INDUSTRIAL AND SYSTEMS ENGINEERING

ISE

Uncovering Thematic Correlations Across Transportation Research Journal Series: Pitting Human Expertise Against Machine Intelligence

MENG ZHAO[1], SHIJIE CHEN[2], AND YANSHUO SUN[2]

[1]Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, 18015 USA

[2]Department of Industrial and Manufacturing Engineering, FAMU-FSU College of Engineering, Florida State University, Tallahassee, FL 32310 US

ISE Technical Report 26T-002

LEHIGH UNIVERSITY

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Uncovering thematic correlations across transportation research journal series: Pitting human expertise against machine intelligence

Meng Zhao[a], Shijie Chen[b], Yanshuo Sun[b,*]

[a] Department of Industrial and Systems Engineering, College of Engineering and Applied Science, Lehigh University, US
[b] Department of Industrial and Manufacturing Engineering, FAMU-FSU College of Engineering, Florida State University, US

ARTICLE INFO

Keywords:
Text classification
Transportation research topics
Journal scope
Inter-journal relations

ABSTRACT

Transportation research plays a significant role in addressing complex societal challenges. The Transportation Research (TR) journal series, comprising six specialized parts A-F, mirrors the thematic breadth of transportation research but presents challenges for certain researchers in clearly delineating thematic boundaries between journals, which could lead to manuscript misalignment and unfortunate desk rejections. Despite the significance of these journals, a systematic analysis of thematic overlaps and distinctions across the TR series using text classification methods remains unexplored. To fill this gap, this study first applies the BERTopic model on 16,341 TR abstracts between 2010 and 2024 to derive the topic distribution of each journal. Three machine learning classifiers and one deep learning algorithm are then trained to classify abstracts accurately into the appropriate TR journal part. Additionally, the journal relationships are analyzed using novel quantitative metrics. A survey inviting 2400 active transportation researchers is conducted to understand the classification performance of human experts. The study finds significant thematic overlaps, especially between TR-B and TR-C, predominantly around driving safety and traffic control, whereas TR-F emerges with highly distinctive thematic clarity. The support vector classifier (SVC) achieves the highest accuracy. When evaluated with the same testing dataset, the SVC significantly outperforms human experts according to the survey results. We publish our machine learning-driven classification tool, which can be used in manuscript submission processes to enhance the accuracy of journal selection.

1. Introduction

Transportation infrastructure and systems are fundamental to economic prosperity and societal well-being, facilitating mobility, trade, and access to essential services. Given their critical role, transportation research has evolved into a highly interdisciplinary field, integrating knowledge from engineering, economics, business, social sciences, and public administration (Sun and Kirtonia, 2020). Among the various academic publication outlets dedicated to transportation research, the *Transportation Research* (TR) journal series, published by Pergamon-Elsevier Science, stands out as one of the most cohesive and comprehensive references in the field (ScienceDirect, 2025a). Established in 1967 as a single journal, the TR has since expanded into six specialized parts (Parts A-F), each addressing distinct dimensions of transportation research (Modak et al., 2019). This decentralized structure enables TR journal series

* Corresponding author.
E-mail addresses: m.zhao@lehigh.edu (M. Zhao), sc20hw@fsu.edu (S. Chen), y.sun@eng.famu.fsu.edu (Y. Sun).

https://doi.org/10.1016/j.tra.2026.104901
Received 8 July 2025; Received in revised form 19 January 2026; Accepted 29 January 2026

to effectively respond to the increasing complexity and diversity of transportation studies, making it a central platform for advancing knowledge and informing transportation policy and practice.

Each part of TR journal series operates independently with a dedicated scope designed to complement the others. For instance, Transportation Research Part B (TR-B) focuses on methodological advancements in transportation systems planning and analysis, emphasizing problem formulation and solution techniques (ScienceDirect, 2025b). By contrast, Transportation Research Part C (TR-C) priorities research on emerging technologies, including novel control algorithms and innovative sensing techniques for transportation systems (ScienceDirect, 2025c). Transportation Research Part D (TR-D), on the other hand, investigates the interactions between transportation and environmental factors, particularly in the context of sustainability and policy implications (ScienceDirect, 2025d). Despite these well-defined scopes, thematic overlap among the six TR parts frequently occurs, making it challenging for early-career researchers to capture the nuanced differences and determine the most appropriate journal for submission. For example, a study on data-driven optimization for e-scooter sharing services could align with TR-B from a methodological standpoint, while also fitting TR-C given its technological implications. Similarly, a paper on low-carbon transportation policies might fall within the scope of either TR-A or TR-D.

Given the complexity of the thematic landscape across the TR journal series, the need for a systematic analysis of their topic structures and interrelationships becomes evident. Such an analysis could guide researchers in selecting the most appropriate journal for their work while offering insights into the evolving landscape of transportation research. A practical benefit is to avoid desk rejection due to scope mismatch (Dwivedi et al., 2022), to the maximum extent. In this context, text classification serves as a powerful tool. If an advanced text classification algorithm can accurately yield the appropriate TR journal for a given paper using metadata such as the title and abstract, it would support the hypothesis that the TR journals maintain complementary scopes. Conversely, poor classification performance may indicate significant thematic overlap, thereby reducing the clarity of the journal boundaries. Meanwhile, topic modeling offers a dual advantage in this regard; it uncovers latent thematic structures in large text corpora, and the identified topics can serve as informative features for text classification tasks. By using topic distributions as input features, classifiers can move beyond surface-level keywords to capture deeper semantic patterns within documents (Rijcken et al., 2022). Although recent studies have combined text classification and topic modeling to extract structured insights from unstructured texts across various domains (Tzika-Kostopoulou et al., 2024; Rijcken et al., 2022; Sayed et al., 2023), none of them have addressed the challenge of distinguishing and comparing the thematic scopes of closely related journals, let alone the TR series of utmost significance in transportation systems research. Currently, classification practices for transportation publications primarily rely on expert judgment; experienced researchers, editors, and reviewers use their experience and domain knowledge to assess whether a manuscript aligns with a journal's scope. However, it remains an open question whether machine intelligence, trained on extensive publication data, can match or even surpass human expertise in assigning manuscripts to the most appropriate journal.

To address this gap, our study leverages large-scale bibliographic data and combines an advanced topic modeling approach and machine learning classifiers to evaluate the thematic correlations among the six TR journals. Specifically, this study includes three main stages. First, we collect and preprocess the bibliographic data from the six TR journals and extract topic representations using a BERTopic model. Second, we train three machine learning classifiers and an advanced deep learning algorithm to assign each abstract to the most appropriate TR journal; based on the classification results, we propose multiple novel quantitative metrics to assess thematic similarities across the TR journals. Finally, we conduct an online survey targeting transportation researchers to evaluate human classification performance and compare it against the best known machine classification approach. Through this study, we aim to answer the following key research questions:

1. How is the thematic structure distributed across the TR journals?
2. Which TR journals exhibit significant topical overlap with each other, and which ones demonstrate distinctive thematic focuses?
3. How accurately can a well-trained text classification model predict the appropriate journal for a given abstract, compared to domain experts?

As a result, our study delivers five key outcomes. First, by applying a BERTopic model to 16,341 abstracts, we distilled 50 research topics and revealed each TR journal's topical focus. Second, a support vector classifier (SVC) reached the highest accuracy of 0.67 in assigning abstracts to the correct TR part; the corresponding precision and recall metrics reflected TR-A's broader scope and TR-B's specialized focus. Third, our proposed metrics quantified the overlap between TR-B and TR-C, the uniqueness of TR-F, and the distinctiveness between TR-C and TR-D. "Driving safety" was identified as the most confusing topic between TR-B and TR-C, while "Air travel" was one of the most distinctive topics between TR-C and TR-D. Fourth, among the 192 respondents in our survey, 21.4% and 85.9% reported editorial and reviewing experience, respectively. However, domain experts achieved a subpar accuracy of 0.38 in our classification test. Using the same testing dataset, SVC, ChatGPT 4o, and a deep learning algorithm TabNet achieved accuracies of 0.62, 0.62, and 0.59, respectively, which all outperformed the respondents. Fifth, we have deployed our approach into a publicly accessible Streamlit application that offers real-time journal recommendations based on manuscript abstracts.

The remainder of this paper is organized as follows. Section 2 presents an overview of the related works in the fields we are focusing upon, including journal relationship identification, topic modeling for text classification, as well as the applications of text classification in transportation. Section 3 describes the collection and processing of article data, and survey design. Section 4 presents the methods, including the topic modeling approach, classification methods, and similarity evaluation metrics we proposed. Section 5 presents the results of our analyses, including topic distributions, machine learning classification performance, journal thematic correlations, and human expert performance. In Section 6, we analyze the implications, provide suggestions, and outline limitations of this study. Section 7 concludes with a summary of key findings and suggestions for future research directions.

2. Literature review

To provide a comprehensive background for this study, the literature review is organized into three main areas. First, we review studies that analyze inter-journal relationships. Second, we review methodological developments in topic modeling, especially its applications in text classification across diverse domains. Finally, we focus on the application of text classification in the transportation research literature.

2.1. Identifying inter-journal relationship

There are multiple ways to characterize how various academic journals interact with each other. For instance, competition, as one type of relationships, arises when journals with similar scopes vie for the same high-impact manuscripts, readership, and citation metrics. Globally, *Nature* and *Science* compete to publish landmark discoveries across the sciences, often issuing competing reports on the same breakthrough within weeks. Another relationship is propagation (influence), which reflects how methodological or conceptual advances in one venue spawn follow-on studies elsewhere. In scientometrics, the pioneering use of altmetrics in *PLOS ONE* inspired journals such as *Journal of Altmetrics* and *Quantitative Science Studies* to explore alternative impact indicators. Varin et al. (2016) modeled cross-citation flows among statistics journals to identify venues that function as central sources of knowledge for the field. The third relationship is specialization (lineage), which captures the creation of spin-off or niche titles from established parents. While there are many examples, in the transportation field, ASCE's *Journal of Transportation Engineering* was reorganized in 2019 into *Part A: Systems* and *Part B: Pavements* to address distinct subdisciplinary needs.

Quantitative analysis of citation flows, co-citation clusters, and temporal publication patterns enables the detection and measurement of these competitive, influential, and genealogical ties to reveal how journals co-evolve and adapt to shifting research landscapes. Most studies in this area use citation-based metrics to capture these relationships. One such metric is h-similarity (Schubert, 2010), which quantifies the similarity between two journals based on the overlap in their cited references. By analyzing the similarity in the journals they cite, h-similarity provides a simple yet effective way to assess how closely related two journals are in terms of their research focus. Colavizza et al. (2018) quantified the similarities of articles cited in four journals by four different indicators, which showed that the articles cited in *Cell* and *European Journal of Operational Research* have the highest and lowest textual similarities, respectively. Besides, the similarity monotonically increased when the cocitation level gets lower, from journal level to bracket level, in the four journals. Katchanov and Markova (2017) analyzed citation distributions for 240 physics journals and identify clusters of closely related titles, revealing how journals co-evolve and differ in influence within their field. A time-heterogeneous log-multiplicative model was built to estimate the cohesion and structural equivalence of journals. As cohesion represented the mutual citation relationships, multiple tight cliques of journals were identified, such as the clusters for management-oriented, method-oriented, and psychology-oriented journals. The structural equivalence analysis discovered the journals that cited the same source journals and that were cited by the same destination journals to share a similar knowledge base. Then, a few feeder journals and knowledge-transfer journals were identified.

In addition, citing discipline analysis emerged to assess the similarity of journals within and across disciplines by further taking discipline profiles into account. Wolfram and Zhao (2014) developed a similarity comparison model that calculated a cocitation frequency matrix that involved the category assignments provided by Web of Science (WoS), and then applied Multidimensional scaling (MDS) analysis and hierarchical cluster analysis. This model clustered 120 journals in five allied and one distinct subjects over three time periods. While the clustering results were impacted by the WoS assignment, some journals were found to be misclassified. The boundaries of certain regions were ambiguous. This procedure was expanded in the field of library and information science by Wang and Wolfram (2015), which grouped forty journals in clusters. The evolution of clustering results was demonstrated. Certain journals exhibited the tendency to bridge two distinct disciplinary areas or gradually shift towards a different discipline over time.

