



Interdisciplinary Analysis of Machine Learning Applications: Focus on Intent Classification

Nabila Khouya^{1*}, Asmaâ Retbi², Samir Bennani³

¹Rime Team, Mohammedia School of Engineers (EMI), Mohammed V University in Rabat, Morocco

¹nabi97ky@gmail.com, ²retbi@emi.ac.ma, ³sbennani07@gmail.com

Abstract

Given the rapid growth of machine learning publications on platforms such as arXiv, there is a need for systematic approaches to understand their objectives and contributions. This study aimed to analyze scientific intentions across domains, identify research trends, and evaluate the impact of external contextual enrichment on automatic intent classification. We perform a cross-domain comparison of research objectives, methodological designs, and application scenarios in machine learning publications, focusing on computer science and biology. We propose IntentBERT-Wiki, an enhanced BERT model enriched with contextual knowledge from Wikipedia, designed for intent classification in scientific documents. Our dataset comprises annotated sentences extracted from arXiv articles, categorized according to established rhetorical role taxonomies. The model's performance is evaluated using standard classification metrics and compared to a baseline BERT model. Experimental results show that IntentBERT-Wiki achieves F1-scores of 95.9% in computer science and 87.4% in biology, with corresponding accuracies of 96.5% and 91.4%, outperforming the baseline. These findings demonstrate that Wikipedia-based contextual enrichment can significantly improve intent classification accuracy, enhance the organization of academic discourse, and facilitate cross-domain knowledge transfers. This study contributes to the understanding of how machine learning research is framed across disciplines and provides a scalable framework for scientific content analysis.

Keywords: academic research, intent classification, language models; machine learning; natural language processing

How to Cite: N. Khouya, A. Retbi, and S. Bennani, "Interdisciplinary Analysis of Machine Learning Applications: Focus on Intent Classification", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 5, pp. 1234 - 1241, Oct. 2025.

Permalink/DOI: <https://doi.org/10.29207/resti.v9i5.6899>

Received: July 4, 2025

Accepted: October 18, 2025

Available Online: October 25, 2025

*This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia*

1. Introduction

When a new field such as machine learning experiences an explosion in the rate of scientific publication, the body of academic output grows considerably, especially on open platforms like arXiv [1], [2]. This abundance of scientific information is both a valuable resource and a significant challenge in terms of access, structuring, and systematic analysis of publications. In such a context, mining relevant information from scientific documents [3] plays an essential role in improving knowledge organization and facilitating the retrieval of scientific works [4]. Among emerging strategies, intention classification [5], [6] in research papers has become a promising perspective for content analysis. The rhetorical intentions of authors—such as proposing a new method, comparing algorithms, or applying an existing method to a novel task—can provide a deeper understanding of the evolution of research trends in machine learning and of how scientific output is structured across disciplines [7].

Although several studies have explored intent detection in diverse contexts such as human–computer interaction [8], dialogue systems [9], biomedical literature [10], and industrial documentation [11], few have focused on the interdisciplinary and systematic classification of intentions in scientific articles [12], [13]. This gap highlights the need for targeted research on academic intent across domains. The present study addresses this limitation by examining scientific discourse within the field of machine learning [14], [15].

In this work, we investigate intent classification in machine learning publications [16] collected from the arXiv repository through an interdisciplinary lens. Using natural language processing (NLP) techniques [17], and particularly the BERT model and its variants, we analyze how rhetorical intentions vary between domains such as computer science, biology, and physics, while emphasizing the recurring discourse patterns characteristic of each discipline.

A relevant contribution in this field is that of Liu et al. [18], who introduced the Chinese Implicit Intent Dataset (CIID), a corpus covering seven categories of implicit intentions across multiple domains. They proposed a prompt-learning strategy that enables models to infer implicit intentions from prompts and responses through “Relevant Clues.” Compared with several pretrained models, including SciBERT, trained on scientific publications to capture domain-specific terminology, their approach achieved state-of-the-art accuracy even with limited training data.

In a related line of work, Qi et al. [19] explored citation intent classification [20] in scientific articles using a multitask learning approach [21]. Their model integrates joint semantic representations from pretrained language models such as SPECTER [22], which combines textual and citation-network information, while leveraging SciSummNet to extract citation-context features. The results demonstrated that incorporating heterogeneous features within a multitask framework improves overall citation intent classification performance.

Together, these studies confirm the effectiveness of advanced NLP approaches—such as prompt learning and multitask learning—supported by models like SciBERT and SPECTER, and external resources like SciSummNet, for intent recognition in scientific discourse. They also highlight the importance of integrating contextual and semantic knowledge to enhance understanding of scholarly communication. Inspired by these works, our study enriches BERT with Wikipedia-based contextual knowledge to capture deeper semantic cues for cross-domain intent classification.

