
Mapping the Landscape of Deep Learning

A Dataset of Indexed Topics & Embeddings from Deep Learning Books *and* Hierarchical Topic Discovery from Embedding Structure

Jeffrey Urban

The University of Texas at Austin

Abstract

This project involves mapping hierarchical concept relationships among terms relevant to deep learning within a corpus of 158 books on deep learning. **Mistral-7B-Instruct-v0.2** is used to identify relevance of terminology and elicit topic names, and an index is created for those topics into terminology within the book pages. The **all-mpnet-base-v2** sentence-transformers model is used to generate embeddings of the titles, pages, and topics. From this dataset, embeddings are dimensionally-reduced using **UMAP**, and clustered using **HDBSCAN** to extract a hierarchical topic structure of deep learning concepts. The resulting topic hierarchy provides a structured view of the landscape of deep learning. The embedding space and topic index provide complementary methods supporting retrieval of topic-relevant book content.

1 Introduction

This project builds a dataset and uses that dataset to explore topic indexing within and among documents, delivering a topic hierarchy of deep learning concepts from a corpus of deep learning books. Hierarchical clustering of topics in large sets of documents and their contents is a powerful tool for information access and summarization (Wang & Wang, 2019). Capturing a graph hierarchy supports exploration of related concepts, both broader (parents) and more nuanced (children). Applications of topic hierarchies include library recommender systems and retrieval-augmented generation by language models.

Towards exploration of topic hierarchies, this project provides a dataset of text and embeddings from books about deep learning, and an index from topics to specific terms on specific book pages.

The dataset comprises:

- 158 book titles with embeddings, authors, and page counts.
- 50794 book pages with embeddings, page text, and page numbers.
- 6217 elicited deep learning related topics with embeddings and counts.
- An index from topics to book + page for 1,902,953 instances of corresponding terms within the corpus of deep learning books.

The topic hierarchy comprises:

- A graph of clustered embeddings, with each node corresponding to the embedding (title, page, or topic) that is most central to the cluster.
- Nodes named by the title, topic, or the inferred main focus of the page corresponding to the most central embedding.
- An index from each node into the book title, page, or topic corresponding to the most central embedding.

This project presents a novel approach to creation of a topic hierarchy, through:

1. A shared embedding space combining embeddings for book titles, page text and topic names.

The project demonstrates how the inclusion of embeddings for individual topic names appears beneficial for dimensionality reduction via HDBSCAN to effectively capture representation structure in the embeddings for the page texts.

2. A unique combination of techniques:

- n-gram (word) sequence capture and indexing of common phrases.
- Mistral-7B-Instruct-v0.2 for identification and elicitation of topics from keyphrases, and complete page texts.
- all-mpnet-base-v2 for generation of embeddings in a common space.
- UMAP for dimensionality reduction of the embedding space, leveraging structure in topic embeddings to provide structure for page embeddings.
- HDBSCAN repeated with different parameter values for clustering.
- Hierarchy extraction via analysis of nesting of HDBSCAN clusters.

This combination of techniques combines strengths of classical and neural methods to delivers a balance of efficiency, thoroughness, and quality in the elicited topic hierarchy.

2 Research Background

Document-level topic indexing from keyphrases and terms was explored by (Medelyan, 2009). (Collins & Beel, 2019) explored embeddings at different levels of granularity including documents, keyphrases and individual terms, for applications in library materials recommendation, similar to this project. (Asmala, 2024) found that embeddings generated from sentence-transformers substantially outperform fixed string keyword searches in improving semantic understanding of texts. (Ordun et al., 2005) also apply UMAP to create a topic modeling graph from texts. As done in this project, (Asyaky & Mandala, 2021) apply HDBSCAN with UMAP dimensionality-reduced word embeddings for clustering, but for grouping short texts. (Zhang et al., 2022) find that clustering

methods can outperform neural models in coherency, in handling diversity, and compute efficiency. The method described here leverages the strengths of both neural and clustering approaches and supports document retrieval using either.

3 Dataset

The dataset is based on a corpus comprising 158 PDF-format books selected for including both words ‘deep’ and ‘learning’ in the title (in either order), as a simple and straightforward heuristic for identifying books about deep learning. A few candidate books were excluded due to parse errors. Among the included books, there is a total of 50794 pages. Among all included books, all pages are included. The complete listing of included books is presented in Appendix A.

3.1 Dataset - Processing

The LlamaIndex toolset facilitates programmatic access to large document collections (Zirnstien, 2023); one such application in document search is presented by (Jo & Lee, 2024). The LlamaIndex SimpleDirectoryReader was used to parse the PDF-format book pages into objects identifying the book title, printed page identifier, and page text.

Example entry from book `info.jsonl`:

```
{"index": 0, "title": "3D Imaging - Multidimensional Signal Processing and Deep Learning", "author": "Patnaik", "page_count": 283}
```

Example entry from book `pages.jsonl`:

```
{"file_name": "3D Imaging - Multidimensional Signal Processing and Deep Learning - Patnaik.pdf", "page_label": "C1", "text": "123\nSmart Innovation, Systems and Technologies\n348\nSrikanta\nPatnaik\nRoumen\nKountchev\nYonghang\nTai\nRoumiana\nKountcheva\nEditors\n3D Imaging\n2014\nMultidimensional \nSignal Processing and \nDeep Learning\nMultidimensional Signals, Video \nProcessing and Applications, Volume 2"}
```

All unique words and all phrases up to 4 words (split on whitespace), excluding text without alphabetic characters, or with symbols besides dash ‘-‘ are identified. These number over 60,000 unique strings. Among these, strings with at least 10 words, approximately 31,000, are kept, and an index is created mapping the string to the book and page identifier.

Example entry from index `words and phrases.json`:

```
{"Learning Deep": [[0, "C1"], [0, "iii"], [6, "45"], ...]}
```

For each of the kept strings, the Mistral-7B-Instruct-v0.2 instruction-tuned language model (Jiang et al., 2023) is used to identify relevance of terminology (via a boolean classification) and elicit a topic name for each terminology string, and return these findings in a JSON structure. Note that most of the input strings are not specifically relevant to deep learning. Significant prompt engineering was necessary to elicit topic names that were relevant to deep learning and reasonable in length. The use of prompt engineering in document access for academic research can be challenging (Giray, 2023). There was a

delicate balance between the model being too hesitant to report that a term was related to deep learning and to report another common use (e.g. ‘CNN’ is ‘Cable News Network’, not related), or too eager to classify an unrelated term as relevant to deep learning (e.g. ‘second’ is ‘second-order’, related). The output topic strings categorized as ‘related-to’ deep learning, were kept. Qualitative inspection revealed that vast majority of input string to topic associations were sensible. One notable exception was that ‘was’ was identified as ‘Wasserstein Loss Function’, while in most cases the meaning of the string was the past tense of ‘is’. Despite the error, this topic was kept, and the complete mapping was kept as-is. Among topics, many were common across input strings, consolidating different formats and spellings for the same topic. An index is created for all 6217 topics to 1,902,953 instances of corresponding original terminology within the book pages. A count of instances is recorded for each topic:

Example entry from sorted term page counts.jsonl:

```
{"topic": "Recurrent Neural Network", "count": 17004}
```

The all-mpnet-base-v2 sentence-transformers model is used to generate embeddings of the titles, pages, and topics. (Ashqar & Multlu, 2023) found that all-mpnet-base-v2 outperforms other sentence embedding models in precision, recall, F1 score, MAP, MRR, and retrieval of relevant keywords, so it appears to be well suited for distilling text into topic names for topic modeling. (Siino, 2024) applied all-mpnet-base-v2 to predict Semantic Textual Relatedness (STR), a basis for topic clustering. (Asmala, 2024) demonstrated a similar application in medical document search all-mpnet-base-v2.

All inference and training was performed on CPU or an NVIDIA RTX 4090 GPU.

3.2 Dataset - Limitations

Identification of relevant topics, both in classification, and in generation of a topic name was dependent on prompt engineering and the Mistral-7B-Instruct-v0.2 model. The dependency on a large language model in the pipeline may result in errors and biases beyond those identified and noted.

3.3 Dataset - Other Methods Explored

Before selecting the complete 4-word phrase set strategy, I had tried named-entity identification via Spacy, as described in (Altinok, 2021) and similar to work in (Kumar et al., 2023). The identified set was at least an order of magnitude smaller than the set identified by the complete set strategy.

Before selecting the Mistral-based method to identify and group deep-learning related topics from the captured phrases, I had tried manually compiling a keeplist of terms, and a mapping from raw terms to a selected topic or manually written topic. This may result in higher quality, but at a huge time expenditure.

Before determining to summarize only pages included in the hierarchy, I had used Mistral to summarize all pages into topics. Given an inference time of approximately 1 second/page, summarization of all pages was achievable at an estimated 14 hours. This option was risky to rely on, given the potential need

for iteration or rework.

3.4 Dataset - Opportunities

The topic index can be used for human or machine page-level access to the corpus of deep learning books. Page embeddings may also be used to search for topics.

4 Hierarchical Topic Extraction (Analysis)

Analysis for hierarchical topic extraction leverages the created dataset described in the previous section. The process extracts a hierarchical topic structure of deep learning concepts. The resulting topic hierarchy provides a structured view of the landscape of deep learning. The embedding space and topic index support retrieval of topic-relevant book content.

4.1 Hierarchical Topic Extraction - Processing

Embeddings for book titles, book pages and topics are reduced to two dimensions in a combined manifold space using UMAP (McInnes et al., 2018). Dimensionality reduction with the combined embeddings qualitatively results in much more clearly-defined clusters vs dimensionality reduction on any of the title, page or topic embedding types separately, suggesting more information is preserved when dimensionality reduction is performed with topic and page embeddings together, so analysis proceeded with the combined embeddings. Since the embeddings are all based on the same embedding model, the more tightly scoped embeddings may help form meaningful structure in the combined space, as seen in the pattern of page embeddings. UMAP results in the appearance of global structural patterns that suggest the model has maintained a decomposition of concepts.

(continues on next page)

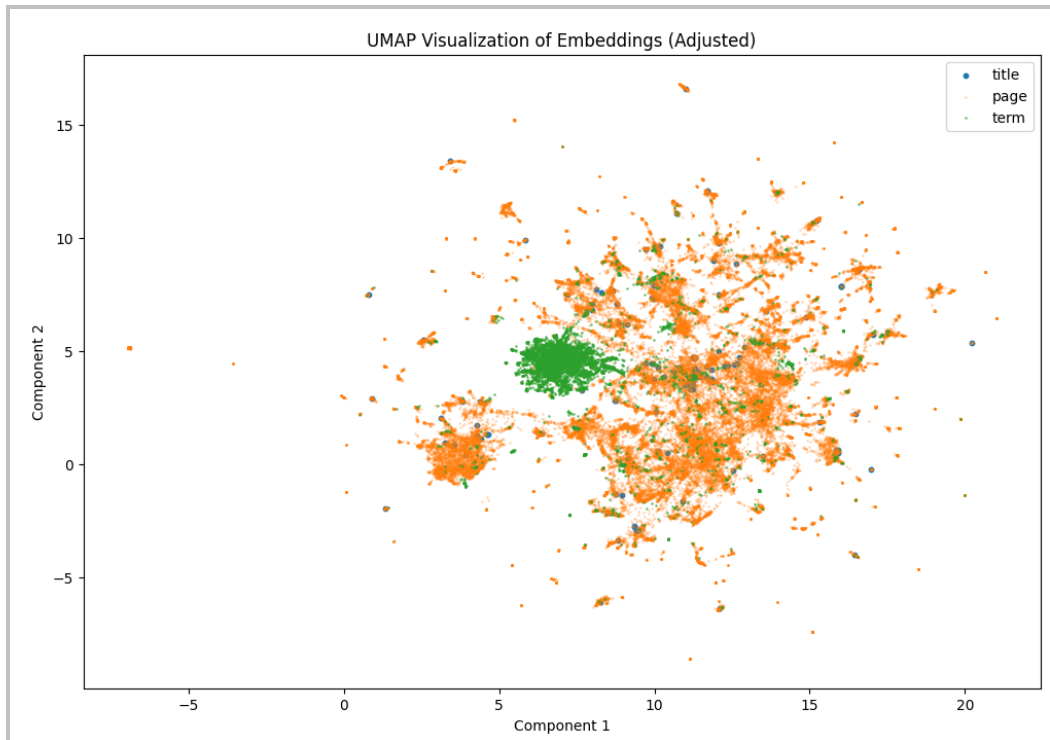


Figure 1: Title, Page and Term Embeddings shown together in the UMAP dimensionally-reduced space.

