



HAL
open science

Quantifying Interdisciplinarity in Scientific Articles Using Deep Learning Toward a TRIZ-Based Framework for Cross-Disciplinary Innovation

Nicolas Douard, Ahmed Samet, George Giakos, Denis Cavallucci

► **To cite this version:**

Nicolas Douard, Ahmed Samet, George Giakos, Denis Cavallucci. Quantifying Interdisciplinarity in Scientific Articles Using Deep Learning Toward a TRIZ-Based Framework for Cross-Disciplinary Innovation. *Machine Learning and Knowledge Extraction*, 2025, 7 (1), pp.7. 10.3390/make7010007 . hal-04920163

HAL Id: hal-04920163

<https://hal.science/hal-04920163v1>

Submitted on 30 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Article

Quantifying Interdisciplinarity in Scientific Articles Using Deep Learning Toward a TRIZ-Based Framework for Cross-Disciplinary Innovation

Nicolas Douard ^{1,2,*} , Ahmed Samet ¹ , George Giakos ² and Denis Cavallucci ¹

¹ National Institute of Applied Sciences (INSA), University of Strasbourg, 24 Boulevard de la Victoire, 67000 Strasbourg, France; ahmed.samet@insa-strasbourg.fr (A.S.); denis.cavallucci@insa-strasbourg.fr (D.C.)

² Department of Electrical and Computer Engineering, Manhattan University, 3825 Corlear Ave, Riverdale, NY 10463, USA; george.giakos@manhattan.edu

* Correspondence: nicolas.douard@insa-strasbourg.fr

Abstract: Interdisciplinary research (IDR) is essential for addressing complex global challenges that surpass the capabilities of any single discipline. However, measuring interdisciplinarity remains challenging due to conceptual ambiguities and inconsistent methodologies. To overcome these challenges, we propose a deep learning approach that quantifies interdisciplinarity in scientific articles through semantic analysis of titles and abstracts. Utilizing the Semantic Scholar Open Research Corpus (S2ORC), we leveraged metadata field tags to categorize papers as either interdisciplinary or monodisciplinary, establishing the foundation for supervised learning in our model. Specifically, we preprocessed the textual data and employed a Text Convolutional Neural Network (Text CNN) architecture to identify semantic patterns indicative of interdisciplinarity. Our model achieved an F1 score of 0.82, surpassing baseline machine learning models. By directly analyzing semantic content and incorporating metadata for training, our method addresses the limitations of previous approaches that rely solely on bibliometric features such as citations and co-authorship. Furthermore, our large-scale analysis of 136 million abstracts revealed that approximately 25% of the literature within the specified disciplines is interdisciplinary. Additionally, we outline how our quantification method can be integrated into a TRIZ-based (Theory of Inventive Problem Solving) methodological framework for cross-disciplinary innovation, providing a foundation for systematic knowledge transfer and inventive problem solving across domains. Overall, this approach not only offers a scalable measurement of interdisciplinarity but also contributes to a framework for facilitating innovation through structured cross-domain knowledge integration.



Received: 13 November 2024

Revised: 7 January 2025

Accepted: 8 January 2025

Published: 12 January 2025

Citation: Douard, N.; Samet, A.; Giakos, G.; Cavallucci, D. Quantifying Interdisciplinarity in Scientific Articles Using Deep Learning Toward a TRIZ-Based Framework for Cross-Disciplinary Innovation. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 7. <https://doi.org/10.3390/make7010007>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: interdisciplinary quantification; deep learning; TRIZ; convolutional neural networks; systematic innovation; semantic analysis

1. Introduction

Interdisciplinary research (IDR) has become increasingly vital in addressing complex global challenges that transcend traditional disciplinary boundaries [1,2]. Defined as the integration of concepts, theories, and methods from multiple disciplines to advance fundamental understanding or solve problems beyond the scope of a single discipline [3], IDR is essential for fostering innovation and tackling multifaceted issues such as climate change, healthcare, and sustainable development.

The importance of IDR is underscored by its potential to generate novel insights and solutions that single-discipline approaches may not achieve. For example, the intersection

of biology and computer science has given rise to bioinformatics, revolutionizing genomic research and personalized medicine [4]. Similarly, the convergence of engineering and environmental science has led to sustainable engineering practices that address ecological concerns while promoting technological advancement [5].

Quantifying the extent and evolution of IDR remains challenging due to conceptual ambiguities and methodological inconsistencies in existing measurement approaches [6–9]. Traditional methods predominantly rely on bibliometric indicators such as citation analysis, co-authorship networks, co-citation patterns, and journal classifications [10–12]. For instance, co-citation analysis examines the frequency with which two documents are cited together, providing insights into the relatedness of different research areas [13]. However, these approaches may not capture the full extent of interdisciplinary integration present in the content of the research itself.

Diversity indices, such as Shannon Entropy [14], Simpson's Index [15], and the Herfindahl–Hirschman Index [16], have been used to quantify interdisciplinarity by measuring the diversity of disciplines cited within a publication. While these metrics offer quantitative measures, they rely heavily on the classification of references into disciplines, which can be subjective and inconsistent across different classification systems. Moreover, they often fail to account for the semantic content and the actual integration of knowledge across disciplines.

Recent advancements in natural language processing (NLP) and machine learning offer new opportunities to analyze text content and capture semantic relationships more effectively [17,18]. Topic modeling techniques, such as Latent Dirichlet Allocation (LDA) [19], have been employed to uncover thematic structures within text corpora; however, these models assume topic independence. Furthermore, topic models often require manual interpretation of the topics, which can introduce subjectivity.

In contrast, deep learning approaches, such as Convolutional Neural Networks (CNNs) and Transformers [20,21], have shown promise in capturing complex semantic patterns in text data. These methods enable the analysis of text content, potentially overcoming the limitations of traditional bibliometric and topic modeling approaches. By leveraging hierarchical feature extraction and contextual embeddings, deep learning models can detect subtle semantic cues indicative of interdisciplinary integration.

We propose a novel text-based deep learning approach to quantify interdisciplinarity in scientific articles and provide insights into how our method can aid in fostering interdisciplinary collaboration and addressing complex engineering problems in the context of the Theory of Inventive Problem Solving (TRIZ) [22]. TRIZ principles emphasize systematic innovation by solving contradictions and transferring knowledge across domains, specifically focusing on engineering applications. In that context, our aim is to facilitate the identification of inventive solutions arising from interdisciplinary research.

To elucidate the various pathways through which interdisciplinary research can foster innovative solutions within the TRIZ framework, we introduce the TRIZ-inspired Frames of Knowledge in Figure 1. This conceptual framework categorizes solutions into four distinct cases:

- Case C1: The solution is within the same industry.
- Case C2 : The solution is in another industry.
- Case C3 : The solution is outside of what exists in all industries.
- Case C4 : The solution does not yet exist.

Each case represents a different level of interdisciplinary integration and innovation scope, providing a structured approach to identifying and leveraging cross-disciplinary solutions within engineering applications. This study focuses on case C2. By targeting interdisciplinary research that bridges distinct industries (Case C2), our method aims to

facilitate the application of TRIZ principles to transfer and adapt innovative solutions from one domain to another, thereby enhancing the systematic problem-solving capabilities advocated by the TRIZ.

