Analyzing Research Trends in Deep Learning using Natural Language Processing and Citation Analysis (2019-2024).

1. Introduction

Deep learning (DL), a subset of machine learning, leverages artificial neural networks with multiple layers to automatically extract complex patterns from data, enabling state-of-the-art performance across various domains (LeCun et al., 2015). This paradigm has significantly impacted fields such as computer vision (Krizhevsky, Sutskever, & Hinton, 2012), natural language processing (Brown et al., 2020), autonomous driving (LeCun et al., 2015), medical diagnostics (Esteva et al., 2017), and precision agriculture (Kamilaris & Prenafeta-Boldú, 2018).

With rapid advancements in computational resources and the proliferation of large-scale datasets, the adoption of deep learning has accelerated, contributing to an exponential growth in scholarly publications. As these methods permeate diverse scientific and industrial sectors, understanding how the focus of DL research evolves over time becomes increasingly important. Analyzing trends in research themes, as well as evaluating the influence of factors such as publication year, topic, and journal on citation impact, offers valuable insights into the trajectory and influence of this rapidly evolving field.

This study aims to examine deep learning research trends between 2019 and 2024 by applying natural language processing (NLP) and topic modeling techniques to a curated dataset of 500 paper titles retrieved from the Scopus database. The project explores key thematic structures and co-occurrence patterns among terms used in different deep learning research. Furthermore, by incorporating citation data, we investigate the scholarly influence of different research directions, thereby offering a quantitative understanding of both emerging and impactful areas within the DL landscape.

2. Background and Related Work

Deep learning has emerged as a cornerstone of modern artificial intelligence, offering unprecedented capabilities in areas such as image classification, speech recognition, and machine translation (LeCun et al., 2015). Its success is underpinned by the increasing availability of big data, high-performance computing, and algorithmic innovations that allow for deeper architectures and more effective training. The resulting surge in research output requires robust, automated techniques to synthesize and interpret trends within the DL corpus.

Text mining and topic modeling have proven effective for mapping thematic structures in scientific literature. Tools such as Latent Dirichlet Allocation (LDA) help uncover hidden topics in large text corpora (Blei, Ng, & Jordan, 2003), while techniques like keyword co-occurrence networks and word cloud visualizations reveal prominent concepts and their interrelations (Griffiths & Steyvers, 2004). For instance, Beliga et al. (2015) demonstrated that frequent keywords in scientific titles often correspond to broader thematic trends, while their citation counts can serve as proxies for impact.

Beyond thematic modeling, citation analysis serves as a fundamental method for evaluating the scholarly influence of research. Citations reflect the extent to which a work has contributed to subsequent studies, acting as a measure of knowledge dissemination and academic recognition (Bornmann & Daniel, 2008). Various factors, such as journal prestige, topic novelty, and publication recency can influence citation patterns (Waltman, 2016). Incorporating citation data into topic analysis enables a more nuanced perspective on not just what themes are emerging, but which are generating substantial academic attention. For example, Yau et al. (2020) integrated citation weighting into topic modeling to differentiate between popular and impactful topics in biomedical research.

By combining text mining with citation metrics, this study offers a comprehensive overview of deep learning research trends, capturing both thematic evolution and scholarly influence over time.

3. Research Questions and Objectives

To systematically understand the thematic evolution and scholarly impact of deep learning research, this study is guided by the following research questions:

1. What are the most prominent research themes in deep learning between 2019 and 2024?
2. Which keywords are most associated with high-impact research in deep learning?

3. How do citation trends differ across identified topics?

In alignment with these questions, the study pursues the following objectives:

- To extract and preprocess deep learning research titles using natural language processing (NLP) techniques for textual analysis.
- To identify and visualize keyword frequency and co-occurrence patterns that reveal dominant research clusters.
- To manually categorize key topics based on keyword groupings and examine their temporal evolution from 2019 to 2024.
- To assess the citation-weighted influence of individual topics and keywords to determine scholarly impact across research domains.

These objectives collectively support a structured approach to uncovering both thematic and impactful dimensions of deep learning scholarship.

4. Methodology

4.1 Data Collection

To construct the dataset for this study, we utilized the Scopus API to retrieve metadata for the top 500 most cited research articles related to deep learning, published between 2019 and 2024. The query used was TITLE-ABS-KEY("deep learning") AND PUBYEAR > 2018, ensuring that all returned records explicitly referenced "deep learning" in their titles, abstracts, or keywords.