The combined embeddings in the UMAP dimensionally-reduced space are clustered using HDBSCAN at three values for `min_cluster_size`. The three values for `min_cluster_size` are hand-tuned to result in coherent-appearing clustering at three levels of granularity. The identified clusters include most of the embeddings, but leave some uncategorized. In each case, these were grouped in an ‘other’ category, for completeness, including the case when a cluster was identified from these excluded embeddings at a lower level of `min_cluster_size`.

HDBSCAN has been noted as more stable than other clustering techniques including DBSCAN and k-means (Zhang et al., 2024). HDBSCAN can accommodate clusters of different densities (Monko & Kimura, 2024), and this may explain the superior qualitative performance observed on this dataset. HDBSCAN performs better on lower-dimensionality data (Asyaky & Mandala, 2021) as generated by UMAP, and the human-interpretability of the graphs aid in intuitively tuning the cluster size. See (Tran et al., 2021) for an in-depth analysis of HDBSCAN performance.

(continues on next page)

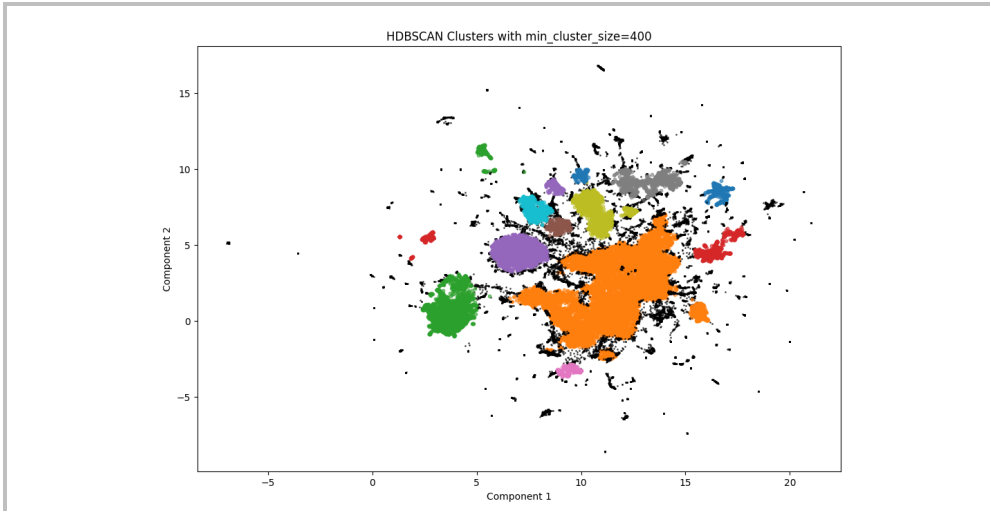


Figure 2: HDBSCAN clusters at min_cluster_size = 400

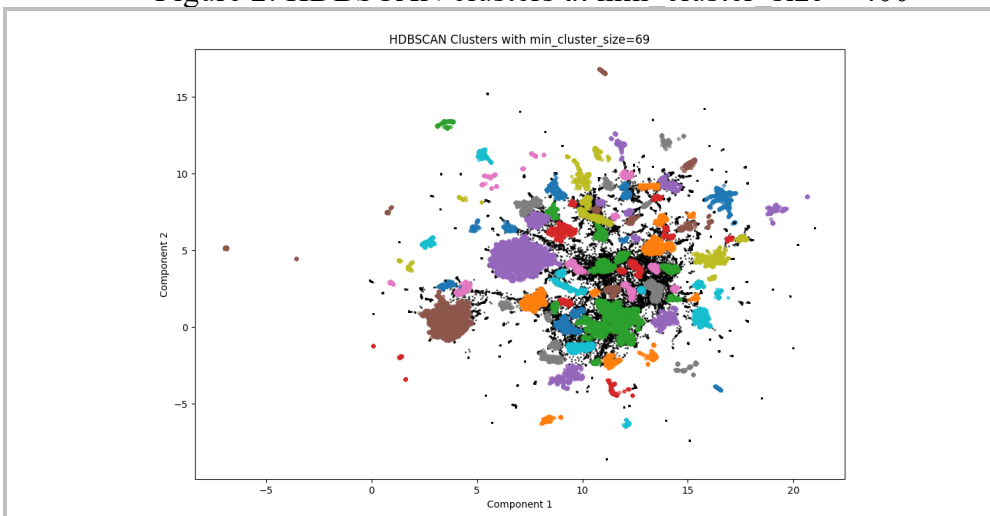


Figure 3: HDBSCAN clusters at min_cluster_size = 69

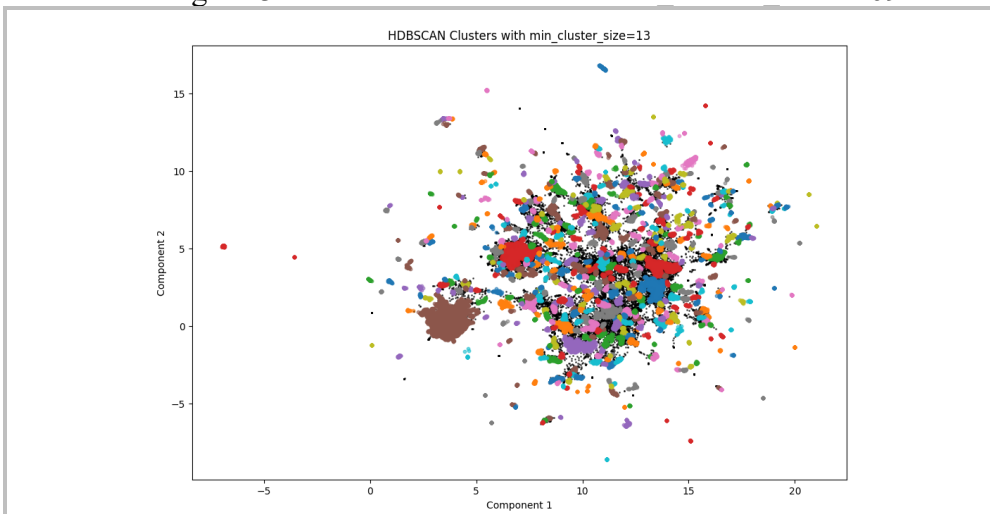


Figure 4: HDBSCAN clusters at min_cluster_size = 13

A hierarchical tree structure is created with each cluster as a node. The clusters from one level of `min_cluster_size` generally align with a cluster at the parent level, but not exactly. Parent-child relationships are determined by identifying the parent cluster (or none, assigning to ‘other’) with the maximum overlap among candidate parents.

Clusters remain anonymous until a meaning is identified and assigned. To identify a meaning to associate, the most central embedding is identified for each cluster, and that embedding’s associated name is used as the node name.

In the case of pages, the dataset provides only the index details (book title, page identifier) and full page text. In order to assign a topic to a node associated with a page, a deep learning topic is elicited for each selected page by the same Mistral-7B-Instruct-v0.2 instruction-tuned language model used in dataset creation. Prompt engineering was ineffective in achieving fully-reliably elicitation and capture of a deep learning topic. Most notable, the generated output consistently refused to follow the requested JSON format, expressed as explicit constraints (Ekin, 2023) with examples in prompts (Marvin et al., 2023), so the prompt is adapted to the most commonly observed format from trials. Parsing the model output invokes a number of clauses to recover as much as practicable. Still, the model fails to generate parsable output for a regular fraction of requests. Repeated requests on the order of 30 repeats gradually resolved the long tail of parse issues, to achieve topic identification for 653 of 676 pages. The remaining 23 pages were manually reviewed to identify the topic. Among these, several are index pages, lists of references, tables, or primarily jumbles of characters, which may be challenging for the model to interpret.

The extracted hierarchy is presented in Appendix B.

All inference and training was performed on CPU or an NVIDIA RTX 4090 GPU.

4.2 Hierarchical Topic Extraction - Limitations

The topic names are selected using the most central embedding within each cluster. This selects what may be a representative instance of the cluster, but does not consider the breadth of topics within the cluster. Eliciting a summary of the cluster terms can help generalize the topic names to better represent the cluster content.

(Ayush & Sia, 2021) explore improvements to UMAP and downstream clustering that can achieve better representation, resulting in improved clustering performance on texts.

4.3 Hierarchical Topic Extraction - Other Methods Explored

Dimensionality reduction was performed on each of the separate embedding types (titles, pages, topics). These separate reductions resulted in poor clustering relative to dimensionality-reduction with the types combined. The embeddings use the same embedding model, so their combination should result in a coherent representation space. It seems that the inclusion of the topic embeddings helps to find a dimensionality reduction that more effectively

separates the page embeddings.

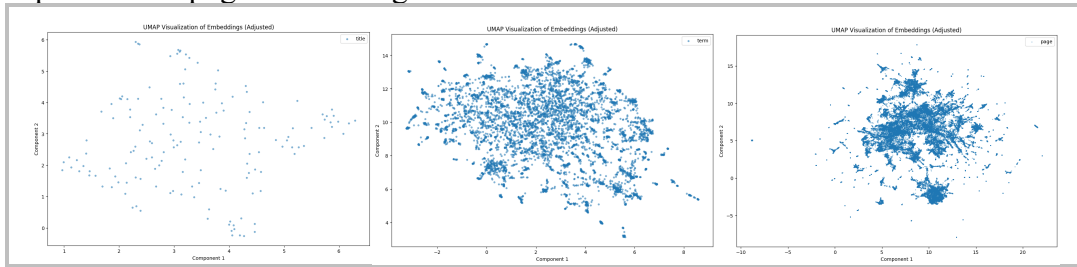


Figure 5: Title, Page and Term Embeddings shown in separate UMAP embedding spaces.

t-SNE is a popular dimensionality-reduction method for textual topic modeling towards document clustering, as used in (Nokolov et al., 2024). (González-Márquez et al. 2022) apply t-SNE to large document libraries, similar to the application described in this project, but stop at visualization.

Dimensionality reduction was performed using t-SNE in addition to UMAP. t-SNE resulted in more diffuse clusters and little to no global structure. This is expected as t-SNE prioritizes local differences, whereas UMAP prioritizes global structure. (Arora et al., 2018) explore guarantees in capturing cluster structure in t-SNE dimensionality reduction.

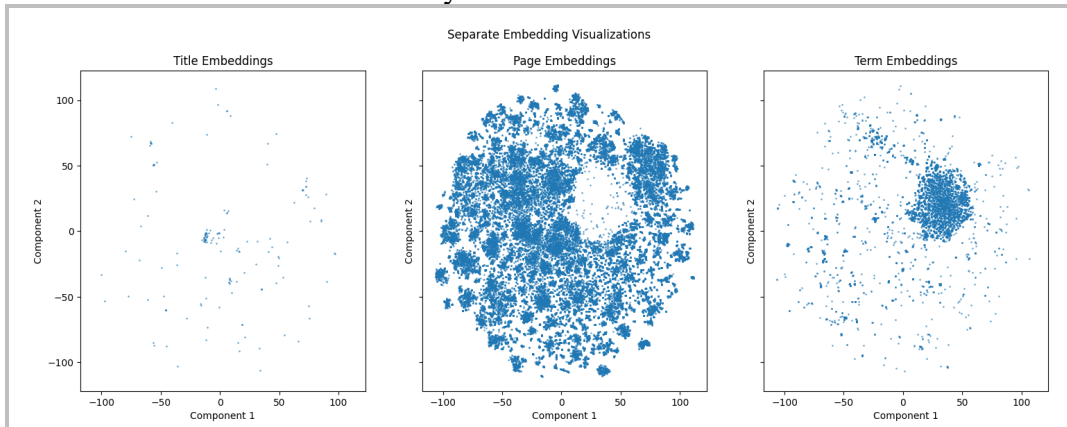


Figure 6: Title, Page and Term Embeddings shown separately in the t-SNE combined embedding space.

t-SNE term embeddings were visualized sized by page count for the term. This appears to possibly suggest more explicit clustering, but not as clearly as other methods.

