ORIGINAL PAPER

DOI: http://dx.doi.org/10.15403/jgld-6740

Unsupervised Machine Learning of Inflammatory Bowel Disease Case Reports Reveals Decades of Evolving Research Themes

Naruaki Ogasawara

Editorial Department, The Japanese Society of Internal Medicine, Tokyo, Japan

Address for correspondence:
Naruaki Ogasawara
Editorial Department, The Japanese Society of Internal Medicine, 3-28-8 Hongo Bunkyo-Ward,
Tokyo 113-8433, Japan
n-ogasawara@naika.or.jp

Received: 03.12.2025
Accepted: 24.12.2025

ABSTRACT

Background & Aims: Inflammatory bowel disease (IBD) case reports provide rich longitudinal insights but have rarely been analyzed using quantitative text-mining approaches. This study applied unsupervised machine learning to PubMed-indexed IBD case reports to identify long-term thematic structures spanning 60 years and evaluate whether major historical milestones in IBD care can be reconstructed from biomedical texts.
Methods: Case reports indexed under the keyword "inflammatory bowel disease" were retrieved from PubMed (1960–2025). Titles, key words, and abstracts were concatenated and preprocessed before TF-IDF vectorization. Non-negative matrix factorization (NMF) was applied to extract latent topics, followed by KMeans clustering using the optimal topic number selected by silhouette evaluation (2–15 topics). Cluster characteristics were summarized using report counts and term frequency-inverse document frequency (TF-IDF) statistics. Top discriminative key words were used to assign data-driven topic labels. All analyses were performed in Python 3.10.5 (PyCharm 2022.1.3) using pandas, numpy, scikit-learn, matplotlib, and seaborn.
Results: A total of 18,458 case reports were analyzed. Across all time periods, two highly stable clusters consistently emerged, corresponding to Crohn's disease and ulcerative colitis. Early decades (1960–1989) emphasized pathology and complication-focused descriptions. Reports from the 1990s showed increasing terminology related to diagnosis and emerging therapies. From 2000 onward, infliximab-related and treatment-focused terms predominated, paralleling the rise of biology. After 2010, clusters reflected diversified therapeutic strategies, including attention to extraintestinal manifestations and biologic or small-molecule therapies.
Conclusions: Unsupervised machine learning successfully reconstructed important historical changes in IBD management, demonstrating that a large case report text corpus captures the evolution of clinical concepts and treatment paradigms over 60 years.

Key words: inflammatory bowel disease – IBD - machine learning – case reports – natural language processing – non-negative matrix factorization.

Abbreviations: AZA: azathioprine; CD: Crohn's disease; IBD: inflammatory bowel disease; ML: machine learning; TF-IDF: term frequency-inverse document frequency; NLP: natural language processing; NMF: non-negative matrix factorization; SASP: salazosulfapyridine; T2T: treat-to-target; UC: ulcerative colitis; 6-MP: 6-mercaptopurine.

INTRODUCTION

Treatment strategies for inflammatory bowel disease (IBD), including ulcerative colitis (UC) and Crohn's disease (CD), have evolved significantly over the past 60 years. From the 1960s to the 1980s, treatment focused on suppressing acute inflammation and controlling symptoms with corticosteroids and salazosulfapyridine (SASP) [1]. In the 1990s, immunomodulatory agents such as azathioprine (AZA) and 6-mercaptopurine (6-MP) were introduced, leading to a shift toward treatment aimed at steroid withdrawal and long-term remission maintenance [2, 3]. In 1998, biologics such as infliximab emerged, ushering in the era of targeted therapy [4]. In recent years, with the emergence of small-molecule drugs such as tofacitinib (2018) [5], in addition to the treat-to-target (T2T) strategy [6], emphasis on mucosal healing, and the diversification of biologics, treatment has become more diverse and precise, accelerating a paradigm shift toward personalized medicine.

While these changes have been well documented clinically in the literature, they have not been adequately documented quantitatively or scientifically. Case reports published in PubMed provide detailed real-world records and represent a valuable yet underutilized source of information. From 1960 to 2025, the PubMed database indexed 144,662 IBD-related publications (including 18,458 case reports), providing a unique opportunity to quantitatively track the evolving research topic.

The evolution of publication numbers is closely related to past treatment developments: 1960–1989: rapid increase during the corticosteroid/SASP era as the IBD concept became established; 1990–1999: only modest growth during the AZA/6-MP and introduction of the treatment pyramid; 2000–2009: sharp increase following infliximab approval (1998) and early biologic adoption; 2010–2024: The most accelerated growth, corresponding with T2T (STRIDE-I, 2015), vedolizumab, ustekinumab, and JAK inhibitors.