Recent findings further show that BERT outperforms other transformer-based architectures such as RoBERTa and DistilBERT for intent recognition tasks in scientific texts [23], establishing it as a reliable baseline for predicting rhetorical roles such as Proposal, Result, or New Method. However, sentences in academic writing often lack explicit rhetorical cues, especially in interdisciplinary contexts. To address this limitation, we developed IntentBERT-Wiki, a model that augments BERT with external contextual knowledge automatically extracted from Wikipedia, enabling better disambiguation of subtle rhetorical variations and improved classification accuracy where domain-specific indicators are sparse.

This study is guided by two main research questions. The first examines the extent to which analyzing scientific intentions in machine learning publications—measured through the distribution of intent categories, their temporal evolution, and their frequency across disciplines—can reveal trends, discourse regularities, and disciplinary particularities. The second explores how external contextual enrichment, through resources such as Wikipedia, can improve the accuracy, precision, recall, and F1-score of intent classification models in

cross-domain contexts compared with a baseline BERT model.

To highlight the novelty of our work, this study makes three principal contributions. First, we construct a cross-domain annotated corpus of scientific intentions extracted from arXiv publications spanning multiple disciplines. Second, we introduce the IntentBERT-Wiki model, an enriched BERT-based architecture that integrates external contextual knowledge from Wikipedia to resolve rhetorical ambiguities. Third, we conduct a comprehensive cross-domain evaluation, demonstrating that IntentBERT-Wiki consistently outperforms the baseline BERT model, particularly in heterogeneous domains such as biology and physics. In addition, an in-depth analysis of disciplinary discourse patterns reveals how rhetorical intentions vary across scientific domains and how these differences can contribute to improved knowledge organization and retrieval.

2. Methods

We investigate intent classification in machine learning papers through a straightforward methodology combining manual annotation with the analysis of arXiv articles. We filter articles using two criteria: their scientific domain (e.g., physics, biology, computer science) and their contribution to machine learning. This approach ensures a focused analysis of works at the intersection of machine learning and other disciplines, allowing for a comparative study of authors’ intentions.

Figure 1 illustrates our structured methodological framework. After selecting articles based on their domain and relevance to machine learning, an annotation phase is conducted to identify the authors’ intentions. We then apply specialized algorithms to classify and analyze these intentions. Finally, a cross-domain comparison reveals both commonalities and specific features of methodological and technological advances in different disciplines.

To complement this process, a detailed schematic of the proposed architecture has been developed. Each step will be explained in detail in the following sections.

Figure 2 presents the complete architecture of our approach, outlining the various stages—from data processing and annotation to the integration of IntentBERT-Wiki and the comparative analysis

To provide a concise overview of the dataset, Table 1 provides a detailed overview of the dataset, showing the distribution of collected articles across eight scientific domains. For each domain, the table reports the total number of articles, the total number of sentences obtained after segmentation, and the average number of sentences per article. This last metric offers an indication of the typical publication length within each discipline. Together, these figures provide valuable context for the methodological workflow outlined in Figures 1 and 2.

