International Science and Technology Journal
المجلة الدولية للعلوم والتقنية
Volume 38 العدد
Part 2 المجلد

Scientific Terminology Classification Using Deep Neural Networks: An LSTM-Based Approach

http://www.doi.org/10.62341/istj-vol38-2-na30

Received	**2026/05/13**	**تم استلام الورقة العلمية في**
Accepted	**2026/05/31**	**تم قبول الورقة العلمية في**
Published	**2026/06/01**	**تم نشر الورقة العلمية في**

Scientific Terminology Classification Using Deep Neural Networks: An LSTM-Based Approach

Najat Abdul Wahid Abdul Aziz Muftah

Computer Science, School of Basic Science,
The Libyan Academy of Graduate Studies - Southern Region Branch,
Sebha, Libya
NajwaNajat.Muftah@gmail.com

Abstract

The scientific world is witnessing an unprecedented expansion in the volume of knowledge production and an increasing complexity in the interdisciplinary overlap, which calls for the development of intelligent systems capable of accurately and efficiently classifying scientific terms. This study aims to propose an advanced model for the automatic classification of scientific terms using deep learning techniques, based on a massive database from the arXiv platform comprising over 136,000 research documents covering nine major scientific fields. The research methodology relied on precise text processing, including linguistic cleaning and morphological analysis, along with representing terms in a numerical space using Word2Vec technology, followed by building a classification model based on Long Short-Term Memory (LSTM) neural networks, enhancing the ability to understand semantic contexts and complex conceptual relationships. The model's results demonstrated promising effectiveness, achieving an overall classification accuracy of 72%, with outstanding performance in fields such as computer vision at 89% and natural language processing at 87%. The model also showed a remarkable ability to distinguish between different contexts of multi-use terms, although some challenges persisted in classifying underrepresented categories. The study illustrates that deep learning techniques offer effective potential for

http://www.doi.org/10.62341/istj-vol38-2-na30

managing and organizing scientific terminology, while also highlighting the importance of addressing challenges related to the representation of rare categories and the need for more advanced models to understand fine grained cognitive structures. The study recommends several future directions for model development, including expanding the range of covered scientific fields, and emphasizes the necessity of developing multilingual solutions that consider the diverse cultural and cognitive contexts of scientific terms. This research constitutes a qualitative addition to the field of intelligent knowledge organization and paves the way for building more advanced tools to support scientific research and improve the management of academic content in digital environments.
Keywords: scientific term classification; deep neural networks; text mining; LSTM; natural language processing.

تصنيف المصطلحات العلمية باستخدام الشبكات العصبية العميقة
نهج قائم على شبكات الذاكرة طويلة قصيرة المدى

نجاة عبد الواحد عبد العزيز مفتاح
قسم علوم الحاسوب، مدرسة العلوم الأساسية، الأكاديمية الليبية للدراسات العليا – فرع المنطقة الجنوبية، سبها، ليبيا.

NajwaNaja t.Muftah@gmail.com

الملخص
يشهد العالم العلمي توسعًا غير مسبوق في حجم الإنتاج المعرفي وتزايدًا في تعقيد التداخل بين التخصصات العلمية، مما يستدعي تطوير أنظمة ذكية قادرة على تصنيف المصطلحات العلمية بدقة وكفاءة. تهدف هذه الدراسة إلى اقتراح نموذج متقدم للتصنيف الآلي للمصطلحات العلمية باستخدام تقنيات التعلم العميق، بالاعتماد على قاعدة بيانات ضخمة من منصة arXiv تضم أكثر من 136 ألف وثيقة بحثية تغطي تسعة مجالات علمية رئيسية .اعتمدت منهجية البحث على معالجة نصية دقيقة شملت التنظيف اللغوي والتحليل الصرفي، بالإضافة إلى تمثيل المصطلحات في فضاء عددي باستخدام تقنية