These patterns suggest that publication trends themselves may mirror therapeutic paradigm shifts. Despite the existence of over 18,000 case reports, few studies have investigated the feasibility of reconstructing known clinical milestones from large-scale biomedical texts using natural language processing (NLP) or unsupervised machine learning (ML). Furthermore, traditional reviews rely on expert interpretation and are unable to quantitatively assess research trends and changes in research focus over time.

In this study, NLP and unsupervised ML [term frequency-inverse document frequency (TF-IDF), non-negative matrix factorization (NMF), k-means, and silhouette-based selection] were used to analyze PubMed-indexed IBD case reports from 1960 to November 2025. This study's primary goal was to determine whether major historical milestones such as the corticosteroid/SASP era, the widespread use of immunomodulatory drugs, the introduction of infliximab, the emergence of T2T, and the arrival of JAK inhibitors (2018), can be reconstructed and quantitatively visualized using NLP and unsupervised ML methods. By leveraging six decades of case reports, this study provides a quantitative framework for mapping the evolution of IBD research.

METHODS

Data Collection and Preprocessing

This study collected case reports related to "inflammatory bowel disease" from the PubMed database in November 2025. Each case report contained an article title, keywords, and abstract. The data for analysis selection procedure included the following steps:

Initial Search: A total of 144,662 records were retrieved using the topic search term "inflammatory bowel disease"

Document Type Filtering: The dataset was further refined by applying the "Case Reports" filter, yielding a dataset of 18,458 records.

Title, keywords, and abstract were concatenated into a single text string per case report. The preprocessing pipeline handled missing values by replacing them with empty strings. All text data underwent conversion to lowercase and tokenization for subsequent analysis.

Text Vectorization Using Term Frequency-inverse Document Frequency

The study applied TF-IDF vectorization to transform textual data into numerical representations utilizing the TfidfVectorizer from the scikit-learn Python library [7], a widely used open-source ML tool in Python. The vectorization process removed common English stop words and excluded terms appearing in more than 95% of documents or fewer than 2 documents (Max_df = 0.95, Min_df = 2). This approach generated a sparse matrix that captured the relative importance of each term across the corpus [8] and has been effectively applied in emotional text classification tasks [9].

The TF-IDF procedure consisted of 1) for each case report, the article title, author-keywords, and abstract were concatenated into a single text document, thereby ensuring that both thematic descriptors and clinical narratives contributed to feature extraction. All text was converted to lowercase prior to analysis to maintain consistency across documents; 2) term frequency was calculated within each document to reflect the local importance of words, while inverse document frequency weighting was applied to down-weight terms commonly shared across many reports and to emphasize more discriminative clinical terminology; 3) the resulting TF-IDF matrix, in which rows corresponded to individual case reports and columns to weighted terms, served as the quantitative input for subsequent unsupervised topic modeling and clustering analyses.

Topic Modeling with Non-negative Matrix Factorization

The analysis implemented topic modeling employing NMF [10] with an optimal number of components (topics) determined by silhouette score evaluation ranging from 2 to 15. The optimal number of components (topics) was selected based on the average silhouette score across 10 independent runs for each topic count ranging from 2 to 15.

The silhouette score is an internal validity metric that measures the quality of clustering by balancing intra-cluster cohesion and inter-cluster separation [11]. For each document, this study compares the average distance within that cluster with the nearest other clusters. Values range from -1 to +1; higher values are better. After projecting TF-IDF documents into the NMF space, KMeans was applied to compute silhouette scores for different numbers of topics. Each configuration was repeated ten times with varying seeds, and the mean score was used to select the optimal value. Because the silhouette score alone may oversimplify thematic structure and risk information loss, the Calinski–Harabasz and Davies–Bouldin indices were additionally employed to assess topic stability across seeds [12].

Clustering with KMeans

The study applied KMeans clustering [13], a widely used unsupervised learning algorithm based on iterative centroid optimization, to the NMF feature matrix to categorize case reports based on their topic distributions. The number of clusters was set equal to the optimal number of topics identified through NMF. The clustering was performed multiple times with random initialization to assess stability. Each report received cluster label assignment, and visualization displayed the distribution of reports across clusters. The clustering algorithm was executed with random initialization (i.e., random state not fixed), meaning results may vary across runs.

Statistical Characterization of Clusters

The analysis calculated the number of reports, mean, and standard deviation of the NMF feature values for each cluster. These metrics provided insights into the internal consistency and density of individual clusters. The study compiled all results into summary tables and exported them as Excel files for further examination.