The Scopus API was accessed using the httr and jsonlite packages in R. To adhere to API rate limits and ensure reliable data retrieval, the data were collected in 25-article increments across 20 sequential API calls. Each request returned the top cited articles in descending order, sorted by citation count, starting from the year 2019 onward.

For each article, key metadata fields were extracted, including the article title, publication year, journal name, and citation count. A custom extraction function was implemented to handle missing fields and maintain data integrity. Articles lacking a valid title or citation count were excluded from the final dataset.

After processing, the resulting dataset contained clean metadata for 500 unique articles, which was saved as a CSV file (top_500_deep_learning_2019_onward1.csv) for subsequent analysis. This dataset served as the foundation for all text mining, topic modeling, and citation-based evaluations conducted in the later stages of the study.

	A	B	C	D
1	**Title**	**Year**	**Journal**	**Citations**
2	PyTorch: An imperative style, high-performance deep learning libr	2019	Advances in Neural Information Processing Systems	29771
3	Highly accurate protein structure prediction with AlphaFold	2021	Nature	22473
4	Exploring the limits of transfer learning with a unified text-to-text t	2020	Journal of Machine Learning Research	11061
5	Generative adversarial networks	2020	Communications of the ACM	9368
6	Physics-informed neural networks: A deep learning framework for	2019	Journal of Computational Physics	8792
7	A survey on Image Data Augmentation for Deep Learning	2019	Journal of Big Data	8470
8	A style-based generator architecture for generative adversarial ne	2019	Proceedings of the IEEE Computer Society Conference on C	7059
9	A Comprehensive Survey on Graph Neural Networks	2021	IEEE Transactions on Neural Networks and Learning System	6265
10	Explainable Artificial Intelligence (XAI): Concepts, taxonomies, op	2020	Information Fusion	5458
11	Generalized intersection over union: A metric and a loss for bound	2019	Proceedings of the IEEE Computer Society Conference on C	4781
12	Review of deep learning: concepts, CNN architectures, challenge	2021	Journal of Big Data	4535
13	nnU-Net: a self-configuring method for deep learning-based biom	2021	Nature Methods	4495
14	A ConvNet for the 2020s	2022	Proceedings of the IEEE Computer Society Conference on C	4397
15	Deep high-resolution representation learning for human pose esti	2019	Proceedings of the IEEE Computer Society Conference on C	4167
16	BioBERT: A pre-trained biomedical language representation mode	2020	Bioinformatics	3985
17	Object Detection with Deep Learning: A Review	2019	IEEE Transactions on Neural Networks and Learning System	3958
18	High-performance medicine: the convergence of human and artifi	2019	Nature Medicine	3849
19	A review of recurrent neural networks: Lstm cells and network arc	2019	Neural Computation	3654
20	Physics-informed machine learning	2021	Nature Reviews Physics	3644
21	Graph neural networks: A review of methods and applications	2020	AI Open	3601
22	Accurate prediction of protein structures and interactions using a	2021	Science	3173
23	Gastric cancer	2020	The Lancet	3151
24	Pointpillars: Fast encoders for object detection from point clouds	2019	Proceedings of the IEEE Computer Society Conference on C	3074
25	Deep learning and process understanding for data-driven Earth sy	2019	Nature	3010
26	SignalP 5.0 improves signal peptide predictions using deep neura	2019	Nature Biotechnology	3007
27	Machine Learning: Algorithms, Real-World Applications and Rese	2021	SN Computer Science	2909
28	Deep High-Resolution Representation Learning for Visual Recogn	2021	IEEE Transactions on Pattern Analysis and Machine Intellige	2891
29	Deepsdf: Learning continuous signed distance functions for shap	2019	Proceedings of the IEEE Computer Society Conference on C	2746

top_500_deep_learning_2019_onwa

Figure 01: Top cited scientific papers in deep learning.

4.2 Text Preprocessing

The titles were preprocessed using standard natural language processing techniques. This involved converting text to lowercase, removing punctuation, numbers, and common English stop words, followed by stemming and whitespace stripping. A document-term matrix was then constructed to capture term frequencies across titles. This clean representation served as the foundation for topic modeling and co-occurrence analysis, ensuring reliable extraction of thematic patterns (Sulea et al., 2017).

