# Enhancing Scientific Research through Knowledge-Informed AI

**Artemis Dampa**[1,2;*]

[1]Institute of Informatics and Telecommunications, "NCSR" Demokritos, Greece
[2]Department of Informatics and Telematics, Harokopio University of Athens, Greece

## 1 Introduction

Artificial Intelligence (AI) is rapidly reshaping the landscape of scientific research [10] by supporting and accelerating several key processes: literature discovery, experimental design, data analysis, and knowledge extraction. As scientific data continues to grow in volume and complexity, there is a growing need for intelligent systems that can assist researchers in navigating this information efficiently.

This work investigates how AI systems enriched with structured or domain-specific expertise (i.e. knowledge-informed) can optimize the organization and indexing of scientific documents, thereby improving literature search. The goal of the current study is to enable researchers to devote more time to high-level cognitive tasks, such as hypothesis generation and creative exploration, while delegating repetitive or computationally intensive steps to intelligent models.

To this end, the current study explores a range of methodologies including machine learning models infused with prior knowledge (e.g. hierarchical taxonomies [1]), knowledge retrieval systems that extract information from scientific texts [5], active learning algorithms [8] that iteratively recommend relevant scientific documents based on minimal annotation or interactive user feedback [4, 3]. Implementing these approaches requires a diverse set of data sources, including open repositories such as ACL Anthology, publicly available datasets (e.g. Forc4cl [1]), and custom datasets developed in collaboration with domain experts.

The core research questions that guide this investigation are:

**RQ1:** How can explicit (e.g., published research and results, equations) and implicit (e.g., personal experience and scientific intuition) prior knowledge of experts be represented and integrated to improve search and retrieval, and classification tasks?

**RQ2:** How can natural language processing (NLP), machine learning, and knowledge representation enable extraction of expert knowledge and its effective use?

**RQ3:** How can the information need of a scientist be effectively modeled in relation to their specific field of interest?

**RQ4:** How can a search and retrieval system be designed to extract and utilize relevant knowledge from various data types (structured, semi-structured, unstructured) to support scientific inquiry?

**RQ5:** How can the efficiency of scientific workflows be quantified and how can AI-based support systems be evaluated for their impact on these workflows?

* Corresponding Author. Email: adampa@iit.demokritos.gr

The ultimate aim is to develop methods for: (a) the efficient representation and management of expert knowledge, enabling its use in recommendation systems, particularly to support new researchers or promote best practices; and (b) effective search and retrieval of relevant literature and data, facilitating the creation of predictive models (e.g., via transfer learning), especially in domains with limited available data. Together, these approaches aim to shorten the research and methodology development lifecycle by promoting the reuse and integration of prior knowledge and data throughout the scientific process.

## 2 Key Contributions

As a foundational step toward knowledge-informed AI that supports scientific research, the current work focuses on organizing and indexing scholarly documents, which is central to literature discovery. By improving how scientific articles are classified and structured, the aim is to enable more effective document retrieval in response to expert queries. This effort directly addresses RQ1 and RQ2.

Specifically, the problem of fine-grained hierarchical multi-label classification of scientific articles [2], using structured taxonomies [1] is studied. Unlike standard hierarchical classification [11], the specific task is characterized by three complexities: (a) documents may be assigned multiple labels; (b) labels may appear at any level of the hierarchy; and (c) incomplete paths are allowed, i.e. there is no requirement for labeled documents to have leaf-only labels.

To handle this, a cascade classification architecture (see Figure 1), inspired by [6] informed by the label hierarchy is implemented and systematically compared against a flat counterpart. This approach integrates multiple strategies to infuse prior knowledge into modeling.

**Document Representation:** Each article is represented using SPECTER2 [9], a transformer-based model pretrained on scientific literature. Inputs include title, abstract, and selected metadata fields (e.g., authors, publisher, venue). This generates semantically rich embeddings capturing high-level scientific discourse.

**Cascade Classification with Hierarchical Sampling:** For each label node in the taxonomy, a dedicated logistic regression classifier is trained as part of a cascading framework. To improve label discrimination and reflect the taxonomy, hierarchy-aware negative sampling is applied, where documents explicitly labeled with the current node are selected as *positives*, while documents from sibling categories or parent-exclusive samples (i.e., labeled with the parent but not with the current or sibling nodes) are selected as *negatives*. In the absence of siblings, sibling categories of the parent node are used to maintain contrastive training. This structure