Validation of Clustering Quality

Clustering quality was evaluated using three established metrics: Silhouette Score [11], Calinski-Harabasz Index [14], and Davies-Bouldin Index [15].

Reproducibility Assessment of Topic Modeling

The analysis conducted five independent runs of the NMF topic modeling process using different random seeds to evaluate model stability. For each topic, the study compared the top 10 keywords across all runs and calculated their overlap with the original keyword set. The average keyword overlap across runs served as a reproducibility score, indicating the consistency of topic identification. Note that this study used the simple number of keyword matches (overlap count) to evaluate NMF reproducibility and does not utilize ratio-based metrics such as the Jaccard Index or percentage overlap.

RESULTS

IBD Publication Trends (1960–2025)

A total of 18,458 IBD-related case reports were identified in PubMed between 1960 and 2025. The annual publication volume showed four distinct phases: a moderate increase from the 1960s to the 1980s, a moderate increase in the 1990s, a notable increase in the 2000s following the introduction of infliximab (1998), and the most rapid increase since 2010, consistent with the adoption of the T2T (2015) strategy and the diversification of biologics and small molecule drugs (Fig. 1, Table I).

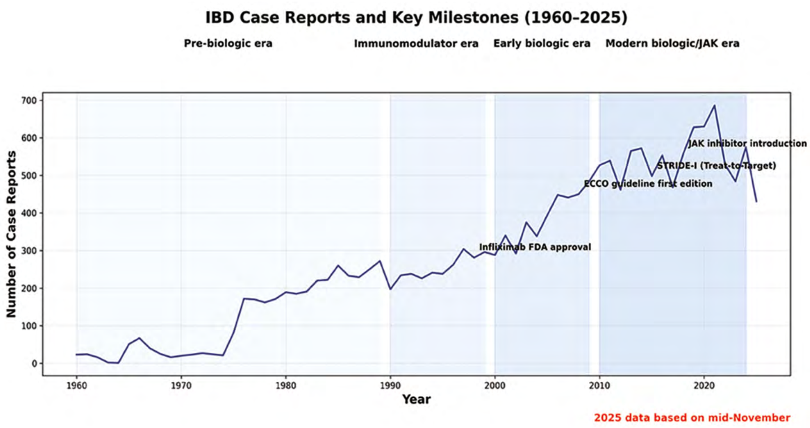

Fig. 1. IBD case reports and key milestones (1960-2025).

Table I. Evolution of IBD treatment strategies (1960–2025)

Era	Main therapeutic agents	Treatment strategy goals	Key words	Major events / Guidelines
1960–1989	Steroids, Sulfasalazine (SASP)	Rapid suppression of acute inflammation, symptom control	Symptomatic therapy, non-specific anti-inflammatory	Establishment of IBD disease concept; treatment approaches were empirical.
1990–1999	Azathioprine (AZA), 6-MP, Methotrexate (MTX)	Steroid-free therapy, maintenance of long-term remission	Immunomodulators, long-term management	Introduction of treatment pyramid concept; initial consensus on hierarchical treatment; Infliximab approved by FDA (1998).
2000–2009	Infliximab (anti-TNF-α antibody)	Start of targeted therapy, disease modification	Biologics, TNF-α inhibition	Approval of infliximab in Japan (2002); ECCO guideline project launched (2006); dramatic improvement in treatment outcomes.
2010–2025	Adalimumab, Vedolizumab, Ustekinumab, JAK inhibitors	Treat-to-target (T2T), mucosal healing, personalized medicine	Diversification of targets, small molecules, pathway specificity	STRIDE-I (2015) introducing T2T strategy; systematic ECCO guidelines; JAK inhibitors introduced (Tofacitinib: 2018, oral targeted therapy).

Topic Structure Identified by NLP and Unsupervised ML

The analysis consistently generated two major themes corresponding to CD and UC across all four time periods (Figs. 2 and 3). K-means clustering (k=2), selected based on the Silhouette and Davies-Bouldin indices, reproduced this dichotomy with high internal consistency. The CD cluster included terms such as Crohn's disease, bowel, inflammatory and infliximab, while the UC cluster included terms such as ulcerative colitis, chronic, associated, gangrenosum and pyoderma. These core terms have been maintained over decades, reflecting the stability of the clinical disease framework (Table II).