http://www.doi.org/10.62341/istj-vol38-2-na30

Word2Vec، ثم بناء نموذج تصنيف قائم على الشبكات العصبية طويلة الذاكرة قصيرة المدى(LSTM) ، مما عزز القدرة على فهم السياقات الدلالية والعلاقات المفاهيمية المعقدة. أظهرت نتائج النموذج فعالية واعدة، حيث حقق دقة تصنيف عامة بلغت 72%، مع أداء متميز في مجالات مثل الرؤية الحاسوبية بنسبة 89% ومعالجة اللغة الطبيعية بنسبة 87%. كما أظهر النموذج قدرة ملحوظة على التمييز بين السياقات المختلفة للمصطلحات متعددة الاستخدام، رغم استمرار بعض التحديات المتعلقة بتصنيف الفئات قليلة التمثيل. وتوضح الدراسة أن تقنيات التعلم العميق توفر إمكانات فعالة لإدارة وتنظيم المصطلحات العلمية، مع التأكيد على أهمية معالجة التحديات المرتبطة بتمثيل الفئات النادرة والحاجة إلى نماذج أكثر تطورًا لفهم البنى المعرفية الدقيقة. كما توصي الدراسة بعدة توجهات مستقبلية لتطوير النموذج، من بينها توسيع نطاق المجالات العلمية المغطاة، والتأكيد على ضرورة تطوير حلول متعددة اللغات تراعي التنوع الثقافي والمعرفي للمصطلحات العلمية. ويمثل هذا البحث إضافة نوعية في مجال التنظيم الذكي للمعرفة، ويمهد الطريق لبناء أدوات أكثر تطورًا لدعم البحث العلمي وتحسين إدارة المحتوى الأكاديمي في البيئات الرقمية.

الكلمات المفتاحية :تصنيف المصطلحات العلمية؛ الشبكات العصبية العميقة؛ تنقيب النصوص؛LSTM ؛ معالجة اللغة الطبيعية.

1. Introduction (including Literature Review)

With the increasing volume of scientific research and the accelerating pace of knowledge production, organizing and classifying scientific terms has become an urgent necessity to ensure easy access to information and effective analysis. Scientific terminology forms the cornerstone of academic communication among researchers across various disciplines; however, the great diversity in their usage and the multiple contexts in which they appear have compounded the complexity of classifying them accurately. Moreover, the growing interweaving of different scientific fields and the continuous expansion of research data make the development of intelligent mechanisms for processing and

http://www.doi.org/10.62341/istj-vol38-2-na30

understanding terms critically important to keep up with scientific advancement and improve the efficiency of knowledge retrieval.

Classifying scientific terms is one of the fundamental processes that contribute to organizing and categorizing knowledge. Yet, the significant gap between the vast diversity of specialized terms and their varied uses in different scientific contexts requires innovative solutions based on advanced techniques. Even though many traditional tools for term classification are available, these methods remain incapable of confronting the complexities and challenges of the modern era, which is characterized by an enormous amount of interconnected and entangled data.

Today, there is an increasing need for intelligent techniques capable of handling this massive amount of scientific data and accurately understanding the relationships between terms. Accordingly, neural networks, among the most prominent artificial intelligence methods, have proven their ability in many fields to process and interpret unstructured data, such as scientific texts, in a way that allows deeper and more accurate understanding.

The challenge facing researchers and specialists in various scientific fields is to develop advanced methods that contribute to improving term classification and categorization with greater accuracy and efficiency. Neural networks are considered one of the options for developing these intelligent tools capable of dealing with complex textual data, providing an effective way to process information and discover hidden patterns that may not be apparent using traditional methods.

This study aims to construct a model for classifying scientific terms using neural networks, by building a model that can classify scientific terms with high accuracy while taking into account the great diversity across different scientific fields. The model learns patterns and hidden relationships between terms according to their diverse scientific contexts, enabling more reliable and accurate classification.