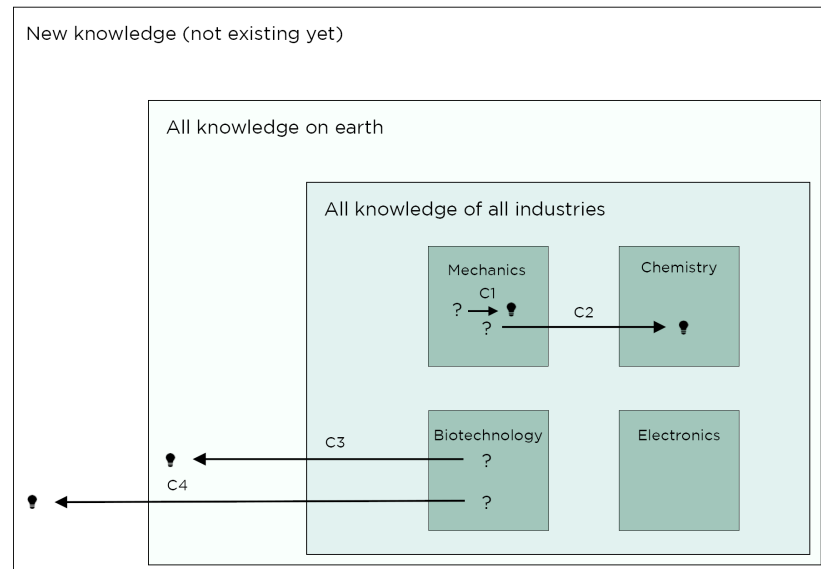


Figure 1. TRIZ-inspired Frames of Knowledge: A conceptual framework illustrating the integration of interdisciplinary research within the TRIZ methodology for systematic innovation. Arrows depict the path taken toward a solution in each distinct case.

Our main contributions are as follows:

1. We introduce a deep learning-based method for quantifying interdisciplinarity by analyzing the titles and abstracts of scientific articles.
2. We demonstrate the effectiveness of our approach by comparing it with other machine learning estimators and traditional metrics.
3. We assess the scale of interdisciplinary research by running predictions on a very large-scale dataset.
4. We quantify the prevalence of interdisciplinarity in the scientific literature by analyzing 136 million abstracts, finding that approximately 25% of the analyzed corpus across specified disciplines is interdisciplinary.
5. We propose the integration of the newly developed interdisciplinarity classifier into a high-level framework toward systematic, cross-disciplinary innovation using the TRIZ, which was theoretically introduced by Douard et al. [23,24].

2. Related Work

Interdisciplinarity measurement has evolved through various methodologies, each with its advantages and limitations.

Bibliometric approaches have been widely used to quantify journal interdisciplinarity, including citation analysis, co-authorship networks, co-citation patterns, and journal classifications. These methods provide valuable quantitative indicators but may not fully capture the interdisciplinary nature of the research content itself.

Diversity indices offer quantitative measures of interdisciplinarity by assessing the distribution of disciplines within the references cited by a publication.

Shannon Entropy is defined as

$$H = - \sum_{i=1}^N p_i \ln(p_i) \quad (1)$$

where p_i is the proportion of references in discipline i , and N is the total number of disciplines. Higher entropy values indicate greater diversity and, by extension, higher interdisciplinarity [14].

Simpson's Index is calculated as

$$D = 1 - \sum_{i=1}^N p_i^2 \quad (2)$$

This index measures the probability that two randomly selected references belong to different disciplines. A higher value of D suggests greater interdisciplinarity [15].

The Rao–Stirling Diversity Index [25] incorporates both the disciplinary diversity and the distances between disciplines:

$$D = \sum_{i=1}^N \sum_{j=1}^N p_i p_j d_{ij} \quad (3)$$

where d_{ij} represents the cognitive distance between disciplines i and j . This index accounts for the relatedness of disciplines, providing a more nuanced measure of interdisciplinarity.

While these indices offer valuable insights, they rely on accurate discipline classification of references and may overlook the depth of integration between disciplines within the content of the publication.

Topic modeling, such as Latent Dirichlet Allocation (LDA) [19], has been employed to identify thematic structures within text corpora. By detecting topics that span multiple disciplines, researchers can infer interdisciplinary content [26]. However, topic models assume that topics are independent and may not capture complex interactions between disciplines. Interpretation of the topics can also be challenging, and the assignment of disciplines to topics may introduce subjectivity.

Nanni et al. [27] investigated text mining methods for the automated identification of interdisciplinary doctoral dissertations by analyzing abstracts. They framed interdisciplinarity detection as a two-step classification process: predicting the main discipline using supervised classifiers and detecting interdisciplinarity by exploiting prediction confidences. Their findings revealed that directly using textual features yielded better performance than relying on main discipline classification results.

Similarly, Pham et al. [28] proposed a metadata-based approach for research discipline prediction using machine learning techniques and distance metrics. They focused on predicting research disciplines associated with projects and measuring interdisciplinarity based on associated metadata. Their framework included feature extraction using topic models, discipline encoding to reduce output dimensionality, and a distance matrix to recommend appropriate disciplines and compute interdisciplinarity.

Advancements in deep learning have opened new avenues for analyzing text content. Models like CNNs and transformers (e.g., BERT) have demonstrated high performance in various NLP tasks [20,21]. These models can capture hierarchical and contextual semantic relationships, making them suitable for detecting interdisciplinarity within texts.

For example, Beltagy et al. [29] introduced SciBERT, a pretrained language model based on BERT, trained on scientific text. SciBERT has shown improved performance on various scientific NLP tasks, highlighting the importance of domain-specific models.

However, deploying large transformer-based models on massive datasets like S2ORC can pose computational challenges. It is our intent to introduce a computationally effective solution enabling filtering literature at scale. Therefore, more computationally efficient models like Text CNNs offer a practical balance between performance and resource require-

ments. The same comment can be made about employing Large Language Models (LLMs) toward this task.

Our work builds upon these methodologies by leveraging deep learning techniques, aiming to capture more comprehensive semantic information indicative of interdisciplinarity. We employ a Text CNN model trained on titles and abstracts from the extensive S2ORC dataset and compare it against multiple other machine learning estimators over the same training set.

To provide a comprehensive overview of the existing methodologies for measuring interdisciplinarity, we present a comparative analysis of the primary approaches in Table 1. This comparison highlights the distinct methodologies, advantages, and limitations associated with each approach, which are supported by relevant scholarly references.

Table 1. Comparison of approaches to measuring interdisciplinarity.

Approach	Methodology	Advantages	Limitations
Bibliometric Indicators	Use citations, co-authorships, co-citations, and journal classifications to infer interdisciplinarity [10–13].	Based on readily available metadata; established in scientometrics; facilitates longitudinal studies.	Dependent on metadata accuracy; may not capture actual content integration; citation practices vary across disciplines.
Diversity Indices (e.g., Shannon Entropy, Simpson’s Index)	Calculate statistical diversity measures from discipline distributions in references [14–16,25].	Quantitative and easily comparable; simple computation and interpretation; applicable across various datasets.	Reliant on accurate discipline classification; may overlook depth of content integration; sensitive to granularity of classification schemes.
Topic Modeling (e.g., LDA)	Uncover thematic structures and topics within text corpora using probabilistic models [19,26–28].	Directly analyzes content semantics; scalable to large datasets; identifies latent thematic connections.	Assumes topic independence, which may not hold true; may miss nuanced interdisciplinary integrations; interpretation of topics can be subjective and challenging.
Machine Learning Approaches	Utilize traditional supervised or unsupervised algorithms (e.g., Random Forest, SVM) often relying on engineered features [30–33].	Often less data-intensive; can be more interpretable; well-established methodologies.	May require careful feature engineering; can struggle with highly complex semantic patterns; performance may plateau without deeper feature extraction.
Deep Learning Approaches	Utilize algorithms like CNNs, RNNs, and Transformers to analyze text content [19,20,29].	Capable of capturing complex semantic patterns; can integrate various data sources (e.g., full texts, abstracts); often achieve higher predictive performance.	Require large labeled datasets for training; computationally intensive; may lack interpretability.

3. Methodology

We utilized the Semantic Scholar Open Research Corpus (S2ORC) [34], which is a large-scale dataset that contains more than 100 million abstracts of academic papers in various disciplines. S2ORC includes metadata, abstracts, and full-text excerpts for open articles, providing a rich resource for text-based analysis.