4.3 Word Cloud Analysis

To explore thematic emphasis within deep learning research titles, two-word clouds were generated. The first visualizes keyword frequency using font size and citation impact using color gradients (from yellow to red, indicating increasing average citation counts). The second provides a traditional frequency-only representation for comparison.

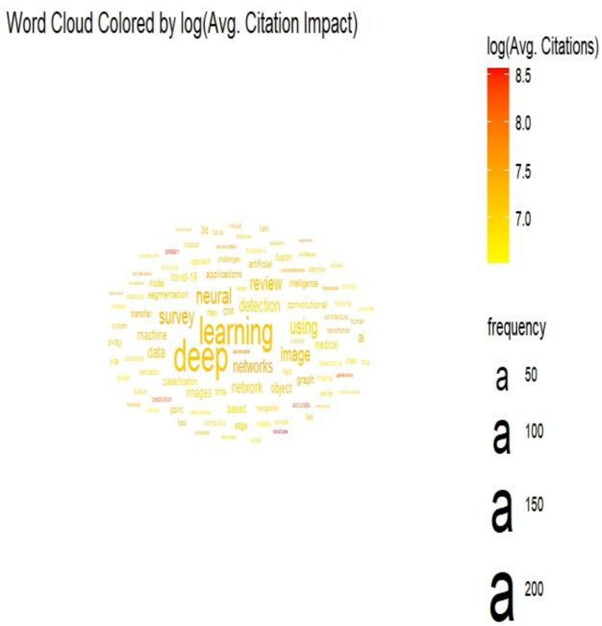

Figure 02: Word cloud with term frequency (size) and citation impact (color).

Figure 03: Word cloud showing term frequency only.

Prominent terms such as "deep," "learning," "neural," and "networks" appear large and intensely colored, reflecting both their frequent mention and high citation impact. This approach goes beyond raw counts by highlighting the influence of specific research directions, allowing for a more meaningful interpretation of topic prominence in the field (e.g., “survey” and “medical” appear both frequent and influential).

4.4 Keyword Co-occurrence Network

A keyword co-occurrence network was constructed to examine how frequent terms are related within deep learning research titles. Nodes represent keywords, with size reflecting average citation impact, and edges denote co-occurrence strength.

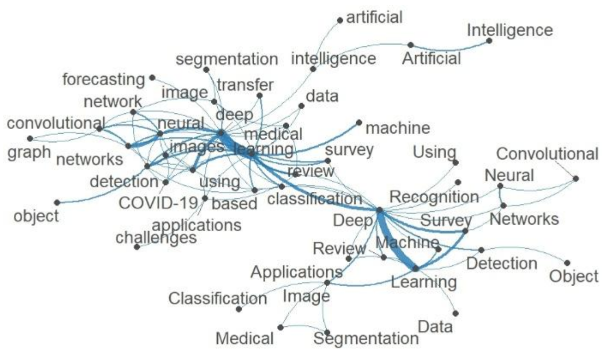

Figure 04: Layout highlighting term clusters and influence.

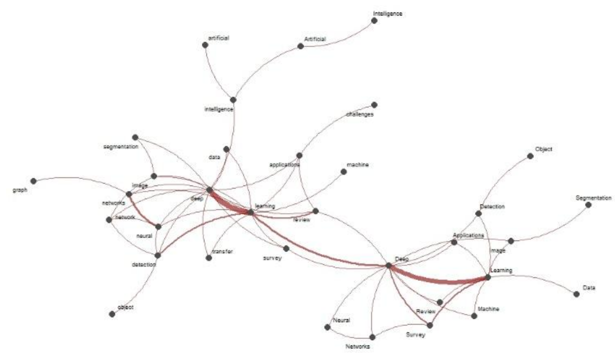

Figure 05: Keyword network with citation-sized nodes and co-occurrence edges.

Central connectors such as "deep," "learning," and "image" bridge multiple clusters, while closely related terms like "neural" + "networks", "detection" + "object", and "review" + "survey" are tightly grouped. Notable thematic clusters emerged around medical applications (e.g., "segmentation," "COVID-19") and technical methods (e.g., "transfer," "convolutional"). Visual clarity was enhanced by reducing node density and applying edge transparency, highlighting influential terms like "learning" and "transfer" more distinctly.