(continues on next page)

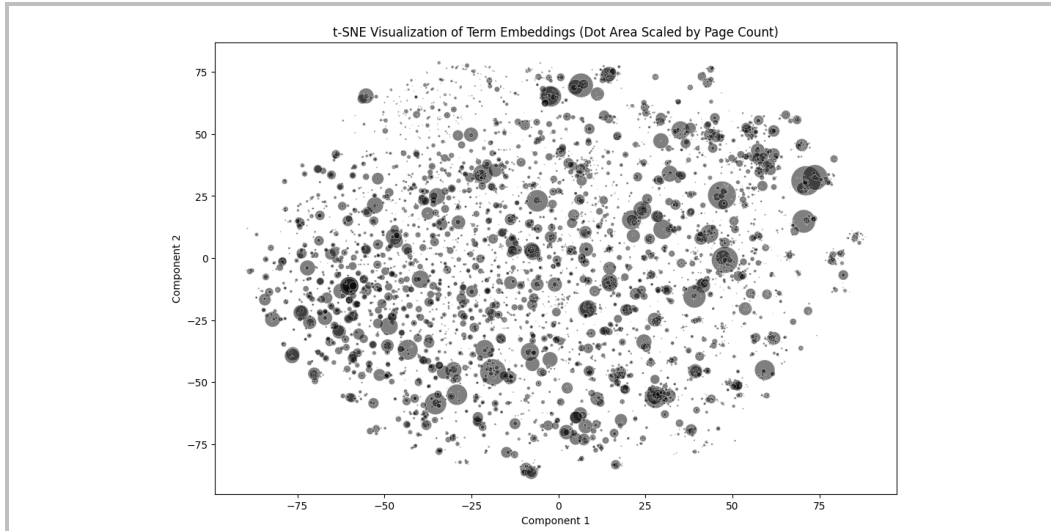


Figure 7: t-SNE Dimensionality-Reduction - Embedding Size by Page Count
 Clustering was performed with DBSCAN in addition to HDBSCAN as utilized. HDBSCAN captured cleaner, more integral clusters vs DBSCAN, whose clusters were less cleanly contained, and which often nearly to entirely surrounded other clusters, suggesting that the separations may not be as meaningful. Tuning on DBSCAN revealed that both the eps and min_samples parameters were effective in exposing what appeared to be hierarchical clustering behavior.

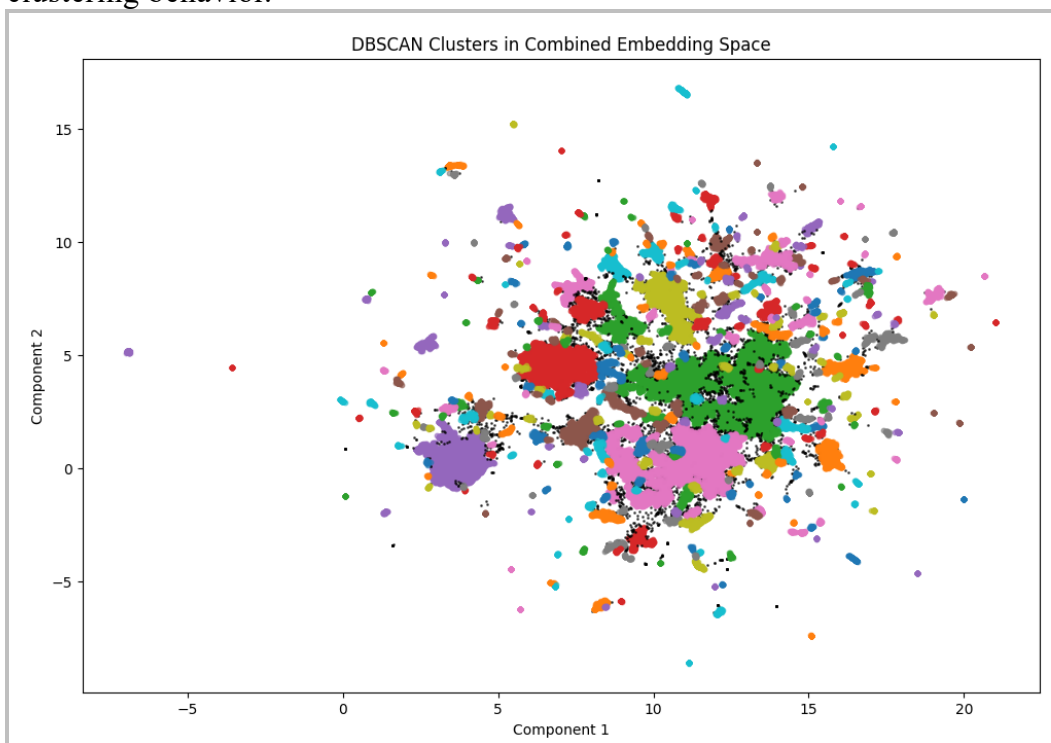


Figure 8: DBSCAN clusters are not cleanly defined

Before selecting different scales of min_cluster_size with HDBSCAN as the driver of parent-child relationships, topic co-occurrence across clusters was

explored, including combinatorially, and via a deep learning model. Initial identified co-occurrences appeared less sensible, so this method was abandoned. Further work on this approach is likely promising, but likely requires modification to become useful.

4.4 Hierarchical Topic Extraction - Further Work

Whereas this project combines source material types and uses a single clustering method with different parameter values, (Espinosa, 2021) explores the combined use of multiple different clustering sources and methods. As explored by (McCleary et al., 2024), fine-tuning of Mistral 7B can improve representation learning on text data. As explored by (Mahata et al., 2018) on a similarly technical corpus of scientific papers, PageRank can capture relationships among phrase embeddings. (Chien & Chueh, 2012) applied Latent Dirichlet allocation (LDA) for hierarchical segmentation in topic modeling, in which documents are segmented into sub-documents before topic extraction. (Dou et al., 2013) provide a user interface for topic hierarchy visualization and exploration. (Gavilanes et al., 2024) demonstrate an end-to-end system for parsing a corpus of documents, extracting topics, automating access, and performing ongoing maintenance.

This project explored and delivered creation of a single topic hierarchy. (Singgalen, 2024) identified that different representation and clustering methods can capture different topic hierarchies corresponding to different perspectives.

5 Conclusion

This project proposed and explored a novel approach to concept mapping and the extraction of a topic hierarchy, while demonstrating its application to topic-driven indexing of technical literature (books) within the field of deep learning.

The project delivers a dataset covering the complete contents of 158 academic books about deep learning, comprising:

- 158 book titles with embeddings.
- 6217 topics relevant to the field of deep learning, with embeddings.
- 50,794 book pages, with embeddings.
- A structured index from all 6217 topics to specific corresponding text and page numbers among the 50,794 book pages, totaling 1,902,952 index entries.

(continues on next page)

This project leverages the deep learning books dataset to develop a novel method for topic hierarchy extraction, based on:

- n-gram (word) sequence capture and indexing of common phrases.
- Mistral-7B-Instruct-v0.2 for identification and elicitation of topics from keyphrases, and complete page texts.
- all-mpnet-base-v2 for generation of embeddings in a common space.
- UMAP for dimensionality reduction of the embedding space, leveraging structure in topic embeddings to provide structure for page embeddings.
- HDBSCAN repeated with different parameter values for clustering.
- Hierarchy extraction via analysis of nesting of HDBSCAN clusters.

5.1 Research Insights

This project qualitatively demonstrated that combining topic embeddings with page embeddings can result in clearer structure and concept separation among the page embeddings.

The project compared techniques and found that:

- UMAP captured more global structure than t-SNE, while preserving sufficient local structure to support clustering.
- HDBSCAN achieved more coherent clusters than DBSCAN or k-means.
- Adjusting the `min_cluster_size` parameter for HDBSCAN resulted in separation of coherent sub-topics within topic clusters at higher levels of granularity.
- An instruction tuned LLM (Mistral-7B-Instruct-v2.0) can be effectively paired with classical machine learning techniques to achieve a combination of efficiency, thoroughness and quality.

5.2 Potential Industry Applications

The topic hierarchy extraction method developed in this project combines a structured index (from specific named topics into specific text on specific pages), with a hierarchical topic representation (of book titles, topics and page content) within a shared representation embedding space.

This combination of access mechanisms provides potentially powerful tools for applications in:

- Library search and recommendation
- Exploration of topic relationships within complex fields
- Retrieval-augmented generation for language models and agents

References

- [1] Giray, L. **Prompt Engineering with ChatGPT: A Guide for Academic Writers**. Annals of Biomedical Engineering, 51, 2629–2633. (2023). <https://doi.org/10.1007/s10439-023-03272-4>
- [2] Sabit Ekin . **Prompt Engineering For ChatGPT: A Quick Guide To Techniques, Tips, And Best Practices**. *TechRxiv*. May 04, 2023. DOI: [10.36227/techrxiv.22683919.v2](https://doi.org/10.36227/techrxiv.22683919.v2)
- [3] Marvin, G., Hellen, N., Jjingo, D., Nakatumba-Nabende, J. (2024). **Prompt Engineering in Large Language Models**. In: Jacob, I.J., Piramuthu, S., Falkowski-Gilski, P. (eds) Data Intelligence and Cognitive Informatics. ICDICI 2023. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-99-7962-2_30
- [4] Ashqar, G., & Mutlu, A. (2023). **A comparative assessment of various embeddings for keyword extraction**. In Proceedings of the 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (pp. 01-06). <https://doi.org/10.1109/HORA58378.2023.10156762>
- [5] Siino, M. (2024). **All-Mpnet at SemEval-2024 Task 1: Application of Mpnet for evaluating semantic textual relatedness**. In Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024) (pp. 379–384). Association for Computational Linguistics.
- [6] Asmala, S. (2024, April 19). **Improving the Search of Medical Articles Using Sentence Transformers** [Master's Thesis, Aalto University]. Computer, Communication and Information Sciences, Computer Science, Big Data and Large-Scale Computing track.
- [7] Altinok, D. (2021). **Mastering spaCy: An end-to-end practical guide to implementing NLP applications using the Python ecosystem**. Packt Publishing.
- [8] Murari Kumar, Krishna Kumar Chaturvedi, Anu Sharma et al. **An Algorithm for Automatic Text Annotation for Named Entity Recognition using spaCy Framework**, 17 May 2023, PREPRINT (Version 1) available at Research Square [<https://doi.org/10.21203/rs.3.rs-2930333/v1>]
- [9] Medelyan, O. (2009). **Human-competitive automatic topic indexing** [Doctoral dissertation, University of Waikato]. Department of Computer Science, Hamilton, New Zealand.
- [10] UMAP: **Uniform Manifold Approximation and Projection for Dimension Reduction**. Leland McInnes, John Healy, James Melville. (2018) [arXiv:1802.03426](https://arxiv.org/abs/1802.03426) [stat.ML]
- [11] **Exploratory Analysis of Covid-19 Tweets using Topic Modeling, UMAP, and DiGraphs**. Catherine Ordun, Sanjay Purushotham, Edward Raff. [arXiv:2005.03082](https://arxiv.org/abs/2005.03082) [cs.SI]
- [12] **Is Neural Topic Modelling Better than Clustering? An Empirical Study on Clustering with Contextual Embeddings for Topics**. Zihan Zhang, Meng Fang, Ling Chen, Mohammad-Reza Namazi-Rad. [arXiv:2204.09874](https://arxiv.org/abs/2204.09874) [cs.CL]
- [13] Asyaky, M. S., & Mandala, R. (2021). **Improving the performance of HDBSCAN on short text clustering by using word embedding and UMAP**. In Proceedings of the 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA) (pp. 1-6). <https://doi.org/10.1109/ICAICTA53211.2021.9640285>
- [14] Arora, S., Hu, W. & Kothari, P.K.. (2018). **An Analysis of the t-SNE Algorithm for Data Visualization**. *Proceedings of the 31st Conference On Learning Theory*, in *Proceedings of Machine Learning Research* 75:1455-1462 Available from <https://proceedings.mlr.press/v75/arora18a.html>.
- [15] Nikolov, N. S., Dowa, A. S., & Alakrot, A. (2024). **Identifying predators in online conversations by clustering based on topic modelling and T-SNE visualization**. International Scientific and Technical Journal.
- [16] González-Márquez, R., Berens, P., & Kobak, D. (2022, March 26). **Two-dimensional visualization of large document libraries using t-SNE**. OpenReview.
- [17] Monko, G., & Kimura, M. (2024). **Enhanced SS-DBSCAN clustering algorithm for high-dimensional data**. Data Science, 0(1), 1-0.
- [18] McCleary, K., Ghawaly, J., & Miele, L. (2024, March 1). **TNM tumor classification from unstructured breast cancer pathology reports using LoRA finetuning of Mistral 7B**. AAAI 2024 Symposium on Clinical Foundation Models.
- [19] D. Mahata, R. R. Shah, J. Kuriakose, R. Zimmermann and J. R. Talburt, "Theme-Weighted Ranking of Keywords from Text Documents Using Phrase Embeddings," *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, Miami, FL, USA, 2018, pp. 184-189, doi: 10.1109/MIPR.2018.00041.