Era-Specific Shifts in Topic Patterns

1960–1989

Case reports from this era primarily described the disease and pathology of IBD. This era's IBD case reports clearly distinguished between CD and UC. In Cluster 1, key words such as "Crohn," "disease," and "bowel" indicated many reports on the lesion site (small intestine and large intestine) of CD. In

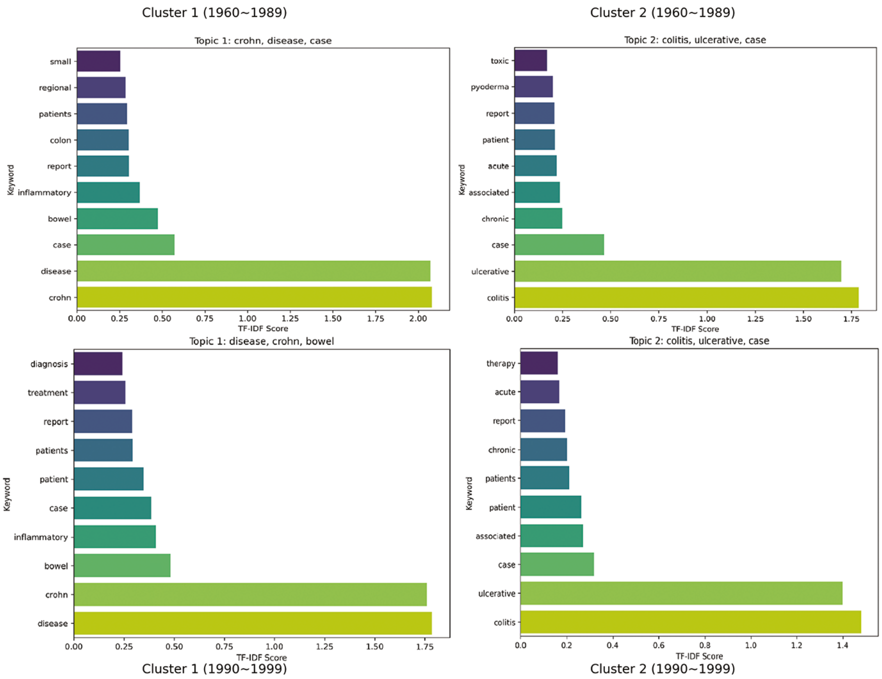

Fig. 2. Analysis of IBD case reports using TF-IDF scores cluster 1 and 2 (1960-1989) (1990-1999).

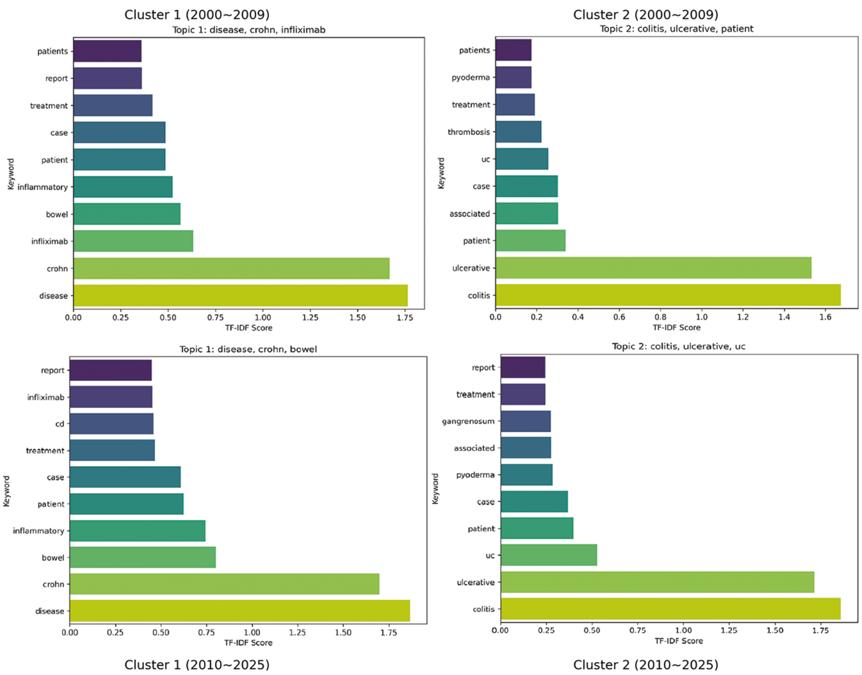

Fig. 3. Analysis of IBD case reports using TF-IDF scores cluster 1 and 2 (2000-2009) (2010-2025).

