Analyzing Deep Learning Research Trends Using Knowledge Mining.

Introduction

Deep learning, a subset of machine learning, utilizes artificial neural networks to automatically learn representations from data, enabling models to perform complex tasks with high accuracy (LeCun, Bengio, and Hinton 2015). It has revolutionized multiple fields, including computer vision (Krizhevsky, Sutskever, and Hinton 2012), natural language processing (Brown et al. 2020), autonomous driving (LeCun et al. 2015), medical diagnostics (Esteva et al. 2017), and agriculture (Kamilaris and Prenafeta-Boldú 2018).

The increasing adoption of deep learning in various fields has resulted in a rapid growth of research publications. This project aims to analyze deep learning research trends over time by mining academic papers, identifying key topics, and understanding how certain factors (e.g., publication year, research topic, journal) influence citation impact.

Research Statement

By applying data mining techniques, this project will extract hidden patterns from research publications to understand:

- How deep does learning research evolved over the years?
- What are the key research trends in deep learning?
- How do factors like publication, journal, and topic affect research impact?

Variables

I have identified some initial variables for this study. The key objective is to extract metadata from each research paper and organize it in a structured format to facilitate further analysis.

Variable	Type	Description
Year of Publication	Independent	The year the paper was published
Number of Citations	Dependent	How many times the paper was cited
Topic	Categorical	The research area (CNN, Transformers, NLP)
Journal Name	Categorical	Where the paper was published
Abstract Keywords	Categorical	Key terms in research abstracts

Method

This project follows a structured knowledge mining approach:

Step 1: Data Collection

The Scopus API was used to retrieve research papers on deep learning. The goal is to cover five years of research to analyze trends over time. If necessary, I may modify the topic selection or incorporate more specific keywords to refine the dataset. Additionally, I may choose to limit the data sources to maintain consistency and relevance in the analysis.

Step 2: Data Preprocessing

The collected research paper titles will be preprocessed to ensure clean and structured text for analysis. This includes removing stopwords, punctuation, numbers, and extra whitespace, as well as converting text to lowercase for consistency. The cleaned text will then be transformed into a Document-Term Matrix (DTM), where each research title will be represented by its word frequencies. This structured format will allow for efficient topic modeling using LDA to extract key research themes (OpenAI 2024).

Step 3: Topic Modeling (LDA)

Latent Dirichlet Allocation (LDA) will be applied to group research papers into three major topics, allowing for the identification of key themes. The extracted topics will be analyzed based on their word distributions, and the topic proportions will be visualized to highlight the most dominant research areas. This structured approach will help uncover trends in deep learning research.

Step 4: Predictive Analysis

A regression model will be built to predict the impact of citation papers using key variables. The model will analyze how year of publication, journal name, and research topic (from LDA) influence the number of citations.