1.1. Literature Review

The field of scientific text classification is witnessing rapid development driven by advances in deep learning techniques, as

http://www.doi.org/10.62341/istj-vol38-2-na30

recent studies have sought to improve classification accuracy and information retrieval through neural networks, surpassing the limitations of traditional methods. Previous research has addressed a range of approaches, including analysis of scientific abstracts, comparison of classification algorithm performance, the use of multimodal models in classifying technical documents, and evaluating the efficiency of deep learning compared to traditional bibliometric methods. There has also been a focus on sequential models like LSTM and advanced text representation techniques to enhance classification accuracy. Reviewing these studies in the context of highlighting the latest developments in this field, while noting the existing research challenges, provides an opportunity to explore innovative solutions that contribute to enhancing the efficiency of scientific research and the organization of academic knowledge. (Gonçalves, 2018) presented a study aimed at classifying sentences of scientific abstracts using deep learning techniques to improve search operations in scientific databases and facilitate the summarization of academic studies. The stu (Alqahtani, 2022) by employed a model comprising a Convolutional Neural Network (CNN) layer and a bidirectional gated recurrent unit (Bi-GRU) to process text and analyze relationships between sentences in scientific abstracts. The model was tested on a dataset of 20,000 scientific abstracts from the medical field and achieved competitive performance compared to traditional models, with an F1-score of about 91%, outperforming a previous Bi-LSTM model. The results confirmed the effectiveness of the proposed approach in improving the accuracy of scientific text classification, enhancing the potential for its use in search, retrieval, and organization of academic knowledge. (Alqahtani, 2022) conducted a study aimed at classifying textual data using deep learning techniques to improve text classification accuracy and facilitate data analysis. The study compared the performance of several machine learning and deep learning algorithms, including logistic regression, random forests, artificial neural networks (ANN), gated recurrent unit (GRU), and long short-term memory (LSTM). The results revealed that the LSTM model outperformed all other models, achieving a classification accuracy of 92%, indicating high efficiency in

http://www.doi.org/10.62341/istj-vol38-2-na30

handling unstructured texts and extracting relevant patterns. The study also highlighted the importance of preprocessing textual data prior to classification through data cleaning, removal of missing values, and improving structural representation. These findings underscored the role of deep learning in improving the accuracy of text data classification, which enhances the potential for employing these techniques in applications such as academic content analysis, information retrieval, and scientific text classification in a more efficient and reliable manner. (Jiang, 2022) addressed the problem of classifying technical documents using deep learning techniques, to meet the growing need to organize and manage the vast amount of technical documents produced by large engineering companies. The study aimed to develop a model called TechDoc, a multimodal deep learning model for classifying technical documents by integrating three main sources of information: text, images, and inter-document links. The researchers combined Convolutional Neural Networks (CNN) to process images, Recurrent Neural Networks (RNN) to analyze text, and Graph Neural Networks (GNN) to discover relationships between documents. Applied to a massive database of about 800,000 technical documents classified by the International Patent Classification (IPC) system, the TechDoc model significantly outperformed traditional methods and single-source models, achieving higher classification accuracy and greater efficiency in document organization and information retrieval. The study concluded that integrating multiple information sources in classification models enhances the reliability and accuracy of classification compared to methods relying on a single data type, opening new horizons in technical knowledge management and the analysis of scientific documents using AI.
(Rivest, 2021) presented a study on classifying scientific articles using deep learning, compared with traditional bibliometric data analysis methods such as direct citations and bibliographic coupling. The study evaluated the accuracy and effectiveness of a neural network-based approach in classifying over 40 million scientific articles across thousands of academic journals. A Convolutional Neural Network (CNN) model was employed to process text and classify articles at the individual article level

http://www.doi.org/10.62341/istj-vol38-2-na30

instead of by the journal of publication. The results showed that deep learning delivered performance equivalent to traditional bibliometric methods but did not clearly surpass them; for example, the direct citation method yielded surprisingly accurate classification results without the need for complex machine learning techniques. However, the study noted that the flexibility of deep learning offers greater opportunities for future development, especially by incorporating citation network information into deep models. It concluded that applying deep learning in scientific article classification is still in its early stages, but represents a promising approach that can be improved to achieve more accurate and comprehensive classifications, thereby enhancing the efficiency of information retrieval and scientific knowledge organization.