Table II. Summary of topic modeling and clustering results across four time periods (1960–2025)

A. Cluster Statistics

Period	Cluster	Topic Label	Num Reports	TF-IDF Mean	TF-IDF SD
1960–1989	1	Crohn, disease, case	2,487	0.0286	0.0316
	2	colitis, ulcerative, case	900	0.0526	0.0534
1990–1999	1	disease, Crohn, bowel	688	0.0553	0.0529
	2	colitis, ulcerative, case	1,829	0.0331	0.0323
2000–2009	1	disease, Crohn, infliximab	2,893	0.0286	0.0277
	2	colitis, ulcerative, patient	957	0.0514	0.0472
2010–2025	1	disease, Crohn, bowel	6,634	0.0226	0.0213
	2	colitis, ulcerative, uc	2,070	0.0423	0.0360

B. Top 10 Key words per Topic (TF-IDF Score)

Period	Topic	Top Key words (TF-IDF Score)
1960–1989	1	Crohn (2.08), disease (2.07), case (0.57), bowel (0.47), inflammatory (0.37), report (0.31), colon (0.30), patients (0.29), regional (0.28), small (0.25)
	2	colitis (1.79), ulcerative (1.70), case (0.47), chronic (0.25), associated (0.24), acute (0.22), patient (0.21), report (0.21), pyoderma (0.20), toxic (0.17)
1990–1999	1	disease (1.78), Crohn (1.76), bowel (0.48), inflammatory (0.41), case (0.39), patient (0.35), patients (0.29), report (0.29), treatment (0.26), diagnosis (0.24)
	2	colitis (1.48), ulcerative (1.40), case (0.32), associated (0.27), patient (0.26), patients (0.21), chronic (0.20), report (0.19), acute (0.17), therapy (0.16)
2000–2009	1	disease (1.76), Crohn (1.67), infliximab (0.63), bowel (0.56), inflammatory (0.52), patient (0.49), case (0.49), treatment (0.42), report (0.36), patients (0.36)
	2	colitis (1.67), ulcerative (1.53), patient (0.34), associated (0.30), case (0.30), uc (0.26), thrombosis (0.22), treatment (0.19), pyoderma (0.18), patients (0.17)
2010–2025	1	disease (1.87), Crohn (1.70), bowel (0.80), inflammatory (0.75), patient (0.63), case (0.61), treatment (0.47), cd (0.46), infliximab (0.45), report (0.45)
	2	colitis (1.86), ulcerative (1.72), uc (0.53), patient (0.40), case (0.37), pyoderma (0.29), associated (0.28), gangrenosum (0.27), treatment (0.25), report (0.24)

C. Clustering Validity Metrics

Period	Optimal k	Silhouette	Calinski–Harabasz	Davies–Bouldin
1960–1989	2	0.5501	3861.97	0.6678
1990–1999	2	0.5776	3687.26	0.6052
2000–2009	2	0.6209	6954.16	0.5563
2010–2025	2	0.6184	15347.02	0.5804

D. Reproducibility (Average Overlap Count)

Period	Topic	Average Overlap Count
1960–1989	1	10
	2	10
1990–1999	1	10
	2	10
2000–2009	1	10
	2	10
2010–2025	1	10
	2	10

Cluster 2, key terms such as “colitis,” “ulcerative,” “chronic,” and “acute” were used, indicating a focus on the pathology and severe complications of UC (toxic megacolon, skin symptoms) (Fig. 2).

1990–1999

The analysis results from this period suggest that case reports on IBD were undergoing a turning point in treatment strategies. Cluster analysis revealed frequent use of keywords such as “treatment,” “therapy,” and “diagnosis,” revealing that in addition to traditional disease concepts and descriptions of pathology, there was increased discussion of diagnosis and treatment. These trends indicate a shift from the pathology-centered reports of 1960-1989 and reflect the characteristics of the era in which treatment options were being explored prior to the introduction of biologics from the 2000s onward (Fig. 2).

2000–2009

There was a notable increase in infliximab and treatment-related terms in case reports during this period. This coincides with the introduction and rapid expansion of biological therapy. In UC, terms such as "treatment," "thrombosis," and "pyoderma" indicated that treatment strategies and complication management in UC were key themes. These trends indicate a shift in the focus of case reports from disease concepts to examinations of treatment efficacy and safety (Fig. 3).

2010–2025

Case reports from this period highlight the diversification of therapeutic strategies for IBD and the growing emphasis on complication management. Cluster 1 is characterized by the presence of "infliximab" and "treatment," reflecting the sustained use of biologic therapies, particularly anti-TNF agents. Cluster 2 includes terms such as "treatment," "pyoderma," and "gangrenosum," indicating that management of extraintestinal manifestations, especially dermatologic complications, was a major focus. These trends align with contemporary treatment paradigms, including T2T strategies, mucosal healing objectives, and the introduction of JAK inhibitors (Fig. 3).