- [20] A. Collins and J. Beel, "Document Embeddings vs. Keyphrases vs. Terms for Recommender Systems: A Large-Scale Online Evaluation," *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, Champaign, IL, USA, 2019, pp. 130-133, doi: 10.1109/JCDL.2019.00027.
- [21] J. -T. Chien and C. -H. Chueh, "Topic-Based Hierarchical Segmentation," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 55-66, Jan. 2012, doi: 10.1109/TASL.2011.2143405.
- [22] W. Dou, L. Yu, X. Wang, Z. Ma and W. Ribarsky, "Hierarchical Topics: Visually Exploring Large Text Collections Using Topic Hierarchies," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2002-2011, Dec. 2013, doi: 10.1109/TVCG.2013.162.
- [23] Wang, J., & Wang, J. (2019). *Hierarchical topic clustering over large collections of documents*. Athens: The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). Retrieved from <https://ezproxy.lib.utexas.edu/login?url=https://www.proquest.com/conference-papers-proceedings/hierarchical-topic-clustering-over-large/docview/2277982988/se-2>
- [24] Calero Espinosa, J. (2021). *Multi-view learning for hierarchical topic detection on corpus of documents*. Universidad Nacional de Colombia.
- [25] Lutao Zhang, Xuesong Su, Yifei Wang, [Mei Wang](#), Xinxin Yang, and Zixuan Xu "HDBSCAN-based semantic clustering model in classifying incidents on security and environmental conservation management", Proc. SPIE 13291, Ninth International Symposium on Advances in Electrical, Electronics, and Computer Engineering (ISAECE 2024), 132915N (16 October 2024); <https://doi.org/10.1117/12.3033910>
- [26] Singgalen, Y. A. (2024). Analyzing social networks and topic clustering in backpacker tourism content reviews using K-means, Fast HDBScan, and Gaussian Mixture with Communalistic. *Journal of Information System Research (JOSH)*, 6(1), 34-48. <https://doi.org/10.47065/josh.v6i1.5969>
- [27] Tran, TH., Cao, TD., Tran, TTH. (2021). HDBSCAN: Evaluating the Performance of Hierarchical Clustering for Big Data. In: Phuong, N.H., Kreinovich, V. (eds) Soft Computing: Biomedical and Related Applications. Studies in Computational Intelligence, vol 981. Springer, Cham. https://doi.org/10.1007/978-3-030-76620-7_24
- [28] Zirnstein, B. (2023, June 26). Extended context for InstructGPT with LlamaIndex [Report]. https://www.researchgate.net/profile/Bruno-Zirnstein/publication/371911146_Extended_context_for_InstructGPT_with_LlamaIndex/links/649bd64e95bbbe0c6efc1345/Extended-context-for-InstructGPT-with-LlamaIndex.pdf
- [29] Jo, M., & Lee, J. (2024). Llama index-based Machine Learning Model for Emergency rescue. *Annual Conference of KIPS*, 705–706. <https://doi.org/10.3745/PKIPS.Y2024M05A.705>
- [30] Gavilanes, J., Bozhilov, Y., Dodeja, U., Valtas, G., & Badrajan, A. (2024). Use of LLM for methods of information retrieval [Thesis, University of Twente].
- [31] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Léo Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, William El Sayed. **Mistral 7B**. (2023) [arXiv:2310.06825](https://arxiv.org/abs/2310.06825) [cs.CL]
- [32] Dalmia, Ayush, Sia, Suzanna. Clustering with UMAP: Why and How Connectivity Matters. (2021). [arXiv:2108.05525](https://arxiv.org/abs/2108.05525) [cs.AI]

Appendix A: Books in Dataset

The dataset consists of the following books, listed as ‘Title - First Listed Author’:

1. 3D Imaging - Multidimensional Signal Processing and Deep Learning - Patnaik
2. 3D Imaging—Multidimensional Signal Processing and Deep Learning - Volume 1 - Jain
3. 3D Imaging—Multidimensional Signal Processing and Deep Learning - Volume 2 - Jain
4. Activation Functions - Activation Functions in Deep Learning with LaTeX Applications - Kutuk
5. Advanced Analytics and Deep Learning Models - Mire
6. Advanced Applied Deep Learning - Convolutional Neural Networks and Object Detection - Michelucci
7. Advanced Deep Learning for Engineers and Scientists - A Practical Approach - Prakash
8. Advanced Deep Learning with Keras - Atienza, Rowel
9. Advanced Methods and Deep Learning in Computer Vision - Davies
10. Advances and Applications in Deep Learning - Aceves-Fernandez
11. Advances in Deep Learning - Wani
12. Applications of Deep Machine Learning in Future Energy Systems - Khooban
13. Applications of Game Theory in Deep Learning - Hazra
14. Applied Deep Learning - Fergus
15. Applied Deep Learning - Michelucci
16. Artificial Intelligence Trends for Data Analytics Using Machine Learning and Deep Learning Approaches - Devi
17. Artificial Intelligence and Brain Research - Neural Networks, Deep Learning and the Future of Cognition - Krauss
18. Artificial Intelligence by Example - Acquire Advanced AI, Machine Learning, and Deep Learning Design Skills - Rothman
19. Artificial Intelligence for Scientific Discoveries - Extracting Physical Concepts from Experimental Data Using Deep Learning - Iten
20. Artificial Intelligence, Machine Learning, and Deep Learning - Campesato
21. Attacks, Defenses and Testing for Deep Learning - Chen
22. Automated Deep Learning Using Neural Network Intelligence - Gridin
23. Automated Software Engineering - A Deep Learning-Based Approach - Satapathy
24. Automatic Speech Recognition - A Deep Learning Approach - Dong Yu
25. Beginning Anomaly Detection Using Python-Based Deep Learning - Sridhar Alla
26. Big Data Analysis and Deep Learning Applications - Proceedings of the First International Conference on Big Data Analysis and Deep Learning - Zin
27. Blockchain of Things and Deep Learning Applications in Construction - Digital Construction Transformation - Elghaish
28. Building Machine Learning and Deep Learning Models on Google Cloud Platform - Ekaba Bisong
29. Computational Mechanics with Deep Learning - Yagawa
30. Computational Methods and Deep Learning for Ophthalmology - Hemanth
31. Computational Methods for Deep Learning - Yan
32. Computational Texture and Patterns - From Textons to Deep Learning - Dana
33. Computer Vision Using Deep Learning - Neural Network Architectures with Python and Keras - Verdhan
34. Concepts and Real-Time Applications of Deep Learning - Srivastava
35. Convergence of Deep Learning In Cyber-IoT Systems and Security - Chakraborty
36. Data Orchestration in Deep Learning Accelerators - Krishna
37. Deep Cognitive Networks - Enhance Deep Learning by Modeling Human Cognitive Mechanism - Huang
38. Deep Learning - A Practical Introduction - Martinez-Ramon
39. Deep Learning - Concepts and Architectures - Pedrycz
40. Deep Learning - Convergence to Big Data Analytics - Khan
41. Deep Learning - Foundations and Concepts - Bishop
42. Deep Learning - Fundamentals, Theory and Applications - Huang
43. Deep Learning - Goodfellow
44. Deep Learning - Kelleher
45. Deep Learning - Recent Findings and Research - Dominguez-Morales
46. Deep Learning Applications in Image Analysis - Roy
47. Deep Learning Approach for Natural Language Processing, Speech, and Computer Vision - Kumar
48. Deep Learning Approaches for Spoken and Natural Language Processing - Kadyan
49. Deep Learning Approaches to Text Production - Narayan
50. Deep Learning Architecture and Applications - Zhang

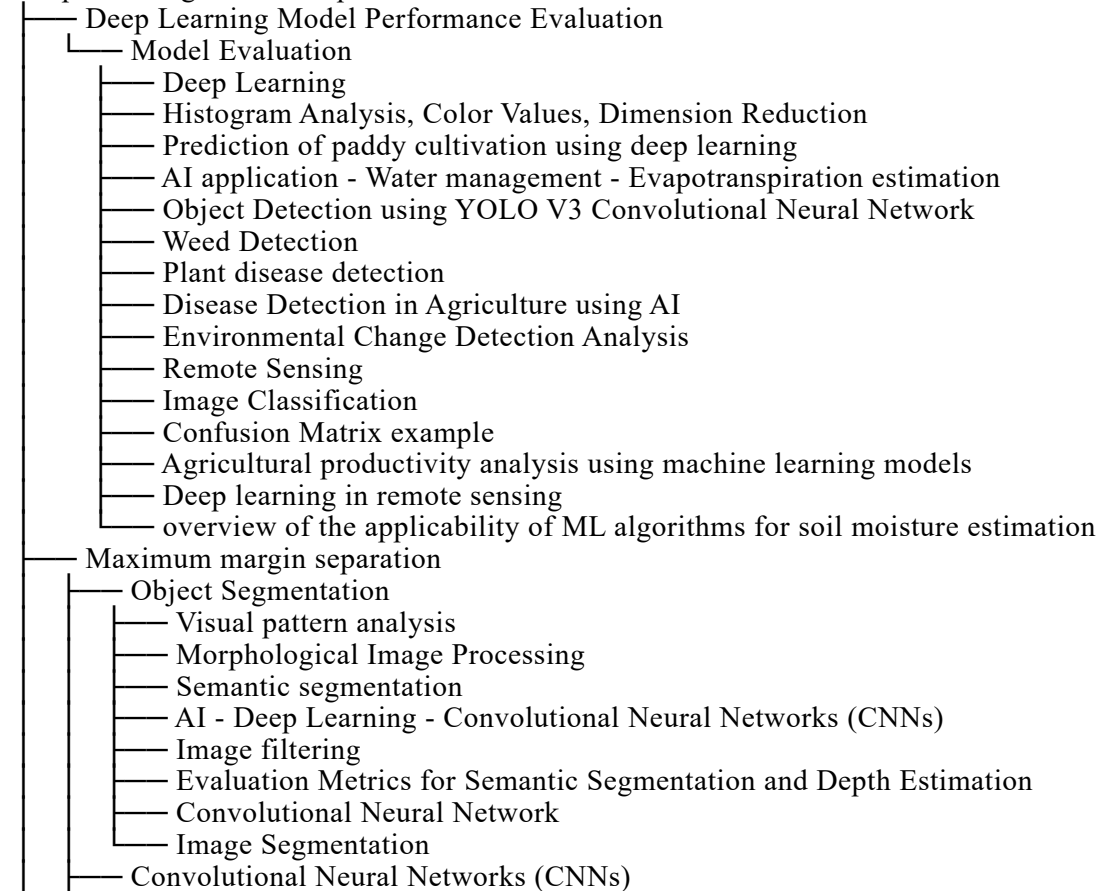
51. Deep Learning Architectures - A Mathematical Approach - Calin
52. Deep Learning Based Speech Quality Prediction - Mittag
53. Deep Learning Classifiers with Memristive Networks - James
54. Deep Learning Foundations - Jo
55. Deep Learning Models - A Practical Approach for Hands-On Professionals - Gamba
56. Deep Learning Networks - Design, Development and Deployment - Singaram
57. Deep Learning Techniques for Music Generation - Briot
58. Deep Learning and Computational Physics - Ray
59. Deep Learning and Edge Computing Solutions for High Performance Computing - Suresh
60. Deep Learning and Its Applications Using Python - Basha
61. Deep Learning and Physics - Tanaka
62. Deep Learning and Reinforcement Learning - Yang
63. Deep Learning at Scale - Mall
64. Deep Learning for Agricultural Visual Perception - Crop Pest and Disease Detection - Wang
65. Deep Learning for Autonomous Vehicle Control - Algorithms, State-Of-the-Art, and Future Prospects - Kuutti
66. Deep Learning for Computer Architects - Brandon Reagen, Robert Adolf, Paul Whatmou
67. Deep Learning for EEG Analysis - DaSilva
68. Deep Learning for Fluid Simulation and Animation - Fundamentals, Modeling, and Case Studies - Giraldi
69. Deep Learning for NLP and Speech Recognition - Kamath
70. Deep Learning for Natural Language Processing - Goyal
71. Deep Learning for Power System Applications - Li
72. Deep Learning for Robot Perception and Cognition - Iosifidis
73. Deep Learning for Social Media Data Analytics - Hong
74. Deep Learning for Sustainable Agriculture - Poonia
75. Deep Learning for Time Series Forecasting - Predict the Future
76. Deep Learning for Unmanned Systems - Koubaa
77. Deep Learning for Video Understanding - Wu
78. Deep Learning in Computational Mechanics - Kollmannsberger
79. Deep Learning in Mining of Visual Content - Zemmari
80. Deep Learning in Multi-step Prediction of Chaotic Dynamics - From Deterministic Models to Real-World Systems - Sangiorgio
81. Deep Learning in Object Detection and Recognition - Jiang
82. Deep Learning in Science - Baldi
83. Deep Learning in Time Series Analysis - Gharehbaghi
84. Deep Learning on Edge Computing Devices - Zhou
85. Deep Learning on Microcontrollers - Learn How to Develop Embedded AI Applications Using TinyML - Gupta
86. Deep Learning to See - Betti
87. Deep Learning with PyTorch - Vishnu Subramanian
88. Deep Learning with Python - Francois Chollet
89. Deep Learning with Relational Logic Representations - Sir
90. Deep Learning, Reinforcement Learning, and the Rise of Intelligent Systems - Uddin
91. Deep Learning-Based Approaches for Sentiment Analysis - Agarwal
92. Deep Learning-Based Face Analytics - Ratha
93. Deep Learning-Based Forward Modeling and Inversion Techniques for Computational Physics Problems - Wang
94. Deep Reinforcement Learning - Dong
95. Deep Reinforcement Learning - Sewak
96. Deep Reinforcement Learning Hands-On - Lapan
97. Deep Reinforcement Learning Processor Design for Mobile Applications - Lee
98. Deep Reinforcement Learning and Its Industrial Use Cases - Mahajan
99. Deep Reinforcement Learning with Guaranteed Performance - A Lyapunov-Based Approach - Zhang
100. Deep Reinforcement Learning with Python - RLHF for Chatbots and Large Language Models - Sanghi
101. Deep Reinforcement Learning with Python - Sanghi
102. Deep learning with Python - learn best practices of deep learning models with PyTorch - Ketkar
103. Deep reinforcement learning - Plaaf
104. Deep reinforcement learning in Unity - Majumder
105. Demystifying Deep Learning - An Introduction to the Mathematics of Neural Networks - Santry
106. Development and Analysis of Deep Learning Architectures - Pedrycz
107. Domain adaptation in computer vision with deep learning - Venkateswara
108. Embedded Deep Learning - Bert Moons