4.5 Manual Topic Grouping

Based on visual inspection of keyword clusters and domain knowledge, three major research themes were manually defined:

1. **Medical & Diagnostic AI** – focusing on health-related applications such as *COVID-19*, *cancer*, *diagnosis*, and *medical imaging*.
2. **Vision & Recognition Models** – covering techniques and architectures for visual understanding, including *convolutional*, *segmentation*, and *transformer* models.
3. **General Deep Learning** – encompassing broader methodological terms like *classification*, *transfer*, *survey*, and *prediction*.

This grouping allowed for better thematic interpretation of the extracted terms and guided downstream trend and citation analyses.

Grouped_Topic	term
Medical & Diagnostic AI	covid-19, cancer, diagnosis, medical, images, X-ray, learning, series, graph, deep
Vision & Recognition Models	architecture, artificial, convolutional, deep, intelligence, neural, segmentation, transformer, network, learning
General Deep Learning	classification, machine, model, transfer, survey, prediction, computing, deep, object,learning

Figure 06: Manually grouped topics according to word distribution.

5. Results

The temporal and citation-based analysis of deep learning literature from 2019 to 2024 revealed key thematic and impact-related trends across the three grouped topics.

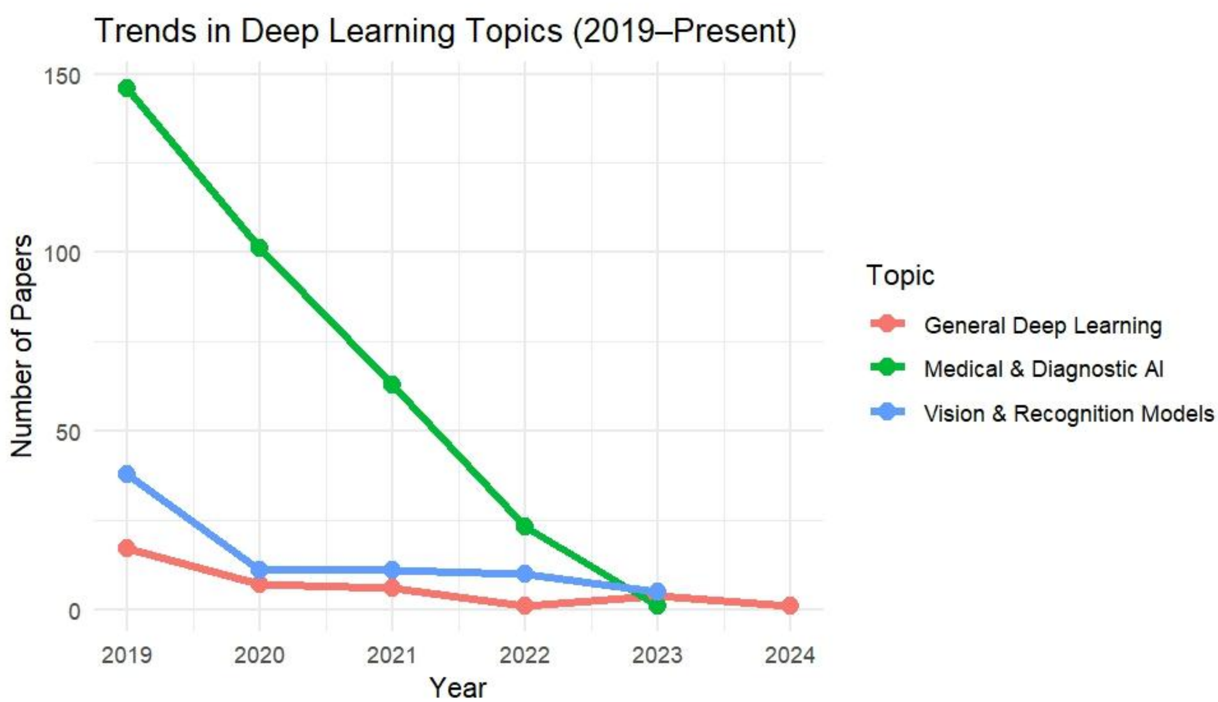

Figure 07: Line graph showing annual paper counts by topic (2019–2024).

As shown in Figure 07, Medical & Diagnostic AI exhibited a sharp rise in publication count in 2019 and peaked in 2020, likely driven by urgent COVID-19-related applications. However, this trend declined rapidly in subsequent years. In contrast, Vision & Recognition Models maintained a steady, albeit modest, output throughout the period. General Deep Learning remained consistently low in volume but showed stability across years.

