Analyzing Research Trend in Deep Learning with Knowledge Mining

A project using NLP, topic modeling, and citation analysis to study deep learning research trends (2019–2024).

Analyzing Research Trend in Deep Learning with NLP

- This study explores how topic modeling and keyword-based NLP techniques can uncover patterns in deep learning research from 2019–2024.

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.
- 500 deep learning paper titles (2019–2024) were collected and cleaned. Each title was tokenized for topic modeling and trend analysis.

	A	B	C	D
1	**Title**	**Year**	**Journal**	**Citations**
2	Application of deep learning algor	1/1/2026	Skeletal Radiology	0
3	The Common Curricular Base and	12/18/2025	Encontros Bibli	0
4	Deep learning in flower quantifica	12/10/2025	Acta Scientiarum - Technolc	0
5	Perceived Information Revisited II	12/9/2025	IACR Transactions on Crypto	0
6	Vision Mamba and xLSTM-UNet for	12/1/2025	Scientific Reports	0
7	Predicting triage of pediatric patie	12/1/2025	International Journal of Em	0
8	Rolling bearing remaining useful l	12/1/2025	Scientific Reports	0
9	SignEdgeLVM transformer model fo	12/1/2025	Discover Computing	0
10	An intelligent ransomware based	12/1/2025	Scientific Reports	0
11	Pixel level deep reinforcement lear	12/1/2025	Scientific Reports	0
12	Artificial intelligence-driven transl	12/1/2025	Journal of Translational Me	0
13	Assessing and developing college	12/1/2025	International Journal of Edu	0
14	Convolutional block attention gate	12/1/2025	BMC Medical Imaging	0
15	Myocardial perfusion imaging SPE	12/1/2025	EJNMMI Physics	0
16	CPHNet: a novel pipeline for anti-H	12/1/2025	Respiratory Research	0
17	Electrochemical ohmic memristor	12/1/2025	Nature Communications	0
18	Prediction of particulate matter PM	12/1/2025	Journal of Air Pollution and	0
19	Identification of enterotype for pat	12/1/2025	Journal of Translational Me	0
20	A large-scale open image dataset	12/1/2025	Scientific Data	0
21	A multi-dilated convolution netwo	12/1/2025	Scientific Reports	0
22	A novel approach for the detection	12/1/2025	Scientific Reports	0
23	A vehicle trajectory prediction moc	12/1/2025	Scientific Reports	0
24	Precise engineering of gene expres	12/1/2025	Genome Biology	0
25	Linear attention based spatiotemp	12/1/2025	Scientific Reports	1
26	Leveraging large language models	12/1/2025	Scientific Reports	0

Data Processing

- 500 deep learning paper titles (2019–2024) were collected and cleaned. Each title was tokenized for topic modeling and trend analysis.

	A	B	C	D
1	**Title**	**Year**	**Journal**	**Citations**
2	Application of deep learning algor	1/1/2026	Skeletal Radiology	0
3	The Common Curricular Base and	12/18/2025	Encontros Bibli	0
4	Deep learning in flower quantifica	12/10/2025	Acta Scientiarum - Technolc	0
5	Perceived Information Revisited II	12/9/2025	IACR Transactions on Crypto	0
6	Vision Mamba and xLSTM-UNet for	12/1/2025	Scientific Reports	0
7	Predicting triage of pediatric patie	12/1/2025	International Journal of Em	0
8	Rolling bearing remaining useful l	12/1/2025	Scientific Reports	0
9	SignEdgeLVM transformer model fo	12/1/2025	Discover Computing	0
10	An intelligent ransomware based	12/1/2025	Scientific Reports	0
11	Pixel level deep reinforcement lear	12/1/2025	Scientific Reports	0
12	Artificial intelligence-driven transl	12/1/2025	Journal of Translational Me	0
13	Assessing and developing college	12/1/2025	International Journal of Edu	0
14	Convolutional block attention gate	12/1/2025	BMC Medical Imaging	0
15	Myocardial perfusion imaging SPE	12/1/2025	EJNMMI Physics	0
16	CPHNet: a novel pipeline for anti-H	12/1/2025	Respiratory Research	0
17	Electrochemical ohmic memristor	12/1/2025	Nature Communications	0
18	Prediction of particulate matter PM	12/1/2025	Journal of Air Pollution and	0
19	Identification of enterotype for pat	12/1/2025	Journal of Translational Me	0
20	A large-scale open image dataset	12/1/2025	Scientific Data	0
21	A multi-dilated convolution netwo	12/1/2025	Scientific Reports	0
22	A novel approach for the detection	12/1/2025	Scientific Reports	0
23	A vehicle trajectory prediction moc	12/1/2025	Scientific Reports	0
24	Precise engineering of gene expres	12/1/2025	Genome Biology	0
25	Linear attention based spatiotemp	12/1/2025	Scientific Reports	1
26	Leveraging large language models	12/1/2025	Scientific Reports	0