Given the scope of TRIZ and our aim to integrate the interdisciplinary model with the systematic, cross-disciplinary innovation framework proposed by Douard et al. [23,24], we focused on papers classified under Biology, Engineering, Physics, Computer Science, Chemistry, Mathematics, and Materials Science. These disciplines were selected due to their representation in the corpus and relevance in engineering applications.

Approximately one million papers were used for initial modeling, with an equal split between interdisciplinary and monodisciplinary papers. Interdisciplinary papers were labeled on the basis of their metadata tags from the Semantic Scholar classification. Papers tagged with more than one discipline were labeled as interdisciplinary, while those tagged with only one discipline were labeled as monodisciplinary. Among the one million papers, approximately 18% are associated with more than one field of study. This proportion declines rapidly: fewer than 0.5% span more than two fields, and only an insignificant 0.02% involve three or more fields.

Semantic Scholar assigns Fields of Study to papers using a machine learning model that analyzes the title and abstract, applying multi-label classification to capture interdisciplinary research. This approach relies on training data derived from venue-level field mappings (assuming each publication venue typically focuses on a relatively narrow range of fields) and human-labeled “gold” sets, combining scalability with reasonable accuracy. Though not perfectly precise, the model’s design and confidence thresholds ensure reliable, broad coverage of English-language scholarship. Our hypothesis is that, at scale, the deep learning model can generalize the concept of interdisciplinarity by learning underlying semantic patterns. The high-level process used to create the training dataset is shown in Figure 2.

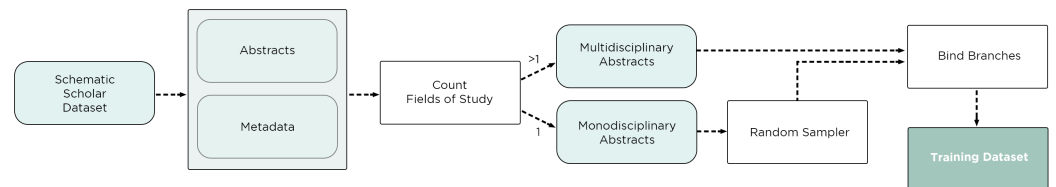


Figure 2. Overview of the process used to create the training dataset, highlighting how the number of fields of study is used to differentiate between multidisciplinary and monodisciplinary abstracts.

To prepare the textual data for input into the neural network, we employed several preprocessing steps, including normalization (converting text to lowercase and removing punctuation), tokenization, lemmatization, removal of stop words, and handling of missing values. We applied Byte Pair Encoding (BPE) [35] to the text data. BPE is a subword tokenization technique that combines the benefits of word-level and character-level representations. It allows the model to handle out-of-vocabulary words and capture meaningful subword patterns, improving the representation of rare and compound words.

Figure 3 presents the inference pipeline used to predict interdisciplinarity. Titles and abstracts undergo BPE for tokenization and then feed into a Text CNN, which outputs a probability score. While standard, this process ensures consistent text representation and is efficient for large-scale analysis, forming the backbone of our approach.

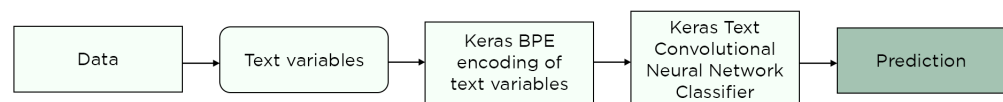


Figure 3. Inference pipeline from raw article text to an estimated interdisciplinarity score, highlighting encoding and classification steps using the Text CNN model.

The Text CNN model architecture is illustrated in Figure 4. The architecture consists of an embedding layer, convolutional layers with varying kernel sizes, a pooling layer, fully connected layers, and an output layer. The convolutional layers apply multiple filters to capture n-grams of different lengths, and the pooling layer uses max-over-time pooling to capture the most salient features. The fully connected layers include dropout regularization, and the output layer uses a sigmoid activation function for binary classification.

To effectively capture varying semantic structures within the text, we selected multiple kernel sizes corresponding to different n-gram lengths. Specifically, kernel sizes of 2 and 3 were chosen to detect bi-grams and tri-grams, respectively. This selection is based on the Text CNN Classifier's capability to process multiple parallel convolutional layers, each with distinct kernel sizes. By incorporating kernel sizes of 2 and 3, the model can identify both short-range dependencies, such as common phrase patterns, and slightly longer contextual relationships that are indicative of interdisciplinary content. This variation allows the Text CNN to better represent the diverse linguistic patterns that signify the integration of multiple disciplines within an abstract.

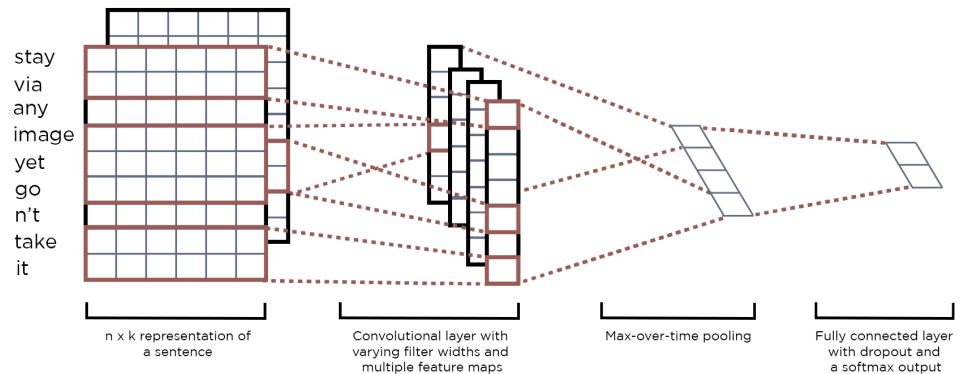


Figure 4. High-level architecture of the Text CNN model utilized for interdisciplinarity detection, as initially introduced in [21]. The model includes multiple convolutional layers with different kernel sizes, a max pooling layer, and a fully connected layers leading to the output.

To enhance the model's performance and generalization capabilities, we employed several training techniques, including the Adam optimizer [36] with a learning rate of 0.001, a binary cross-entropy loss function, a batch size of 64, and an early stopping based on validation loss. Hyperparameters were tuned using random search with 5-fold cross-validation on the validation set, exploring various learning rates, dropout rates, kernel sizes, number of filters, and batch sizes [37].

The model was implemented using the Keras library with a TensorFlow backend [38]. We utilized mixed-precision training to speed up computation and reduce memory usage. The model is highly efficient, as it can run on a CPU, which is a key factor in its scalability for processing large-scale datasets. This efficiency ensures that our approach remains computationally feasible even when applied to extensive corpora, such as the 136 million abstracts analyzed in this study. Additional technical details regarding the model architecture and hyperparameters are available in Appendix A, and technical validation details are provided in Appendix B.

We used multiple evaluation metrics to assess model performance, including accuracy, precision, recall, F1 score, Matthews Correlation Coefficient (MCC) [39], and ROC-AUC score.

Other estimators considered for benchmarking consist of Gradient Boosted Trees Classifier [30], SVM Classifier [31] (Nyström Kernel), Random Forest Classifier [32], and Extra Trees Classifier [33].

We propose that this interdisciplinary classifier and its inference outputs be used as a first step in the framework proposed by Douard et al. [23,24]. Subsequent steps could involve discipline-specific classifiers to narrow down to a specific pairing objective (e.g., biology and engineering-related disciplines in the context of biomimicry). From there, topic modeling could be employed to map the fields of study of interest. This would in turn enable the capacity to derive, for each article, a primary and a secondary topic.

The frequency at which topics are paired could inform the creation of an interdisciplinary graph, bridging concepts across disciplines.

