to this study, as it involved a bibliometric analysis of publicly available bibliographic data and did not involve human participants or the collection of identifiable personal information.

Footnotes

Authorship Contributions

Concept: H.D., T.A., S.M., Design: H.D., M.T., Data Collection or Processing: H.D., M.T., Analysis or Interpretation: H.D., M.T., Literature Search: H.D., T.A., S.M., M.T., Writing: H.D., M.T.

Conflict of Interest: The authors declared that there were no conflicts of interest.

Financial Disclosure: The authors declared that this study received no financial support.

Declaration of AI Use: The article was written solely by the authors. OpenAI's ChatGPT model was used to translate the manuscript into English, and Anthropic's Claude AI model assisted in reviewing the translated text to ensure linguistic accuracy and professional quality.

References

1. Fukuzawa M, Gotoda T. History of endoscopic submucosal dissection and role for colorectal endoscopic submucosal dissection: A Japanese perspective. Gastrointest Interv. 2012;1:30-5.
2. Moutzoukis MK, Manolakis A, Kapsoritakis A, Christodoulou D, Argyriou K. Endoscopic submucosal dissection: Challenges, innovations, and the road ahead. World J Gastrointest Endosc. 2026;18:113918.
3. Elkholy S, Awad A, Haggag H, Essam K, Yousef K, El Sherbiny M. Different endoscopic submucosal dissection techniques: a tailored approach. VideoGIE. 2025;10:277-84.
4. Pal S, Bhaduri G. Endoscopic submucosal dissection for early gastrointestinal malignancies: Current state and future perspectives. World J Gastrointest Endosc. 2025;17:109144.
5. Ziachehabi A, Worm M, Liu DHW, Pimingstorfer P, Langer R. Endoscopic submucosal dissection (ESD) of upper gastrointestinal carcinomas: an integrated clinical and pathological perspective. J Clin Med. 2025;14:8817.
6. Zheng X, Xu L. Endoscopic submucosal dissection vs. endoscopic mucosal resection for early gastric cancer: a meta-analysis. Front Med (Lausanne). 2025;12:1702512.
7. Bourke MJ, Neuhaus H, Bergman JJ. Endoscopic submucosal dissection: Indications and application in Western endoscopy practice. Gastroenterology. 2018;154:1887-900.
8. Matsui N, Akahoshi K, Nakamura K, Ihara E, Kita H. Endoscopic submucosal dissection for removal of superficial gastrointestinal neoplasms: a technical review. World J Gastrointest Endosc. 2012;4:123-36.
9. Mendes RR, Barreiro P, Mascarenhas A, Franco AR, Carvalho L, Chagas C. Endoscopic submucosal dissection for resections larger than 10 cm: outcomes from a Portuguese center. GE Port J Gastroenterol. 2023;31:33-40.
10. Bronswijk M, Rasschaert G, Hayashi Y, Yamamoto H. Colorectal endoscopic submucosal dissection: outcomes from a European multicenter cohort. Endoscopy. 2023;55:130-9.
11. Yamamoto H. Endoscopic submucosal dissection of early cancers and large flat adenomas. Clin Gastroenterol Hepatol. 2005;3:S74-6.
12. Odagiri H, Yasunaga H. Complications following endoscopic submucosal dissection for gastric, esophageal, and colorectal cancer. Ann Transl Med. 2017;5:189.
13. Kuroha M, Shiga H, Kanazawa Y, et al. Factors associated with fibrosis during colorectal endoscopic submucosal dissection. Digestion. 2021;102:590-8.
14. Bhatt YM, Weilert F, Fredrick RT, et al. EUS-guided treatment of gastric fundal varices with combined injection of coils and cyanoacrylate glue: A large U.S. experience over 6 years. Gastrointest Endosc. 2016;83:1164-72.

15. Barbour JA, O'Toole P, Suzuki N, Dolwani S. Learning endoscopic submucosal dissection in the UK: Barriers, solutions and pathways for training. Frontline Gastroenterol. 2020;12:671-6.
16. Akintoye E, Kumar N, Obaitan I, Alayo QA, Thompson CC. Peroral endoscopic myotomy: a meta-analysis. Endoscopy. 2016;48:1059-68.
17. Hatta W, Ogata Y, Ogasawara K, et al. Recent advances in endoscopic submucosal dissection for gastric cancer: focusing on expanded indications and technological innovations. J Gastric Cancer. 2026;26:76-91.
18. Vajjala S, Majumder B, Gupta A, Surana H. Practical natural language processing: A comprehensive guide to building real-world NLP systems. O'Reilly Media; 2020.
19. Rogers EM. Diffusion of innovations. 5th ed. New York: Free Press; 2003.
20. Li K, Rollins J, Yan E. Web of Science use in published research and review papers 1997-2017: a selective, dynamic, cross-domain, content-based analysis. Scientometrics. 2018;115:1-20.
21. Google. Google Colaboratory [cloud computing platform]. 2025. Available from: https://colab.research.google.com
22. Meddeb A, Ben Romdhane L. Using topic modeling and word embedding for topic extraction in Twitter. Procedia Comput Sci. 2022;207:790-9.
23. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv. 2013.
24. Lloyd SP. Least squares quantization in PCM. IEEE Trans Inf Theory. 1982;28:129-37.
25. Ketchen DJ Jr, Shook CL. The application of cluster analysis in strategic management research: an analysis and critique. Strateg Manag J. 1996;17:441-58.
26. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9:2579-605.
27. Blei DM, Carin L, Dunson D. Probabilistic topic models. IEEE Signal Process Mag. 2010;27:55-65.
28. Manning CD, Raghavan P, Schütze H. Introduction to information retrieval. Cambridge: Cambridge University Press; 2008.
29. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res. 2003;3:993-1022.
30. Skeppstedt M, Ahltorp M, Kucher K, Lindström M. From word clouds to Word Rain: Revisiting the classic word cloud to visualize climate change texts. Inf Vis. 2024;23:217-38.
31. Blei DM, Lafferty JD. Correlated topic models. Adv Neural Inf Process Syst. 2006;18:147-54.
32. Pearson K. Note on regression and inheritance in the case of two parents. Proc R Soc Lond. 1895;58:240-2.
33. Waskom ML. Seaborn: Statistical data visualization. J Open Source Softw. 2021;6:3021.
34. Hunter JD. Matplotlib: A 2D graphics environment. Comput Sci Eng. 2007;9:90-5.
35. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53-65.
36. Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell. 1979;1:224-7.
37. Mimno D, Wallach HM, Talley E, Leenders M, McCallum A. Optimizing semantic coherence in topic models. In: Barzilay R, Johnson M, editors. Proc 2011 Conf Empirical Methods Nat Lang Process. Association for Computational Linguistics; 2011.
38. Python Software Foundation. Python (Version 3.9.18) [programming language]. 2024. Available from: https://www.python.org/
39. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. Nature. 2020;585:357-62.
40. McKinney W. Data structures for statistical computing in Python. In: Proc 9th Python Sci Conf; 2010.
41. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-30.
42. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. Proceedings of the 9th Python in Science Conference. 2010:92-6.
43. Mueller A. WordCloud: a little word cloud generator in Python. Python Package Index (PyPI). 2023.