Data Processing

- Each title was tokenized for topic modeling and trend analysis.

```r
# Create corpus from the Title column
corp <- corpus(data, text_field = "Title")

# Tokenize and clean
toks <- tokens(corp, remove_punct = TRUE, remove_numbers = TRUE) %>%
  tokens_remove(stopwords("en"))

# Create document-feature matrix
dfm <- dfm(toks)

# Trim rare terms (appear in only 1 doc)
dfm_trimmed <- dfm_trim(dfm, min_termfreq = 2)

dfm_trimmed

# Convert dfm to topicmodels-compatible format
dtm <- convert(dfm_trimmed, to = "topicmodels")

# Optional: Remove empty documents (just in case)
row_totals <- apply(dtm, 1, sum)
dtm <- dtm[row_totals > 0, ]

# Set number of topics
k <- 5

# Fit LDA model
lda_model <- LDA(dtm, k = k, control = list(seed = 1234))
```

52:1 (Top Level) R Script

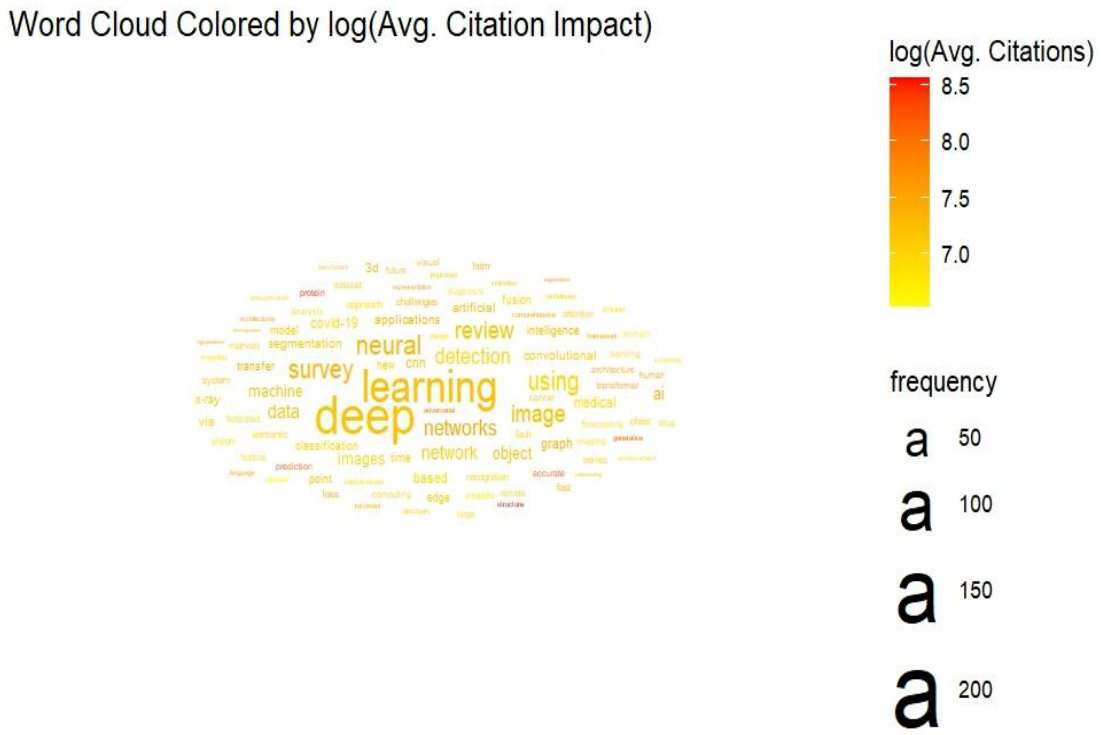
```
> words_per_topic
  1   2   3   4   5
416 416 416 416 416
>
```