Figure 5 illustrates our multi-phase pipeline for uncovering cross-domain connections at scale. We began with around 100 million abstracts from the Semantic Scholar dataset and used an interdisciplinary classifier to narrow the corpus to about 25 million articles that cross multiple fields. Next, discipline-specific classifiers (for engineering and biology) and topic modeling (e.g., BERTopic [40]) extracted coherent themes and assigned each abstract a set of topic probabilities. By embedding and interlinking these topics in a graph, we systematically uncovered synergies—particularly between engineering and biology. Finally, we applied TRIZ contradiction formalism to highlight how solutions in one domain address challenges in the other, revealing novel thematic overlaps and potentially facilitating systematic cross-disciplinary innovation.

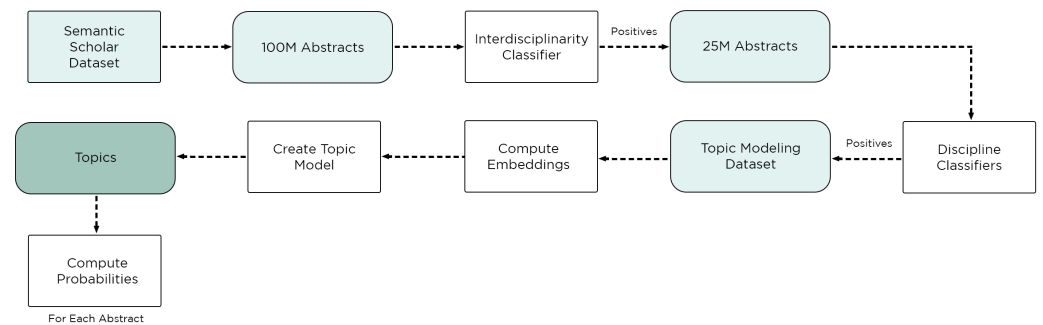


Figure 5. Framework for interdisciplinary topic pairing and graph construction using classifiers and topic modeling. Example over 100 M engineering and physics abstracts as input.

By integrating our interdisciplinarity quantification method into the TRIZ-based framework for cross-disciplinary innovation, we suggest a systematic approach to leveraging interdisciplinary knowledge for inventive problem solving. Specifically, the interdisciplinarity scores produced by our Text CNN model could serve as an initial filtering mechanism to identify scientific articles potentially rich in cross-disciplinary content. These selected articles might then be subjected to TRIZ analytical processes, such as contradiction identification and the application of inventive principles, to extract and adapt innovative solutions across domains. Technically, this integration involves using the classifier’s output to prioritize literature for TRIZ analysis, thereby potentially streamlining the knowledge transfer process and enhancing the effectiveness of systematic innovation within the TRIZ methodology.

4. Results

The performance comparison of the classification models is shown in Table 2. Our Text CNN model outperformed all baseline models in terms of LogLoss, F1 score, and max MCC. Comparison was done over 26,999 holdout samples that neither model has been exposed to. In addition, examining precision (0.80) and recall (0.83) provides insights into the model’s balanced performance, indicating that it can effectively identify interdisciplinary articles while maintaining a relatively low rate of false positives. This suggests that the model’s learning process captures semantic cues related to interdisciplinarity. Some misclassifications occurred when abstracts lack explicit cross-field terminology, pointing to opportunities for refining the training data. Furthermore, while this study uses a binary classification approach, the model could be extended to a multi-class or multi-label framework, enabling it to classify articles into multiple disciplines simultaneously and offering more granular insights into interdisciplinary research.

Additionally, the max MCC value of 0.64 for the Text CNN model indicates a strong overall correlation between predicted and actual classifications, outperforming the baseline models' MCC of 0.59. The precision–recall curve (Figure 6) demonstrates that the Text CNN maintains a high balance between precision and recall across various thresholds, highlighting its effectiveness in accurately identifying both interdisciplinary and monodisciplinary articles.

Table 2. Performance comparison of classification models using holdout data.

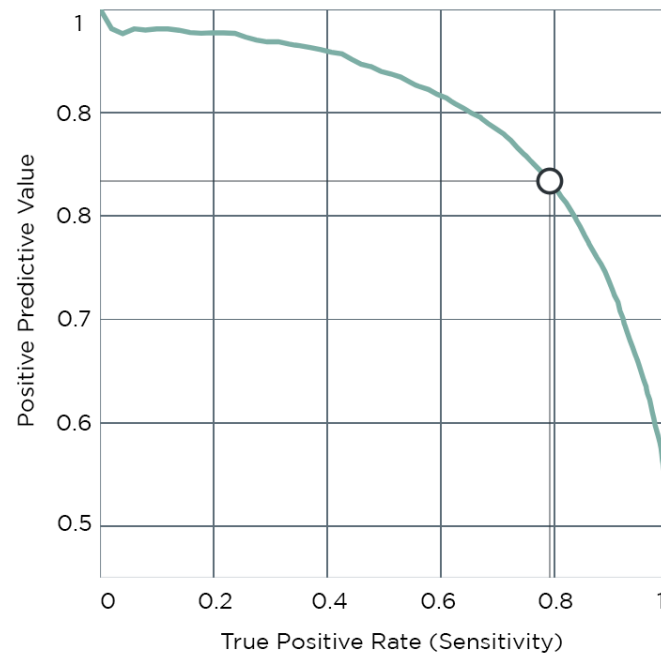
Model	LogLoss	Precision	Recall	F1 Score	Max MCC
Text CNN	0.41	0.80	0.83	0.82	0.64
Boosted Trees	0.45	0.74	0.86	0.80	0.59
Random Forest	0.46	0.76	0.82	0.79	0.57
SVM	0.46	0.75	0.83	0.79	0.57
Extra Trees	0.46	0.74	0.84	0.79	0.57

The confusion matrix for the Text CNN model over holdout data is presented in Table 3. The model correctly classified 21,908 out of 26,999 papers.

Table 3. Confusion matrix for text CNN model over holdout data.

	Predicted Negative	Predicted Positive
Actual Negative	10,655	2845
Actual Positive	2246	11,253

The precision–recall curve (Figure 6) and ROC curve (Figure 7) indicate a good balance between precision and recall and strong discriminative ability, respectively.



○ Classification threshold

Figure 6. Precision–recall Curve for the Text CNN model. The area under the curve (AUC) indicates the model's ability to balance precision and recall across different thresholds.

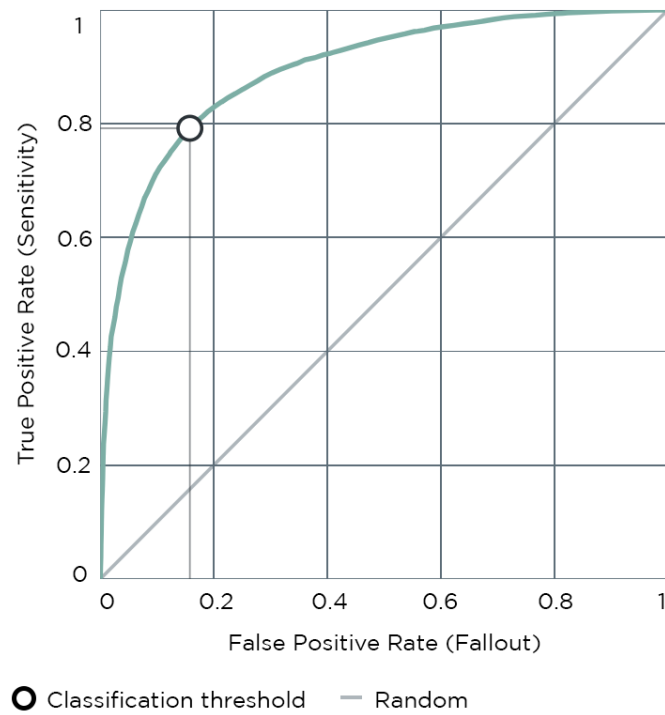


Figure 7. ROC Curve for the Text CNN model. The area under the ROC curve (AUC) indicates discriminative ability.