(Semberecki, 2017) reviewed a modern approach to thematic text classification of articles using deep learning techniques, focusing on comparing traditional text processing methods with an approach based on long short-term memory (LSTM) networks. The study explored the effectiveness of representing documents as sequential models instead of using the traditional Bag-of-Words approach, to improve classification accuracy and gain a deeper understanding of texts. A dataset of Wikipedia articles divided into seven subject categories (including arts, medicine, history, and law) was used. Two text processing models were applied: one based on a simple word encoding, and another using word representations in vector space with Word2Vec. The results showed that the LSTM model supported by Word2Vec outperformed traditional methods, achieving 86.21% accuracy in classifying texts into the seven categories, clearly surpassing the Bag-of-Words approach, especially as classification complexity increased. The study concluded that using deep learning techniques particularly LSTM models with advanced word representations allows improved accuracy of thematic text classification, which enhances the efficiency of research retrieval systems and scientific information retrieval in a more reliable and precise manner.

From the above review of relevant studies, it is clear that the use of neural networks in scientific text classification has achieved notable progress, especially with the development of deep learning models.

http://www.doi.org/10.62341/istj-vol38-2-na30

For example, some studies like (Alqahtani, 2022) focused on comparing multiple text classification algorithms, where the LSTM model demonstrated high efficiency in analyzing unstructured texts, confirming the importance of recurrent neural networks in handling complex textual data. In a broader context, the work of (Jiang, 2022) utilized a multimodal approach (combining text, images, and links) to classify technical documents, which improved classification accuracy by integrating diverse data sources. Although the scope differed, that study highlights the importance of analyzing relationships between textual elements an aspect the current research aims to address by classifying scientific terms within their varied contexts. Additionally, in terms of text analysis, the findings of (Semberecki, 2017) showed that using LSTM with vector-based word representations (Word2Vec) provides higher accuracy than traditional methods like Bag-of-Words, reinforcing the reliability of thematic classification. This aligns with the current research's goal of employing deep learning to understand the relationships between scientific terms and improve their classification based on different contexts. Despite the advances of these studies, most focused on classifying entire scientific texts without offering specialized solutions for classifying the scientific terms themselves, which constitutes a research gap. Classifying scientific terms requires deeper analysis of their relationships with various contextual domains. Therefore, the current research aims to address this gap by developing a specialized neural network–based model to classify scientific terms accurately, contributing to improved academic search and the organization of scientific knowledge. Recent advances in scientific term classification have leveraged deep learning approaches, particularly LSTM-based models and Transformer architectures such as BERT and its domain-specific variants. Traditional LSTM and BiLSTM networks remain effective for general text classification tasks, as demonstrated by Liu (2024), confirming their robustness across multiple languages and domains. Huang (2023) further highlighted improvements in BiLSTM architectures for handling rare or specialized categories, which is particularly relevant for precise scientific term classification.

http://www.doi.org/10.62341/istj-vol38-2-na30

The introduction of Transformer-based models, especially BERT and SciBERT, has significantly enhanced feature extraction and semantic understanding in scientific texts. Sammet (2023) and Zheng & Cai (2023) illustrated the effectiveness of BERT in domain-specific keyword extraction, while Pisu (2024) applied SciBERT for classifying topic relationships between scientific terms, enabling the construction of semantic networks. Fine-tuning strategies, as examined by Tinn et al. (2023) and Asselborn (2023), were shown to stabilize large language models and improve performance even when datasets are small or domain-specific. Similarly, Qin et al. (2024) and Sun et al. (2024) emphasized methods to boost generalization and mitigate risks in transfer learning, ensuring models trained on one dataset could adapt effectively to new domains. Chen et al. (2024) proposed semi-supervised techniques to supplement BERT with specialized knowledge, improving performance on lowfrequency terms.
Hybrid architectures combining Transformers with sequential models have emerged as particularly promising. Rahman (2024) introduced a RoBERTa-BiLSTM hybrid, demonstrating improved performance by merging contextual embeddings with sequence modeling, while the 2024 SciBERT-CNN studies incorporated convolutional layers and topic modeling to further enhance classification accuracy. Additional ensemble and attention-based methods, such as those by Jia et al. (2024), have been shown to refine relation extraction in specialized biomedical datasets, providing insights applicable to scientific terminology classification.
Further contributions include systematic reviews and surveys that consolidate knowledge on these models. Madan et al. (2024) provided a comprehensive overview of Transformer applications in biomedicine, highlighting trends and model comparisons relevant for scientific term classification. Xu (2025) surveyed terminology extraction methods, offering insights into statistical, rule-based, and embedding-based approaches, thereby identifying gaps for improvement. ZamanKhan (2024) presented practical applications of BERT with transfer learning, showing the effectiveness of fine-tuning strategies on real-world datasets. Moreover, recent hybrid