Lastly, economic indices have recently been applied to analyze journal relationships. Wandelt et al. (2025) conducted a comprehensive bibliometric review of transportation journals. They first applied word-frequency counts on abstracts and then reduced dimensionality using t-SNE. Next, they performed k-means clustering to uncover 37 thematic clusters. Journal homogeneity was quantified via the Herfindahl-Hirschman Index. However, their approach differs from ours in two primary ways. First, their approach is fully unsupervised, with a primary focus on analyzing annual paper counts, co-authorship patterns, and journal scopes. By contrast, our focus is on analyzing journal relationships through supervised classifications and our newly developed metrics, which utilize the true journal labels. Second, while Wandelt et al. (2025) used t-SNE and k-means clustering to identify thematic clusters, we use BERTopic, a dedicated topic modeling method, to capture the topic distribution of each journal. This topic information is further integrated into our classification approach.

Overall, multiple citing network-based, discipline-based, or economic metrics were applied to assess the similarity of journals and identify their relationships. However, none of them have considered the textual and semantic information by leveraging topic modeling and text classification methods.

2.2. Topic modeling for text classification

Topic modeling has been an effective technique for uncovering latent thematic structures in text corpora over the last decades (Bi et al., 2024). The early topic modeling was laid by probabilistic approaches, which sought to model the generative processes underlying textual data. Probabilistic Latent Semantic Indexing (pLSI) was one of the first methods to explicitly model the probability distributions of topics within documents. However, pLSI lacked a generative process for new documents, limiting its applicability. To address the limitation, Latent Dirichlet Allocation (LDA) was proposed, assuming that documents are mixtures of topics and topics are distributions over words (Blei et al., 2003). After LDA, Non-Negative Matrix Factorization (NMF) emerged as a deterministic alternative, leveraging matrix factorization techniques to identify latent topics (Kuang et al., 2015). While NMF offered simplicity and interpretability, it lacked the probabilistic grounding of LDA. More recently, the advent of transformer-based language models, such as BERT (Devlin et al., 2019), has revolutionized topic modeling. BERTopic, introduced in 2022, has been the most advanced topic modeling approach (Grootendorst, 2022). It utilizes contextual embeddings generated from pre-trained language models to capture richer semantic representations of text and operates mainly in three steps. First, it converts documents into embeddings using Sentence-BERT (Reimers and Gurevych, 2019); second, it reduces the dimensionality of these embeddings and clusters them; third, it extracts topics by applying a modified version of TF-IDF. More details are introduced in Section 4.2.

When used in conjunction with text classification, topic modeling serves as a feature engineering approach that transforms raw text into topic distributions. These topic distributions or vectors can then be used as inputs for supervised learning methods such as logistic regression, support vector machines, or neural networks. This approach has demonstrated competitive performance across multiple domains, including biomedical research (Rijcken et al., 2022), social media analysis (Sayed et al., 2023), and recommendation systems (Xia et al., 2019). Compared to traditional bag-of-words or term frequency inverse-document-frequency (TF-IDF) representations, topic-based features offer a more abstract and noise-resistant summary of document content, often leading to improved classification accuracy and interpretability. Despite these advantages, the application of topic modeling for text classification and subsequent inter-journal relationship characterization in the transportation research area remains underexplored.

2.3. *Text classification in transportation*

In transportation studies, text classification has most often been applied to domain-specific reports and social-media sources to extract insights for safety and service evaluation. For example, Zhang et al. (2018) developed a text-mining pipeline that tokenizes unstructured crash reports and employs a logistic regression classifier to identify secondary crashes, namely, incidents triggered by earlier collisions, which demonstrates substantial gains in detection accuracy over traditional spatio-temporal rules. Similarly, Gong et al. (2024) constructed a comprehensive lexicon of urban-rail terminology and applied rule-based matching combined with sentiment and semantic analysis to classify millions of social-media posts about service attributes (e.g., punctuality, cleanliness), which can monitor passenger opinions across ten Chinese cities.

One study aligns more closely with the methodological challenges of classifying large transportation corpora. Tzika-Kostopoulou et al. (2024) conducted a systematic literature analysis of big-data applications in transportation. They assembled a corpus of 2671 peer-reviewed articles (2012–2022) and applied TF-IDF-weighted LDA topic modeling. Naive Bayes and a convolutional neural network were then used to categorize publications into eight coherent themes, achieving over 91% classification accuracy. However, Tzika-Kostopoulou et al. (2024) did not aim to analyze thematic relationships between journals.

While classification studies on reports and social media unlock domain-specific insights and literature-level analyses organize broad research landscapes, semantic-driven text classification has yet to be used for uncovering the relationships that bind transportation journals. When two journals share closely overlapping thematic scopes, their articles naturally become more difficult to distinguish purely by topic, which is an issue no existing study has systematically examined.

2.4. *Review findings*

Discerning boundaries between academic journals and understanding research themes covered by a target journal are important yet challenging tasks for researchers. One good reason, among others, is that desk rejection due to mismatches with a journal's scope is a common and frustrating experience for authors. Recent editorial retrospectives have provided in-depth, journal-specific overviews of individual TR parts. For instance, Jiang et al. (2020) offered a bibliometric analysis of TR-B over four decades (1979–2019), while Cao et al. (2021) and Chen et al. (2022) reviewed the first 25 years of TR-D and TR-E, respectively. These studies synthesized the historical evolution, key themes, and methodological developments within each journal. However, a cross-journal perspective of modeling the thematic structure and overlaps across all six TR parts (A-F) is still lacking. Existing studies on inter-journal relationships have largely relied on citation-based analyses and disciplinary mapping, which do not consider semantic patterns within the articles and the associated journal labels. Although topic modeling has been widely used to uncover latent themes and support text classification in various fields, this integration in transportation research is underexplored. Although text classification has proven effective for parsing text in reports and social media, and for discovering literature trends in big-data transportation research, it has not been applied to investigating inter-journal thematic relationships or to guiding content-based journal selection. To address this gap, this study proposes a novel semantic-driven classification approach to analyze the thematic structure and relationships of the TR journal series.

3. Data

Our data collection consisted of three primary components, namely (1) the collection of research paper abstracts for topic modeling and text classification (Section 3.1), (2) the processing of collected abstracts (Section 3.2), and (3) the implementation of an online survey targeting domain experts for assessing human classification performance (Section 3.3).

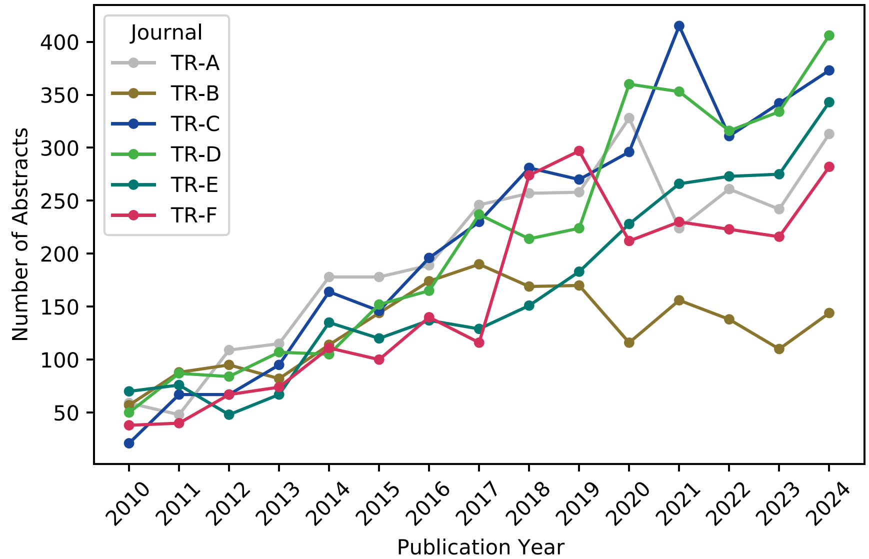


Fig. 1. Number of valid abstracts from each TR journal by year.

3.1. Research paper metadata collection

Research paper metadata were sourced from the Web of Science (WoS) Core Collection, a globally recognized database known for indexing high-quality scientific publications (Clarivate, 2025). The search procedure began with specifying each TR journal title within the "Publication Titles" field and applying a publication year filter spanning from 2010 to 2024. The initial query yielded 18,953 records across all TR journals. Attributes of the WoS database records include article title, abstract, authorship details, issue number, publication year, journal title, author emails, and page count. To focus exclusively on research articles, a filtering process to exclude non-article records was implemented. Specifically, only records under document types "Article," "Review," "Article," or "Proceedings Paper" were kept; other document types, including "Editorial Material" and "Correction," were removed. Additionally, some records lacking abstracts were removed, as abstracts are essential for subsequent text analyses. This filtering resulted in a dataset comprising 16,341 valid abstracts for further modeling. Fig. 1 shows the annual abstract counts for each TR journal (i.e., TR A-F) from 2010 to 2024. Overall, the data reveal a clear upward trend in publication volumes across all journals, with evident surges during the COVID-19 pandemic. This pattern is also observed in other research fields such as life science and biomedicine, where shifts in research priorities led to increased academic writing and publication output (Rousseau et al., 2023).

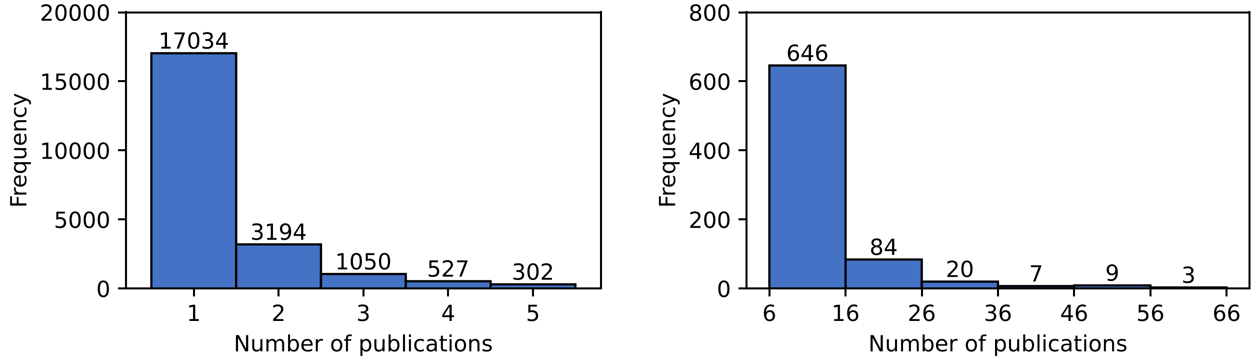
3.2. Abstract preprocessing

To prepare the research paper abstracts for downstream modeling, the following processing steps were performed:

- Truncating. Statements about copyright, publisher, license type, and access right, such as "This is an open access article under the CC BY license," were removed from abstracts of open-access articles.
- Lowercasing. Lowercasing was used to convert all characters in the abstracts to lowercase so that the same words in different cases are treated identically.
- Tokenization. Tokenization was used to split each abstract into a sequence of individual units or "tokens," typically a single word or two-word phrase (bigrams), to enable subsequent text analyses.
- Stopword removal. We first removed common stopwords (e.g., "the," "a," and "and") using NLTK's standard English stop-word list (Bird et al., 2009). Next, we manually identified and excluded domain-specific words (e.g., "study," "paper," and "propose") that frequently appear in scientific abstracts but add little thematic value.
- Word filtering. We calculated the frequency of each word and retained only those present in more than 5% but fewer than 95% of abstracts, following Sun and Kirtonia (2020). This step eliminated both very rare (e.g., "embankment") and overly common words (e.g., "model") that lack discriminative value.

Following the above preprocessing steps, each abstract was transformed into a clean, standardized token representation, which consisted of 176,255 tokens and served as the input for the BERTopic model for topic extraction and subsequent text classification.

Note that, although lemmatization is a standard component of text preprocessing and is often used to collapse different inflected forms into a single base form, in this study it did not improve performance. In preliminary experiments, we compared preprocessing pipelines with and without an additional lemmatization step and found that lemmatization slightly degraded both the coherence of the BERTopic topics and the accuracy of the subsequent journal-classification models. This is consistent with recommendations for BERTopic and other transformer-based topic models, where aggressive normalization (such as lemmatization) can remove informative



(a) Developing Scholars (<= 5 publications)   (b) Established Scholars (>5 publications)

Fig. 2. Frequency distribution of publication counts over 2010 and 2024.

morphological cues. We therefore retain the original surface forms and do not apply lemmatization in the finally adopted preprocessing pipeline.