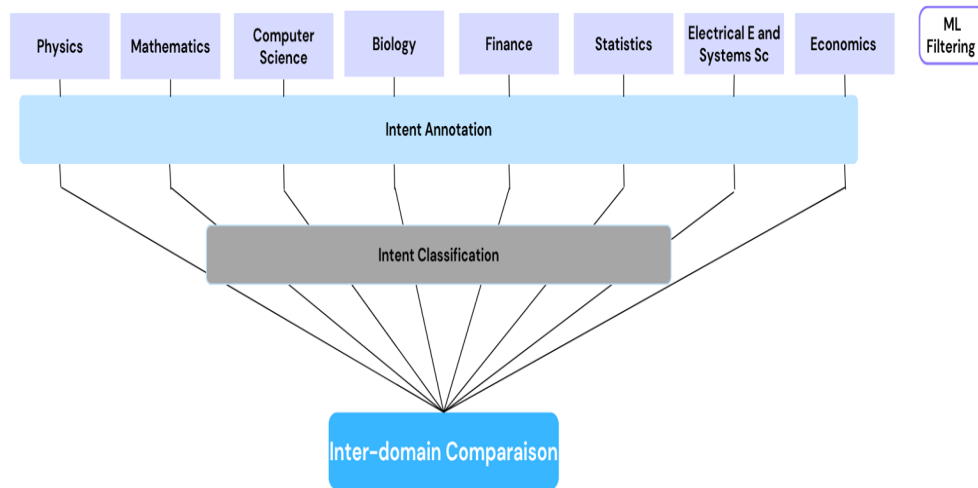


Figure 1. Cross-Domain Comparison in Machine Learning

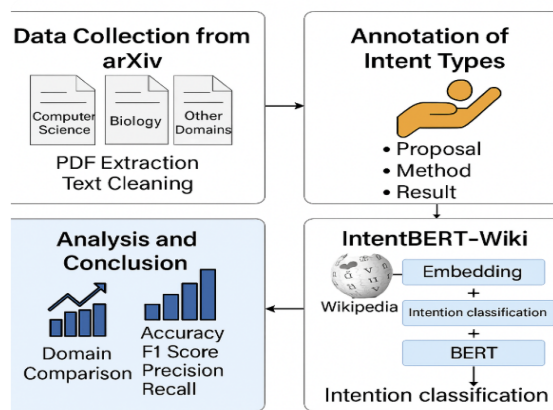


Figure 2. Schema of the Proposed Architecture

Table 1. Distribution of articles and sentences per domain

Scientific Domain	Number of Articles	Total Sentences	Avg. per Article
Biology	100	26178	261.78
Economics	100	75049	750.49
Electrical E	100	50542	505.42
Finance	100	72846	728.46
Computer S	100	85634	856.34
Mathematics	100	76635	766.35
Physics	100	63307	633.07
Statistics	100	70805	708.05
Total	800	520996	-

2.1 Identification of intentions

To build datasets devoted to intention exploration of academic papers within machine learning, we designed a fine-grained categorization across arXiv domain categories, including physics, biology, and finance. Such a structure enables meaningful cross-domain comparative analysis, as it allows to see the underlying intentions between scientific contributions. We also evaluated each article based on criteria that helped us determine what the authors' intentions were and how these have affected the development of machine learning across various domains.

At the heart of our synthesis, we focused on the stated intentions in the papers. Annotation categories included description of new methods, experimental verification, discussion of implications and challenges, and comparison of techniques. It helps interpret each study's aim and highlight innovation patterns within different disciplines.

One major issue with this approach is to classify their intention, such as those discussed in Section 2: new approach presentation, Section 3: model improvement, or Section 4: performance analysis in context. This approach helps us understand research trajectories, but also the ways in which researchers organise their work to meet the specific challenges of an application domain.

Comprehending the contributions of machine learning better is also facilitated by reviewing results reported on articles. In light of explicit intentions, we are better able to interpret the technological and methodological development and identify cross-references across various disciplines. Moreover, by examining the intentions that are presented in the publications, we can find the key focuses of machine learning research. It focuses on the trends in innovation, the problems that contemporary innovators have to solve specific to a number of scientific fields and the preferred innovation methods in science.

Table 2 presents the different categories of intentions used to annotate arXiv publications. It provides a concise description of the identified intent types and their role in structuring and analyzing machine learning trends across diverse scientific disciplines. The intent categories were adapted from established rhetorical role taxonomies in scientific discourse analysis and refined to suit the specific context of machine learning publications on arXiv. Each label was operationalized with explicit lexical patterns to ensure reproducibility. For clarity, we also provide illustrative sentences from our annotated dataset for selected categories, such as *Proposal*: "We propose a novel framework for

multimodal sentiment analysis”, *Result*: “Our approach achieves a 7% improvement in F1-score over the baseline”, and *Limitation*: “Our method struggles with extremely noisy datasets”. These examples demonstrate how each intention type appears in scientific writing and was identified during the annotation process.

Table 2. Labels and Descriptions Used in Annotation

Entity Type	Description
Proposal	The article proposes a new approach, concept, or innovative solution in machine learning.
New Method	Presents a new algorithm, architecture, or machine learning technique.
Result	Highlights experimental results demonstrating the effectiveness or improvement of an approach.
Validation	Evaluates the robustness, reproducibility, or relevance of results through complementary experiments or benchmarks.
Demonstration	Illustrates the application or functioning of a method using examples or use cases.
Analysis	Compares, explores, or examines different approaches in machine learning.
Limitation	Highlights the constraints, weaknesses, or scenarios where a method reaches its limits.
Synthesis	Presents a state of the art, a literature review, or a structured summary of advances in a given field.
Discussion	Analyzes the implications, challenges, limitations, or impacts of a methodology.

This fine-grained intent identification methodology, combined with our two-dimensional filtering strategy based on scientific domain and the expressed intent in publications, allows us to deeply explore how researchers formulate and structure their contributions in machine learning. It aims to refine our understanding of the underlying motivations and objectives behind academic work. By doing so, it highlights emerging trends, methodological orientations, and key contributions across various scientific domains.

2.2 Dataset

In order to construct a new dataset focusing on intentions present in the academic publications, in this case intentions related to the research, we created the following Python script which is able to efficiently download and filter the published articles from arXiv based on two main categories: domain of the scientific area and explicit research intent concerning the machine learning. This two-step filtering allowed us to pinpoint representative datasets from different domains (physics, biology, finance, etc.), and to make sure that reasoning about each publication abounds in machine learning.