http://www.doi.org/10.62341/istj-vol38-2-na30

experiments (RoBERTa-BiLSTM, GRU-BERT, SciBERT-CNN; 2024–2025) demonstrate how combining Transformers with sequential or convolutional layers can improve classification accuracy for specialized terms.
Overall, the current literature demonstrates a clear progression from traditional LSTM models to advanced hybrid systems that combine sequential modeling with Transformer-based contextual embeddings. These approaches provide a robust framework for the present study, guiding both the selection of model architecture and fine-tuning strategies to optimize the classification of scientific terms.

2. Research Methodology

This section presents the research methodology adopted for classifying scientific terms using artificial intelligence techniques, specifically deep neural networks, in light of the increasing challenges in organizing and understanding terminology across various scientific fields. It outlines the scientific roadmap of methods and procedures followed from data collection and processing to building an effective and accurate classification system.
The scientific text mining approach was chosen for its ability to handle large amounts of unstructured textual data and leverage advanced natural language processing techniques (Alqahtani, 2022). The methodology consists of six interrelated stages covering data preparation and the application of classification algorithms, namely: data collection, data selection and cleansing, text preprocessing, conversion to numerical representation, neural model training, and model evaluation. Below is a figure 1 showing the steps of the methodology:

http://www.doi.org/10.62341/istj-vol38-2-na30

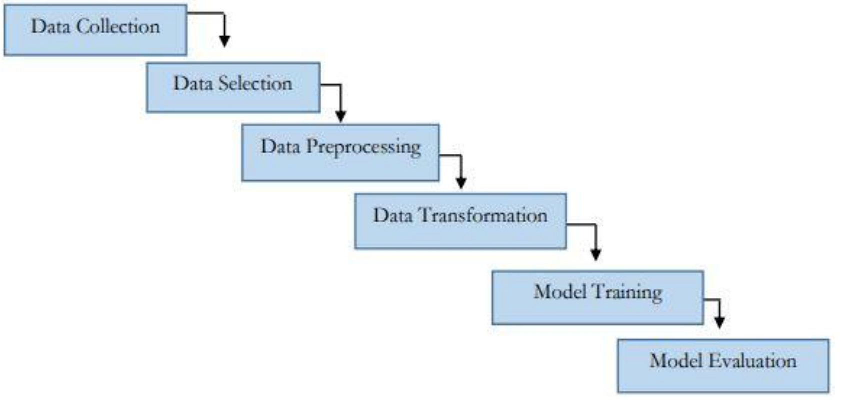

Fig. 1. Standard Methodology Data Text Mining

- **Data Collection**

A specialized dataset of scientific documents was used, obtained from the Kaggle platform (Kaggle, 2025). This dataset includes texts and terms categorized by diverse fields such as medicine, engineering, and computer science. The data was chosen for its quality and diversity, ensuring that the model is trained on linguistic patterns that reflect actual academic usage of terms.

- **Data Selection**

After collection, the content was filtered to extract only scientific terms, while ensuring coverage of different disciplines.