Preliminary Results

So far, 25 research papers have been collected from Scopus. Below is a sample of the dataset:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Title	Authors	Year	DOI	Journal	Citations	Link						
2	The Common Curricular Base and improvement in the cu	Gasque K.	12/18/2025	10.5007/1!	Encontros	0	https://api.elsevier.com/content/abstract/scopus_id/85213370716						
3	Perceived Information Revisited II Information-Theoretic	Ito A.	12/9/2025	10.46586/t	IACR Tran:	0	https://api.elsevier.com/content/abstract/scopus_id/85215327334						
4	The application of artificial intelligence in the field of me	Dehbozor	12/1/2025	10.1186/s1	BMC Psycl	0	https://api.elsevier.com/content/abstract/scopus_id/85218427662						
5	Intrusion Detection System for Network Security Using Nc	Manivann	12/1/2025	10.1007/s4	Internatio	0	https://api.elsevier.com/content/abstract/scopus_id/85218426412						
6	Statistical batch-based bearing fault detection	Jorry V.	12/1/2025	10.1186/s1	Journal of	0	https://api.elsevier.com/content/abstract/scopus_id/85218424625						
7	Author Correction: An automated deep learning pipeline	Cai L.	12/1/2025	10.1038/s4	npj Precis	0	https://api.elsevier.com/content/abstract/scopus_id/85218423954						
8	SVEA: an accurate model for structural variation detectio	Qiu T.	12/1/2025	10.1186/s1	Journal of	0	https://api.elsevier.com/content/abstract/scopus_id/85218423462						
9	Predicting mother and newborn skin-to-skin contact usi	Safarzade	12/1/2025	10.1186/s1	BMC Pregr	0	https://api.elsevier.com/content/abstract/scopus_id/85218419515						
10	Coal burst spatio-temporal prediction method based on	Yang X.	12/1/2025	10.1007/s4	Internatio	0	https://api.elsevier.com/content/abstract/scopus_id/85218417560						
11	Advanced prognostic modeling with deep learning: asse	Raji C.G.	12/1/2025	10.1186/s1	Journal of	0	https://api.elsevier.com/content/abstract/scopus_id/85218417489						
12	DenseNet-ABiLSTM: Revolutionizing Multiclass Arrhythmi	Saranya K	12/1/2025	10.1007/s4	Internatio	0	https://api.elsevier.com/content/abstract/scopus_id/85218417030						
13	In vivo electrophysiology recordings and computational	Gedela N.	12/1/2025	10.1186/s4	Bioelectro	0	https://api.elsevier.com/content/abstract/scopus_id/85218415290						
14	Prediction of cognitive conversion within the Alzheimerâ	Yang S.	12/1/2025	10.1186/s1	Alzheimer	0	https://api.elsevier.com/content/abstract/scopus_id/85218411294						
15	Design and developing a robot-assisted cell batch micro	Guo X.	12/1/2025	10.1038/s4	Microsyst	0	https://api.elsevier.com/content/abstract/scopus_id/85218411260						
16	IoT integrated and deep learning assisted electrochemic	Lahari S.A	12/1/2025	10.1038/s4	npj Clean	0	https://api.elsevier.com/content/abstract/scopus_id/85218408928						
17	Automatic magnetic resonance imaging series labelling	Gomis-Ma	12/1/2025	10.1186/s4	Journal of	0	https://api.elsevier.com/content/abstract/scopus_id/85218408786						
18	A deep learning model to predict dose distributions for b	Hou X.	12/1/2025	10.1007/s1	Discover (	0	https://api.elsevier.com/content/abstract/scopus_id/85218408326						
19	A Hybrid LECNN Architecture: A Computer-Assisted Early	GÃ¼raksÄ	12/1/2025	10.1007/s4	Internatio	0	https://api.elsevier.com/content/abstract/scopus_id/85218407436						
20	Correction to: Physics-informed deep generative learning	Brown E.E	12/1/2025	10.1038/s4	Nature Co	0	https://api.elsevier.com/content/abstract/scopus_id/85218406555						
21	Multimodal text-emoji fusion using deep neural network:	Kusal S.	12/1/2025	10.1186/s4	Journal of	0	https://api.elsevier.com/content/abstract/scopus_id/85218356016						
22	The role of IoT and XAI convergence in the prediction, exp	Abrokwah	12/1/2025	10.1007/s4	Discover I	0	https://api.elsevier.com/content/abstract/scopus_id/85218353824						
23	Enhanced bearing RUL prediction based on dynamic tem	Jin Z.	12/1/2025	10.1007/s4	Autonomo	0	https://api.elsevier.com/content/abstract/scopus_id/85218353759						
24	Mammalian piRNA target prediction using a hierarchical	Zhang T.	12/1/2025	10.1186/s1	BMC Bioin	0	https://api.elsevier.com/content/abstract/scopus_id/85218352799						
25	Multi-classification of breast cancer histopathological i	Yusuf M.	12/1/2025	10.1186/s4	Journal of	0	https://api.elsevier.com/content/abstract/scopus_id/85218351457						
26	Hyperspectral and Multispectral Image Fusion Based on	Luo Y.	12/1/2025	10.1007/s1	Sensing ai	0	https://api.elsevier.com/content/abstract/scopus_id/85218349381						

References

Brown, Tom, et al. 2020. "Language Models Are Few-Shot Learners." Advances in Neural Information Processing Systems (NeurIPS).

Esteva, Andre, et al. 2017. "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks." Nature 542(7639): 115-118.

Kamilaris, Andreas, and Francesc Prenafeta-Boldú. 2018. "Deep Learning in Agriculture: A Survey." Computers and Electronics in Agriculture 147: 70-90.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." Paper presented at the Advances in Neural Information Processing Systems Conference, Lake Tahoe, NV.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." Nature 521(7553): 436-444.

OpenAI. 2024. "ChatGPT-generated text on Data Preprocessing." Accessed [Date]. https://chat.openai.com