The script was used to crawl and structure an equal number of articles by domain in a balanced manner so

that there would be enough diversity to be able to have a comparative analysis of research intentions. Restricting extractions to articles which satisfied both criteria at the same time, we constructed datasets of 100 articles per discipline. This in turn permits a comprehensive investigation of writing fashions and methodological stances of researchers. This systematic approach is critical to know how machine learning is conceived and applied across the variety of scientific knowledge domains.

The tool uses the arXiv API for targeted searches, and BeautifulSoup [24] to parse XML responses and extract key metadata for articles. When detected, the documents are downloaded and saved in a central location for processing and marking. This automatic structuring guarantees a consistent and coherent extraction of the research intents expressed by the authors.

After collection, there is a preprocessing step to prepare it for analysis. Each PDF is transformed into raw, plain text and then cleaned by removing unwanted characters and normalizing sentence structures. It segments analysis content in a controlled manner to alleviate potential memory and computation constraints offered by other NLP models (e.g. SpaCy [25]). This requires to extract conveniently and accurately declared intentions and to annotate and classify them.

Automating retrieval and structuring helps identify challenges in each domain and spot cross-domain trends. This methodology can help discussing the dynamics of research in machine learning, and interdisciplinary cross-fertilization, and induce novel scientific, methodological synergies.

2.3 Data pre-processing

After we have managed to scrape structured content from the PDF files and converted it into a format we can use within our data processing workflow, our second step consists in identifying the intentions that are expressed in the publications. This is necessary in order to organise and analyse the scientific discourse around machine learning and to get insights into the goals the researchers are working towards.

We classify intents in a few major categories in order to capture the breadth of academic inputs. These can be reports of new methods or techniques, new experimental results, or comparative analyses and help to sharpen the questions, and the debate, of what can be achieved, and what remains to be done. Through this we can explore how machine learning is thought about and utilized across scientific fields.

We extract and structure this information by means of a rule-based text extraction and NLP models applied to the cleaned texts. A list of relevant text segments is identified, annotated, and arranged in a DataFrame, which forms the basis of our analyses. This systematic organization allows intentions to be compared between

different disciplines and helps to highlight the current trends in machine learning research.

This methodological approach provides a sound data preparation that will allow for a profound investigation of academic intentions from various fields of study. In revealing the motivations of researchers, such an approach also paves the way to more profound comprehension of research dynamics and trends in machine learning.

2.4 IntentBert-Wiki: our proposed algorithm

The task of scientific intent classification has recently received some attention in the natural language processing (NLP) community [26], and has been addressed in the context of tasks related to the rhetorical structure of academic papers. As BERT [27] has accomplished state-of-the-art results on a variety of NLP benchmarks, our previous study [18] found that BERT also outperforms RoBERTa [28] and DistilBERT [29] on intent recognition tasks in the context of the machine learning literature. Such results have confirmed BERT as a strong baseline frame of both accuracy and generalization capabilities to predict and classify rhetorical roles such as Proposal, Result or New Method in connection to scientific text.

To enhance this capability, we propose a new architecture called IntentBERT-Wiki, which integrates external contextual knowledge from Wikipedia [30] into the BERT language modeling framework. Unlike conventional approaches that rely solely on the input sentence, IntentBERT-Wiki incorporates external semantic cues to provide enriched representations capable of resolving subtle variations in author intent. These cues are obtained by querying Wikipedia for each input sentence and extracting a short summary, which is then encoded into a vector representation using a sentence transformer. To ensure relevance, each query to Wikipedia is automatically generated from the input sentence by extracting its main content words and named entities, removing stop words and generic terms. We retrieve the top five candidate pages using the Wikipedia API, then compute the cosine similarity between the MiniLM embedding of the original sentence and the embedding of each candidate summary. The page with the highest similarity score above a predefined threshold ($\tau = 0.35$) is selected; otherwise, a null vector is assigned. This mechanism ensures that only semantically relevant summaries contribute to the external context, and that noisy or unrelated Wikipedia content does not bias the model's representations. It is important to note that Wikipedia is not used as a source of scientific content, but purely as an automatic contextual resource to enrich semantic representations. This contextual vector, combined with BERT's internal encoding, improves the model's ability to capture discourse-level signals relevant to intent classification.

From an ethical and reproducibility perspective, we acknowledge that Wikipedia, while a rich and openly

accessible resource, may contain cultural, linguistic, or editorial biases that could influence NLP models.

Additionally, to further mitigate potential bias, we cross-check the contextual information retrieved from Wikipedia with curated scientific databases such as PubMed, IEEE Xplore, or DBpedia. Leveraging domain-specific ontologies where available strengthens the accuracy and reliability of external enrichment, ensuring that noisy or biased content does not propagate through the model and that the semantic representations remain consistent with established scientific knowledge. To mitigate these risks, our approach systematically filters and validates retrieved content using semantic similarity thresholds, ensuring that only highly relevant and contextually appropriate information is retained. Furthermore, all preprocessing scripts, Wikipedia retrieval parameters, and model configurations are documented and shared to facilitate reproducibility and allow independent verification of results.