3.3. Survey design and implementation

We survey to evaluate the classification by human experts in the transportation community. The survey consisted of two sections: (1) researchers' background (Q1 - Q9 in Table A.5), and (2) abstract classification (Q10 - Q19 in Table A.5). The first section gathered information on participants' professional experience and engagement with TR journals. Specifically, we queried participants' editorial experience (e.g., "Q1. Have you ever served as an editor-in-chief, associate editor, or editorial board member for a Transportation Research journal (Parts A-F)?"), reviewing experience (e.g., "Q4. Have you ever served as a reviewer for a Transportation Research journal (Parts A-F)?"), and authoring experience (e.g., "Q7. How many papers have you submitted in total to any TR journals in the past three years, as an author or co-author?"). Additionally, we asked participants to self-rate their familiarity with the scopes of TR Parts A-F. These questions help us explore how participants' experience and familiarity with TR Parts A-F relate to their classification performance. The second section (Q10-Q19) presented participants with ten abstracts randomly, consisting of five that were easy to classify and five that were difficult to classify. The classification difficulty of each abstract is determined through a procedure described in Section 4.4. As a result, approximately 30% of all abstracts (approximately 11,000) were identified as hard-to-classify (or hard for brevity hereafter) and 300 sampled abstracts (150 easy and 150 hard) were added to the survey question bank.

We next detail how survey participants were selected from TR authors. In the following sampling, each author was assumed to have a unique email address and an email address was interchangeable with an author; otherwise, additional name disambiguation efforts were necessary (Sanyal et al., 2021). We observed that, over the years from 2010 to 2024, the number of papers associated with an email or author varies substantially. Specifically, nearly 90% of the authors published a single paper, while prolific authors (top 0.1% by productivity) published more than 40. To effectively visualize the long-tailed distribution, we divided all authors into two categories depending on whether the total publication count is over five. Fig. 2 thus presents the distribution of total publication count per author in two separate plots. Therefore, to ensure a representative sample of transportation experts for this survey, we first limited our sampling pool to individuals who (1) had authored at least two publications from 2010 in any of the TR journals (Part A-F) cumulatively, and (2) had at least one publication within the period from 2021 to 2025. These criteria ensure that participants were active contributors to the field and had sufficient familiarity with TR journal scopes. Eventually, around 3000 authors were eligible for the survey. Then, we employed a two-stage stratified sampling strategy based on author productivity and journal output (number of research articles published yearly). In the first stage, we selected one TR journal (from Parts A-F) with selection probability weighted by the total number of research articles that a journal published, namely total journal output between 2021 and 2025. In the second stage, within the selected journal, we sampled an author with probability weighted by their total number of publications in that journal over the same period (i.e., productivity). To avoid duplication, once an email address was selected, it was removed from the pool before the next iteration.

This survey was distributed and administered online via Qualtrics. Each invitation contained a unique Qualtrics link tied to a specific email address, which prevented multiple submissions from the same invitee. No survey links were distributed through public websites or social media. Because access to the survey was tightly controlled, we considered the likelihood of bot attacks to be very low. For this reason, we did not implement CAPTCHA verification, explicit attention-check questions, or IP-based restrictions. As shown in Table 1, a total of 2400 authors were invited over three batches spanning April and May of 2025. To improve response rates, a reminder was sent approximately one week after the initial invitation to those who did not complete the survey. When the survey was concluded on May 17, 2025, 270 invitees had provided at least one response, corresponding to an overall response rate of 11.3% (270/2,400). Among these respondents, 192 fully completed the survey, corresponding to a completion rate of 71.1% (192/270) among respondents and an overall completion rate of 8.0% of all invited authors (192/2,400). We then examined completion times and found no unusual outliers. We also checked for duplicate or patterned responses (e.g., identical responses across all items) and

Table 1
Three batches of the survey distribution.

Survey batch	Sample size	Initial request	First reminder	Total responses received
1	300	April 15, 2025	April 23, 2025	47
2	1000	April 25, 2025	May 2, 2025	103
3	1100	May 5, 2025	May 11, 2025	120

found none. In addition, the response timeline did not show any sudden spikes that would suggest automated or suspicious activity. Based on these checks, we believe the quality of the survey data is sufficient for our analysis. This process resulted in a dataset of expert assessments, which served as a benchmark for evaluating the performance of our machine learning classification approach.

4. Research methods

4.1. Overall research framework

Our proposed analysis framework mainly involves topic modeling, text classification, and classification difficulty determination. Specifically, a BERTopic model is employed to generate a soft assignment over a given number of latent topics for each preprocessed abstract and a topic distribution matrix, described in Section 4.2. Subsequently, features are constructed by concatenating topic distribution vectors and term-frequency vectors to predict the TR journal label for each abstract, detailed in Section 4.3.1. Based on the misclassification pattern, we propose new similarity metrics to explore journal relationships in Section 4.3.2. Finally, we determine the difficulty of each abstract to facilitate survey design in Section 4.4.

4.2. Topic identification with the BERTopic model

The BERTopic model is applied for identifying topics for all collected abstracts (Grootendorst, 2022). Let the corpus of preprocessed abstracts be $D = \{d_1, d_2, \dots, d_N\}$, where each abstract d_i is represented as a token sequence $d_i = (t_{i,1}, t_{i,2}, \dots, t_{i,n_i})$, with each token $t_{i,j} \in V$ denoting a word or a bigram in the vocabulary V. We define a contextual embedding function:

$$\phi : D \to \mathbb{R}^L, \quad \phi(d_i) = u_i, \tag{1}$$

where $u_i \in \mathbb{R}^L$ is the L-dimensional sentence-level embedding of document d_i, produced by a pretrained Sentence-Transformer model (e.g., all-MiniLM-L6-v2) (Reimers and Gurevych, 2019). As L is on the order of hundreds (such as 384 for all-MiniLM-L6-v2), to obtain a lower-dimensional representation suited for clustering, we learn a projection:

$$\psi : \mathbb{R}^L \to \mathbb{R}^K, \quad \psi(u_i) = y_i, \tag{2}$$

where $y_i \in \mathbb{R}^K$ is the K-dimensional embedding obtained via Uniform Manifold Approximation and Projection (UMAP), and K is chosen to match the desired number of latent topics, such as 50. Hyperparameters of the projection function ψ, namely UMAP, were selected via grid search to maximize clustering coherence.

On the reduced embeddings, we apply Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). Let Z be a discrete topic variable taking values in $\{1, \dots, K\}$. For each document d_i, HDBSCAN yields a posterior membership vector

$$\mathbf{h}_i = \left(P(Z = k \mid d_i)\right)_{k=1}^{K} \in [0,1]^K, \tag{3}$$

with $\sum_{k=1}^{K} P(Z = k \mid d_i) = 1$. Collecting these for all abstracts produces the topic-distribution matrix

$$H = \left[h_{i,k}\right]_{i=1,\dots,N}^{k=1,\dots,K} \in [0,1]^{N\times K}. \tag{4}$$

To derive representative tokens for each topic k, we employ class-based TF-IDF (c-TF-IDF). Let $M = |V|$ denote the vocabulary size and let $T \in \mathbb{N}^{N\times M}$ be the raw term-frequency matrix with entries $T_{i,v} = \mathrm{tf}(w_v \mid d_i)$. Define the aggregated frequency of term w_v in topic k as

$$\mathrm{tf}_k(w_v) = \sum_{i=1}^{N} h_{i,k}\, T_{i,v}, \tag{5}$$

and let $\mathrm{df}(w_v) = \left|\{ i : T_{i,v} > 0\}\right|$. Then the c-TF-IDF score is:

$$\mathrm{cTFIDF}(w_v, k) = \mathrm{tf}_k(w_v) \times \log\left(\frac{N}{1 + \mathrm{df}(w_v)}\right). \tag{6}$$

Ranking terms w_v by $\mathrm{cTFIDF}(w_v, k)$ yields the top descriptors for topic k.

Table 2
Notation for similarity and distinctiveness quantification.

Notation	Definition
N_j	Number of abstracts in journal j
$W_{j,k}$	Total weight assigned to topic k in journal j by the BERTopic model
$\widetilde{W}_{j,k}$	Average weight of topic k in journal j: $W_{j,k}/N_j$
$M_{p\to q,k}$	Total topic misclassification weight from journal p to journal q on topic k by a classifier
$\widetilde{M}_{p\to q,k}$	Misclassification rate from journal p to journal q: $M_{p\to q,k}/N_p$
$s^k_{p,q}$	Topic-level similarity between journals p and q on topic k
$S_{p,q}$	Overall similarity between journals p and q across all topics
$c^k_{j\to j}$	Total topic correct classification weight in topic k for journal j
$u^k_{p,q}$	Topic-level distinctiveness between journals p and q on topic k
$U_{p,q}$	Overall distinctiveness between journals p and q across all topics
$r^k_{p,q}$	Topic-level misclassification rate to all journals outside the pair (p,q) on topic k
$R_{p,q}$	Overall misclassification rate leaking to journals outside the pair (p,q)

4.3. Classification and journal relationship

4.3.1. Feature representation and text classification

To construct the feature representation for classification, we represent the lexical content of each abstract with vectors. Specifically, we identify a vocabulary $V' = \{w_1, w_2, \ldots, w_{M'}\}$ consisting of the M' most frequent tokens across the corpus. Each abstract d_i is then encoded as a term-frequency (tf) vector:

$$\mathbf{t}_i = \left(\mathrm{tf}\left(w_1 \mid d_i\right), \mathrm{tf}\left(w_2 \mid d_i\right), \ldots, \mathrm{tf}\left(w_{M'} \mid d_i\right)\right)^\top \in \mathbb{R}^{M'}, \tag{7}$$

where $\mathrm{tf}(w \mid d_i)$ denotes the raw frequency count of token w in d_i. Then, we concatenate the topic-distribution vector $\mathbf{h}_i$ in Eq. (3) and the term-frequency vector $\mathbf{t}_i$ in Eq. (7) for each document d_i, which yields a combined feature vector $\mathbf{x}_i$:

$$\mathbf{x}_i = \begin{bmatrix}\mathbf{h}_i \\ \mathbf{t}_i\end{bmatrix} \in \mathbb{R}^{K+M'}. \tag{8}$$

By stacking the feature vectors across all abstracts, we construct a design matrix $\mathbf{X} \in \mathbb{R}^{N\times(K+M')}$, which is then standardized column-wise to have zero mean and unit variance. This normalized matrix is subsequently used to train four classifiers, namely, multinomial logistic regression (MLR), support vector classifier (SVC), random forest (RF), and TabNet, to map each abstract's combined lexical and topical features to its corresponding journal label. In the case of MLR, the method estimates the probability that an abstract belongs to each journal by fitting a linear function to the standardized features, which facilitates direct probabilistic interpretation of the classification output. The SVC seeks an optimal hyperplane in this high-dimensional feature space, maximizing the margin between abstracts from different journals through the use of kernel functions, which enhances its ability to capture non-linear decision boundaries (Plakandaras et al., 2019). Meanwhile, the RF ensemble constructs multiple decision trees via bootstrap aggregation, with each tree growing on a randomly selected subset of the data and a randomly selected subset of features at each split; during prediction, the forest aggregates individual tree outputs by majority voting, which reduces variance and improves generalization performance (Roy et al., 2025). TabNet is a deep learning algorithm designed for tabular data that leverages sequential attention and trainable feature masks to iteratively focus on the most relevant input features. By comparing these approaches, we assess their effectiveness in identifying the most appropriate TR journal for a given abstract.

4.3.2. Similarity and distinctiveness quantification

To assess thematic similarity and distinctiveness between journals, we define a set of topic-level and journal-level quantities derived from the BERTopic's probabilistic topic assignment and the classifier's misclassification pattern, with the notation in Table 2.