Understanding Text Data with Wordclouds

- Initial wordclouds revealed frequent terms like 'deep', 'learning', 'neural', and 'medical'. These guided our manual topic group creation.

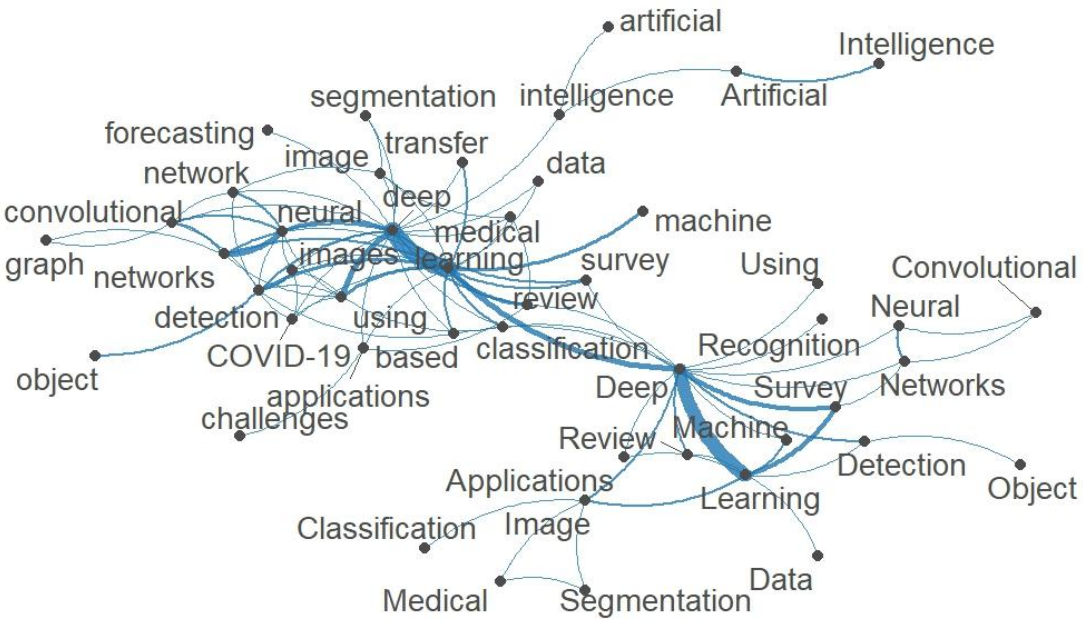
Understanding Text Data with Wordclouds

- This visualization highlights frequently occurring keywords from deep learning paper titles. Word size indicates frequency, while color intensity (yellow to red) reflects average citation impact.