Our architecture retains the core BERT encoder, but augments each sentence with its corresponding vector derived from Wikipedia. This hybrid representation is then passed through a classification layer to produce intent predictions. If no summary is available for a given sentence, a null vector is used as a fallback mechanism. This design ensures robustness and consistency across scientific domains.

The IntentBERT-Wiki algorithm proceeds in sequential phases, including context extraction from Wikipedia, construction of joint representations, model training, and evaluation using standard classification metrics such as accuracy, precision, recall, and F1-score. Through k-fold cross-validation, our model consistently demonstrates better generalization in interdisciplinary fields such as biology and computer science. These results confirm the value of external knowledge enrichment in modeling scientific intent.

The proposed IntentBERT-Wiki algorithm integrates pretrained contextual representations from BERT with external knowledge extracted from Wikipedia to enhance intent classification across scientific domains. The process can be described in four main phases.

In the initial state, all sentences $s \in S$ and their corresponding intent labels L are extracted from the annotated dataset D . Each sentence s is associated with a contextual vector $C(s)$, which is initialized to zero. During Phase 1 – *Knowledge Embedding Extraction*, the algorithm retrieves relevant summaries from Wikipedia for each sentence. For every $s \in S$, a query is sent to Wikipedia to obtain a short description $W(s)$. When a summary is found, it is encoded using the MiniLM model to generate an external context vector $C(s)$. If no summary is available, the contextual vector remains null.

In Phase 2 – *Representation Enrichment*, each sentence s is tokenized using the BERT tokenizer. The pooled embedding $B(s)$ generated by BERT represents the

internal semantic features of the sentence. The two vectors $B(s)$ and $C(s)$ are then concatenated to form an enriched embedding $E(s)$, which incorporates both linguistic and external contextual knowledge. The third phase – *Training and Evaluation* applies k-fold cross-validation on the dataset D . For each fold, the model is trained using the enriched representations $E(s)$ and their corresponding intent labels. The trained model is then evaluated on the validation fold to measure predictive performance.

Finally, in Phase 4 – *Performance Computation*, evaluation metrics including accuracy, precision, recall, and F1-score are computed for each fold. The process is repeated until all folds have been processed, ensuring a robust estimation of model performance. To complement the description of the algorithm previously provided, Figure 3 illustrates the overall architecture of the proposed IntentBERT-Wiki model. The process begins with the preparation of the annotated dataset, followed by the extraction of contextual knowledge from Wikipedia and the generation of corresponding embeddings. These external embeddings are then combined with BERT representations to form enriched sentence vectors, which are used for model training, evaluation, and performance measurement.

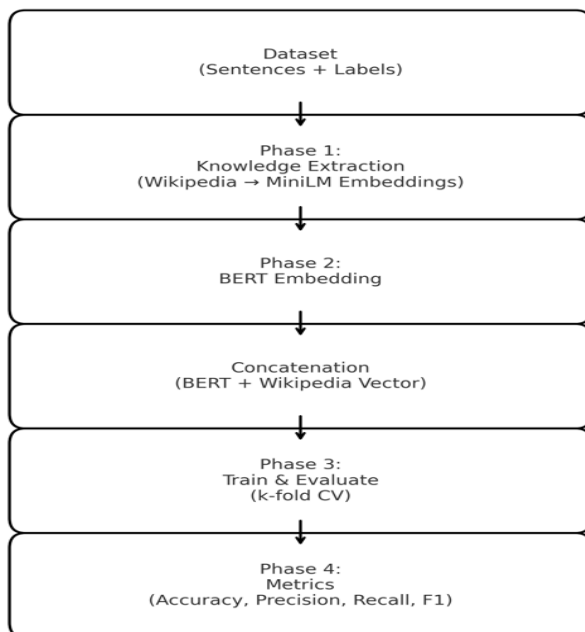


Figure 3. IntentBERT-Wiki architecture

To ensure transparency and reproducibility, the dataset, preprocessing scripts, and the implementation of the proposed IntentBERT-Wiki algorithm will be made available after the completion of the extended version of this research. This release will include detailed documentation to facilitate further experimentation and validation.

3. Results and Discussions

In this section, the explosion of scientific knowledge calls for methods capable of synthesizing information

more effectively. Our research applies machine learning techniques to data from various scientific domains on arXiv, revealing interesting trends across disciplines.

The comparative analysis of intent classification tasks highlights varying performance levels depending on the domain, as shown in table 3. The computer science domain ranks highest, with particularly strong metrics: an accuracy of 0.9653, an F1-score of 0.9585, a precision of 0.9575, and a recall of 0.9601.

The fields of physics and mathematics also show strong performance, although slightly lower than that of computer science. These results are linked to the relatively structured nature of their data, which partially favors machine learning models while still preserving a degree of variability. Discipline-specific adaptations could further improve the quality of intent classification in these areas.