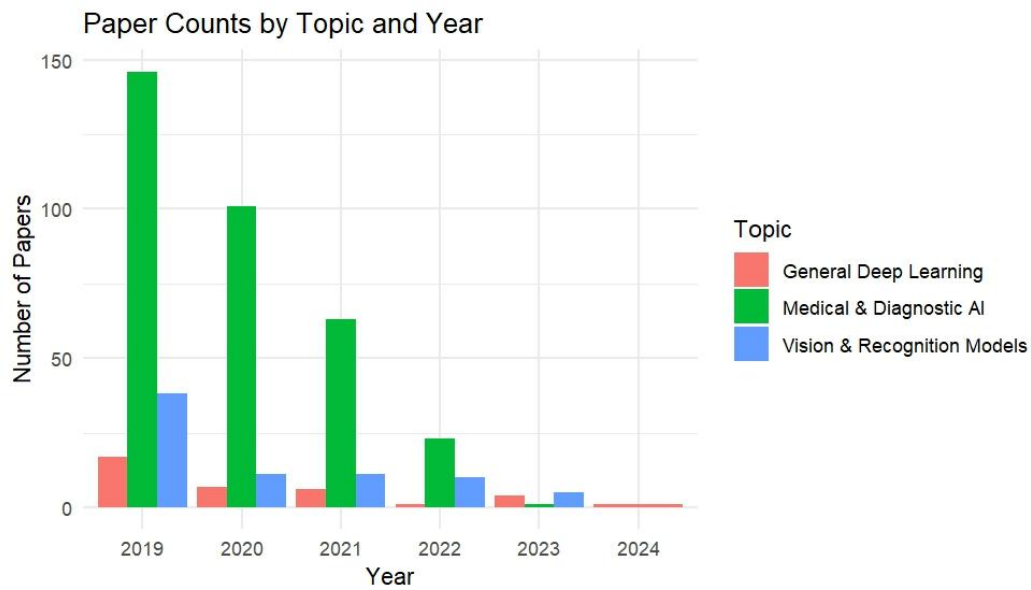

Figure 08: Bar chart of topic-wise publication frequency across years.

Figure 08 further reinforces these trends, displaying absolute paper counts by topic and year. Medical AI clearly dominated the early phase, while other topics exhibited smaller but steadier contributions.

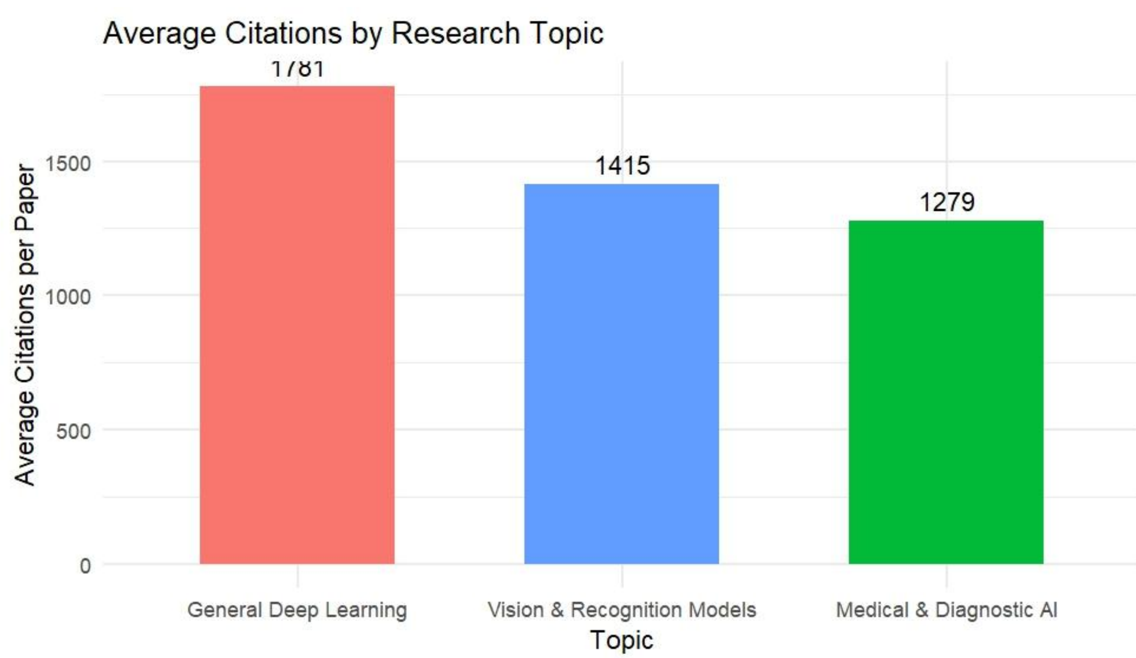

Figure 09: Average citations per paper for each topic group.