The learning curves showing training and validation loss and accuracy over epochs are presented in Figures 8 and 9. The convergence of the loss and consistency between training and validation accuracy suggest satisfactory generalization without overfitting.

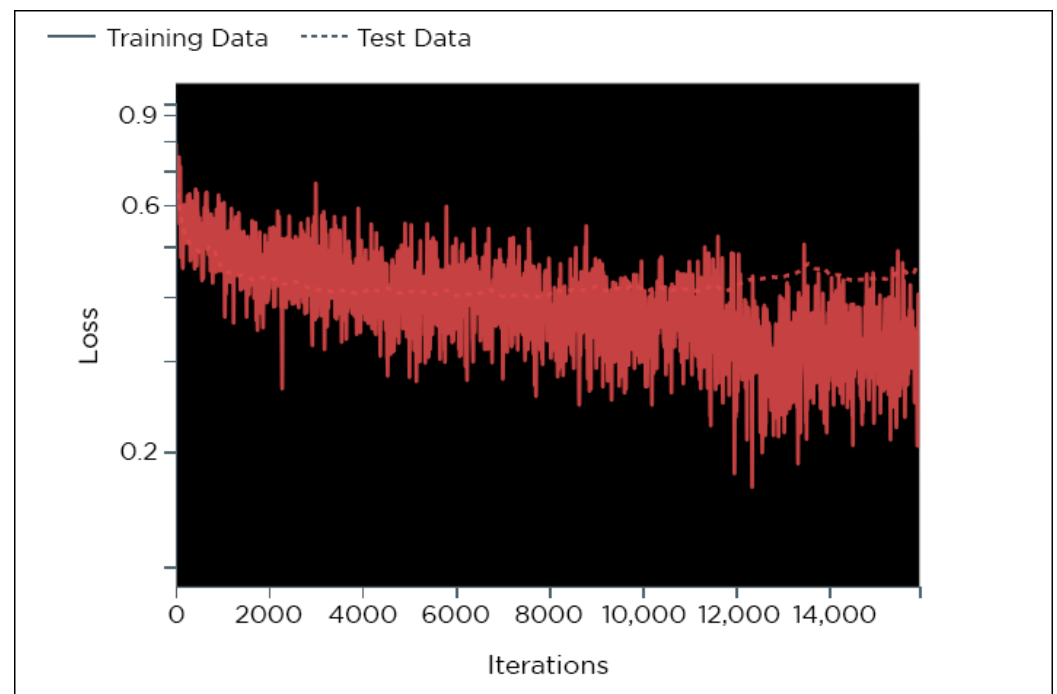


Figure 8. Learning curve showing training and validation loss over epochs for the Text CNN model.

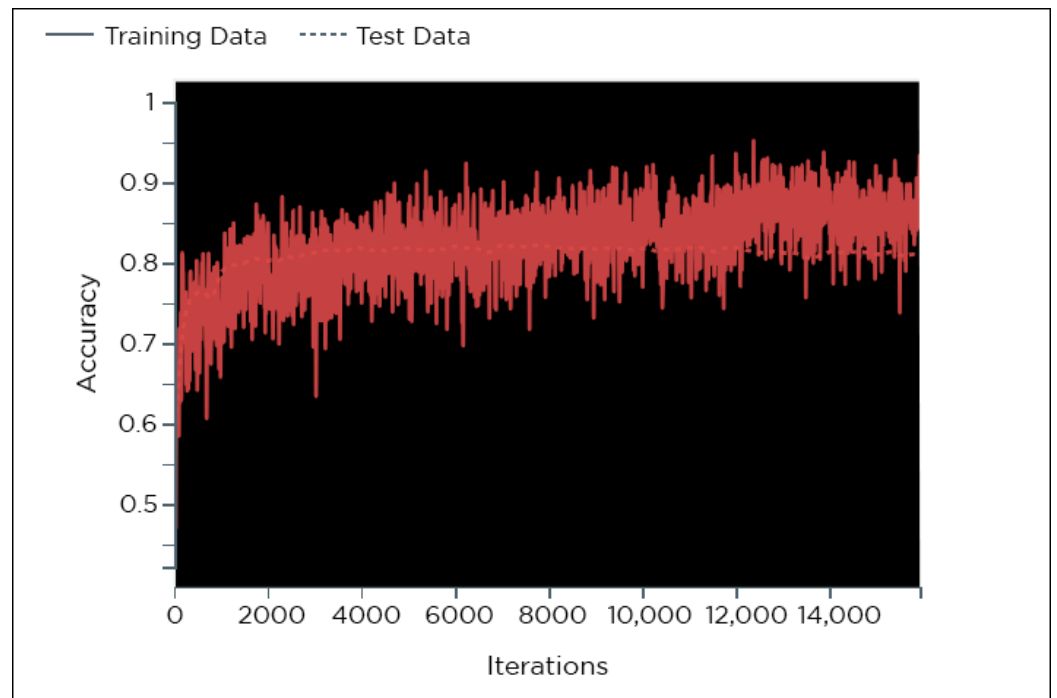


Figure 9. Learning curve showing training and validation accuracy over epochs for the Text CNN model.

To further understand the model's confidence in its predictions, Figure 10 presents the density distribution of predicted probabilities for both the interdisciplinary and monodisciplinary classes. This visualization reveals how the model assigns probabilities, providing insights into areas where the model is more or less confident in its classifications.

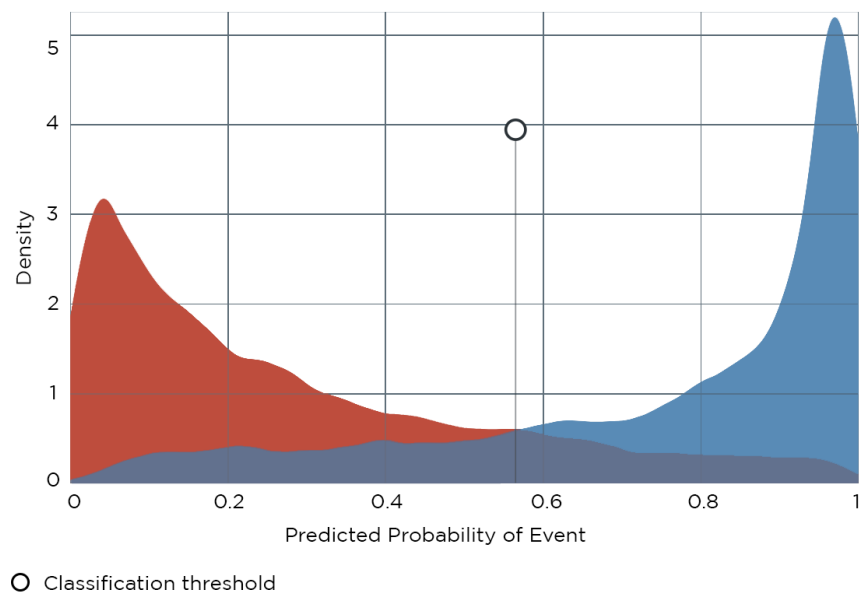


Figure 10. Density plot of predicted probabilities for interdisciplinary and monodisciplinary classes. The plot illustrates the distribution of predicted probabilities, highlighting the confidence levels of the Text CNN model's classifications.

To illustrate the ability of our Text CNN model in evaluating interdisciplinarity, we provide several examples of research abstracts classified at different levels. These abstracts are based on real-world studies and include citations; however, they have been paraphrased to comply with copyright policies. Please note that these examples are intended solely for

illustrative purposes and do not comprehensively reflect the model's overall performance or capabilities.