Let N_j denote the number of abstracts in journal j. For each topic k, the BERTopic model assigns a topic weight $H_{d,k}$ to abstract d, and summing these weights over all abstracts in journal j yields the total topic weight $W_{j,k}$. Dividing this by N_j gives the average prominence of topic k within journal j, which we denote as

$$\widetilde{W}_{j,k} = W_{j,k}/N_j. \tag{9}$$

To capture thematic confusion between two journals, we define the total topic misclassification weight $M_{p\to q,k}$ as the total weight of topic k assigned by the classifier to journal q from abstracts that actually originate in journal p. Normalizing by N_p gives the misclassification rate $\widetilde{M}_{p\to q,k} = M_{p\to q,k}/N_p$, indicating how frequently journal p's abstract on topic k is mistakenly attributed to journal q.

We then define the confusion rate, namely the topic-specific similarity score, between journals p and q as:

$$s^k_{p,q} = \frac{\widetilde{M}_{p\to q,k} + \widetilde{M}_{q\to p,k}}{\widetilde{W}_{p,k} + \widetilde{W}_{q,k}}. \tag{10}$$

This quantity reflects the proportion of topic k shared between journals p and q in terms of mutual misclassification. If $s^k_{p,q}$ is higher than $\frac{1}{6}$, the confusion level is higher than a random guesser, which will be shown in Eq. (19).

To obtain an overall similarity score, we weight $s_{p,q}^{k}$ by the average prominence of topic k in journals p and q, and then sum across all topics to get the journal similarity:

$$S_{p,q} = \sum_{k} \left(\frac{\widetilde{W}_{p,k} + \widetilde{W}_{q,k}}{2} \cdot s_{p,q}^{k} \right). \tag{11}$$

A higher $S_{p,q}$ suggests that the classifier tends to confuse the two journals on topics that are central to both, indicating thematic proximity.

To quantify distinctiveness, we turn to correctly classified abstracts. Let $c_{j\to j}^{k}$ represent the total topic k weight from journal j's abstracts that are correctly labeled by the classifier. For a journal pair (p, q), we define purity rate, namely the topic-level distinctiveness, as:

$$u_{p,q}^{k} = \frac{c_{p\to p}^{k} + c_{q\to q}^{k}}{\widetilde{W}_{p,k} + \widetilde{W}_{q,k}}, \tag{12}$$

which captures how well topic k separates journals p and q in terms of correct classifications. A high value of $u_{p,q}^{k}$ implies the topic is distinctive across the pair. Likewise, if $u_{p,q}^{k}$ is lower than 1/6, the distinctiveness is worse than a random guesser, which will be shown in Eq. (20).

The overall journal distinctiveness score is computed analogously:

$$U_{p,q} = \sum_{k} \left(\frac{\widetilde{W}_{p,k} + \widetilde{W}_{q,k}}{2} \cdot u_{p,q}^{k} \right). \tag{13}$$

While $S_{p,q}$ measures thematic closeness, $U_{p,q}$ reveals the clarity of boundaries between journals. Analyzing both allows us to trace where the two journals align and diverge.

Next, we explore the relationship between $S_{p,q}$ and $U_{p,q}$. In the general multi-journal setting, where abstracts may be assigned to any of a journal set $\{p, q, r, \dots\}$. In other words, a abstract must fall into exactly one of three outcomes: (i) be confused between p and q, (ii) be correctly classified to its source journal, or (iii) be sent to some third journal $r \notin \{p, q\}$. At the topic level these are quantified by $s_{p,q}^{k}$ and $u_{p,q}^{k}$, and by introducing

$$r_{p,q}^{k} = \frac{\sum_{r\notin\{p,q\}} \left(\widetilde{M}_{p\to r,k} + \widetilde{M}_{q\to r,k} \right)}{\widetilde{W}_{p,k} + \widetilde{W}_{q,k}}, \tag{14}$$

to capture misclassification outside the pair. Since these three rates exhaust all possibilities on topic k, they satisfy the equation

$$s_{p,q}^{k} + u_{p,q}^{k} + r_{p,q}^{k} = 1. \tag{15}$$

To lift these to journal-level summaries, we weight each by the joint prominence of topic k, namely, $\frac{1}{2}(\widetilde{W}_{p,k} + \widetilde{W}_{q,k})$, and sum over k, which yields the journal similarity $S_{p,q}$ and distinctiveness $U_{p,q}$, and introduces

$$R_{p,q} = \sum_{k} \frac{\widetilde{W}_{p,k} + \widetilde{W}_{q,k}}{2} \, r_{p,q}^{k}, \tag{16}$$

with the equation

$$S_{p,q} + U_{p,q} + R_{p,q} = 1. \tag{17}$$

$R_{p,q}$ measures the fraction of topic weight on which the classifier "leaks" into any journal other than p or q.

The tripartite decomposition has immediate implications. A high $S_{p,q}$ indicates strong pairwise overlap (namely, many p-papers look like q on their shared topics), while a high $U_{p,q}$ signals that journals p and q occupy clear, non-overlapping thematic niches. When $R_{p,q}$ is large, much of the confusion is diffuse, spread across the wider set of journals rather than concentrated between the two, reflecting a richly interconnected multi-journal landscape.

In particular, under the simpler binary scenario, when only two journals p and q are considered, we have $r_{p,q}^{k} \equiv 0$ for all k, and hence $R_{p,q} = 0$. The relationship reduces to

$$U_{p,q} = 1 - S_{p,q}, \tag{18}$$

illustrating that in pairwise comparisons, distinctiveness and similarity are exact complements, whereas in the full multi-journal context the introduction of $R_{p,q}$ is essential to capture the full spectrum of misclassification behaviors.

4.4. Classification difficulty assessment

We design an approach to determine the classification difficulty of each abstract for the survey in Section 3.3. We first randomly split the full dataset into 11,438 (or 70%) training and 4903 (30%) testing samples. The four classifiers in Section 4.3.1 are evaluated. The best classifier is adopted as the judge for determining the difficulty, either "hard-to-classify" or "easy-to-classify." Using the best classifier, we applied ten-fold cross-validation on the full set of abstracts. Each abstract served as a test instance once, allowing us to observe whether it was correctly or incorrectly classified. Abstracts misclassified during their test fold were labeled as "hard-to-classify," while those consistently classified correctly were labeled as "easy-to-classify."

5. Results

This section presents our results. In Section 5.1, we examine the topic distributions across journals. Section 5.2 evaluates the classification performance of selected machine learning methods and a deep learning algorithm. Section 5.3 delves into the inter-journal thematic relationships by quantifying similarities and dissimilarities among the TR parts. Section 5.4 demonstrates the classification performance of domain experts based on our survey and generative AI. Finally, Section 5.5 discusses the deployment of our classification approach.

5.1. Journal-specific topic distributions

First, we visualize the 50 topics identified by the BERTopic model in Figs. 3 and 4. Note that the choice of the fixed number 50 is consistent with other studies, such as Sun and Yin (2017), Sun and Kirtonia (2020). Each subplot represents a single topic, where the size of each token (word or bigram) reflects its relative frequency within that topic. The topic number is serial and does not carry semantic meaning. Topic titles in the figures are added to aid interpretation and reference. By design, the BERTopic produces a miscellaneous topic ("T49"), which is excluded from the topic distribution analysis. It is unsurprising that several interrelated topics share a major theme, such as T5 to T7 on public transit, and T21 to T24 on supply chains. Since lemmatization is not conducted in the text preprocessing, different morphological variants remain as distinct tokens in the vocabulary and in the topic visualizations, so Figs. 3 and 4 may display multiple forms of the same underlying word (e.g., *ship* and *shipping* in T0, *driver* and *driving* in T10, or *price* and *pricing* in T18).

Then, we show the topic distribution for each TR journal. An abstract is represented by a weighted distribution over topics, reflecting the extent to which it engages with each thematic area. For each journal, we sum the topic-assignment weights of all its abstracts and then divide by the total number of abstracts, which yields the mean topic distribution per paper (namely the value of $\widetilde{W}_{j,k}$ in Eq. (9)). Six journals are analyzed in two groups. The first group includes TR-B, TR-C, and TR-E. Their weighted topic distributions are shown in Fig. 5. Both TR-B and TR-C dedicate their greatest share of topical weight to "T32 Traffic Control," "T10 Driving Safety," and "T5 Bus Transit." By contrast, TR-E displays different focuses from TR-B and TR-C. Its top three topics, namely "T21 Supply Chain," "T16 Vehicle Routing," and "T0 Maritime", reflect TR-E's orientation toward logistics research. Additionally, TR-E has considerable interest in "T5 Bus Transit" but less interest in "T32 Traffic Control," the leading topic of TR-C.

The second group consists of TR-A, TR-D, and TR-F. Their topic distributions are presented in Fig. 6. TR-A has a broadly interdisciplinary portfolio that aligns closely with its focus on policy, planning, and management of transport systems. The most prominent topic of TR-A is "T2 Cycling," indicating a substantial interest in infrastructure and policy instruments to promote cycling. The next few topics, "T10 Driving Safety," "T32 Traffic Control," "T37 Travel Behavior," and "T43 Government Policy," reflect the journal's emphasis on operational safety, planning, and regulatory frameworks. TR-D's leading topics, such as "T12 Electric Vehicles," "T2 Cycling," "T10 Driving Safety," "T29 Air Quality," and "T30 Emission" align precisely with its scope on exploring sustainable mobility, emissions mitigation, and environmental policy in transportation systems. By contrast, TR-F's distribution is overwhelmingly dominated by "T10 Driving Safety," with secondary emphasis on "T2 Cycling" and "T32 Traffic Control," reflecting its core concern with human factors, risk perception, and behavioral interventions in traffic safety.

Topic distributions between Groups 1 and 2 are quite different. For instance, a cluster of supply chain-related topics (T21-T24) accounts for a major portion (15.8%) for TR-E, while they account for a tiny portion for TR-D (3.3%) or TR-F (1.4%). The prominence of a handful of topics, notably "T2 Cycling," "T32 Traffic Control," "T37 Travel Behavior," and "T10 Driving Safety," across nearly all TR parts indicates how core research questions unite the transportation field. These significant topics transcend individual journal scopes. By contrast, niche topics appear more concentrated and journal-specific. For instance, "T21 Supply Chains" and "T16 Vehicle Routing" account for 11.2% and 10.5%, respectively, of TR-E's content, but they are absent from the top five of every other TR part. 60.0% of "T21 Supply Chains" abstracts are published by TR-E. Here, we note that the topic visualizations in Figs. 3 and 4 should be consulted when interpreting a specific topic, even though simplified topic titles (such as "Cycling") are used above for ease of presentation and brevity.

5.2. Classification performance evaluation

As stated in Section 4.3.1, we evaluated three classical classifiers (namely, MLR, SVC, and RF) and a deep learning algorithm called TabNet (Arik and Pfister, 2021) for journal prediction. The entire abstract dataset was shuffled and split into training and testing subsets using a 70%-30% ratio. Each classifier was trained and evaluated using topic-distribution vectors ($\mathbf{h}_t$) generated by the BERTopic model and the raw term-frequency vectors ($\mathbf{t}_t$) of 2000 most frequent tokens (i.e., $M' = 2,000$). Among the four methods, SVC achieved the highest accuracy at 0.67, closely followed by MLR at 0.65, RF at 0.63, and TabNet at 0.59. A naive six-class random guess yielded a baseline accuracy of 0.17, and SVC represented a nearly fourfold improvement over random guessing. These results reflect that the extracted features are informative for text classification. Notably, although TabNet is a state-of-the-art deep learning algorithm designed for tabular data, it achieved a lower accuracy than the other three classifiers, even after careful hyperparameter tuning using Bayesian optimization.

Given the superior overall performance of SVC, we examined its results in greater detail. Table 3 reports the precision, recall, and F1-score for each journal part. Precision and recall measure the proportion of correct predictions among all abstracts predicted for a given journal, and among all actual abstracts of that journal, respectively. The F1-score, defined as the harmonic mean of precision and recall, provides a balanced measure of performance. SVC achieved the highest F1-score (0.85) on TR-F, supported by equally

Fig. 3. Token distributions of topics T0 to T24.

high precision and recall. This indicates that TR-F abstracts are characterized by a highly distinctive vocabulary, with terms like “collision,” “risk perception,” and “driver behavior” serving as strong indicators for classification. TR-D also stands out with an F1-score of 0.73, reflecting its clear environmental and policy-oriented lexicon (such as “emissions,” “air quality,” and “sustainability”) that the SVC model captures well. By contrast, the model performed poorly on TR-B, yielding the lowest recall (0.43) and F1-score (0.49). These results suggest that journals with tightly defined, domain-specific vocabularies (e.g., TR-F) are easier for the model to classify accurately, whereas TR parts (e.g., TR-B) with adjacent methodological or thematic scopes present a greater challenge for SVC.