109. Explainable AI - Interpreting, Explaining and Visualizing Deep Learning - Samek
110. Explainable Deep Learning AI - Benois-Pineau
111. Feature and Dimensionality Reduction for Clustering with Deep Learning - Ros
112. Fog Computing, Deep Learning and Big Data Analytics-Research Directions - Prabhu
113. Fundamentals and Methods of Machine and Deep Learning - Singh
114. Generative Deep Learning - Foster
115. Generative adversarial networks and deep learning - Raut
116. Geometric Deep Learning Grids, Groups, Graphs, Geodesics, and Gauges - Bronstein
117. Geometry of Deep Learning - A Signal Processing Perspective - Ye
118. Guide to Deep Learning Basics - Logical, Historical and Philosophical Perspectives - Skansi
119. Hacking artificial intelligence - a leader's guide from deepfakes to breaking deep learning - Gibian
120. Handbook of Deep Learning Applications - Balas
121. Hands-On Reinforcement Learning with Python - Master Reinforcement and Deep Reinforcement Learning Using OpenAI Gym and TensorFlow - Sudharsan Ravichandiran
122. Hyperparameter Optimization in Machine Learning - Make Your Machine Learning and Deep Learning Models More Efficient - Agrawal
123. Intelligent Autonomous Drones with Cognitive Deep Learning - Build AI-Enabled Land Drones with the Raspberry Pi 4 - Blubaugh
124. Interpretability in Deep Learning - Somani
125. Learn Computer Vision Using OpenCV - With Deep Learning CNNs and RNNs - Gollapudi
126. Learning Deep Brain Stimulation Management through Clinical Cases - Beudel
127. Malware Analysis Using Artificial Intelligence and Deep Learning - Stamp
128. Many-Sorted Algebras for Deep Learning and Quantum Technology - Giardina
129. Mathematical Aspects of Deep Learning - Grohs
130. Mathematical Engineering of Deep Learning - Liquet
131. Mathematical Introduction to Deep Learning - Methods, Implementations, and Theory - Jentzen
132. Modern Deep Learning Design and Application Development - Ye
133. Modern Deep Learning for Tabular Data - Ye
134. Multi-faceted Deep Learning - Benois-Pineau
135. Multimodal Scene Understanding - Algorithms, Applications and Deep Learning - Yang
136. Navigating Mathematical Basics - A Primer for Deep Learning in Science - Liquet
137. Neural Networks and Deep Learning - Charu C. Aggarwal
138. Order Analysis, Deep Learning, and Connections to Optimization - Jahn
139. Practical Computer Vision Applications Using Deep Learning with CNNs - Ahmed Fawzy Gad
140. Practical Deep Learning for Cloud, Mobile & Edge - Anirudh Koul
141. Principles of Artificial Neural Networks - Basic Designs to Deep Learning - Graupe
142. Privacy-Preserving Deep Learning - A Comprehensive Survey - Kim
143. Sensor Data Analysis and Management - The Role of Deep Learning - Suresh
144. Session-Based Recommender Systems Using Deep Learning - Ravanmehr
145. Shallow and Deep Learning Principles - Scientific, Philosophical, and Logical Perspectives - Sen
146. State-of-the-Art Deep Learning Models in TensorFlow - Modern Machine Learning in the Google Colab Ecosystem - Paper
147. Synthetic Data for Deep Learning - Generate Synthetic Data for Decision Making and Applications with Python and R - Gursakal
148. Synthetic Data for Deep Learning - Nikolenko
149. The Deep Learning Revolution - artificial intelligence meets human intelligence - Terrence J. Sejnowski
150. The Development of Deep Learning Technologies - Research on the Development of Electronic Information Engineering Technology in China - Chinese Academy of Engineering
151. The Modern Mathematics of Deep Learning - Berner
152. The Principles of Deep Learning Theory - An Effective Theory Approach to Understanding Neural Networks - Roberts
153. The Principles of Deep Learning Theory - Hanin
154. Time Series Algorithms Recipes - Implement Machine Learning and Deep Learning Techniques with Python - Kulkarni
155. Trends in Deep Learning Methodologies - Algorithms, Applications, and Systems - Piuri
156. Understanding Deep Learning - Prince
157. Variational methods for machine learning with applications to deep networks - Cinelli
158. Visual and Text Sentiment Analysis through Hierarchical Deep Learning Networks - Chaudhuri

Appendix B: Topic Hierarchy

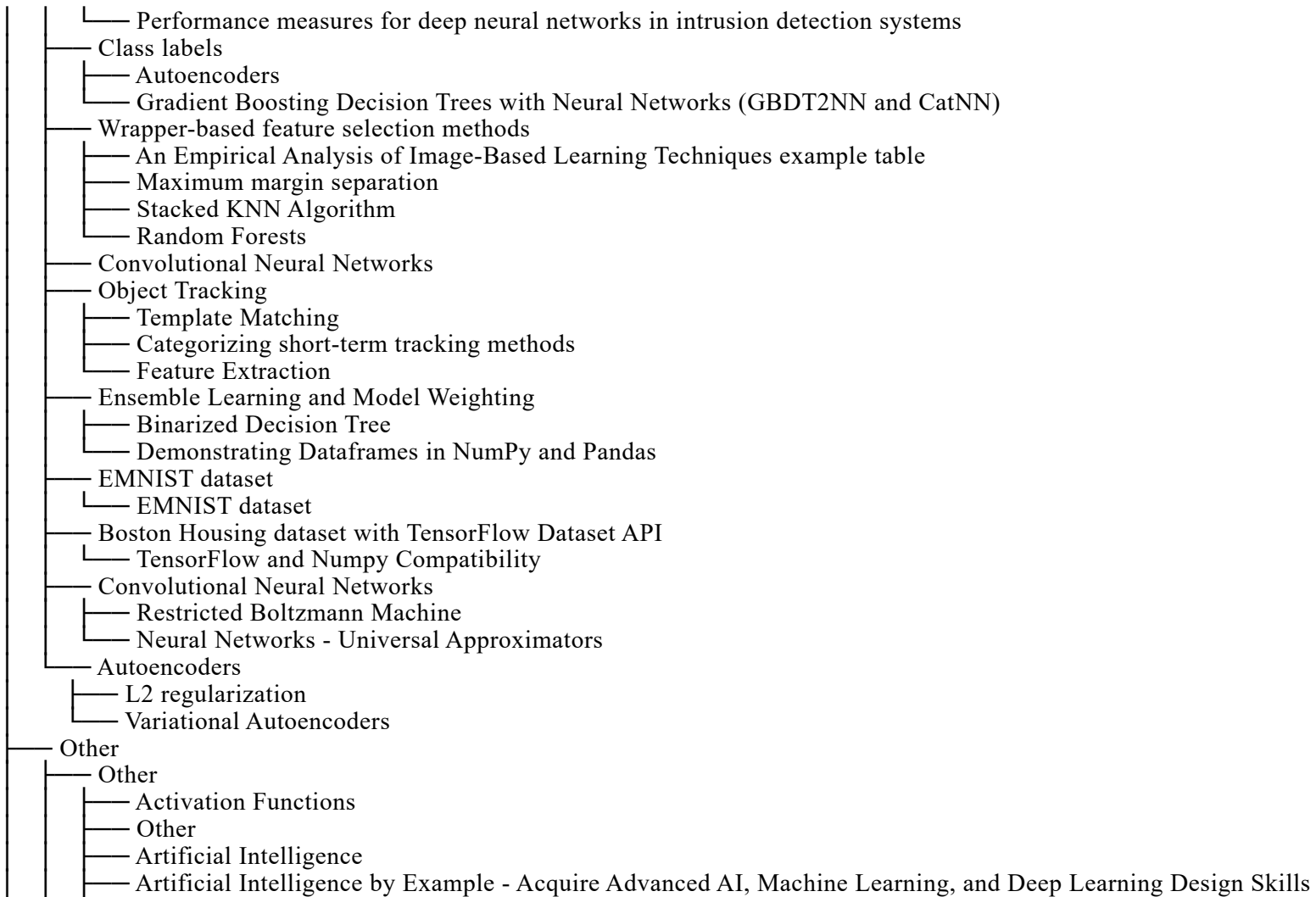
Complete topic hierarchy:

Deep Learning Books Corpus



- Convolutional Neural Network
- Deep Learning
- Deep Learning in Neutrino Physics and Antimatter Research
- Deep Learning Networks
- self-supervised learning
- Transfer Learning
- Learning deep neural networks from corrupted datasets
- Data Augmentation
- Transfer Learning
- shift register sequences
- Stat
- Cerebellum and cognitive codes
- Deep learning libraries and frameworks
- Automated Deep Learning
- Machine learning
- Convolutional neural networks (CNNs)
- Variational Inference
- Generative Adversarial Networks
- CGAN
- CycleGAN
- Texture Style Transfer
- Style Transfer using Gram Matrix in PyTorch
- StackedGAN code example
- Convolutional Neural Networks (C
- Generative AI
- text-to-image models
- DCGAN Discriminator
- Multilayer Perceptrons (MLPs) and Backpropagation
- Vanishing Gradient Problem in GANs
- Deep Learning as Sparsity-Enforcing Algorithms
- Conditional Vector Optimization Problems
- Deep Learning - Activation Functions
- Gradient Descent and Cost Function Minimization in Neural Networks

- Clustering - Unsupervised Learning
- Deep Neural Networks
- Backpropagation
- Activation Functions in Deep Learning
- Cross-entropy loss function in deep learning
- Stochastic Gradient Descent (SGD)
- Stochastic Gradient Descent (SGD) for Training Deep Models
- Adagrad and RMSProp with Momentum
- Dropout
- Neural models and applications in machine learning
- Curse of Dimensionality
- Decision
- Regularization Techniques
- Regularization in Deep Learning
- RMSProp
- Mathematical Neural Network
- A Proof in Linear Algebra
- Riesz Representation Theorem
- Information Compression in Neural Networks
- Backpropagation
- Index
- Fisher Matrix
- Theoretical Results for Generalization Gap in Two
- Backpropagation Algorithm
- Conditional Probabilities
 - Machine Learning - Probabilistic Graphical Models - Undirected Models
 - Optimization algorithms
 - Backpropagation Algorithm
- Pre-processing
 - Pre-processing
- Inductive bias in deep learning
 - Bayesian Inference with Finite-width Neural Networks
- Convolution



- Deep Learning
- Deep Cognitive Networks
- Deep Learning Approaches to Text Production
- Deep Reinforcement Learning
- Hopfield network learning algorithm
- Lipschitz Continuity
- Deploying deep learning networks in Edge devices
- Modeling and simulating human brain-environment interaction using neural networks
- List of Contributors
- Camera Control for Image Acquisition in AI/DL Systems
- Graph Database Technology for Distributed Energy Networks
- Data augmentation
- Chest X-ray analysis
- Bayesian Network
- Hashing methods for face verification and search
- License Plate Recognition
- Spectral theory of linear operators
- convolutional_neural_networks
- 3D Digital Technology in Traditional Arts and Crafts
- Rule-based Neural Network for Resource Allocation in Cloud Environment (RBNNOM)
- Multi-timescale Hidden Markov Models (MTS-HMM)
- Application of machine-learning-based vision algorithms for detecting surface defects in industrial products
- Mining Frequently Occurring Patterns in Big Data using Hadoop
- Discrete SIR and SIS Systems
- Perspective Invariants
- Image processing or analysis
- Autoencoders for feature generation
- Automatic Essay Grading
- Mosaic image analysis
- Mean Square Error (MSE) and Root Mean Square Error (RMSE) for Regression Problems
- References
- Deep Learning in Medical Image Classification
- 2D Convolution in Deep Learning

- Atmospheric Interference Removal in Radar-based Deformation Measurement
- NLP and Language Modeling
- Semi-supervised Learning
- Friedman Functions
- Natural Language Processing (NLP)
- Tensor slicing in Numpy
- Deep learning in biomedical image analysis
- Classification Analysis
- Rheumatoid Arthritis Symptom Prediction
- Generative Adversarial Networks (GANs)
- ECG database classification results
- Bellman equations
- Image Segmentation
- Long-Term Dependencies in Recurrent Neural Networks
- Authoring, Feedback, Errata, Piracy
- Text Classification, Convolutional Neural Networks
- Image Blurring
- Residual connections in deep learning, specifically the interpretation as a branching operation
- Image Style Transfer
- Image Source Locating
- Reinforcement learning
- Fuzzy Deep Learning
- {m_input:l_btc, m_true:m_true_btc}
- Cross-validation
- Convolutional Neural Networks (CNNs)
- Pairwise scatter plot
- Predicting
- Inverse Variance Portfolio Optimization
- Deep learning
- Health Survey summaries
- COVID-19 epidemic prediction using deep learning
- Back Propagation
- Transparency in AI

- Probabilistic anomaly detection using Variational Autoencoders (VAEs) on time-series data
- 3D object detection and tracking
- Limited Supervision Learning
- Synthetic Data for Fraud Detection
- Affordance Mapping in AI or Deep Learning contexts
- Event Segmentation in Computer
- Deep Plug-and-Play IR
- Index
- Deep learning in nanophotonics design optimization
- Meta-Learning
- Dynamic Model Deployment
- Deep learning
- Latent Variables in Deep Learning
- Performance Comparison of Capsule Networks
- Principal Component Analysis (PCA)
- Artificial Neural Networks
- False Data Injection Attacks
- LPV (Linear Parameter-Varying) controller design for learning-based control systems
- Memory Allocation for Multi-dimensional Arrays in DLneuroBLAS_mem.c
- Power Electronics for AI/DL Systems
- Device Fingerprinting
- Real World Use Cases, Smart Agriculture
- gated RNNs
- Balancing Data
- Limitations of Quadratic Discriminant Analysis (QDA) in Object Detection and Classification
- Training CNN
- User Interested Test Rate (UITR) Analysis
- Medical Image Analysis
- Directly connected neurons/connection between neurons
- Core Probability Theory
- encoding process in a diffusion model
- Neuromorphic Computing
- Normalization using Softmax Function

- Reward Prediction Error and Dopamine Neurons
- Computer Vision and Pattern Recognition
- References
- Graph Nets and Learning Symbolic Physics
- Hyperparameter tuning in deep learning
- Neural Network Based Link Prediction on Social Networks using Centrality Based Features
- Graph Embedding Adversarial Attacks
- backdoor attack framework Dyn-Backdoor for Deep Learning-based Link Prediction
- Robust Adversarial Deep Reinforcement Learning
- Differentially Private GAN for Synthetic Data Generation
- Interpreting bias in deep learning models using biased neurons
- Error rates
- Preprocessing and Building Taxonomy of Source Code Comments for Machine Learning Experiments
- padding approach
- Nonlinear Dimension Reduction (NDR) algorithm
- Model Pruning
- Bio-inspired meta-heuristic algorithms for object-oriented testing
- Mathematical Engineering of Deep Learning
- Cross-stitch networks
- Data normalization
- Anomalies
- Teacher Evaluation
- Video Analytics, Deep Learning on Videos
- Sign Language Recognition
- Membership Inference Attacks on Deep Learning Models
- Understanding classifications in video data using deep neural networks and explaining predictions with deep Taylor/Layer-wise Relevance Propagation (LRP)
- Transfer Learning
- List Extend Method
- OneHotEncoder
- Theano library
- Variable Interaction Analysis for Credit Risk Assessment
- Cloud MLE

- Deep Learning, NURBS, B-spline basis functions
- Reproducing Kernel Hilbert Space, Representer Theorem
- connectionism and distributed representations in neural networks
- Linear Models
- Speech Emotion Recognition (SER)
- Recurrent Neural Networks (RNNs)
- IoHT for noninvasive oral submucous fibrosis detection
- Microcontroller (Arduino UNO)
- CPS Security
- Deep Learning for Question Answering System in Bengali Language
- Bidirectional Long Short-Term Memory (biLSTM)
- Women Safety Applications
- Application - Industry-Specific - Use of Digital Twins in Car Manufacturing Industry
- Human Activity Recognition
- Deep Learning Model Comparison: MLP, RNN
- Synthetic Data for Crowd Analysis
- Applying attention to tabular data
- Key-Value Memory Networks
- Autonomous driving and multimodal localization for embedded systems
- Entropy Regularization
- Mathematical Engineering of Deep Learning
- Feature Encoding
- Using Convolutional Neural Networks to Forecast Sport
- Perceptron for Two-Cluster Classification
- Knowledge Preservation in Neural Network Distillation
- ConvAC
- Variance and Effective Sample Size in Sampling Algorithms
- Gibbs Sampling
- Maximum Likelihood Estimation (MLE) of Multivariate Conditional Factor Analysis (MCFAs)
- Computer Vision
- Hopfield Model and Spinglass
- Music Information Retrieval
- Deep Reinforcement Learning (DRL)

- Mathematical Modeling
- Bangla handwritten character recognition
- Vehicle Speed Detection System, Curve Detection
- Deep learning for hyperspectral image analysis
- Phishing Detection
- Human vs Machine Face Recognition Performance
- Performance Evaluation of Proposed Method in UML Diagram Creation using Natural Language Processing
- machine learning - time series prediction
- Phase-contrast microscopy data
- Geochemical modeling using Phreeqc for calculating ion activity and ion strength in aqueous solutions
- Policy Gradient Methods
- Multi-Task Learning, Prediction of Speech Quality, Dimensions, MOS
- Window Size Optimization in Running Window Averaging for Biometric Authentication
- Remote Sensing Example for Deep Learning
- Hankel matrices and their use in convolutions
- Smart Agriculture using Blockchain technology
- Malaria Parasite Enumeration and Classification using Convolutional Neural Networking
- Neural Networks for Learning Emergent Metric in Superstring Theory
- Word2vec and machine translation
- AI-powered chatbots
- Facebook App Development
- Deep Q-Learning
- Grasp detection with RGB-D
- Robotic Simulator
- Robot Arm Reset
- Permaculture and Agroecology
- Loading and Preprocessing Data for LSTM Model
- Deep Learning
- Multi-Agent Reinforcement Learning (MARL)
- Deep Learning Based Drones Detection and Formation Control
- Animal Breed Identification
- Image Registration and Stereo Visual Servoing
- A* algorithm

- Multi-thread Parallel Training and Frame Prediction in A3C Algorithm
- Recurrent Neural Networks (RNNs)
- Vehicle detection in nighttime using deep learning or AI techniques
- 3D Object Recognition
- Deep Learning
- Generative Adversarial Networks (GANs)
- Object Detection using Convolutional Neural Networks (CNNs)
- Deep Learning-Based Descriptors for Object Instance Search
- Thermal-to-visible synthesis and matching methods performance comparison
- Hopfield Networks
- Artificial Intelligence and Life
- Deep Learning with Relational Logic Representations
- lossless model compression via lifting
- Saturations in ILP algorithms
- Real-time human pose estimation
- YOLOv8, Object Detection, Billiard Game
- end-to-end deep network for single image blind using autoencoder and GAN
- Newborn Face Recognition
- Deep Learning
- Convolutional Neural Network
- Military and Civilian Use of Unmanned Systems
- Navigation Mesh
- Polar to Cartesian conversion
- Comparison of DBN and MLP for soft sensor design
- Object Detection with Limited Labeled Data
- References
- Monocular Depth Estimation
- Hierarchical Cascaded Processing
- Linear Decoding
- Autoencoder Feature Selection
- Reverse-engineering of Neural Architectures
- LSTM series
- Rainfall-runoff forecasting using LSTMs in hydrology

- Explainable deep learning in medical diagnosis
- security
- Machine learning algorithms for online television video
- Assistive Wearables for Vision, Hearing and Speech Impairments
- Image-to-Image Translation using Generative Adversarial Networks (GAN)
- Softmax Function
- Markov Logic Network
- Hidden Markov Models
- Malware Categorization
- Backpropagation algorithm
- Linear Algebra
- one-hot disentangled
- Synthetic Data
- Feature Detection and Extraction for Human Activity Recognition
- Conditional Generative Adversarial Networks (cGANs)
- weighting
- log
- Forecasting gradient
- convolutional neural networks
- created tensor
- Many autoencoders or
- Reinforcement Learning
 - Unsupervised deep learning and analysis of harmonic variation patterns
 - Sliding Mode Control
 - Deep Learning
 - Frequency control of power grids under cyber attacks
- Quantum Gravity
 - Multi-agent Systems
 - Transformers
- Convolutional Neural Network (CNN)
 - Software defect prediction
 - minimizing cross entropy
 - Vulnerability Prediction based on Software Metrics