Example 1: High Interdisciplinarity

Paraphrased Abstract: Li et al. (2022) [41] examined organ-on-a-chip devices that integrate micro-manufacturing and tissue engineering to replicate the essential physiological environments and functions of human organs. These devices are utilized to predict drug responses and evaluate the effects of environmental factors on organs. Precision control of micro-scale reagents is achieved through micro-fluidic technology, leading to its widespread application in organ-on-chip systems for mimicking specific or multiple organs *in vivo*. Enhanced with various sensors, these models demonstrate significant potential in simulating the human environment. In this review, the typical structures and recent research advancements of several organ-on-a-chip platforms are introduced, and innovations in models for pharmacokinetics/pharmacodynamics, nanomedicine, continuous dynamic monitoring in disease modeling, and other applications are discussed.

The abstract was classified as highly interdisciplinary, with a probability score of 0.90. It integrates methodologies from micro-manufacturing and tissue engineering to develop organ-on-a-chip devices, employs micro-fluidic technology for precise control of reagents, and incorporates sensor technologies to simulate the human environment. Additionally, it encompasses applications in pharmacokinetics/pharmacodynamics, nanomedicine, and disease modeling. This convergence of engineering, biology, chemistry, and medical sciences demonstrates a high level of interdisciplinarity, enabling comprehensive simulation and the analysis of complex physiological systems.

Example 2: Moderate Interdisciplinarity

Paraphrased Abstract: Vincent et al. (2006) [42] explored the theory and practice of biomimetics—the transfer of ideas from biology to technology. They adapted TRIZ, a Russian problem-solving methodology, to enhance this transfer process. Their analysis revealed only a 12% similarity between biology and technology in problem-solving principles. While technology primarily manipulates energy, biology focuses on information and structure, factors often overlooked in technological applications.

The abstract was classified as moderately interdisciplinary, with a probability score of 0.72. It bridges biology, specifically biomimetics, with engineering problem-solving techniques like TRIZ. By integrating biological concepts into technological innovation, the study demonstrates a moderate level of interdisciplinarity. The model appropriately classified it as moderately interdisciplinary.

Example 3: Low Interdisciplinarity

Paraphrased Abstract: Patil et al. (2022) [43] conducted a critical review on optimizing cutting parameters during CNC milling of EN24 steel using tungsten carbide coated inserts. The study emphasizes that by optimizing the feed rate, speed, and depth of cut, one can enhance the material removal rate, surface roughness, and tool wear. The researchers employed optimization techniques like the Taguchi method and Response Surface Methodology (RSM), and they used Analysis of Variance (ANOVA) to analyze the impact of machining parameters on performance metrics.

The article was classified as having low interdisciplinarity, with a probability score of 0.18. This work is rooted within the field of mechanical engineering, specifically focusing on CNC milling optimization. It utilizes established methods and concepts within the same discipline without significant integration of ideas from other fields. The model correctly classified it as having low interdisciplinarity.

In order to estimate the proportion of interdisciplinary research in the scientific literature, we scaled up our analysis to process 136 million abstracts, specifically focusing on

disciplines within the scope of TRIZ. Abstracts were sourced using the Semantic Scholar API and subsequently filtered to a subset encompassing TRIZ-related disciplines, including biology, engineering, physics, computer science, chemistry, mathematics, and materials science. To classify papers as interdisciplinary, we utilized a thresholding approach based on the Matthews Correlation Coefficient (MCC) [39], which is a metric that considers the balance between true and false positives and negatives. By varying the classification threshold and selecting the value that maximizes the MCC, we aimed to achieve an effective balance between precision and recall in our classification process. This approach allowed us to identify papers that exhibit characteristics of interdisciplinarity with a reasonable degree of confidence. Consequently, our analysis suggests that approximately 25% of the literature across the specified disciplines can be characterized as interdisciplinary. This proportion represents the subset of papers that exceeded the MCC-derived threshold, indicating a notable presence of interdisciplinary integration within the extensive body of scientific research examined.

5. Discussion

Our approach demonstrates that deep learning models can effectively capture complex semantic patterns indicative of interdisciplinarity, exceeding the performance of traditional machine learning techniques used in previous studies [27,28].

In comparison to recent approaches, our deep learning method demonstrates improved performance. For instance, Nanni et al. [27] (2016) used text mining and machine learning classifiers on dissertation abstracts, reporting accuracy improvements from direct textual features but not achieving the F1 score levels found in our study. Similarly, Pham et al. [28] achieved improvements in discipline prediction using metadata and topic modeling. However, their approach achieved a correctly predicted discipline percentage (CPDP) of around 58% to 69% depending on the dataset. In contrast, our Text CNN model, trained on large-scale abstracts from S2ORC, attained an F1 score of 0.82, indicating a more robust capture of interdisciplinary cues directly from semantic content.

This shows that our method, which leverages modern deep learning architectures, not only surpassing traditional classifiers but also offering measurable gains over recent text and metadata-driven approaches. By quantifying interdisciplinarity and identifying areas of strong cross-disciplinary interaction, we provide a systematic method to uncover opportunities for inventive problem solving as advocated by TRIZ principles.

This interdisciplinary quantification directly supports TRIZ-based innovation by identifying cross-disciplinary research that offers inventive solutions. For instance, in developing advanced medical devices, our Text CNN model can detect studies that integrate biomedical engineering and materials science. TRIZ principles can then be applied to these identified interdisciplinary insights to resolve specific contradictions, such as enhancing device biocompatibility without compromising mechanical strength.

Similarly, in sustainable energy technologies, our model can pinpoint research combining electrical engineering, environmental science, and data analytics. This allows TRIZ practitioners to systematically apply inventive principles to address challenges like optimizing energy efficiency while minimizing environmental impact. These examples demonstrate how our interdisciplinarity quantification method serves as a practical tool for facilitating TRIZ-based systematic innovation.

While our approach shows promising results, it has some limitations. Its reliance on metadata tags for labeling can introduce biases due to inaccuracies. In addition, understanding the model decision-making process is challenging. To improve interpretability in interdisciplinary detection, future work could explore and incorporate techniques such

as attention mechanisms [44], layer-wise relevance propagation [45], saliency maps [46], integrated gradients [47], and LIME [48].

Future research directions include the refinement of labels by incorporating expert annotations or semi-supervised learning to improve label quality. Developing methods to enhance the interpretability of deep learning models in the context of interdisciplinarity is also essential. Applying the methodology to other corpora can validate generalizability. Creating tools that integrate our interdisciplinarity quantification with TRIZ methodologies can support engineers in inventive problem solving. Especially, STEM articles classified as interdisciplinary can be used as a foundation to build upon in the context of a subsequent model bridging concepts across disciplines.

Building upon recent advancements in interdisciplinary research quantification, our study aligns with methodologies such as those proposed by Likhareva et al. [49], who employed SciBERT-CNN models augmented with topic modeling to enhance multi-label text classification. While their approach leverages domain-specific embeddings and topic extraction to address class imbalances and semantic nuances, our Text CNN model focuses on scalable deep learning techniques using titles and abstracts from the extensive S2ORC dataset. Future work could explore integrating SciBERT's domain-specific capabilities and topic modeling into our framework to further refine the semantic analysis and improve classification accuracy. Additionally, combining these methodologies may enable a more nuanced understanding of interdisciplinary interactions, thereby enhancing the robustness and applicability of interdisciplinarity quantification within the TRIZ-based innovation framework.

Recent developments, such as the Set-CNN model proposed by Zhou et al. [50], highlight the benefits of semantic extension in enriching text representations for classification tasks. Similar to our deep learning approach, Set-CNN employs convolutional neural networks to capture intricate semantic patterns, which can enhance the identification of interdisciplinary content in scientific abstracts. Integrating semantic extension and multi-channel convolutions from Set-CNN into our framework could further improve our model's ability to detect subtle cross-disciplinary integrations.