Cluster Validity and Reproducibility

Model output demonstrated high reproducibility. All clustering runs yielded identical topic structures (overlap 10/10), with silhouette scores consistently above 0.55 and low Davies–Bouldin indices. The robustness of these metrics indicates that the extracted themes were not sensitive to sampling variation, supporting the reliability of the comparisons (Table II).

DISCUSSION

This study demonstrated that NLP and unsupervised ML can reconstruct research themes in IBD over a 60-year period using case reports indexed in PubMed. Despite the heterogeneity of case reports, this study's analytical approach, combining TF-IDF, NMF-based topic modeling, k-means clustering, and silhouette-driven cluster selection, was able to capture clinically meaningful differences and the evolution of long-term themes aligned with major events in IBD research.

This approach extends prior bibliometric and text-mining studies by focusing exclusively on case reports and by explicitly evaluating topic stability and reproducibility across multiple random initializations.

IBD case reports from 1960 to 1989 focused on describing pathology and lesion location, reflecting early efforts to define IBD pathology. In the 1990s, with the introduction of immunomodulatory agents, terminology shifted toward diagnosis and treatment, the feature not present in the 1960-1989 terminology (Fig. 2, Table II). In the 2000s, the introduction of biologics and increased interest in their therapeutic efficacy and complications led to a significant increase in infliximab-related terminology. IBD case reports from 2010 to 2025 emphasized the diversification of treatments, including new biologics, small molecules, targeted therapeutic strategies, and T2T, and detailed descriptions of extraintestinal manifestations such as pyoderma gangrenosum were also included. These trends reflect the contemporary shift toward personalized medicine and comprehensive comorbidity management.

Two reproducible topic structures emerged, corresponding to CD and UC. The stability of these clusters across iterations suggests that linguistic distinctions between CD and UC are deeply embedded in clinical reporting, despite changes in terminology, therapeutic approaches, and publication practices over time.

High silhouette scores, low Davies–Bouldin indices, and reproducibility of top key words indicate robust topic cohesion. The alignment of computationally derived transitions with clinical milestones-such as immunomodulators in the 1990s and biologics in the 2000s-supports the validity of this approach. Large-scale text mining can complement traditional reviews by revealing longitudinal patterns not easily captured manually.

This study used NLP and unsupervised ML, a technology for extracting meaning from text, to analyze a large number of IBD case reports [16]. It demonstrates its potential for practical application in clinical practice and research, going beyond simple bibliometric analysis [17, 18]. Case reports can contain rare complications, unexpected treatment responses, and signs of safety that appear earlier than in large-scale clinical trials. By structuring this information using NLP and unsupervised ML, the methods can provide a mechanism for discovering important patterns that would be difficult to find through traditional manual analysis or the analysis of a small number of case reports [19].

From a clinical perspective, the temporal changes demonstrated in this study, such as the rapid increase in case reports in the 2000s following the approval of infliximab in 1998, the development of guidelines and treat-to-target strategies since 2010, and the introduction of JAK inhibitors in 2018-reflect major paradigm shifts in IBD management [3, 4]. These findings suggest that NLP and unsupervised ML methods can support clinical decision-making by automatically structuring medical information. For example, it may be possible to detect side effects, including immune-related adverse events of new drugs or complications associated with specific treatments before they are fully incorporated into official guidelines [20, 21].

From a translational perspective, this analysis could help inform future diagnosis and treatment by providing data on how new drugs, treatment strategies, and disease characteristics-such as symptoms and severity—affect clinical practice and identify key areas of focus [22, 23]. For example, trends observed in large volumes of case reports could be used to identify groups with higher rates of side effects or unusual symptoms, which could be used to monitor drug safety and develop hypotheses for research and clinical trials [24, 25]. In fact, similar analysis methods have been used to analyze symptoms, predict treatment outcomes, and check safety in other chronic diseases, demonstrating their broad applicability.

Future refinements to the current methodology could further enhance the usefulness of this approach: first, by incorporating full-text articles rather than just abstracts, and second, by integrating case reports published in databases other than PubMed database.

Several recent studies have applied NLP and unsupervised ML techniques to biomedical literature, including that related to IBD and other chronic diseases. For example, Li et al. [26] used latent Dirichlet allocation (LDA) in 2022 to identify thematic trends in IBD-related publications but focused primarily on original articles rather than case reports. Although several studies have analyzed gastroenterology literature, most rely on bibliometric or network-analysis approaches rather than NLP-based topic modeling [27, 28].