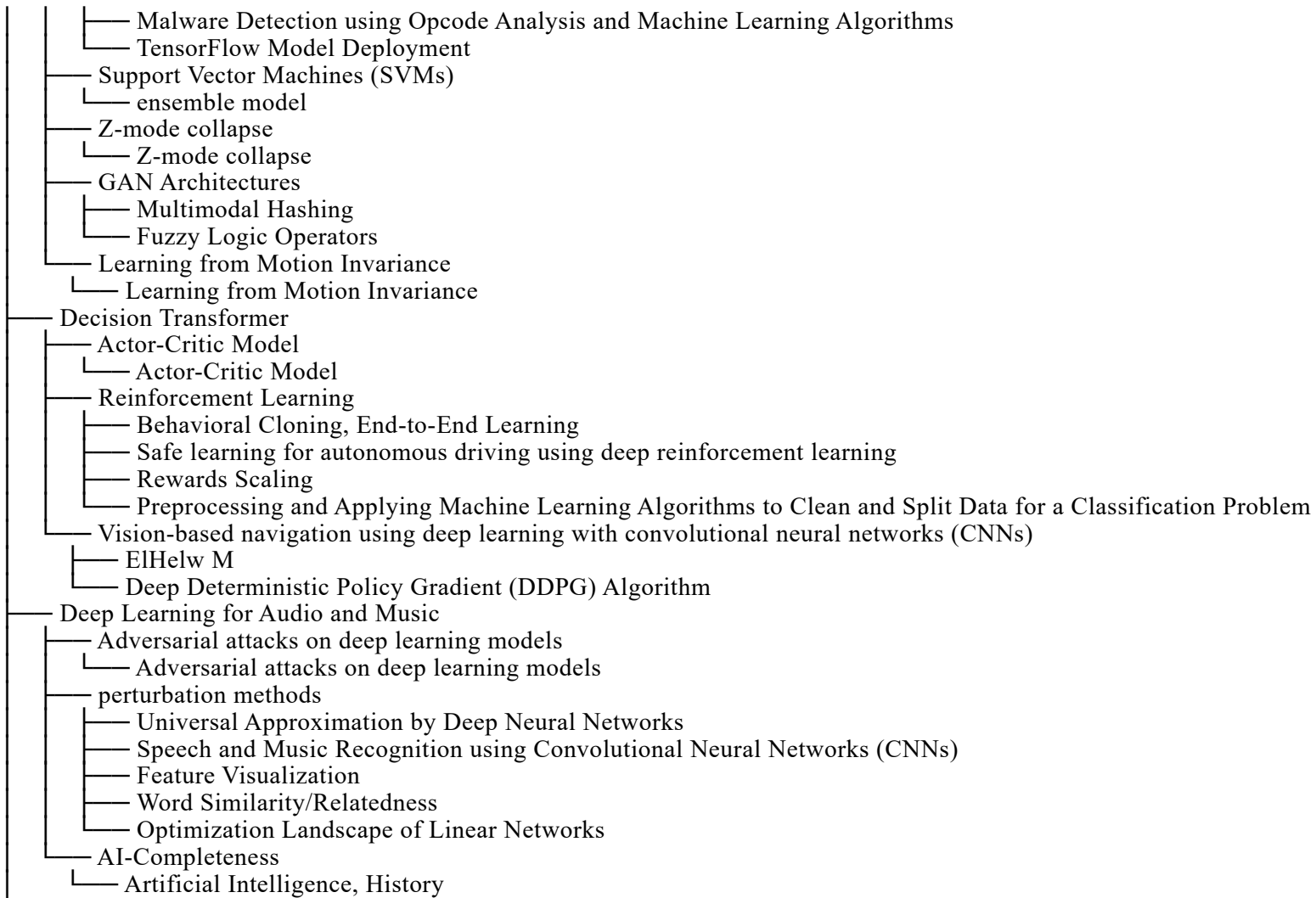
- Anomaly Detection
 - Anomaly Detection
- deep learning
 - Blockchain in Construction Industry
 - AI prediction algorithms using blockchain data
 - Blockchain and Artificial Intelligence
 - Deep Learning in Construction Site Management
 - 5D BIM and Cost Management in AEC
- Image Preprocessing Techniques
 - Image Preprocessing Techniques
- Energy-efficient deep learning
 - Mapping in DNN accelerators
 - Model compression
 - Sparse Architectures
 - Acknowledgments
 - Hardware-Algorithm Co-optimizations for Quantized Neural Networks (QNNs)
 - Smart Camera
 - Binary Neural Networks on Chip
- Speech Quality Analysis using ITU-T P Suppl. 23 Datasets and CNN-SA-AP model
 - Speech Quality Analysis using ITU-T P Suppl. 23 Datasets and CNN-SA-AP model
- Neuro-Memristive Systems
 - Neuro-memristive Systems Engineers
 - ReRAM (Resistive Random Access Memory)
- Transformer architecture
 - Deep Learning for Speech Recognition and Non-Speech Audio Processing
 - Algorithmic Composition in Computer Music
 - Progressive GAN Training
 - Feature Extraction
- TinyML Ecosystem
 - Deep Learning on Microcontrollers
- Large-Scale Agricultural Pest and Disease Datasets
 - pest used in agriculture
 - Wheat Disease Detection with Arbitrary Leaf Orientations

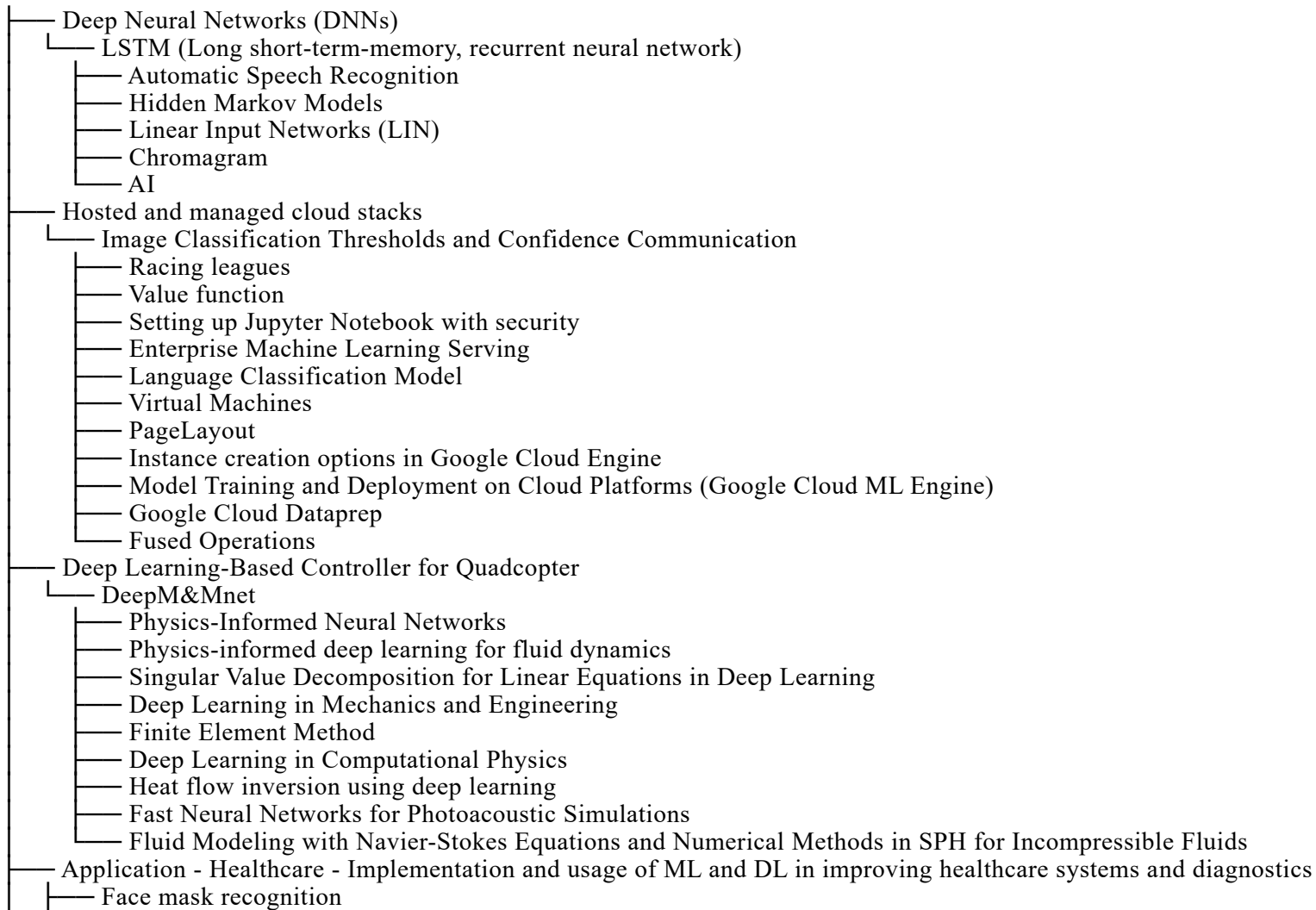
- CNNs for IED detection
 - Deep Learning for EEG Analysis
 - LSTM Model Training and Accuracy
 - Deep Reinforcement Learning
- Feature Aggregation for Long-Range Temporal Modeling
 - Dynamic Network Precision Selection
 - Weakly Supervised Action Localization
 - Video Captioning
- Neural Predictors for Solar Radiation
 - Chaotic Systems in AI and Deep Learning
 - Memristor based synapses
- Graph Neural Networks (GNNs)
 - Deep-
- Deep face recognition with partial data
 - Face Recognition Datasets
- Deep Reinforcement Learning Processor Design for Mobile Applications
 - Deep Reinforcement Learning Processor Design for Mobile Applications
- Unsupervised domain adaptation
 - Unsupervised domain adaptation
- Data Statistics
 - Deep Clustering
 - Computer Vision - table of contents
 - Probability Calculations
 - Hydraulic conductivity in fluid mechanics
 - Deep Clustering Techniques
 - Missing Data Model
- Hyperparameter Optimization
 - Human-in-the-loop Reinforcement Learning
 - Neural Architecture Search
 - Weight Clustering
 - Grid Search
 - Feature extraction
 - Automatic Machine Learning