Table 3. Model Performance Metrics across Different Academic Domains

Domain	Accuracy	F1-score	Precision	Recall
Physics	0.9494	0.9267	0.9301	0.9248
Mathematics	0.9326	0.9225	0.9320	0.9161
Computer Sc	0.9653	0.9685	0.9575	0.9601
Biology	0.9142	0.8740	0.8780	0.9275
Finance	0.9611	0.9371	0.9526	0.7887
Economics	0.9686	0.9640	0.9722	0.9565
Statistics	0.9567	0.9521	0.9543	0.9521
Electrical/sys	0.9430	0.8891	0.9283	0.8769

In contrast, biology, economics, and finance display more modest performance levels, with respective accuracy scores of 0.9142, 0.9686, and 0.9611. This can be explained by the semantic complexity and stylistic diversity of writing within these fields. Biology, for example, frequently contains highly specialized terminology (e.g., gene or protein names), long and compound sentence structures combining experimental procedures with results, and domain-specific abbreviations that are not easily interpretable without specialized resources such as ontologies. Finance exhibits strong variability in reporting styles and is influenced by contextual market events, which introduce noise into the data. Economics often blends qualitative reasoning with quantitative analysis, creating mixed rhetorical patterns that complicate intent recognition.

These observations highlight that biology represents one of the most heterogeneous and linguistically complex domains. The coexistence of sub-disciplines (e.g., molecular biology, genetics, biochemistry) generates diverse rhetorical patterns, making it difficult for the model to generalize intent cues across contexts. Additionally, biological abstracts often merge methodological descriptions with experimental results, reducing the clarity of intent boundaries. From a practical standpoint, adapting the model to such heterogeneous domains may require domain-specific fine-tuning, integration of biomedical ontologies, or the use of hybrid representations that combine linguistic and knowledge-based features. These strategies could

strengthen the model's robustness and interpretability in complex scientific fields.

The fields of statistics and electrical/systems engineering show heterogeneous results. Statistics achieve high performance (accuracy of 0.9567), likely due to the homogeneity of quantitative data. In contrast, the electrical and systems engineering domain suffers from a less consistent writing style and frequent inclusion of hardware-specific terminology, which may reduce the model's ability to identify intent with high precision.

Overall, this interdisciplinary analysis highlights the importance of data quality and the need for domain-specific adaptations. Disciplines with well-structured and established corpora—such as computer science or statistics—achieve better results. Conversely, less homogeneous fields require more contextualized approaches to optimize the performance of intent classification models.

Building on BERT's strong baseline performance established in our previous experiments [27], IntentBERT-Wiki leverages external contextual knowledge from Wikipedia to address limitations in domains where rhetorical cues are sparse. This enrichment enhances the model's capacity to capture subtle intent signals, leading to expected performance gains in heterogeneous disciplines such as biology and electrical engineering, while maintaining strong results in structured fields like computer science and statistics.

Beyond these results, IntentBERT-Wiki can be adapted to other scientific repositories, such as PubMed, IEEE Xplore, or Scopus, by fine-tuning the model on domain-specific corpora. Furthermore, its architecture is compatible with multilingual transformer models, enabling its extension to multilingual corpora for cross-lingual intent classification. Such adaptations would enhance its applicability to a broader range of research communities and facilitate global knowledge discovery.

4. Conclusions

In this study, we introduced IntentBERT-Wiki, an enriched variant of BERT designed to improve intent classification in scientific articles. By combining contextual representations from BERT with semantic vectors generated from automatically extracted Wikipedia summaries, our approach adds a layer of general knowledge useful for fine-grained interpretation of authors' intentions. This fusion enhances the understanding of the scientific context expressed in each sentence.

The results confirm the relevance of this integration, particularly in structured domains such as computer science and more heterogeneous ones like biology, by offering improved robustness in intent analysis. Using Wikipedia as an open knowledge source proved both accessible and effective, providing linguistic enrichment without relying on specialized databases.

This study was guided by two research questions. First, we examined the extent to which analyzing scientific intentions can reveal trends, discourse regularities, and disciplinary specificities. Our findings highlight clear cross-domain patterns that enhance the understanding of research dynamics in machine learning. Second, we evaluated whether external contextual enrichment could improve intent classification performance. Results show that IntentBERT-Wiki significantly outperforms the baseline BERT model in all domains, particularly where rhetorical cues are sparse, confirming the value of contextual integration.

For future work, we plan to fully automate the intent annotation process, which currently relies on manually defined keywords. This automation, through supervised or reinforcement learning models, would make our system more scalable and replicable on a larger scale. We also intend to explore other open sources, such as Wikidata or scientific abstracts, to increase the coverage and relevance of the contextual vectors.