In contrast, this study appears to be one of the first to systematically apply TF-IDF-based NMF topic modeling and k-means clustering specifically to IBD case reports spanning several decades. By incorporating silhouette-driven model selection, multiple validation metrics, and keyword overlap-based reproducibility assessment, our approach extends previous work methodologically and analytically.

Furthermore, previous studies have treated medical literature as a uniform corpus [29, 30]. However, this study showed that while case reports have diverse writing styles, they consistently maintain disease-specific language patterns (CD and UC), while at the same time reflecting changes in treatment methods and concepts over time. This dual pattern of stability and plasticity has not been clearly demonstrated in previous IBD-related text mining studies.

In recent years, studies using ML have begun to appear in internal medicine journals [31], reflecting growing interest in data-driven approaches. This study's methodological framework offers a practical approach for monitoring long-term scientific trends, identifying emerging clinical challenges, and supporting literature surveillance in rapidly advancing fields. Although limited by its reliance on PubMed data and abstract-level information, the approach remains robust, reproducible, and adaptable to other chronic diseases. Nevertheless, the findings of this study should be interpreted considering several limitations.

This study has several limitations. First, the data used was only from PubMed, so case reports published in region-specific journals may not be included. Second, all case reports were given the same weight regardless of their quality or number of cases, so ambiguous diagnoses may have affected the classification of topics. Furthermore, only paper titles, key words, and abstracts were analyzed, and detailed clinical information was not included. It would be desirable to utilize full-text data in the future.

CONCLUSIONS

This study applied NLP and unsupervised ML to systematically characterize research themes in PubMed-indexed case reports on IBD. The results revealed stable diagnostic structures alongside distinct era-specific shifts. By consistently identifying two core topics corresponding to CD and UC, the analysis highlights how linguistic patterns in clinical reporting have preserved disease fundamentals while reflecting major therapeutic milestones, from pathologycentered descriptions in the 1960s–1980s, to diagnosis, and treatment-oriented terminology in the 1990s, to biologic-driven advances in the 2000s, and the diversification of targeted therapies and complication management after 2010. These findings show that large-scale text mining can uncover transitions that parallel the real-world evolution of IBD care. Future work incorporating full-text data and broader databases may further enhance precision, enabling more comprehensive assessments of how clinical language evolves alongside medical innovation.

Conflicts of interest: None to declare.

Author's contributions: N.O. conceived and designed the study, collected and analyzed the data, interpreted the results, and drafted the manuscript. The author revised the manuscript, approved the final version to be published, and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

REFERENCES

1. Manu P, Rogozea LM, Dumitrașcu DL. Pharmacological Management of Inflammatory Bowel Disease: a Century of Expert Opinions in Cecil Textbook of Medicine. Am J Ther. 2022;29(5):e500–e506. doi: 10.1097/MJT.0000000000001555
2. de Boer NKH, Peyrin-Biroulet L, Jharap B, Sanderson JD, Meijer B, Atreya I, et al. Thiopurines in Inflammatory Bowel Disease: New Findings and Perspectives. J Crohns Colitis. 2018;12(5):610–620. doi: 10.1093/ecco-jcc/jjx181
3. Magro F, Cordeiro G, Dias AM, Estevinho MM. Inflammatory Bowel Disease - Non-biological treatment. Pharmacol Res. 2020;160:105075. doi: 10.1016/j.phrs.2020.105075
4. Nadpara N, Reichenbach ZW, Ehrlich AC, Friedenberg F. Current Status of Medical Therapy for Inflammatory Bowel Disease: The Wealth of Medications. Dig Dis Sci. 2020;65(10):2769–2779. doi: 10.1007/s10620-020-06471-4
5. Jefremow A, Neurath MF. Novel Small Molecules in IBD: Current State and Future Perspectives. Cells. 2023;12(13):1730. doi: 10.3390/cells12131730
6. West J, Tan K, Devi J, Macrae F, Christensen B, Segal JP. Benefits and Challenges of Treat-to-Target in Inflammatory Bowel Disease. J Clin Med. 2023;12(19):6292. doi: 10.3390/jcm12196292
7. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825–2830.
8. Manning CD, Raghavan P, Schütze H. Introduction to Information Retrieval. Cambridge: Cambridge University Press; 2008.
9. Alfarizi A, et al. Emotional Text Classification Using TF-IDF and LSTM. Jurnal Informatika. 2022.
10. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature. 1999;401(6755):788–791. doi: 10.1038/44565
11. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53–65. doi: 10.1016/0377-0427(87)90125-7
12. Vardakas G, Pavlopoulos J, Likas A. Revisiting Silhouette Aggregation. arXiv [Preprint]. 2024. arXiv:2401.05831.