Interestingly, TR-A and TR-B revealed opposite patterns in precision and recall. For TR-A, the model achieved high recall (0.65) but lower precision (0.55), indicating that while it successfully captures many true TR-A abstracts, it also misclassifies a number of others as TR-A. This pattern is consistent with TR-A’s relatively broad aims and scope, which span multiple areas of transportation policy, planning, operations, and practice and therefore share vocabulary with several other TR parts. This is likely due to TR-A’s broad, interdisciplinary vocabulary characteristics. We note that TR-A obtained its focus “Policy and Practice” in 1992, before which it was “General”. On the other hand, TR-B showed higher precision (0.57) but low recall (0.43), suggesting the conservative behavior

Fig. 4. Token distributions of topics T25 to T49.

that the confidence of identified TR-B abstracts is high, but many TR-B abstracts are missed. This could be caused by TR-B's more specialized and distinctive technical vocabulary. These contrasting patterns highlight a key insight that journals with broad thematic scopes tend to yield higher recall but lower precision, while those with tightly focused, domain-specific content achieve higher precision at the expense of recall.

Fig. 7a presents the confusion matrix achieved by the SVC, illustrating the distribution of misclassifications among TR parts. The classifier demonstrates strong discrimination for TR-D and TR-F, as mentioned earlier. However, Fig. 7b shows the breakdown of the low recall at 0.43 of TR-B; more than half of all TR-B abstracts are "lost" to other classes (29% to TR-C, 16% to TR-E, and smaller fractions elsewhere), which has two interrelated implications. First, a high false-negative rate reflects SVC's difficulty in fully capturing TR-B's focus. Second, this distribution of false-negative samples is consistent with the reality that many papers declined by TR-B could be subsequently accepted by TR-C or TR-E after revisions. The misclassifications from TR-B to TR-C and TR-E indicates the close connection between methodological and technological topics among those journals. In contrast to the spillover of TR-B, TR-A often absorbs a fair portion (over 4%) of abstracts from every other journal, which is very unique. This finding is also consistent with the low precision for TR-A in Table 3. In other words, TR-A's expansive scope spans diverse research areas. Overall, the uneven bidirectional confusion quantifies thematic proximity and reflects how journal scopes shape classification errors.

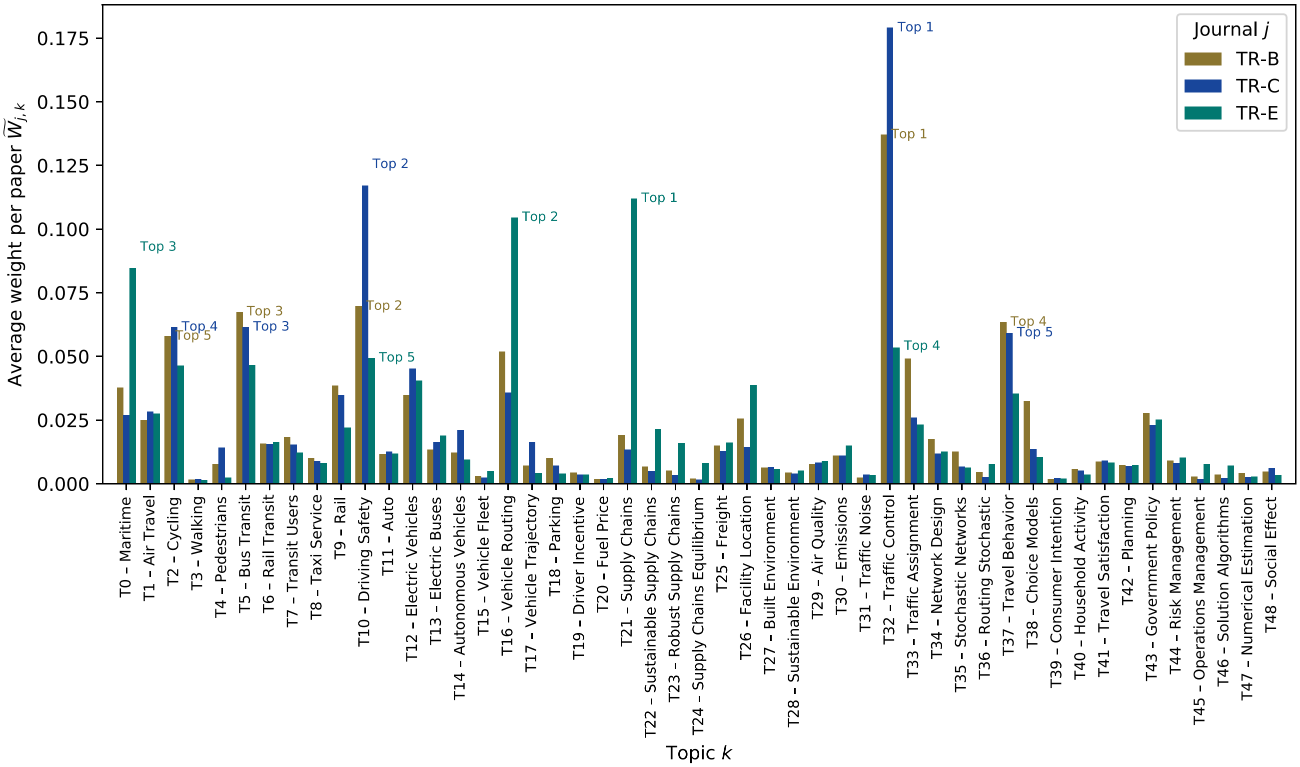


Fig. 5. Weighted topic profiles of journals TR-B, TR-C, and TR-E.

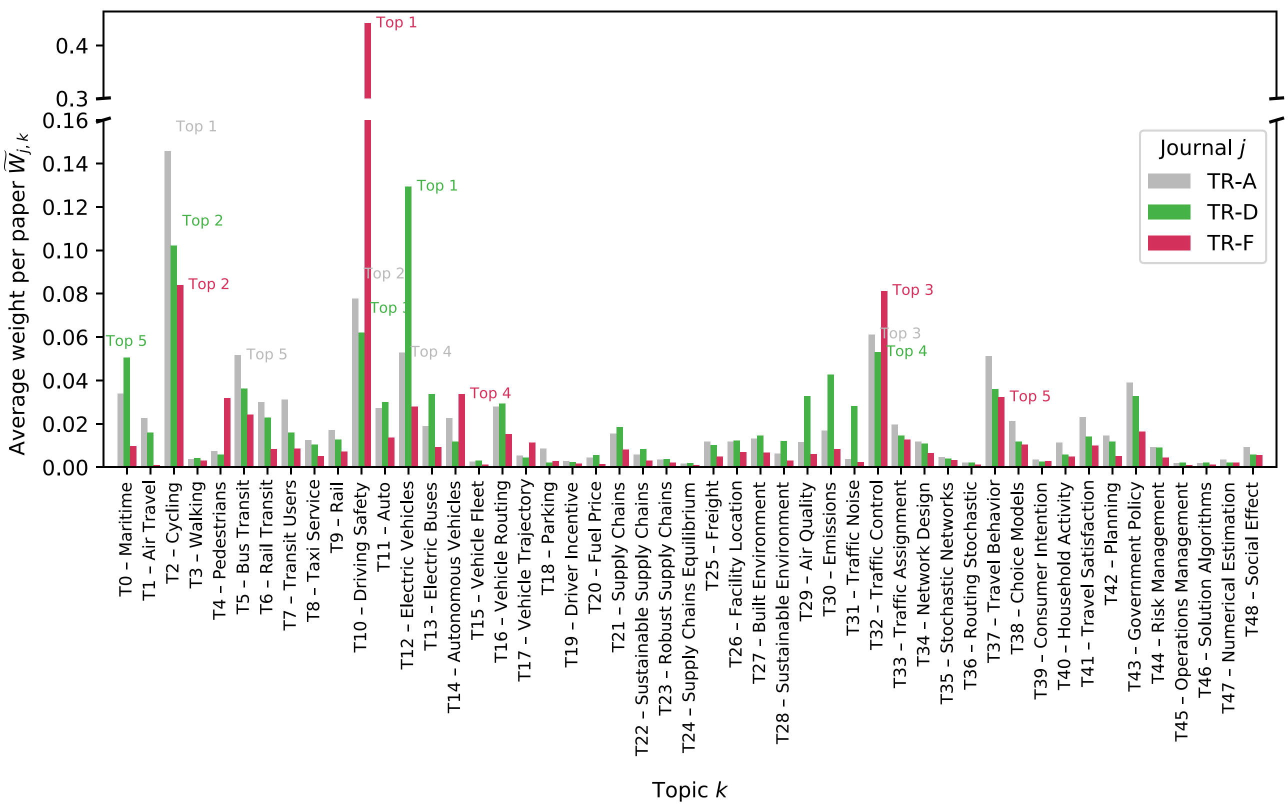
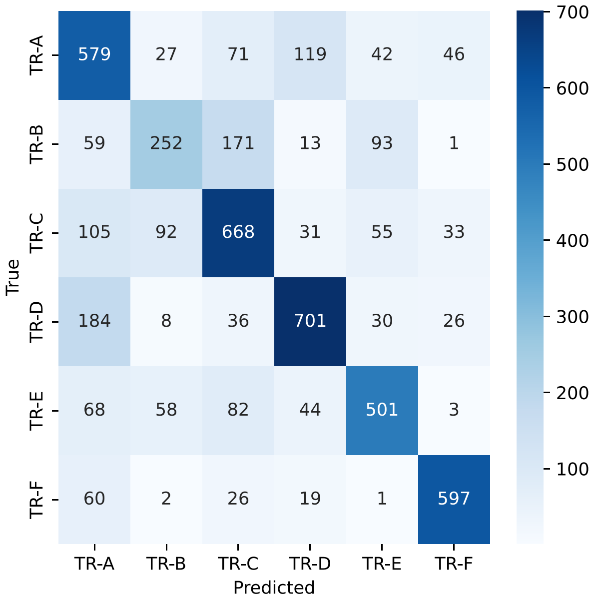


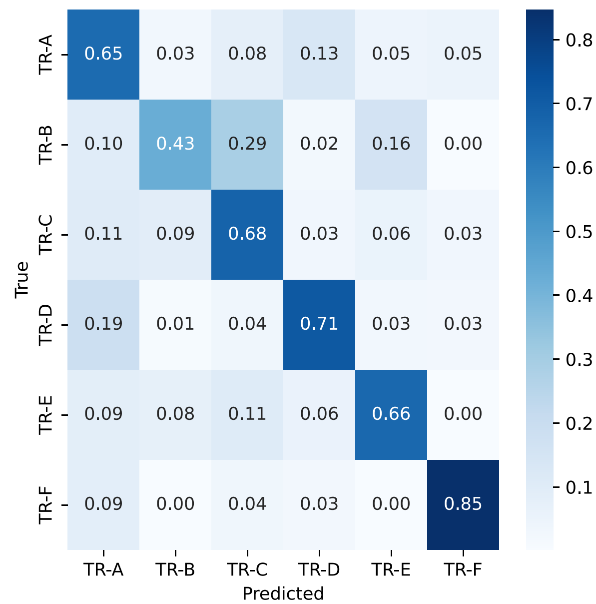
Fig. 6. Weighted topic profiles of journals TR-A, TR-D, and TR-F.

Table 3
Per-journal classification metrics using SVC.