Ultimately, IntentBERT-Wiki represents a step forward in the fine-grained understanding of scientific publications and paves the way for new methods of academic discourse analysis supported by accessible and integrable general knowledge.

Beyond the scope of this study, IntentBERT-Wiki has broader implications for the organization and discovery of scientific knowledge. The ability to automatically identify the rhetorical intentions of academic texts can support applications such as intent-aware academic search engines, research synthesis platforms, and intelligent recommendation systems. These systems could help researchers retrieve more relevant studies, trace methodological evolution across disciplines, and enhance evidence-based decision-making. In the long term, intent-based text understanding may contribute to the development of intelligent infrastructures for managing and interconnecting scientific knowledge across domains.

Acknowledgements

The authors declare no conflict of interest.

References

- [1] B. S. Lubis and A. Yus, "Mapping knowledge and research trend on technology adoption in higher education: A bibliometric analysis," *Educ. Inf. Technol.*, Jun. 2024, doi: 10.1007/s10639-024-12801-0.
- [2] Cornell University Library, "arXiv.org e-Print archive," [Online]. Available: <https://arxiv.org/>. Accessed: Jun. 20, 2025.
- [3] H. Gbada, K. Kalti, and M. A. Mahjoub, "Deep learning approaches for information extraction from visually rich documents: Datasets, challenges and methods," *IJDAR*, vol. 28, no. 1, pp. 121–142, Mar. 2025, doi: 10.1007/s10032-024-00493-8.
- [4] H. Zhu and L. Lei, "The research trends of text classification studies (2000–2020): A bibliometric analysis," *Sage Open*, vol. 12, no. 2, Art. no. 21582440221089963, Apr. 2022, doi: 10.1177/21582440221089963.
- [5] C. B. Chandrakala, R. Bhardwaj, and C. Pujari, "An intent recognition pipeline for conversational AI," *Int. J. Inf.*