Furthermore, TRIZ formalism can be used to build a topic model on text excerpts narrowed to contradictions and associated parameters [51]. This enables the frequency of association between parameters to be used as a parameter of interest in the creation of a discipline-specific interdisciplinary graphs as opposed to having pairing relying solely on semantic content. For example, volume and weight could be paired in the same article.

A comprehensive integration would involve several key steps. First, the trained Text CNN model would be utilized to classify a large corpus of scientific articles, identifying those with high interdisciplinarity scores. For these interdisciplinary articles, topic modeling techniques would extract the primary and secondary disciplines involved, effectively pairing the domains. TRIZ contradiction analysis would be applied to these articles to uncover inherent conflicts and inventive principles that can be leveraged. Finally, using the insights gained, innovative solutions would be generated by adapting principles from one discipline to address problems in another following the TRIZ methodology.

6. Conclusions

In this study, we presented a text-based deep learning approach to quantify interdisciplinarity within the TRIZ framework, specifically focusing on engineering applications. By leveraging semantic analysis and employing a Text CNN model, we addressed limitations of previous citation-based measures and content-based approaches that relied solely on metadata and did not analyze semantic content.

Our model achieved higher performance compared to traditional machine learning models, demonstrating the effectiveness of deep learning in capturing complex semantic patterns indicative of interdisciplinarity. The findings highlight the significant presence of interdisciplinary research within the corpus analyzed, emphasizing the importance of cross-disciplinary collaboration in advancing knowledge and innovation in engineering.

By expanding on the methodologies of previous studies, our research contributes to the field by offering a scalable, deep learning-based solution for quantifying interdisciplinarity. This approach not only enhances detection accuracy but also provides actionable insights for promoting interdisciplinary collaboration.

Our interdisciplinarity quantification method facilitates the systematic application of TRIZ principles by identifying and utilizing cross-disciplinary research, supporting innovative problem solving in engineering. A robust analysis of interdisciplinary work can help identify emerging fields and aid strategic decision making to foster innovation.

Future Perspectives

While our model performs well in quantifying interdisciplinarity, several key areas for future research could further enhance its effectiveness. Extending the analysis beyond abstracts to include full-text articles would provide a more comprehensive view of interdisciplinary connections, capturing richer contextual details often absent in abstracts alone. Developing discipline-specific models could also improve accuracy, allowing the model to capture field-specific language and nuances, thereby refining its analysis across diverse domains.

Additionally, combining semantic analysis with bibliometric network analysis—such as co-authorship or citation networks—could deepen insights into the structural and social dynamics of interdisciplinary research. This combined approach would not only highlight thematic connections but also reveal collaborative patterns that influence interdisciplinary knowledge exchange. Together, these directions offer pathways to a more robust and nuanced understanding of interdisciplinarity in the scientific literature.

Future studies could also explore the interpretability of deep learning models in the context of interdisciplinarity detection, aiding in understanding the underlying factors that contribute to interdisciplinary integration.

Author Contributions: Conceptualization, N.D. and D.C.; methodology, N.D.; software, N.D.; validation, A.S.; formal analysis, N.D.; investigation, N.D.; resources, G.G.; data curation, N.D.; writing—original draft preparation, N.D.; writing—review and editing, A.S., G.G. and D.C.; visualization, N.D.; supervision, D.C. and G.G.; project administration, G.G. and D.C.; funding acquisition, D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was completed during a PhD contract funded by the AIARD (Artificial Intelligence Aided Research and Development) industrial chair. The AIARD industrial chair is co-financed by the Grand Est region, the Eurométropole de Strasbourg, and ten companies, a list of which is available on the chair's website www.aiard.eu (accessed on 7 January 2025).

Institutional Review Board Statement: We adhered to all ethical guidelines and licensing requirements associated with the use of the Semantic Scholar Open Research Corpus (S2ORC). The dataset is publicly available under the CC BY-NC 4.0 license, and we used it solely for non-commercial research purposes.

Data Availability Statement: All data used in this study have been generated using the tools and methods outlined in this article. Should any difficulties arise when attempting to recreate these experiments, all data used, including annotations made during the project, can be sent upon request.

Acknowledgments: This research was conducted as part of the Collaborative Doctoral Program in Applied Artificial Intelligence for Industrial Applications, which is jointly offered by the University of Strasbourg and the National Institute of Applied Sciences (INSA) in France in partnership with Manhattan College's q(CINE) laboratory in New York, USA. We extend our heartfelt appreciation to the CSIP group at the ICube Laboratory, INSA of Strasbourg and to all our colleagues for their unwavering support, collaboration, and invaluable ideas throughout this project.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Technical Details

Model Architecture and Hyperparameters

The Text Convolutional Neural Network (Text CNN) model was configured with the following hyperparameters:

- **BPE Encoding Maximum Length:** 300
- **CNN Filters:** (256, 256)
- **CNN Kernel Sizes:** (2, 3)
- **CNN Pooling Transformation:** max
- **Learning Rate:** 0.001
- **Dropout Rate:** 0.2
- **Batch Size:** 64
- **Optimizer:** Adam with a learning rate of 1×10^{-3}
- **Activation Function:** PReLU for convolutional layers, sigmoid for the output layer
- **Hidden Units:** None
- **Loss Function:** Binary Crossentropy
- **Early Stopping Patience:** 2
- **Trainable Embeddings:** True
- **Dropout Type:** normal
- **Hidden Weight Initializer:** $he_{uniform}$
- **Hidden Bias Initializer:** zeros
- **Output Bias Initializer:** mean

The BPE encoding was configured with the following parameters:

- **Maximum Length:** 300
- **Padding:** post

The model was trained for six epochs with early stopping based on validation loss.

Appendix B. Technical Validation

In addition to the metrics reported, we conducted robustness checks by evaluating the model on different random seeds and subsets of the data. The results were consistent across runs, indicating the stability of the model's performance.

We also performed an error analysis to understand the types of misclassifications made by the model. A significant portion of errors occurred in cases where the interdisciplinarity was subtle. This suggests that improving label quality and incorporating additional contextual information could help further enhance the performance of the model.

References

1. Van Noorden, R. Interdisciplinary research by the numbers. *Nature* **2015**, *525*, 306–307. [[CrossRef](#)] [[PubMed](#)]
2. Wagner, C.S.; Roessner, J.D.; Bobb, K.; Klein, J.T.; Boyack, K.W.; Keyton, J.; Ràfols, I.; Börner, K. Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *J. Inf.* **2011**, *5*, 14–26. [[CrossRef](#)]
3. National Academy of Sciences; National Academy of Engineering; Institute of Medicine. *Facilitating Interdisciplinary Research*; The National Academies Press: Washington, DC, USA, 2005.

4. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)] [[PubMed](#)]
5. Allen, D.T.; Shonnard, D.R. *Sustainable Engineering: Concepts, Design and Case Studies*; Prentice Hall: Saddle River, NJ, USA, 2011.
6. Hu, J.; Zhang, Y. Measuring the interdisciplinarity of Big Data research: A longitudinal study. *Online Inf. Rev.* **2018**, *42*, 681–696. [[CrossRef](#)]
7. Rafols, I.; Meyer, M.S. How cross-disciplinary is bionanotechnology? Explorations in the specialty of molecular motors. *Scientometrics* **2007**, *70*, 633–650. [[CrossRef](#)]
8. Wang, Q.; Schneider, J.W. Consistency and validity of interdisciplinarity measures. *Quant. Sci. Stud.* **2018**, *1*, 239–263. [[CrossRef](#)]
9. Porter, A.L.; Roessner, J.D.; Cohen, A.S.; Perreault, M. Interdisciplinary research: Meaning, metrics and nurture. *Res. Eval.* **2006**, *15*, 187–195. [[CrossRef](#)]
10. Leydesdorff, L.; Rafols, I. Indicators of the Interdisciplinarity of Journals: Diversity, Centrality, and Citations. *J. Inf.* **2011**, *5*, 87–100. [[CrossRef](#)]
11. Morillo, F.; Bordons, M.; Gómez, I. Interdisciplinarity in science: A tentative typology of disciplines and research areas. *J. Assoc. Inf. Sci. Technol.* **2003**, *54*, 1237–1249. [[CrossRef](#)]
12. Porter, A.L.; Porter, A.L.; Cohen, A.S.; Roessner, J.D.; Perreault, M. Measuring researcher interdisciplinarity. *Scientometrics* **2007**, *72*, 117–147. [[CrossRef](#)]
13. Small, H.G. Co-citation in the scientific literature: A new measure of the relationship between two documents. *J. Am. Soc. Inf. Sci.* **1973**, *24*, 265–269. [[CrossRef](#)]
14. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [[CrossRef](#)]
15. Simpson, E.H. Measurement of Diversity. *Nature* **1949**, *163*, 688. [[CrossRef](#)]
16. Hirschman, A.O. The Paternity of an Index. *Am. Econ. Rev.* **1964**, *54*, 761–762.
17. Tshitoyan, V.; Dagdelen, J.; Weston, L.; Dunn, A.; Rong, Z.; Kononova, O.V.; Persson, K.A.; Ceder, G.; Jain, A. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **2019**, *571*, 95–98. [[CrossRef](#)]
18. Otter, D.; Medina, J.R.; Kalita, J.K. A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 604–624. [[CrossRef](#)] [[PubMed](#)]
19. Blei, D.M. Probabilistic topic models. *Commun. ACM* **2012**, *55*, 77–84. [[CrossRef](#)]
20. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), Minneapolis, MN, USA, 2–7 June 2019.
21. Kim, Y. Convolutional Neural Networks for Sentence Classification. In Proceedings of the Conference on Empirical Methods in Natural Language Processing 2014, Doha, Qatar, 25–29 October 2014.
22. Altshuller, G.S.; Shulyak, L.A. *And Suddenly the Inventor Appeared: TRIZ, the Theory of Inventive Problem Solving*; Technical Innovation Center, Inc.: Worcester, MA, USA, 1996.
23. Douard, N.; Samet, A.; Giakos, G.C.; Cavallucci, D. Bridging Two Different Domains to Pair Their Inherent Problem-Solution Text Contents: Applications to Quantum Sensing and Biology. In Proceedings of the TRIZ Future Conference 2022, Warsaw, Poland, 27–29 September 2022.
24. Douard, N.; Samet, A.; Giakos, G.C.; Cavallucci, D. Navigating the Knowledge Network: How Inter-Domain Information Pairing and Generative AI Can Enable Rapid Problem-Solving. In Proceedings of the TRIZ Future Conference 2023, Offenburg University, Offenburg, Germany, 12–14 September 2023.
25. Stirling, A. A general framework for analysing diversity in science, technology and society. *J. R. Soc. Interface* **2007**, *4*, 707–719. [[CrossRef](#)] [[PubMed](#)]
26. MacLeod, M.; Merz, M.; Mäki, U.; Nagatsu, M. Investigating Interdisciplinary Practice: Methodological Challenges (Introduction). *Perspect. Sci.* **2019**, *27*, 545–552. [[CrossRef](#)]
27. Nanni, F.; Dietz, L.; Faralli, S.; Glavas, G.; Ponzetto, S.P. Capturing Interdisciplinarity in Academic Abstracts. *D-lib Mag.* **2016**, *22*, 9. [[CrossRef](#)]
28. Pham, H.-S.; Poelmans, H.; Ali-Eldin, A.M.T. A Metadata-Based Approach for Research Discipline Prediction Using Machine Learning Techniques and Distance Metrics. *IEEE Access* **2023**, *11*, 61995–62012. [[CrossRef](#)]
29. Beltagy, I.; Lo, K.; Cohan, A. SciBERT: A Pretrained Language Model for Scientific Text. In Proceedings of the Conference on Empirical Methods in Natural Language Processing 2019, Hong Kong, China, 3–7 November 2019.
30. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
31. Cortes, C.; Vapnik, V.N. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
32. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
33. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [[CrossRef](#)]
34. Lo, K.; Wang, L.L.; Neumann, M.; Kinney, R.M.; Weld, D.S. S2ORC: The Semantic Scholar Open Research Corpus. In Proceedings of the Annual Meeting of the Association for Computational Linguistics 2020, Online, 5–10 July 2020.
35. Sennrich, R.; Haddow, B.; Birch, A. Neural Machine Translation of Rare Words with Subword Units. *arXiv* **2015**, arXiv:1508.07909.

36. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
37. Bergstra, J.; Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
38. Manaswi, N. Understanding and Working with Keras. In *Deep Learning with Applications Using Python: Chatbots and Face, Object, and Speech Recognition with TensorFlow and Keras*; Springer: Berlin/Heidelberg, Germany, 2018.
39. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 6. [[CrossRef](#)]
40. Grootendorst, M.R. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794.
41. Li, Z.; Hui, J.; Yang, P.; Mao, H. Microfluidic Organ-on-a-Chip System for Disease Modeling and Drug Development. *Biosensors* **2022**, *12*, 370. [[CrossRef](#)] [[PubMed](#)]
42. Vincent, J.F.V.; Bogatyreva, O.; Bogatyrev, N.; Bowyer, A.; Pahl, A. Biomimetics: Its practice and theory. *J. R. Soc. Interface* **2006**, *3*, 471–482. [[CrossRef](#)]
43. Patil, S.; Rao, P.S.; Prabhudev, M.S.; Khan, M.Y.; Anjaiah, G. Optimization of Cutting Parameters During CNC Milling of EN24 Steel with Tungsten Carbide Coated Inserts: A Critical Review. *Mater. Today Proc.* **2022**, *62*, 3213–3220. [[CrossRef](#)]
44. Kitada, S. Improving Prediction Performance and Model Interpretability through Attention Mechanisms from Basic and Applied Research Perspectives. *arXiv* **2023**, arXiv:2303.14116.
45. Montavon, G.; Binder, A.; Lapuschkin, S.; Samek, W.; Müller, K. Layer-Wise Relevance Propagation: An Overview. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*; Springer: Berlin/Heidelberg, Germany, 2019.
46. Ismail, A.A.; Bravo, H.C.; Feizi, S. Improving Deep Learning Interpretability by Saliency Guided Training. *Neural Inf. Process. Syst.* **2021**, *34*, 26726–26739.
47. Ancona, M.; Ceolini, E.; Oztireli, C.; Gross, M.H. A unified view of gradient-based attribution methods for Deep Neural Networks. *arXiv* **2017**, arXiv:1711.06104.
48. Zafar, M.R.; Khan, N.M. Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 525–541. [[CrossRef](#)]
49. Likhareva, D.; Sankaran, H.; Thiyagarajan, S. Empowering Interdisciplinary Research with BERT-Based Models: An Approach Through SciBERT-CNN with Topic Modeling. *arXiv* **2024**, arXiv:2404.13078.
50. Zhou, Y.; Li, J.; Chi, J.; Tang, W.; Zheng, Y. Set-CNN: A text convolutional neural network based on semantic extension for short text classification. *Knowl. Based Syst.* **2022**, *257*, 109948. [[CrossRef](#)]
51. Guarino, G.; Samet, A.; Cavallucci, D. PaTRIZ: A framework for mining TRIZ contradictions in patents. *Expert Syst. Appl.* **2022**, *207*, 117942. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.