13. Lloyd S. Least squares quantization in PCM. IEEE Trans Inf Theory. 1982;28(2):129–137. doi: 10.1109/TIT.1982.1056489
14. Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat. 1974;3(1):1–27. doi: 10.1080/03610927408827109
15. Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell. 1979;PAMI-1(2):224–227. doi: 10.1109/TPAMI.1979.4766909
16. Lasko TA, Denny JC, Levy MA. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. PLoS One. 2013;8(6):e66341. doi: 10.1371/journal.pone.0066341
17. Jensen P, Jensen L, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet. 2012;13(6):395–405. doi: 10.1038/nrg3208
18. Harpaz R, Callahan A, Tamang S, Low Y, Odgers D, Finlayson S, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. Drug Saf. 2014;37(10):777–790. doi: 10.1007/s40264-014-0218-z
19. Lieber T, Gosselt HR, Kools PC, Kruijssen OC, Van Lierop SNC, Härmark L, et al. Natural language processing for automated triage and prioritization of individual case safety reports for case-by-case assessment. Front Drug Saf Regul. 2023;3:1120135. doi: 10.3389/fdsfr.2023.1120135
20. Golder S, Xu D, O'Connor K, Wang Y, Batra M, Hernandez GG. Leveraging Natural Language Processing and Machine Learning Methods for Adverse Drug Event Detection in Electronic Health/Medical Records: A Scoping Review. Drug Saf. 2025;48(4):321–337. doi: 10.1007/s40264-024-01505-6
21. Eguia H, Sánchez-Bocanegra CL, Vinciarelli F, Alvarez-Lopez F, Saigí-Rubió F. Clinical Decision Support and Natural Language Processing in Medicine: Systematic Literature Review. J Med Internet Res. 2024;26:e55315. doi: 10.2196/55315
22. Baxi V, Edwards R, Montalto M, Saha S. Digital pathology and artificial intelligence in translational medicine and clinical practice. Mod Pathol. 2022;35(1):23–32. doi: 10.1038/s41379-021-00919-2
23. Bjerrum JT, Wang YL, Seidelin JB, Nielsen OH. IBD metabonomics predicts phenotype, disease course, and treatment response. EBioMedicine. 2021;71:103551. doi: 10.1016/j.ebiom.2021.103551
24. Fusaroli M, Raschi E, Poluzzi E, Hauben M. The evolving role of disproportionality analysis in pharmacovigilance. Expert Opin Drug Saf. 2024;23(8):981–994. doi: 10.1080/14740338.2024.2368817
25. Salas M, Petracek J, Yalamanchili P, Aimer O, Kasthuril D, Dhingra S, et al. The Use of Artificial Intelligence in Pharmacovigilance: A Systematic Review of the Literature. Pharmaceut Med. 2022;36(5):295–306. doi: 10.1007/s40290-022-00441-z
26. Li K, Feng C, Chen H, Feng Y, Li J. Trends in Worldwide Research in Inflammatory Bowel Disease Over the Period 2012–2021: A Bibliometric Study. Front Med (Lausanne). 2022;9:880553. doi: 10.3389/fmed.2022.880553
27. Ogasawara N. Collaborative networks in gastroenterology research: a co-authorship network analysis (2000–2023). Arq Gastroenterol. 2024;61:e24083. doi: 10.1590/S0004-2803.24612024-083
28. Ogasawara N. Evolving trends and emerging themes in gut microbiota research: a comprehensive bibliometric analysis (2015–2024). Arq Gastroenterol. 2025;62:e25023. doi: 10.1590/S0004-2803.24612025-023
29. Zweigenbaum P, Jacquemart P, Grabar N, Habert B. Building a text corpus for representing the variety of medical language. Stud Health Technol Inform. 2001;84(Pt 1):290–294.
30. Shaikina AA, Funkner AA. Medical corpora comparison using topic modeling. Procedia Comput Sci. 2020;178:244–253. doi: 10.1016/j.procs.2020.02.034
31. Ogasawara N, Matsunaga K, Isomoto H, Shimizu W. Internal Medicine Year in Review 2023. Intern Med. 2024;63(23):3137–3140. doi: 10.2169/internalmedicine.4396-24