- Technol., vol. 16, no. 2, pp. 731–743, Feb. 2024, doi: 10.1007/s41870-023-01642-8.
- [6] Y. Wang, L. Zhou, W. Zhang, F. Zhang, and Y. Wang, “A soft prompt learning method for medical text classification with simulated human cognitive capabilities,” *Artif. Intell. Rev.*, vol. 58, no. 4, Art. no. 118, Jan. 2025, doi: 10.1007/s10462-025-11121-0.
- [7] V. Desai, S. Wadhwa, A. A., and B. Bajaj, “Text-based intent analysis using deep learning,” *IJISRT*, vol. 5, no. 7, pp. 267–274, Jul. 2020, doi: 10.38124/IJISRT20JUL342.
- [8] C. Wang, S.-J. Cai, B.-X. Shi, and Z.-H. Chong, “Visual topic semantic enhanced machine translation for multi-modal data efficiency,” *J. Comput. Sci. Technol.*, vol. 38, no. 6, pp. 1223–1236, Dec. 2023, doi: 10.1007/s11390-023-1302-6.
- [9] M. D. Oliveira, A. I. Tavares, R. Miranda, R. Trindade, and A. C. L. Vieira, “A novel policy dialogue to build sustainable and resilient health systems: Findings from PHSSR Portugal,” *Health Res. Policy Syst.*, vol. 23, no. 1, Art. no. 58, May 2025, doi: 10.1186/s12961-025-01329-5.
- [10] A. Hao, H. Yuan, S. C. Hui, and J. Su, “Effective type label-based synergistic representation learning for biomedical event trigger detection,” *BMC Bioinformatics*, vol. 25, no. 1, Art. no. 251, Jul. 2024, doi: 10.1186/s12859-024-05851-1.
- [11] T. Sellhorn, “Machine learning und empirische Rechnungslegungsforschung: Einige Erkenntnisse und offene Fragen,” *Schmalenbachs Z. betriebswirtsch. Forsch.*, vol. 72, no. 1, pp. 49–69, Mar. 2020, doi: 10.1007/s41471-020-00086-1.
- [12] R. Nacheva, M. Czaplowski, and P. Petrov, “Data mining model for scientific research classification: The case of digital workplace accessibility,” *Decision*, vol. 51, no. 1, pp. 3–16, Mar. 2024, doi: 10.1007/s40622-024-00378-z.
- [13] R. Lenart-Gansiniec, W. Czakon, and N. Meyer, “Antecedents of researchers’ behavioral intentions to use crowdsourcing in science: A multilevel approach,” *Rev. Manag. Sci.*, vol. 19, no. 5, pp. 1411–1445, May 2025, doi: 10.1007/s11846-024-00797-3.
- [14] M. A. Ayanwale, R. R. Molefi, and S. Oyeniran, “Analyzing the evolution of machine learning integration in educational research: A bibliometric perspective,” *Discov. Educ.*, vol. 3, no. 1, Art. no. 47, May 2024, doi: 10.1007/s44217-024-00119-5.
- [15] Z. Zhong, H. Guo, and K. Qian, “Deciphering the impact of machine learning on education: Insights from a bibliometric analysis using bibliometrix R-package,” *Educ. Inf. Technol.*, vol. 29, no. 16, pp. 21995–22022, Nov. 2024, doi: 10.1007/s10639-024-12734-8.
- [16] O. Ovtšarenko, “Opportunities of machine learning algorithms for education,” *Discov. Educ.*, vol. 3, no. 1, Art. no. 209, Nov. 2024, doi: 10.1007/s44217-024-00313-5.
- [17] S. B. Yoon, J. Lee, H.-C. Lee, C.-W. Jung, and H. Lee, “Comparison of NLP machine learning models with human physicians for ASA physical status classification,” *npj Digit. Med.*, vol. 7, no. 1, Art. no. 259, Sep. 2024, doi: 10.1038/s41746-024-01259-6.
- [18] S. Liu, L. Li, M. Fang, C.-C. Hung, and S. Yang, “A prompt learning-based intent recognition method on a Chinese implicit intent dataset CIID,” *Neural Process. Lett.*, vol. 55, no. 8, pp. 11017–11034, Dec. 2023, doi: 10.1007/s11063-023-11362-6.
- [19] R. Qi et al., “Multi-task learning model for citation intent classification in scientific publications,” *Scientometrics*, vol. 128, no. 12, pp. 6335–6355, Dec. 2023, doi: 10.1007/s11192-023-04858-4.
- [20] T. Ghosal, K. K. Varanasi, and V. Kordoni, “A deep multi-tasking approach leveraging cited–citing paper relationships for citation intent classification,” *Scientometrics*, vol. 129, no. 2, pp. 767–783, Feb. 2024, doi: 10.1007/s11192-023-04811-5.
- [21] N. El-Rashidy, Z. Tarek, A. M. Elshewey, and M. Y. Shams, “Multitask multilayer-prediction model for predicting mechanical ventilation and the associated mortality rate,” *Neural Comput. Appl.*, vol. 37, no. 3, pp. 1321–1343, Jan. 2025, doi: 10.1007/s00521-024-10468-9.
- [22] C. Zhou et al., “A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT,” *Int. J. Mach. Learn. Cybern.*, Nov. 2024, doi: 10.1007/s13042-024-02443-6.
- [23] N. Khouya, A. Retbi, and S. Bennani, “Contextual revelation of intentions in machine learning articles on arXiv: A new dimension in academic content analysis,” in *Proc. 2024 IEEE 12th Int. Symp. Signal, Image, Video and Communications (ISIVC)*, 2024, pp. 1–6, doi: 10.1109/ISIVC57473.2024.10577900.
- [24] H. M. Tanrikulu and H. Pabuccu, “The effect of data types on the performance of machine learning algorithms for cryptocurrency prediction,” *Comput. Econ.*, Mar. 2025, doi: 10.1007/s10614-025-10919-y.
- [25] S. Ferro, R. Giovanelli, M. Leeson, M. De Bernardin, and A. Traviglia, “A novel NLP-driven approach for enriching artefact descriptions, provenance, and entities in cultural heritage,” *Neural Comput. Appl.*, Aug. 2025, doi: 10.1007/s00521-025-11449-2.
- [26] D. Khurana, A. K. Koli, K. Khatter, and S. Singh, “Natural language processing: State of the art, current trends and challenges,” *Multimed. Tools Appl.*, vol. 82, no. 3, pp. 3713–3744, Jan. 2023, doi: 10.1007/s11042-022-13428-4.
- [27] S. Gorla, S. S. Tangeda, L. B. M. Neti, and A. Malapati, “Telugu named entity recognition using BERT,” *Int. J. Data Sci. Anal.*, vol. 14, no. 2, pp. 127–140, Aug. 2022, doi: 10.1007/s41060-021-00305-w.
- [28] S. Chopra, P. Agarwal, J. Ahmed, S. S. Biswas, and A. J. Obaid, “RoBERTa and BERT: Revolutionizing mental healthcare through natural language,” *SN Comput. Sci.*, vol. 5, no. 7, Art. no. 889, Sep. 2024, doi: 10.1007/s42979-024-03202-8.
- [29] M. Jojoa, P. Eftekhari, B. Nowrouzi-Kia, and B. Garcia-Zapirain, “Natural language processing analysis applied to COVID-19 open-text opinions using a DistilBERT model for sentiment categorization,” *AI Soc.*, vol. 39, no. 3, pp. 883–890, Jun. 2024, doi: 10.1007/s00146-022-01594-w.
- [30] F. Petroni et al., “Improving Wikipedia verifiability with AI,” *Nat. Mach. Intell.*, vol. 5, no. 10, pp. 1142–1148, Oct. 2023, doi: 10.1038/s42256-023-00726-1.