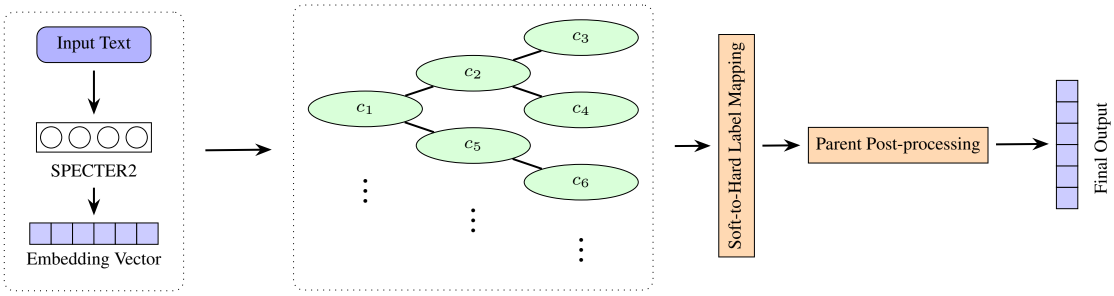


**Figure 1.** Diagram of the hierarchical multi-label classification process. The figure illustrates the stages of document representation, node-specific classifiers $c_i$ training, soft-to-hard label mapping, hierarchy-enforcing post-processing, and generation of final output.

| **Method** | **Micro** | | | **Macro** | | | **Weighted** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Prec** | **Rec** | **F1** | **Prec** | **Rec** | **F1** | **Prec** | **Rec** | **F1** |
| Flat baseline | **0.803** | 0.601 | **0.687** | **0.625** | 0.392 | **0.467** | **0.790** | 0.601 | **0.673** |
| Hierarchical | 0.668 | **0.679** | 0.673 | 0.521 | **0.413** | 0.446 | 0.665 | **0.679** | 0.664 |

**Table 1.** Evaluation of flat baseline and hierarchical method, using micro, macro, and weighted precision, recall, and F1 score.

ensures that each classifier learns to make fine-grained distinctions between closely related categories, reducing misclassification among semantically similar labels.

**Soft-to-Hard Label Mapping:** Each classifier produces a confidence score (in $[0, 1]$). These are converted to discrete label predictions using one of three strategies: (a) thresholding-based selection; (b) top-k label selection; and (c) LLM-guided selection.

**Hierarchy-Enforcing Post-Processing:** To ensure structural consistency, one of several hierarchy-enforcing policies is applied: from no changes to predictions, to recursively adding ancestor labels, to enforcing full parent inclusion, or requiring at least one parent label.

These strategies allow us to flexibly adjust the degree to which hierarchy is enforced.

An experimental evaluation, conducted on a curated set of NLP research papers from the ACL Anthology, demonstrates that the hierarchical approach improves recall and reduces false negatives compared to the flat counterpart, as presented in Table 1. Although the flat approach is superior in terms of precision and F1, integrating prior knowledge from domain taxonomies can improve the relevance of retrieved documents that could otherwise be missed, which constitutes a key asset for scientific discovery tasks where missing relevant work might be more costly than reviewing a few false positives.

Statistical significance testing confirms these trends, and the findings underscore that the method of hierarchical information infusion can significantly influence the classification outcome.

In summary, the contribution of current work lies in demonstrating that structured prior knowledge, when effectively integrated (e.g. via the proposed post-processing), enhances the classification of scientific literature. This lays the groundwork for AI systems to support researchers in navigating the ever-expanding scientific landscape.

## 3 Future Directions

Building on the current focus of enhancing scientific workflows through knowledge-informed AI, future work will further explore both methodological advancements and expanded application areas.

The current findings indicate that incorporating hierarchical knowledge improves recall in scientific document classification, but not precision or F1, highlighting challenges in balanced performance. As a next step, the plan is to investigate more efficient representations of hierarchical information, moving beyond rigid parent-child constraints. This includes exploring graph neural networks and attention-based models that can better model inter-label dependencies across levels of the taxonomy [12], and integrating knowledge-informed features, such as metadata-derived embeddings and citation network information [7], to enrich the input representations and improve generalization. These enhancements aim to strengthen the model in leveraging structure without sacrificing prediction quality.

To support dynamic literature discovery (addressing RQ3 and RQ4), future work will implement active learning frameworks that recommend documents, iteratively, based on the evolving information needs of researchers. This involves (a) developing adaptive user modeling techniques that infer a researcher's intent based on their queries, prior interactions, domain-specific terminology, and current stage in the research process; and (b) implementing active learning-based retrieval systems that iteratively refine document recommendations by incorporating minimal user feedback, enabling the system to surface increasingly relevant literature over time.

Finally, to address RQ5, future work will explore methods for quantitatively evaluating the efficiency and effectiveness of AI-supported research processes, defining task-specific metrics for different stages of the scientific workflow (e.g. time saved in literature review) and collecting user-centered feedback from domain experts to assess how AI influences productivity and innovation outcomes.