In terms of citation influence, Figure 09 shows that General Deep Learning achieved the highest average citations per paper (1781), followed by Vision & Recognition Models (1415) and Medical AI (1279). This suggests that although fewer in number, general DL papers tend to make broader foundational or theoretical contributions.

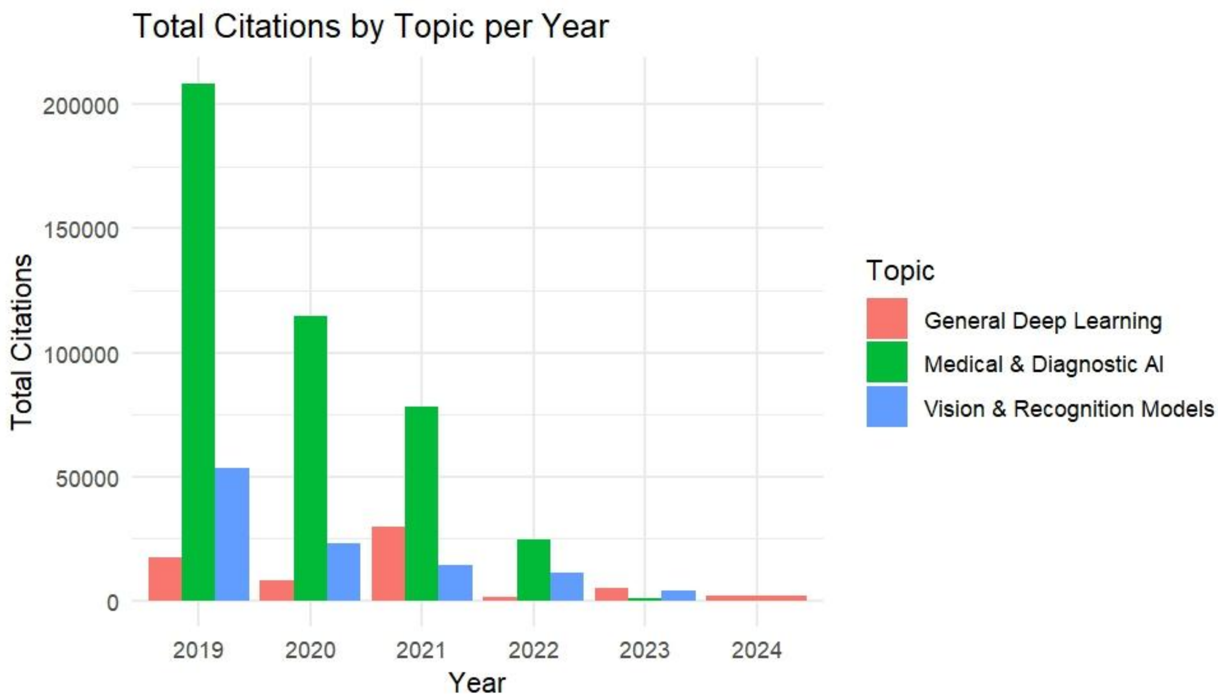

Figure 10: Total citations by topic and year.

Figure 10 depicts the total citations per topic per year, where Medical AI accumulated the highest citation volume in 2019 and 2020. However, as the field diversified and matured, General DL and Vision papers sustained more balanced citation contributions over time.

These results confirm that Medical AI was a dominant short-term focus, especially during the pandemic, whereas General Deep Learning maintained long-term scholarly influence, and Vision-based models contributed consistently throughout the period.

6. Discussion

The results indicate that deep learning research has evolved into distinct subfields, each exhibiting different temporal patterns and citation impacts. A notable surge in Medical & Diagnostic AI occurred in 2020, likely driven by the global urgency to address the COVID-19 pandemic. However, despite its high publication volume, this topic showed relatively lower average citations per paper, suggesting a tradeoff between short-term topical relevance and long-term scholarly influence.