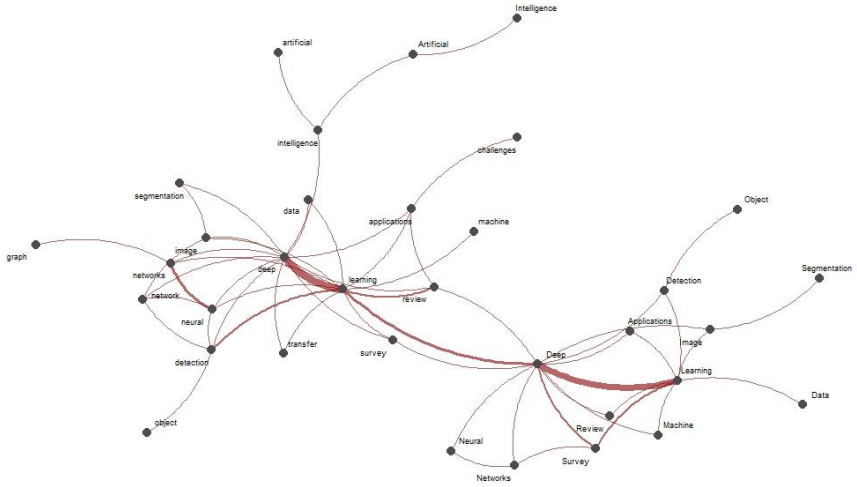
Keyword Co-occurrence Network

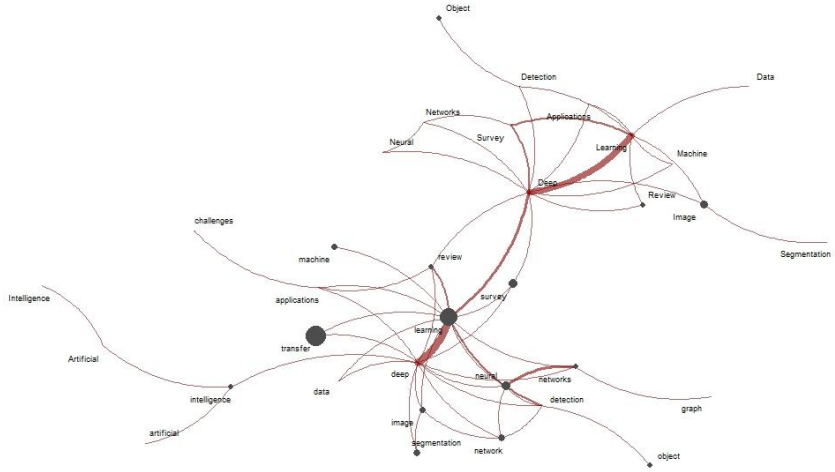
- Visualized how key terms are co-occuring. Terms like 'image', 'segmentation', and 'covid-19' formed meaningful clusters related to medical AI and vision tasks.



Citation-Weighted Network

- Word node sizes and colors were scaled by citation count, showing impactful terms like 'transfer', 'networks', and 'survey'.