## 4 Acknowledgments

This work was supported by the European Union's Digital Europe Programme (DIGITAL) under grant agreement No 101146490. I would like to thank my supervisors G. Giannakopoulos, I. Varlamis and V. Karkaletsis for their valuable guidance throughout this work.

## References


[1] Raia Abu Ahmad, Ekaterina Borisova, and Georg Rehm, 'Forc4cl: A fine-grained field of research classification and annotated dataset of nlp articles', in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, eds., Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, p. 7389–7394, Torino, Italia, (May 2024). ELRA and ICCL.

[2] Raia Abu Ahmad, Ekaterina Borisova, and Georg Rehm, 'Forc@nslp2024: Overview and insights from the field of research classification shared task', in *Natural Scientific Language Processing and Research Knowledge Graphs*, eds., Georg Rehm, Stefan Dietze, Sonja Schimmler, and Frank Krüger, p. 189–204, Cham, (2024). Springer Nature Switzerland.

[3] Evangelos Axiotis, Andreas Kontogiannis, Eleftherios Kalpoutzakis, and George Giannakopoulos, 'A personalized machine-learning-enabled method for efficient research in ethnopharmacology. the case of the southern balkans and the coastal zone of asia minor', *Applied Sciences*, **11**, 5826, (06 2021).

[4] Soumen Chakrabarti, Martin van den Berg, and Byron Dom, 'Focused crawling: a new approach to topic-specific web resource discovery', *Computer Networks*, **31**(11), 1623–1640, (1999).

[5] Gautier Izacard, Mathilde Caron, Lucas Hosseini, Sebastian Riedel, Piotr Bojanowski, Armand Joulin, and Edouard Grave, 'Unsupervised dense information retrieval with contrastive learning', *Transactions on Machine Learning Research*, (2022).

[6] Aris Kosmpoulos, Georgios Paliouras, and Ion Androutsopoulos, 'The effect of dimensionality reduction on large scale hierarchical classification', in *Information Access Evaluation. Multilinguality, Multimodality, and Interaction*, eds., Evangelos Kanoulas, Mihai Lupu, Paul Clough, Mark Sanderson, Mark Hall, Allan Hanbury, and Elaine Toms, pp. 160–171, Cham, (2014). Springer International Publishing.

[7] Malte Ostendorff, Nils Rethmeier, Isabelle Augenstein, Bela Gipp, and Georg Rehm, 'Neighborhood contrastive learning for scientific document representations with citation embeddings', in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, eds., Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, pp. 11670–11688, Abu Dhabi, United Arab Emirates, (December 2022). Association for Computational Linguistics.

[8] Burr Settles, 'Active learning literature survey', (2009).

[9] Amanpreet Singh, Mike D'Arcy, Arman Cohan, Doug Downey, and Sergey Feldman, 'Scirepeval: A multi-format benchmark for scientific document representations', in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, eds., Houda Bouamor, Juan Pino, and Kalika Bali, p. 5548–5566, Singapore, (December 2023). Association for Computational Linguistics.

[10] Yongjun Xu, Xin Liu, Xin Cao, Changping Huang, Enke Liu, Sen Qian, Xingchen Liu, Yanjun Wu, Fengliang Dong, Cheng-Wei Qiu, Junjun Qiu, Keqin Hua, Wentao Su, Jian Wu, Huiyu Xu, Yong Han, Chenguang Fu, Zhigang Yin, Miao Liu, Ronald Roepman, Sabine Dietmann, Marko Virta, Fredrick Kengara, Ze Zhang, Lifu Zhang, Taolan Zhao, Ji Dai, Jialiang Yang, Liang Lan, Ming Luo, Zhaofeng Liu, Tao An, Bin Zhang, Xiao He, Shan Cong, Xiaohong Liu, Wei Zhang, James P. Lewis, James M. Tiedje, Qi Wang, Zhulin An, Fei Wang, Libo Zhang, Tao Huang, Chuan Lu, Zhipeng Cai, Fang Wang, and Jiabao Zhang, 'Artificial intelligence: A powerful paradigm for scientific research', *The Innovation*, **2**(4), 100179, (2021).

[11] Alessandro Zangari, Matteo Marcuzzo, Matteo Rizzo, Lorenzo Giudice, Andrea Albarelli, and Andrea Gasparetto, 'Hierarchical text classification and its foundations: A review of current research', *Electronics*, **13**(7), (2024).

[12] Yu Zhou, Di Jin, Jianguo Wei, Dongxiao He, Zhizhi Yu, and Weixiong Zhang, 'Generalized taxonomy-guided graph neural networks', in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, ed., Kate Larson, pp. 2616–2624. International Joint Conferences on Artificial Intelligence Organization, (8 2024). Main Track.