Interestingly, the observed decline in the explicit use of the term "deep learning" in recent years may not reflect reduced interest in the field but rather a shift toward more specific terminology. Terms such as "convolutional," "transfer," and "network" are all foundational to deep learning and have become more prominent. This trend suggests that researchers increasingly refer to specialized techniques or architectures, implying maturation of the field where deep learning is now an assumed framework rather than an explicitly stated term.

Moreover, citation dynamics must be interpreted considering publication age. Older papers have had more time to accumulate citations, which naturally inflates the citation counts for studies published earlier in the timeframe (e.g., 2019–2021). In contrast, papers from 2023 and 2024 have had limited exposure and citation opportunities, explaining their lower citation numbers despite potentially high relevance or quality. This time-lag effect is especially relevant in citation-weighted analyses and must be considered when comparing topic impact across years.

Finally, the prominence of terms like "transfer," "survey," and "networks" in both frequency and citation metrics underscores their foundational role in deep learning research. Vision-related topics maintained consistent attention, reflecting their broad applicability in areas such as robotics, healthcare imaging, and autonomous systems.

Overall, this study demonstrates the effectiveness of combining NLP with citation analysis to extract both thematic structure and scholarly influence from scientific literature. Manual topic grouping further enhanced interpretability, confirming the value of hybrid human-machine approaches in bibliometric analysis.

7. Conclusion

This study applied natural language processing, keyword co-occurrence analysis, and citation-based metrics to examine 500 deep learning research titles published between 2019 and 2024. Through manual and computational techniques, three dominant topic groups: Medical & Diagnostic AI, Vision & Recognition Models, and General Deep Learning were analyzed based on their temporal and citation trends. The findings highlight how deep learning research has diversified and evolved, with notable shifts in terminology and impact. The methodology presented here offers a scalable framework for tracking thematic and influential trends in scientific literature, supporting both academic inquiry and strategic research planning.

Disclaimer

Artificial intelligence tools primarily ChatGPT, along with Grok 3 and GitHub Copilot were used during various stages of this research project to assist with writing, coding, and analysis. These models supported tasks such as code structuring, paraphrasing text, formatting results, and exploring alternative approaches to specific problems. However, their use was supplementary; the research design, decision-making, and final outputs were developed and validated by the author.

During the data collection phase, initial attempts to retrieve research papers from the Scopus API using AI-generated code were unsuccessful, yielding only 25 recent entries from 2025. The AI was unable to directly solve the issue when given a simple command. A more effective strategy was to first ask the AI model about all possible methods for retrieving academic papers and then guide it toward implementing a specific approach. This iterative process of asking about options before issuing direct instructions significantly improved its performance.

In topic modeling, the AI models provided a foundation but could not produce complete or interpretable outputs without additional adjustments. The author referred to reference code from Dr. Ho and manually revised the R scripts to generate meaningful topic structures. Throughout the project, AI assistance was most effective when used interactively, breaking down tasks into smaller steps, validating intermediate outputs, and correcting errors as needed.

This study is based on a sample of 500 research paper titles retrieved from the Scopus database and focuses on title-level metadata only. Therefore, the findings may not fully capture the complexity or scope of the full texts. All interpretations, categorizations, and thematic groupings reflect the author's analytical decisions and domain expertise, guided but not replaced by AI support.

8. References

Beliga, S., Meštrović, A., & Martinčić-Ipšić, S. (2015). An overview of graph-based keyword extraction methods and approaches. Journal of Information and Organizational Sciences, 39(1), 1–20.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3(Jan), 993–1022.

Bornmann, L., & Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. Journal of Documentation, 64(1), 45–80.

Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. Proceedings of the National Academy of Sciences, 101(suppl 1), 5228–5235.

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70–90.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

Sulea, O. M., Zampieri, M., Vela, M., van Genabith, J. (2017). Exploring the use of text classification in the legal domain. arXiv preprint arXiv:1710.09306.

Waltman, L. (2016). A review of the literature on citation impact indicators. Journal of Informetrics, 10(2), 365–391.

Yau, C., Tzeng, T. R., & Yau, H. L. (2020). Topic modeling with citation weighting for knowledge discovery. Journal of Informetrics, 14(1), 101003.