	Precision	Recall	F1-score
TR-A	0.55	0.65	0.60
TR-B	0.57	0.43	0.49
TR-C	0.63	0.68	0.66
TR-D	0.76	0.71	0.73
TR-E	0.69	0.66	0.68
TR-F	0.85	0.85	0.85



(a) Original confusion matrix



(b) Row-normalized confusion matrix

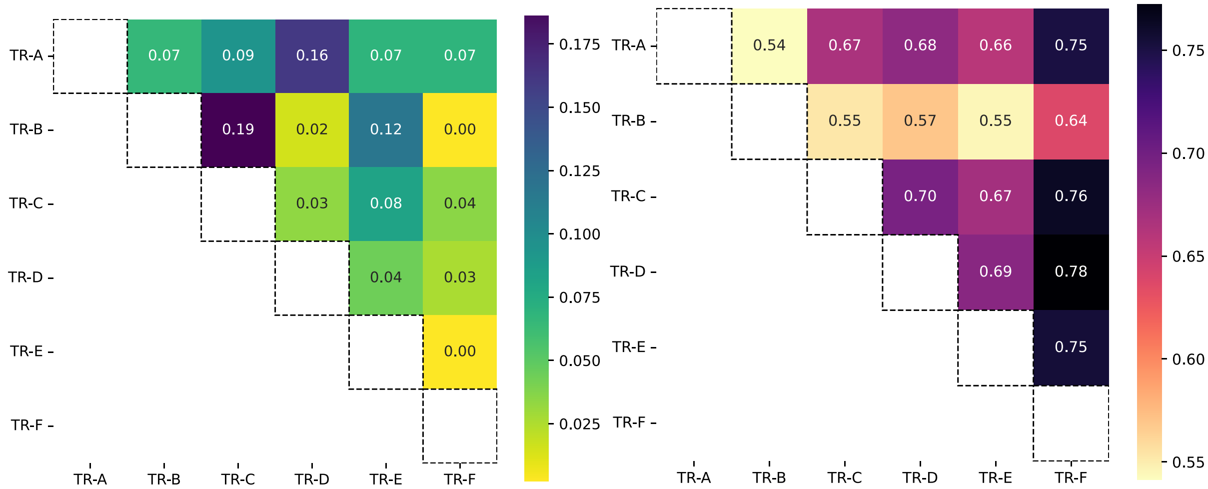
Fig. 7. Confusion matrix for abstract classification.

5.3. *Thematic characteristics across transportation research journals*

Leveraging the proposed metrics in Eqs. (11) and (13), we examine thematic interactions across journals using the misclassification distributions of SVC. The results are presented in Fig. 8. Fig. 8a presents the degree of similarity $S_{p,q}$ between a pair of TR parts (p,q). The highest similarity is observed between TR-B and TR-C. Both journals frequently address similar topics, such as highway traffic modeling. The second strongest similarity occurs between TR-A and TR-D, suggesting a thematic intersection centered on sustainability, policy evaluation, and environmental impact. The third similarity arises between TR-B and TR-E; the shared concern for network design, optimization models, and freight transportation systems creates a substantial area of thematic overlap.

As shown in Eq. (17), a lower similarity score $S_{p,q}$ does not imply a higher dissimilarity score $U_{p,q}$ in a multi-journal scenario. To more accurately capture thematic distinctions, Fig. 8b presents dissimilarities $U_{p,q}$ for all journal pairs. Clearly, significant divergence is found between TR-F and others, among which TR-D, TR-C, and TR-E demonstrate the most significant divergence. This separation is rooted in fundamentally different research focuses and expected. TR-F is dedicated to human factors, psychological modeling of driver behavior, and traffic safety interventions; these themes are relatively unique.

Building on the quantitative metrics proposed in Eqs. (10) and (12), Fig. 9 drills down into the specific topics driving similarity and distinctiveness among journal pairs. Only significant topics with a total weight across all journals exceeding the average threshold of 0.12 are presented, thus excluding minor topics such as "T3 Walking" and "T20 Fuel Price." The journal with a higher weight among the journal pair is indicated for each bar. As an example, Fig. 9b shows the topics of highest confusion rates $s^k_{p,q}$ (defined in Eq. (10)) for TR-B and TR-C. Driving safety, bus transit, and government policy are the most major thematic overlaps between TR-B and TR-C. For instance, X. Li and his team have published multiple TR-B papers whose dominant topic is identified as "T10 Driving Safety," mostly in an automated driving environment. An example paper (Li, 2022) involves car following and traffic oscillation. TR-C also publishes a significant volume of research on related topics. Notably, some research teams, such as those led by M. Abdel-Aty (Anik et al., 2024) and X. Wang (Yuan et al., 2018), are among the most prolific contributors to this subject area and journal. Moreover, several groups, including those led by Z. Zheng (Wang et al., 2025), have actively published driving safety studies in both TR-B and TR-C, reflecting the strong thematic overlap between the two journals. By contrast, Fig. 9b highlights the topics with top purity rates



(a) Similarity (b) Distinctiveness

Fig. 8. Relationships among journals.

$u^k_{p,q}$ (defined in Eq. (12)) for TR-C and TR-D, indicating their thematic niches. Air travel, traffic assignment, and vehicle routing define their thematic boundary, as those topics are often attributed to TR-C, but not as often to TR-D. The aims & scope statement of TR-D (ScienceDirect, 2025d) reads "on all modes of transportation, including maritime and air transportation…" About 100 out of 3194 (or 3.3%) TR-D papers have a dominating topic of "T1-Air Travel," which is only half of the number for TR-C, whose journal output is nearly the same (shown earlier in Fig. 1). Substantial research on airport slot scheduling, air traffic flow, airport surface movement, and urban air mobility has appeared in TR-C; by contrast, those related TR-D papers are focused on fuel, emissions, and air pollution.

Notably, by way of comparison to interpret confusion rate $s^k_{p,q}$ and purity rate $u^k_{p,q}$, a random guesser over six journals would, on any given topic k, confuse journal p with journal q at a rate of

$$\hat{s}^k_{p,q} = \frac{\frac{W_{p,k}}{6} + \frac{W_{q,k}}{6}}{W_{p,k} + W_{q,k}} = \frac{1}{6} \approx 0.17, \tag{19}$$

where $\frac{W_{p,k}}{6}$ is the expected topic-k weight from p that a random guesser would assign to q (and vice versa), and the denominator normalizes by the total topic weight of journal pair (p,q). Similarly, the random-guess purity baseline is

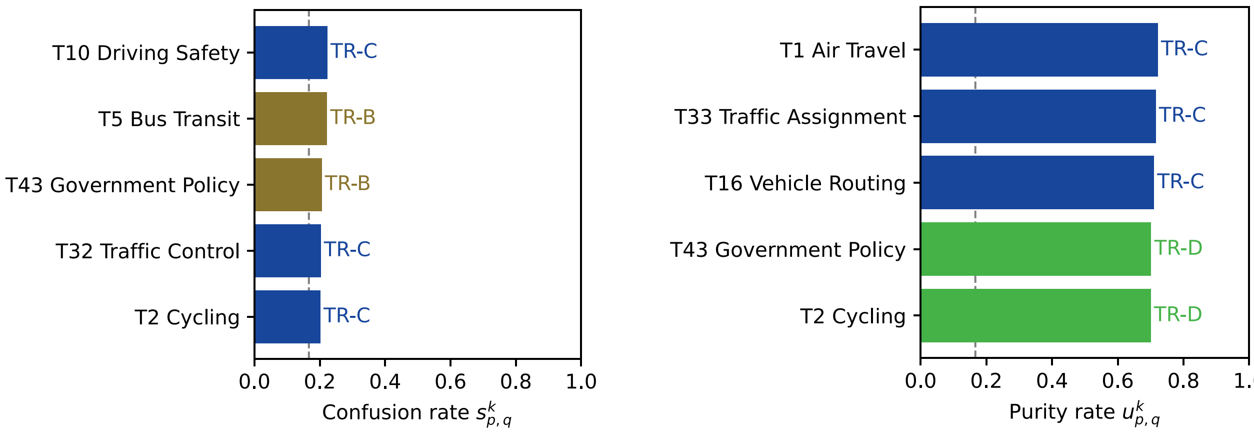
$$\hat{u}^k_{p,q} = \frac{\frac{W_{p,k}}{6} + \frac{W_{q,k}}{6}}{W_{p,k} + W_{q,k}} = \frac{1}{6} \approx 0.17, \tag{20}$$

because a random guesser also has a $\frac{1}{6}$ chance of correctly labeling a p-paper as p (and likewise for q), yielding the same numerator and hence the same baseline value. Clearly, a confusion rate above 0.17 indicates the classifier is more prone to mix up p and q than a random guesser; a purity rate over 0.17 indicates the classifier can more precisely assign abstracts between p and q than a random guesser. The grey lines in Fig. 9a and b indicate the baseline values. Therefore, the confusion rates exceeding 0.20 in Fig. 9a reveal a higher confusion than a random guesser for those topics between TR-B and TR-C. By contrast, purity rates around 0.70 in Fig. 9b sit well above the 0.17 random baseline, reflecting that those topics serve as strong, distinctive ones for TR-C versus TR-D.

5.4. Performance comparison with domain experts and generative AI

In this section, we evaluate the relative performance of human and machine classification by comparing our proposed approach, domain experts, and generative AI.

We first discuss expert survey results. As indicated in Section 3.3, 192 respondents completed the survey, and 41 respondents (21.4%) reported their editorial experience with one or more TR journals. Reviewing experience was more common, with 165 respondents (85.9%) indicating that they have served as peer reviewers for at least one TR journal. Fig. 10 illustrates a detailed breakdown of editorial and reviewing roles across TR Parts A-F, as well as each journal's share of publications (output) from 2021 to 2024. Clearly, TR-F accounts for only 6.1% of editors and 10.2% of reviewers but represents 14.5% of published papers, whereas TR-B accounts for 22.4% of editors and 15.8% of reviewers despite only 8.4% of publications. Several factors may drive the apparent overrepresentation and underrepresentation in survey responses. First, respondents could list services on every editorial board or reviewer pool to which they belonged, whereas each paper counts only once toward its home journal, so larger, more interconnected boards (e.g., TR-B) naturally inflate tallies of "served" individuals. Indeed, 100% of TR-B reviewers reported cross-part experience, whereas 17.8% of TR-F reviewers serve exclusively on TR-F. Second, broadly scoped parts draw on a wider reviewer network than subject-specific outlets, so specialized journals (e.g., TR-F), despite similar review loads, appear underrepresented in reviewer and editorial respondents. Finally, differences in response rates, shaped by perceived prestige, workload, or familiarity with the survey



(a) Confusing topics for TR-B and TR-C (b) Unique topics for TR-C and TR-D

Fig. 9. Contributing topics for journal similarity and distinctiveness.

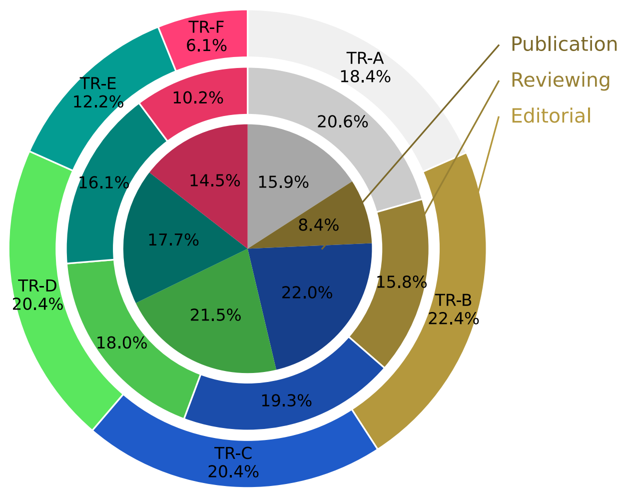


Fig. 10. The publication, reviewing, and editorial distribution across TR journals.

topic, can further skew these distributions. Moreover, the respondent pool is geographically very diverse, including participants from North America, South America (e.g., Chile and Colombia), Europe (e.g., Italy and Finland), Asia (e.g., Japan and India), and other regions, reflecting the global reach of the TR journal series.