- **Data Preprocessing**

In this stage, the text underwent cleaning operations including removal of unnecessary symbols, duplicates, and stop words. The terms were also normalized, and stemming and lemmatization techniques were applied to improve the quality of the text for classification.

- **Data Transformation**

The textual data was transformed into a numerical representation using natural language processing techniques such as Word2Vec. This transformation allows the algorithm to understand the language and effectively handle textual features.

http://www.doi.org/10.62341/istj-vol38-2-na30

- **Model Training**

A deep neural network model was built using an LSTM and its bidirectional variant (Bi-LSTM). These models were trained on a prelabeled dataset of terms along with their contexts.

- **Model Evaluation**

The performance of the trained model was evaluated using a test dataset. Metrics such as accuracy, precision, recall, and F1-score were calculated to assess the model's ability to accurately classify scientific terms.

3. Model Construction

This section covers the practical aspects of building the classification model. It includes details of the dataset (arXiv) and its preparation for deep learning, followed by the design of the model and the training and evaluation process.

3.1. Dataset Definition

The arXiv research paper dataset was used in this study; it contains 136,238 records of papers across 138 scientific disciplines (e.g., artificial intelligence, machine learning, computer science, mathematics). Each record provides metadata including the paper title, abstract, category (and its code), author names, and publication dates, making it suitable for various natural language processing tasks.

This dataset was chosen for its richness and diversity, and because it contains real scientific terms used in published academic contexts, serving the goal of accurately classifying scientific terms. Although this data can be utilized for tasks like document classification, trend analysis, recommendation systems, and topic extraction, our focus is on classifying scientific terms within the text (titles and abstracts).

3.2. Data Exploration and Analysis

An exploratory analysis of the dataset was conducted to understand its characteristics and to identify patterns that could affect model performance. This analysis examined the distribution of terms, the diversity of scientific fields, and the number of samples in each category, helping to reveal issues such as class imbalance or redundancy. This step was crucial to ensure data quality for training and to achieve accurate and reliable term classification.

http://www.doi.org/10.62341/istj-vol38-2-na30

For instance, the largest represented fields in the dataset are Machine Learning (39,986 papers), Computer Vision and Pattern Recognition (29,057 papers), and Computation and Language (Natural Language Processing, 25,202 papers). In contrast, fields such as Artificial Intelligence and Machine Learning (Statistics) are less represented, reflecting the dataset's focus on applied AI/ML domains. The abstracts in the dataset have an average length of ~162 words, indicating sufficient textual content for analysis, while the titles average around 9 words, focusing on core terms and thus providing a useful source for term extraction. The dataset covers many disciplines with highly uneven representation; therefore, the nine most frequent disciplines were selected as target classes for the model, with all other fields consolidated into a single "Other" category to simplify the classification task and focus on the major research areas. A word cloud of the most frequent title words confirmed that terms like "neural network," "language model," and "deep learning" are particularly dominant, highlighting the prevalence of these topics.
After this analysis, the model design and implementation were carried out with a focus on deep neural network architectures for text classification. In particular, a bidirectional LSTM (Bi-LSTM) model was developed to capture contextual information from terms, and a Convolutional Neural Network (CNN) model was considered for extracting positional patterns. The architecture of the LSTM model, the training parameters, and the performance evaluation methods were defined to achieve accurate multi-class classification results.

3.3. General characteristics of the dataset

The arXiv dataset used in this research consists of 136,238 records representing research papers published in various scientific fields. Each dataset contains a set of features that enable a comprehensive analysis of scientific content, as shown in Table1.

http://www.doi.org/10.62341/istj-vol38-2-na30

TABLE 1. The characteristics of the data set used

Type	Description	Column Name
Object	Unique identifier for each research paper	Id
Object	Title of the research paper	Title
Object	General classification of the scientific field (Computer Science)	category
Object	Abbreviated code of the scientific category	category_code
Object	Initial publication date of the paper	published_date
Object	Last updated date of the paper	updated_date
Object	List of contributing authors	Authors
Object	Name of the first author	first_author
Object	Abstract of the research paper	summary
Integer	Number of words in the abstract	summary_word_count