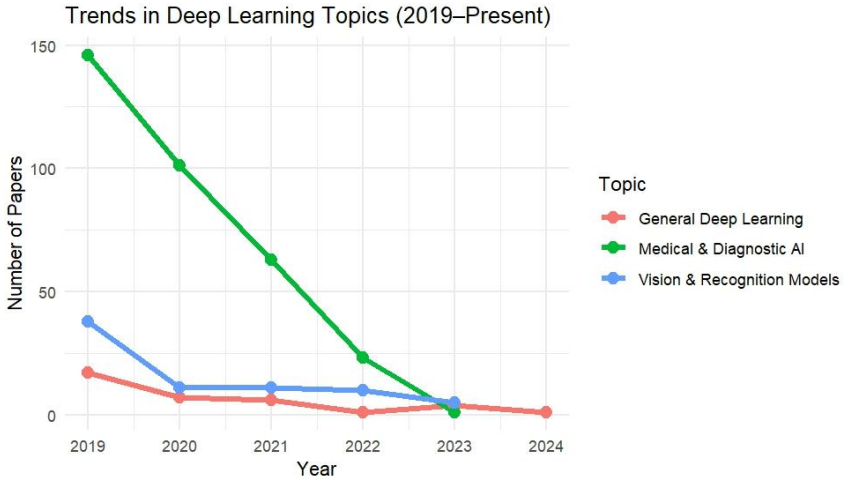
Manual Topic Grouping

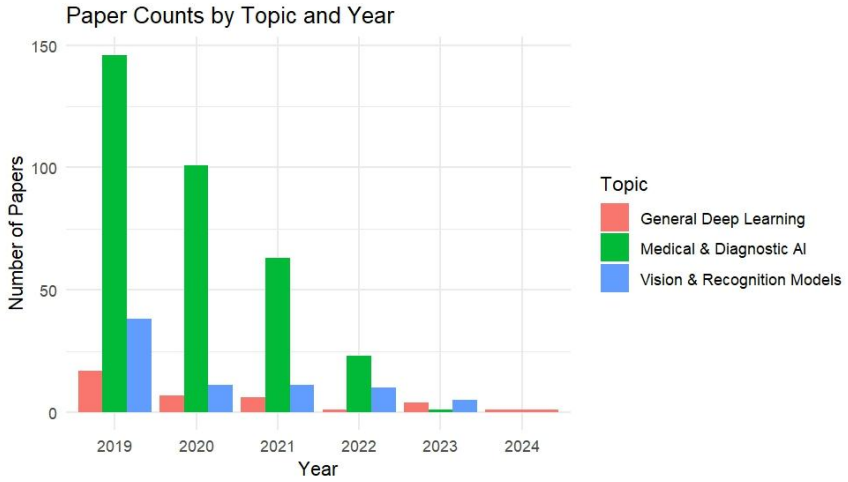
- Three themes were created: Medical & Diagnostic AI, Vision & Recognition Models, and General Deep Learning, using clustered keyword sets.

	A	B
1	Grouped_Topic	Words
2	Medical & Diagnostic AI	covid-19, cancer, diagnosis, medical, images, X-ray, learning, series, graph, deep
3	Vision & Recognition Models	architecture, artificial, convolutional, deep, intelligence, neural, segmentation, transformer, network, learning
4	General Deep Learning	classification, machine, model, transfer, survey, prediction, computing, deep, object,learning
5		

Topic Trends Over Time

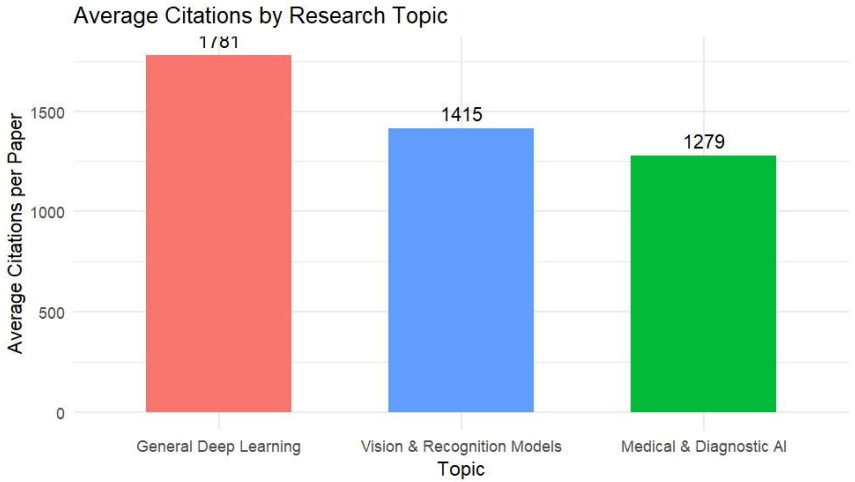
- Paper counts for each topic were tracked yearly. Medical AI peaked early (COVID-19), while Vision topics showed steady interest.