Next, we discuss the completion time and classification performance of respondents. Fig. 11 shows three components of the survey results, namely a histogram showing the distribution of abstract classification accuracy (top), another histogram showing the distribution of survey completion time (right), and a box plot illustrating the relationship between accuracy and completion time (central). The classification accuracy over all respondents was 0.38. Expert classification performance differed by the difficulty level, as expected. Specifically, participants achieved a higher accuracy on the easy-level abstracts (0.49) than the hard-level abstracts (0.26). This low accuracy does not well align with participants' self-reported familiarity, where over 90% of respondents indicated that they were either very familiar or moderately familiar with the scopes of the TR journals and believed that selecting an appropriate journal for submission is not difficult. The average survey completion time was 8.9 minutes (standard deviation = 6.15), suggesting that most participants spent a reasonable amount of time engaging with the classification task. The box plot reveals a slight upward trend in completion time with increasing classification accuracy. We also conducted a correlation analysis between both and found a positive correlation ($p < .05$), indicating that participants who spent more time on the task tended to perform better. For example, participants in the top quartile of completion time (more than 10.2 minutes) had an average accuracy of 0.41, compared to 0.33 for those in the bottom quartile (less than 4.8 minutes). We also found that the respondents indicating editorial or reviewing experience do not show a significant difference in accuracy or completion time.

Then, we compare the performance of our classification approach with two generative AI agents, namely ChatGPT 4o and DeepSeek V3, used as independent classifiers to examine the potential of generative AI to perform domain-specific classification tasks through textual understanding. In this experiment, we provided AI agents with the official scope statements of the six journals (TR-A through TR-F) and asked them to assign each abstract to the most appropriate journal based solely on journal scope statements. The prompt was given as follows:

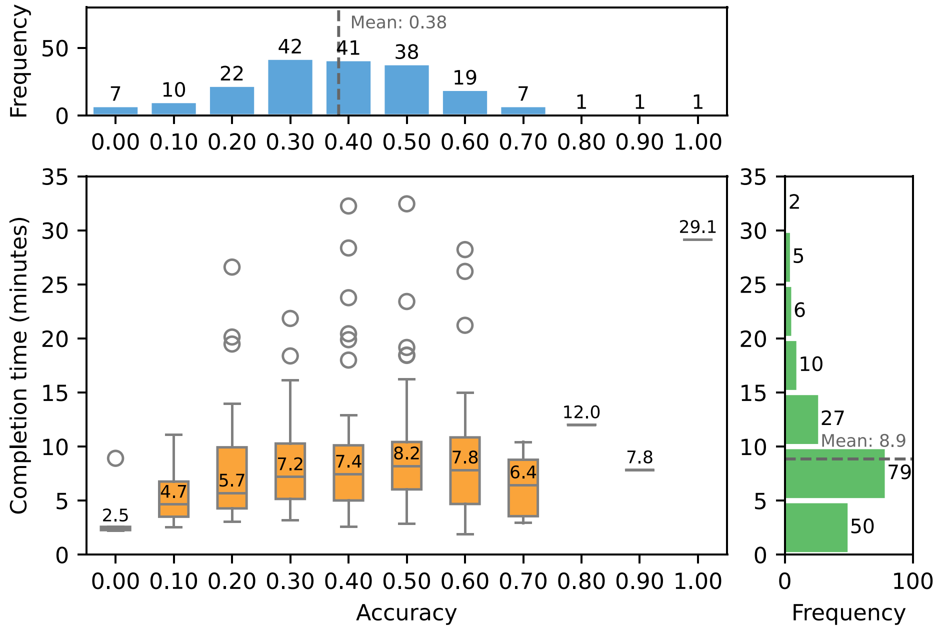


Fig. 11. Distributions of accuracy and survey completion time.

Table 4
Accuracy comparison of domain experts, the best machine classifier, and two generative AI agents.

Classifier	Easy abstracts	Hard abstracts	Overall
Domain experts	0.49	0.26	0.38
SVC	0.91	0.33	0.62
ChatGPT 4o	0.80	0.44	0.62
DeepSeek V3	0.76	0.44	0.60

> "There are six transportation research journals, abbreviated as TR-A to TR-F. The scopes of the six journals are given as follows [text omitted]. You are then given ten abstracts. Please identify the most relevant and suitable journal for each abstract based on your understanding of journal scopes and research topics reflected in each abstract. No additional information beyond what is provided here should be used."

The classification task was run ten times. For each time, five easy and five hard abstracts were randomly drawn. The predicted journal labels were then compared against the actual publication journal for each abstract.

Finally, for a fair comparison, the 300 abstracts in the survey questions are used for the testing dataset. In this way, we ensure the ratio of difficult abstracts to easy abstracts is 5:5, consistent with expert surveys. All non-testing samples are used for training the SVC.

The results are reported in Table 4. While hard abstracts are challenging for both AI agents, they perform well on easy abstracts. ChatGPT 4o achieves a slightly higher accuracy (0.62) than DeepSeek V3, which is also close to the performance of SVC (0.62) and significantly higher than that of domain experts (0.38). We should note that neither ChatGPT 4o nor DeepSeek V3 accesses the training dataset, which should be weighed in a fair accuracy comparison. Despite of this, both AI agents significantly outperform others for hard abstracts.

5.5. Research deployment

We have deployed a preliminary web application using Streamlit at https://tr-journal-recommender.streamlit.app to enable the public to access our classification method anytime. Fig. 12 presents the interface of the live Streamlit app. This app takes an abstract submitted by a user and returns a TR part that has the most thematic relevance. As shown at the bottom of Fig. 12, we provide a complementary online test for researchers who are interested in conducting abstract classifications based on their experience.

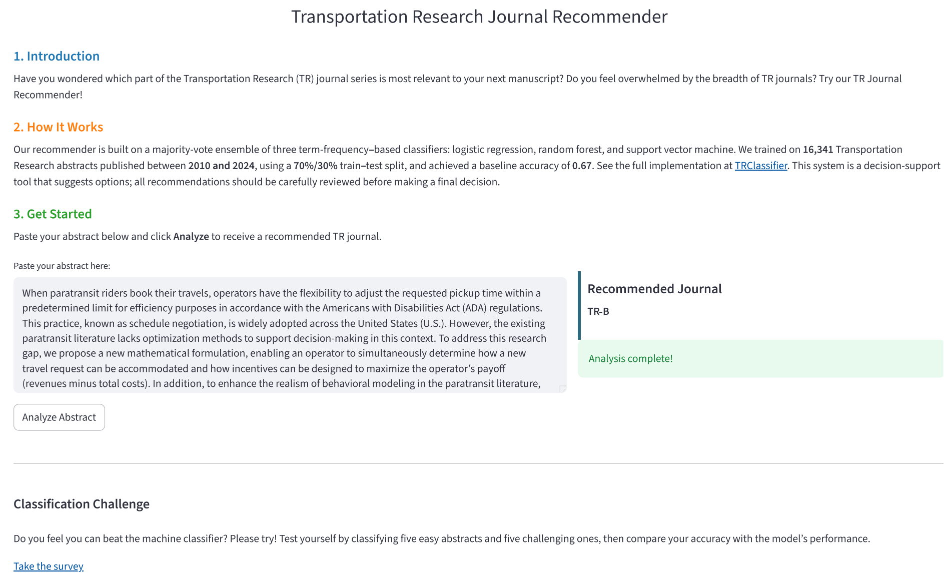


Fig. 12. Interface of the deployed Streamlit app.

6. Discussions

6.1. *Implications for transportation research and publication practice*

This study is of significant value to the transportation research community by offering important insights into the thematic structures that shape scholarly activities across various journals. One primary benefit of our analysis is the illumination of research landscapes, enabling transportation researchers to identify areas of scholarly concentration and potential research gaps effectively. By clearly articulating underrepresented areas, such as "T19 Driver Incentives" and "T39 Consumer Intention," scholars can strategically position their research endeavors to maximize impact. Additionally, revealing dominant thematic focuses, such as TR-E's substantial emphasis on "T16 Vehicle Routing" and "T21 Supply Chains," equips researchers and policy makers with a precise understanding of critical priority areas within the broader transportation field.

Furthermore, our approach offers considerable benefits in clarifying journal relationships and thematic overlaps, providing researchers with guidance regarding journal selection and submission strategies. For example, the notable thematic intersections between *Transportation Research Part B* and *Part C* highlight potential pathways for cross-journal collaboration and thematic integration. Similarly, *Transportation Research Part B* and *Part E* both emphasize vehicle routing problems. There is no wonder that both journals share editors, reviewers, and authors. Our approach, once generalized, can guide authors towards alternate publication outlets if their primary target journal poses issues related to scope alignment.

Reflecting on the broader implications, the insights generated from our topic-based analysis hold practical implications for early-career researchers, offering guidance to enhance manuscript fit and potentially reduce desk rejection rates (Dwivedi et al., 2022). During the course of this study, a distinguished scholar, who also serves as editor-in-chief of a Springer journal, shared an internal list of research topics that the journal does not consider for publication. These exclusions are not evident from the publicly available scope statement on the journal's website, making them difficult to discern even for experienced authors. However, our developed classifier can potentially "learn" such undisclosed preferences from historical publication records and offer valuable guidance to novice authors. Editors and reviewers can likewise leverage these insights to refine journal scopes, anticipate emerging research trends, and promote thematic special issues.

6.2. *Theoretical implications*

This study advances the theory of topic-model-based journal recommendation in several ways. First, by integrating BERTopic's transformer-based topic representations with supervised classifiers, we formalize manuscript-journal matching as learning decision boundaries in a high-dimensional semantic space rather than as matching on surface keywords or citation profiles. In a journal family with intentionally overlapping scopes, our results show that this representation supports substantially more accurate assignment than domain experts, indicating that journal "scope" can be operationalized as a learnable function of latent semantic structure.

Second, the similarity-distinctiveness metrics derived from the classifier's probabilistic outputs (e.g., the $S_{p,q}$, $U_{p,q}$, and $R_{p,q}$ measures) provide a new lens on inter-journal relationships. Unlike citation-based indices, these metrics decompose the classifier's behavior into pairwise topical overlap, boundary clarity, and leakage to third journals, all anchored in article-level topic distributions. This yields a general framework for quantifying how sharply journals are separated in topic space, for identifying which topics are intrinsically ambiguous versus genuinely distinctive, and for characterizing the thematic granularity of multi-part journal series such as Transportation Research.

Overall, these contributions move research on topic modeling and journal similarity from purely descriptive mapping toward a classification framework that explains how manuscripts are matched to journals.

6.3. Recommendations

We make two recommendations to bridge the gap between model insights and TR editorial practice. First, manuscript submission platforms, such as Editorial Manager, could embed a similar classifier directly into the abstract-upload workflow, such that when an author enters an abstract, the system automatically evaluates its thematic alignment and flags whether the chosen TR part best matches the content or if an alternative part might yield a more suitable thematic fit. Realizing this capability will require collaboration among TR editorial boards to advocate for the feature and to work with the publisher Elsevier on its technical integration.

Second, journal editors-in-chief could leverage the quantitative inter-journal similarity metrics produced by our approach to undertake a periodic review of each journal's scope statement, for instance, on a triennial cycle. By systematically comparing thematic overlaps and divergences, editorial teams can refine their published aims and scope to maintain reasonable boundaries, anticipate emerging research frontiers, and mitigate author confusion about target venues.

6.4. Limitations to overcome

Despite its promise, our research has several limitations that might hinder its practical application and thus should be addressed through future work. First and foremost, by basing classifications and recommendations solely on abstract text, the classifier misses other critical components present in full-length articles, including detailed methodologies, extensive case studies, or appendices, which carry substantial weight in reviewers' recommendations and editors' decisions. Extending the model to ingest full-text content and adding related features (e.g., number of equations and proofs, and availability of hypotheses for testing) would likely improve classification performance while accessing and parsing over 15,000 articles is practically challenging. Other factors, such as review turnaround time or editorial reputation, lie outside the scope of a purely text-driven classifier. This further explains that the best-performing machine classifier achieves an accuracy slightly above 0.6.

Further, the present model does not incorporate authors' academic disciplines and affiliations and assumes away their impact on authors' journal choices. For instance, since TR-C is not indexed by the SSCI (Social Sciences Citation Index), authors affiliated with social science departments may prefer submitting to TR-A when a paper fits the scopes of two journals. A researcher in an accredited school of business, such as one co-author of Sun et al. (2021), favors TR-E , among others, not because of content alignment but because TR-E employs a double-anonymous review policy that is increasingly recognized in the operations management community.