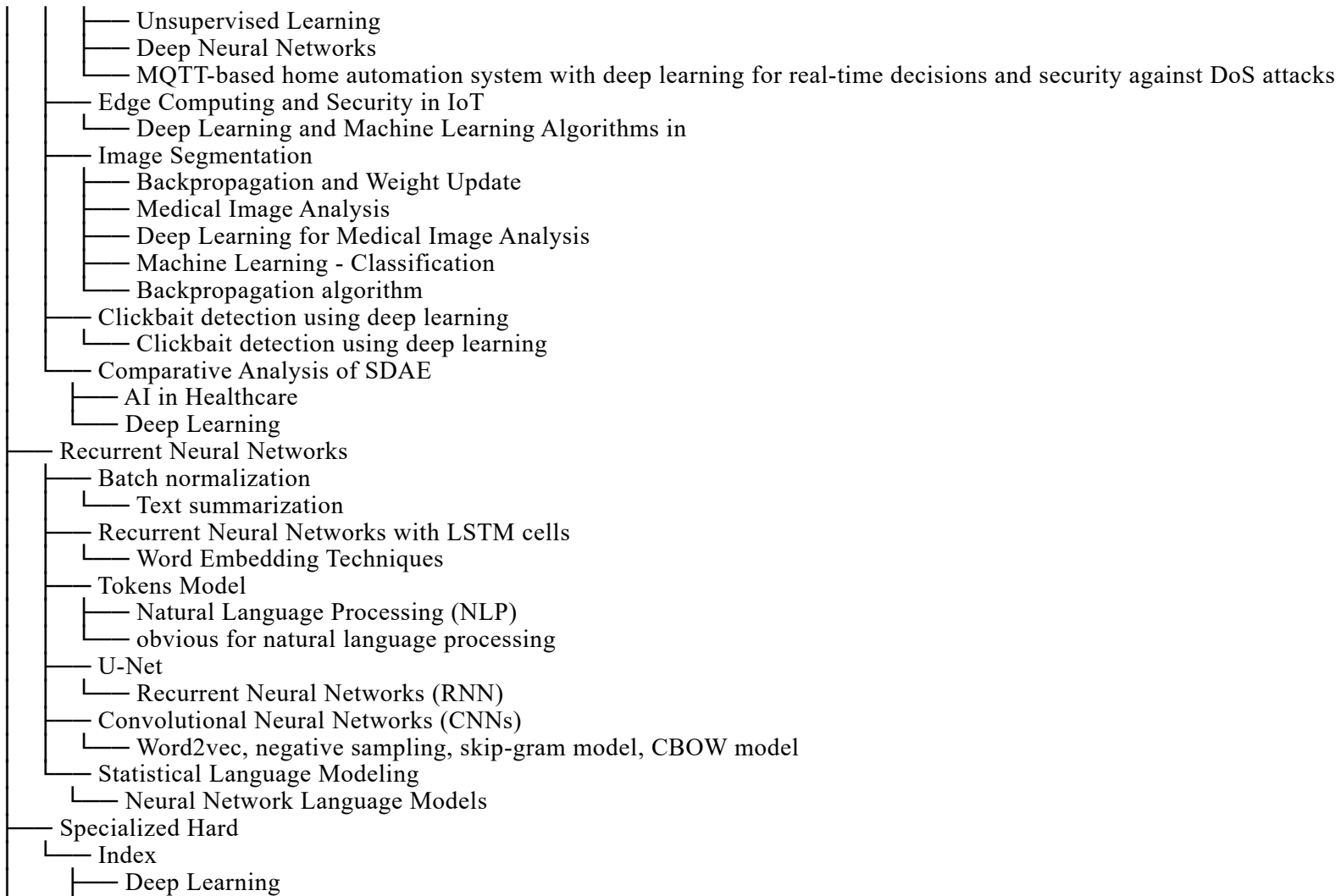
- Convolutional Neural Network
- Tensor Concatenation
- Deep Brain Stimulation (DBS)
 - Deep Brain Stimulation (DBS)
- Malware Detection
 - Remote consultations and data privacy
 - Adversarial Samples and Decision Boundaries
 - Convolutional Neural Networks (CNN)
 - Person re-identification
 - Feature selection in deep reinforcement learning for malware detection
 - Adversarial attacks on machine learning and deep learning-based malware detectors
- Almost Periodic Functions
 - Activation Functions for Neural Networks
 - Quantum Computing
- Secure Multi-Party Computation
 - References
 - Fidelity of IP protection in Federated Learning
 - Query-Efficient Adversarial Attack (QEAA) for Vertical
 - Benzene Exposure and Childhood Leuk
- an underactuated system
 - Convolutional Neural Network
 - Lyapunov Function
 - Squared least regression-based polynomials as path models in robotics
 - Robustness to periodic input disturbance in robotics control
- LBP descriptors
 - DeepFakes
 - Facial Recognition
- Occlusion augmentation
 - Database Size Estimation for Web Applications
 - Landmark detection for viewpoint estimation
 - 3D morphable model
- Intelligent Depth Perception in Human-Computer Interaction
 - 3D scene simulation in sand table therapy

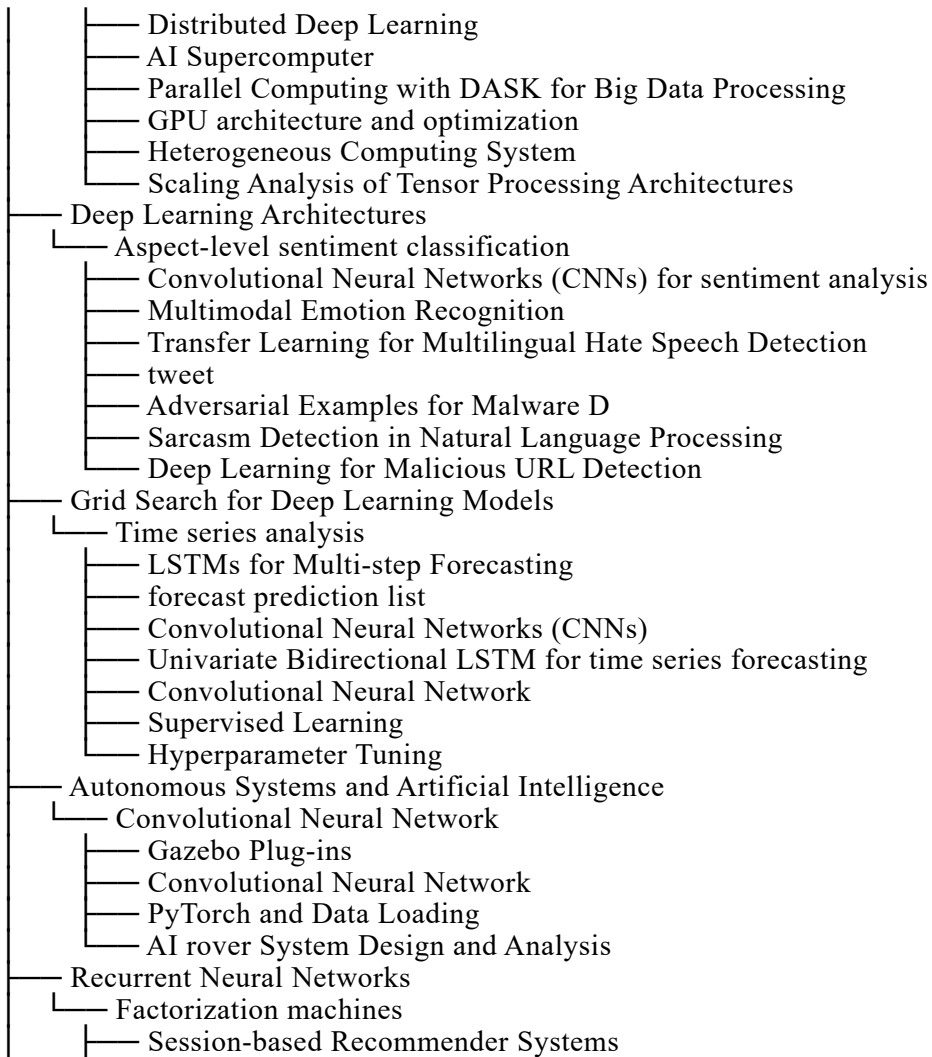
- └─ Computational Thinking Education
- └─ Pooling Layer
 - └─ Author Index
 - └─ No topic relevant to AI or deep learning found in the given text.
 - └─ Deep Learning Algorithms
- └─ Genetic Algorithms
 - └─ Genetic Algorithm
 - └─ Genetic Algorithm
 - └─ Backpropagation
 - └─ Genetic Algorithms (GA)
- └─ Scene Understanding
 - └─ Scene Understanding
- └─ 3D structure prediction for small molecules
 - └─ Training of Convolutional Neural Networks
- └─ Spectrum Poisoning Attack
 - └─ Source Detection/Anomaly Detection
 - └─ Autoencoder for communication systems
 - └─ Redundant Manipulator Control with Recurrent Neural Network for Periodic Input Disturbance
- └─ dbscan
 - └─ Partitioning Davies-Bouldin Index (PDBI)
 - └─ Data Preprocessing
- └─ Adam optimization algorithm
 - └─ Adam optimization algorithm
- └─ AI in the browser with TensorFlow.js and ml5.js
 - └─ AI in the browser with TensorFlow.js and ml5.js
- └─ Machine Learning Algorithms
 - └─ Autoregressive (AR) models and ARIMA
 - └─ Meanpressure predictions vs actuals plot
- └─ Biclustering
 - └─ Detecting Fake Reviewer Groups in Social Media
- └─ Few-shot learning in decoder models
 - └─ Deep Learning Models
- └─ Gaussian Mixture Models with Ridge Regression for Zero-Shot Learning

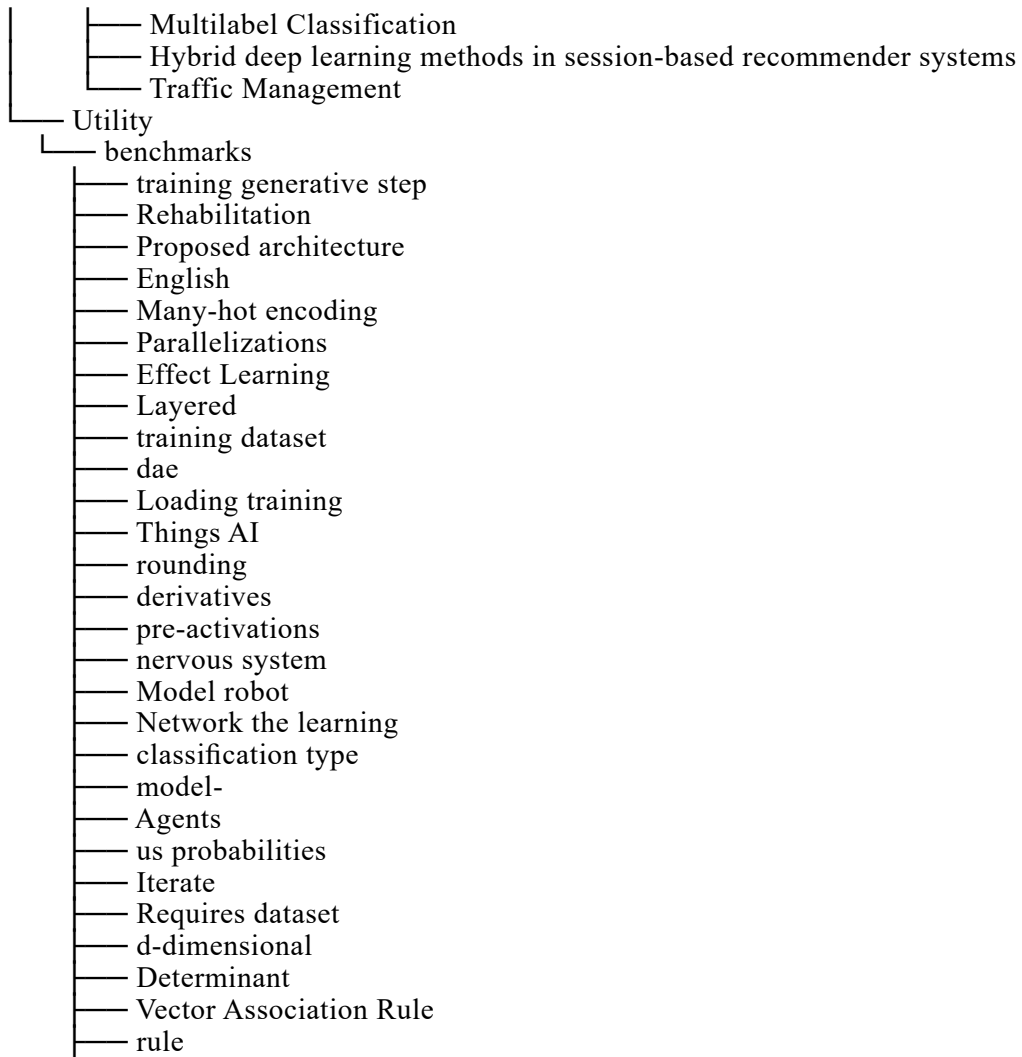
- Sim
- Energy Management System for Multi-microgrids
 - Convolutional Neural Network
 - Time-varying decision-making variables and energy consumption optimization using Hybrid TSKFS&MADRL algorithm
 - Reinforce
 - Physical layer security for beyond 5G communications
 - Control architectures
- If neuro-fuzzy
 - If neuro-fuzzy
- Fuzzy Logic
 - Crisp Logic
- Jacobian Matrix Computation
 - IMU (Inertial Measurement Unit) Data Processing for Robotics and Computer Vision
 - Location Based Personal Task Reminder System
 - Levenberg-Marquandt Algorithm
- Variable order structures in vector optimization
 - Gradient Descent
 - ANN approximations results for one-dimensional functions
 - Momentum GD process
 - Softplus function properties
 - GELU
- Deep Learning
 - Deep Learning
- Automatic Spelling Correction, Machine Learning
 - Preparing Data for Chatbot Input using RBMs and PCA
- LLGC algorithm
 - Feature Selection for Malware Detection in Android Apps
 - Malware detection
 - Cluster Analysis, Malware Families
 - Adversarial attacks on machine learning models
 - Disassembled code analysis
 - Malware Evolution Detection
 - Multi-class Classification of Binaries Based on Static Features











- reinforcement of training
- Learning the embedding
- Deepfake