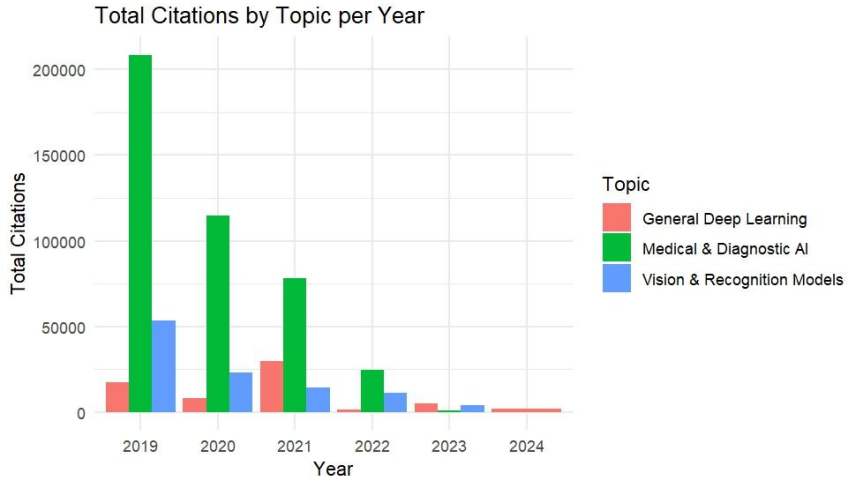




Citation-Weighted Topic Impact

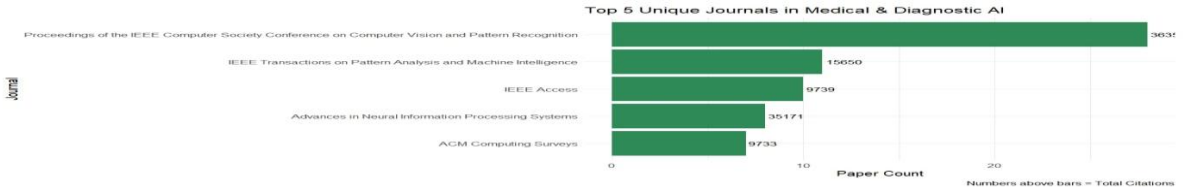
- General DL had highest average citations per paper. Medical AI had high volume but lower per-paper influence.

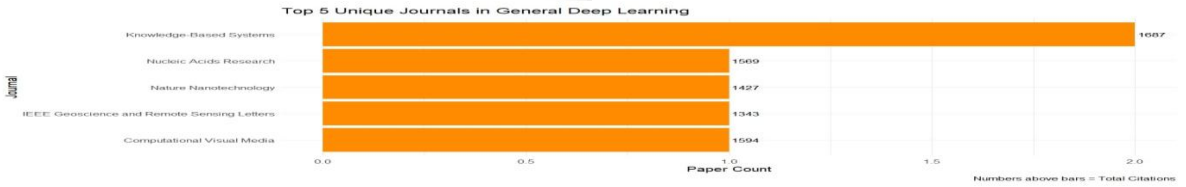


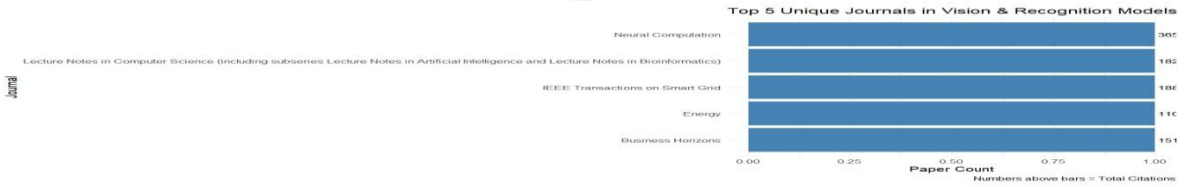


Journal vs. Topic Mapping

- Each journal was uniquely assigned to one dominant topic. Top 5 per topic showed specialization in medical, vision, or general DL areas.







Conclusion

- Using NLP, topic modeling, and citation analysis, we identified evolving interests and influential subfields in deep learning research.