Finally, an editor-in-chief has the discretion to determine whether a submission is in the scope of his/her managed journal. A chief editor's or the editorial board's strategic special-issue plans, or desire to cultivate emerging subfields all influence how research articles are aligned with different journals. Such human judgments, which are complex, and context-dependent, are not captured by machine learning classifiers. Therefore, automated recommendations are not designed to replace editorial judgments; instead, they are valuable copilots to inform authors in the paper submission process.

7. Conclusions

This paper has introduced a novel semantic-driven approach to the characterization of thematic structures across selected transportation journals using the BERTopic model, machine learning classifiers, and new inter-journal relation characterization metrics. This study thus has addressed the research gap that there are no systematic studies investigating the thematic relationships of transportation research journals based on text classification.

The key findings of this study include:

(1) By applying the BERTopic model to the large corpus of abstracts, we successfully derived 50 research topics and identified each journal's thematic emphasis. For example, TR-B and TR-C demonstrated strong focuses on methodological advances in traffic control, driving safety, and bus transit.

(2) Among all machine classifiers tested, the SVC achieved the highest accuracy of 0.67 in assigning abstracts correctly to their corresponding TR journals. As evidenced by precision and recall metrics, TR-A has a broader and more diverse topical coverage, in contrast with TR-B's specialized methodological focus.

(3) Our newly proposed semantic similarity metrics effectively quantified thematic overlaps and distinctiveness among journals. Notably, significant overlaps were identified between TR-B and TR-C around "Driving safety," whereas clear distinctions were evident between TR-C and TR-D, particularly regarding "Air travel." TR-F displayed a uniquely differentiated thematic identity, which was supported by its very high recall value (0.85).

(4) The survey involving 192 respondents facilitated the comparison between our approach versus traditional human judgment. Domain experts achieved a relatively low accuracy of 0.38, substantially lower than the machine classification accuracies, namely, SVC (0.62 on the same testing dataset), ChatGPT 4o (0.62), and TabNet (0.59). This clearly demonstrated the superiority of machine intelligence trained on extensive bibliographic data for accurate thematic classification.

(5) As a practical product of our study, we developed and deployed a publicly accessible Streamlit application. This tool provides on-demand journal recommendations based on abstract inputs, mitigating the practical issue of manuscript misalignment and desk rejections due to scope mismatch.

Future work may focus on the following enhancements to improve this framework:

- Extend the classifier to cover a wider range of transportation journals, such as *Transportation Science*, *Journal of Transport Geography*, *Journal of Air Transport Management*, and *Transportation*. Incorporating abstracts from other journals into the training set will deepen the model's thematic coverage and, by virtue of a larger, more diverse dataset, improve overall classification performance.
- Although the current Streamlit interface is user-friendly, more metrics can be introduced to interpret the classification results, such as thematic relevance of the submitted abstract to topics and journals. Increasing model transparency would help editors and authors understand the basis for recommendations, which increases trust in the system.
- We also received valuable feedback by email from several participants, including preeminent researchers and chief editors. One important suggestion we agree with is to allow Top-2 journal labels per abstract, instead of a single label, as some submissions are expected to be in the scope of two TR journals.

Data availability

Data will be made available on request.

CRediT authorship contribution statement

Meng Zhao: Writing – original draft, Methodology, Formal analysis, Data curation; **Shijie Chen:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization; **Yanshuo Sun:** Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Conceptualization.

Acknowledgement

The corresponding author is partially supported by the National Science Foundation (Grant Nos. 2200384, 2100745, and 2055347).

Appendix A. Online survey questions

Table A.5 lists questions asked in the transportation expert survey.

Table A.5
Overview of survey questions.

ID	Survey question (may be shortened or rephrased)	Question type
Q1	Have you ever served as an editor (editor-in-chief, associate editor, or editorial board member) for a Transportation Research (TR) journal (Parts A–F)?	Yes-or-No
Q2	Which TR Part journal/journals have you served as an editor (if you answered Yes to Q1)?	Multiple choice (multiple)
Q3	While you serve as an editor, how many papers do you handle each year (if you answered Yes to Q1)?	Multiple choice (single)
Q4	Have you ever served as a reviewer for a TR journal?	Yes-or-No
Q5	For which TR Part (A–F) have you served as a reviewer (if you answered Yes to Q4)?	Multiple choice (multiple)
Q6	How many TR journal submissions have you reviewed in total in the past 3 years (if you answered Yes to Q4)?	Multiple choice (single)
Q7	How many papers have you submitted to any TR journals in total in the past 3 years, as an author or co-author?	Multiple choice (single)
Q8	How easy do you feel when choosing which TR journal for submission?	Multiple choice (single)
Q9	How would you rate your familiarity with the scope of TR Parts A–F?	Multiple choice (single)
Q10–Q19	Which TR Part does this abstract belong to?	Multiple choice (single)

Reference

Anik, B.M. T.H., Islam, Z., Abdel-Aty, M., 2024. A time-embedded attention-based transformer for crash likelihood prediction at intersections using connected vehicle data. Transp. Res. Part C Emerg. Technol. 169, 104831.

Arik, S.Ö., Pfister, T., 2021. TabNet: attentive interpretable tabular learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35, pp. 6679–6687.

Bi, H., Gao, H., Li, A., Ye, Z., 2024. Using topic modeling to unravel the nuanced effects of built environment on bicycle-metro integrated usage. Transp. Res. Part A Policy Pract. 185, 104120.

Bird, S., Klein, E., Loper, E., 2009. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc.
Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3 (Jan), 993–1022.
Cao, J., Li, S., Noland, R.B., Ge, Y.-E., 2021. The first 25 years of transportation research Part D: transport and environment. Transp. Res. Part D Transp. Environ. 100, 103078.
Chen, S., Meng, Q., Choi, T.-M., 2022. Transportation research Part E-logistics and transportation review: 25 years in retrospect. Transp. Res. Part E Logist. Transp. Rev. 161, 102709.
Clarivate, 2025. What is Web of Science Core Collection? https://clarivate.libguides.com/webofscienceplatform/woscc. [Online; May 1, 2025].
Colavizza, G., Boyack, K.W., van Eck, N.J., Waltman, L., 2018. The closer the better: similarity of publication pairs at different cocitation levels. J. Assoc. Inf. Sci. Technol. 69 (4), 600–609.
Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171–4186.
Dwivedi, Y.K., Hughes, L., Cheung, C. M.K., Conboy, K., Duan, Y., Dubey, R., Janssen, M., Jones, P., Sigala, M., Viglia, G., 2022. How to develop a quality research article and avoid a journal desk rejection. Int. J. Inf. Manage. 62, 102426.
Gong, S.-H., Teng, J., Duan, C.-Y., Liu, S.-J., 2024. Framework for evaluating online public opinions on urban rail transit services through social media data classification and mining. Res. Transp. Bus. Manage. 56, 101197.
Grootendorst, M., 2022. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.
Jiang, C., Bhat, C.R., Lam, W. H.K., 2020. A bibliometric overview of transportation research Part B: methodological in the past forty years (1979–2019). Transp. Res. Part B Methodol. 138, 268–291.
Katchanov, Y.L., Markova, Y.V., 2017. The "space of physics journals": topological structure and the journal impact factor. Scientometrics 113 (1), 313–333.
Kuang, D., Choo, J., Park, H., 2015. Nonnegative matrix factorization for interactive topic modeling and document clustering. Partitional Clustering Algorithms , 215–243.
Li, X., 2022. Trade-off between safety, mobility and stability in automated vehicle following control: an analytical method. Transp. Res. Part B Methodol. 166, 1–18.
Modak, N.M., Merigó, J.M., Weber, R., Manzor, F., de Dios Ortúzar, J., 2019. Fifty years of transportation research journals: a bibliometric overview. Transp. Res. Part A Policy Pract. 120, 188–223.
Plakandaras, V., Papadimitriou, T., Gogas, P., 2019. Forecasting transportation demand for the US market. Transp. Res. Part A Policy Pract. 126, 195–214.
Reimers, N., Gurevych, I., 2019. Sentence-BERT: sentence embeddings using siamese BERT-Networks. arXiv preprint arXiv:1908.10084.
Rijcken, E., Kaymak, U., Scheepers, F., Mosteiro, P., Zervanou, K., Spruit, M., 2022. Topic modeling for interpretable text classification from EHRs. Front. Big Data 5, 846930.
Rousseau, R., Garcia-Zorita, C., Sanz-Casado, E., 2023. Publications during COVID-19 times: an unexpected overall increase. J. Informetr. 17 (4), 101461.
Roy, S., Nahmias-Biran, B.-h., Hasan, S., 2025. Spatial transferability of machine learning based models for ride-hailing demand prediction. Transp. Res. Part A Policy Pract. 193, 104413.
Sanyal, D.K., Bhowmick, P.K., Das, P.P., 2021. A review of author name disambiguation techniques for the PubMed bibliographic database. J. Inf. Sci. 47 (2), 227–254.
Sayed, M.A., Braşoveanu, A. M.P., Nixon, L. J.B., Scharl, A., 2023. Unsupervised topic modeling with BERTopic for coarse and fine-grained news classification. In: International Work-Conference on Artificial Neural Networks. Springer, pp. 162–174.
Schubert, A., 2010. A reference-based Hirschian similarity measure for journals. Scientometrics 84 (1), 133–147.
ScienceDirect, 2025a. About the journal - Transp. Res. Part A. https://www.sciencedirect.com/journal/transportation-research-part-a-policy-and-practice. Accessed: May 1, 2025.
ScienceDirect, 2025b. About the journal - Transp. Res. Part B. https://www.sciencedirect.com/journal/transportation-research-part-b-methodological. Accessed: May 1, 2025.
ScienceDirect, 2025c. About the journal - Transp. Res. Part C. https://www.sciencedirect.com/journal/transportation-research-part-c-emerging-technologies. Accessed: May 1, 2025.
ScienceDirect, 2025d. About the journal - Transp. Res. Part D. https://www.sciencedirect.com/journal/transportation-research-part-d-transport-and-environment. Accessed: May 1, 2025.
Sun, L., Yin, Y., 2017. Discovering themes and trends in transportation research using topic modeling. Transp. Res. Part C Emerg. Technol. 77, 49–66.
Sun, Y., Kirtonia, S., 2020. Identifying regional characteristics of transportation research with transport research international documentation (TRID) data. Transp. Res. Part A Policy Pract. 137, 111–130.
Sun, Y., Kirtonia, S., Chen, Z.-L., 2021. A survey of finished vehicle distribution and related problems from an optimization perspective. Transp. Res. Part E Logist. Transp. Rev. 149, 102302.
Tzika-Kostopoulou, D., Nathanail, E., Kokkinos, K., 2024. Big data in transportation: a systematic literature analysis and topic classification. Knowl. Inf. Syst. 66 (8), 5021–5046.
Varin, C., Cattelan, M., Firth, D., 2016. Statistical modelling of citation exchange between statistics journals. J. R. Stat. Soc. Ser. A Stat. Soc. 179 (1), 1–63.
Wandelt, S., Zheng, C., Sun, X., 2025. A review of trends and patterns in transportation journals from 2000 to 2024. Transp. Res. Interdiscip. Persp. 31, 101406.
Wang, C., Jia, D., Zheng, Z., Ngoduy, D., Wang, W., Wang, S., 2025. Segment-then-refine: a general calibration framework incorporating intra-driver heterogeneity into car-following models. Transp. Res. Part C Emerg. Technol. 176, 105144.
Wang, F., Wolfram, D., 2015. Assessment of journal similarity based on citing discipline analysis. J. Assoc. Inf. Sci. Technol. 66 (6), 1189–1198.
Wolfram, D., Zhao, Y., 2014. A comparison of journal similarity across six disciplines using citing discipline analysis. J. Informetr. 8 (4), 840–853.
Xia, L., Luo, D., Zhang, C., Wu, Z., 2019. A survey of topic models in text classification. In: 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE, pp. 244–250.
Yuan, J., Abdel-Aty, M., Wang, L., Lee, J., Yu, R., Wang, X., 2018. Utilizing bluetooth and adaptive signal control data for real-time safety analysis on urban arterials. Transp. Res. Part C Emerg. Technol. 97, 114–127.
Zhang, Z., He, Q., Gao, J., Ni, M., 2018. A deep learning approach for detecting traffic accidents from social media data. Transp. Res. Part C: Emerg. Technol. 86, 580–596.