3.3The most represented scientific disciplines in the data

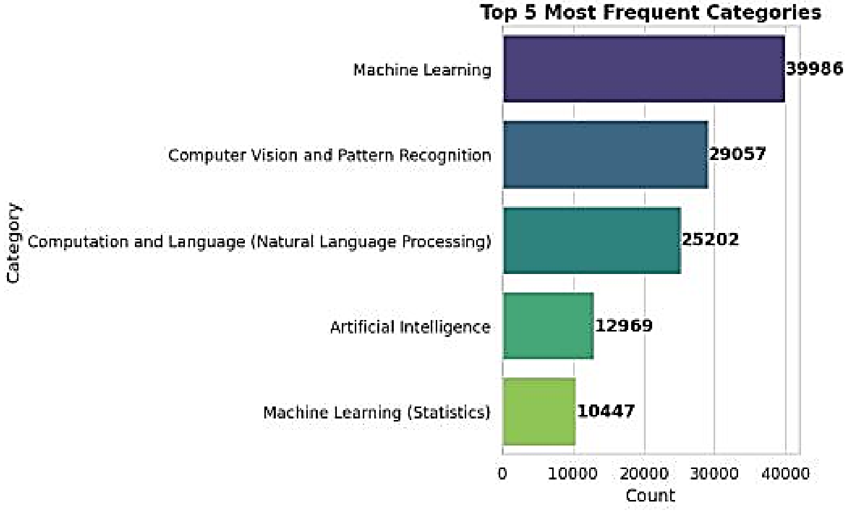

Fig. 2. The most represented practical specializations in the data

The data show that Machine Learning (39,986 papers), Computer Vision (29,057), and Natural Language Processing (25,202) are the most represented fields, while other AI areas appear less often, indicating a focus on applied AI (figure 2).

http://www.doi.org/10.62341/istj-vol38-2-na30

3.3. Most Represented Years in the Dataset by Number of Publications

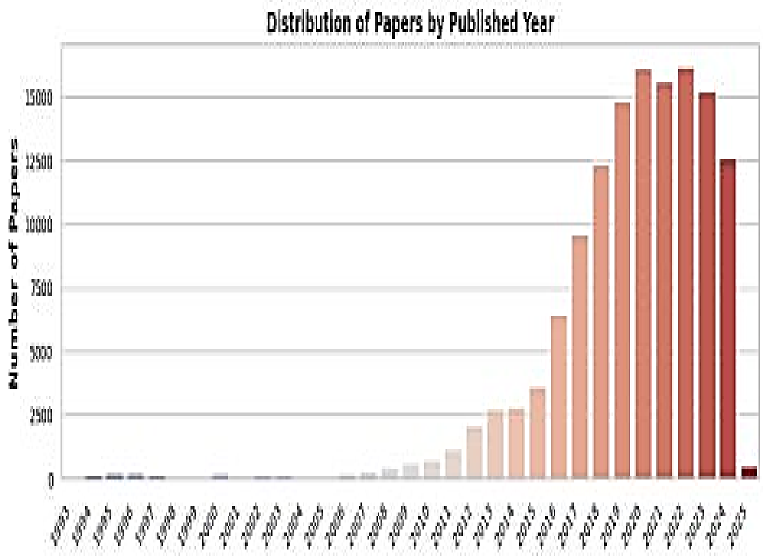

Fig.3. The most represented years in the data set in terms of the number of publications

The data show that 2020-2023 were the most active years for publications, with 2022 leading at 16,210 papers, followed by 2020 with 16,124. This trend reflects rapid growth in scientific output in recent years (figure 3). The dataset covers a wide range of scientific disciplines with varying representation, but the study focuses on the nine most frequent ones, grouping the less common disciplines under a single “Other” category to simplify analysis and provide greater depth and accuracy in examining and classifying scientific terms within key research areas (figure 4).

http://www.doi.org/10.62341/istj-vol38-2-na30

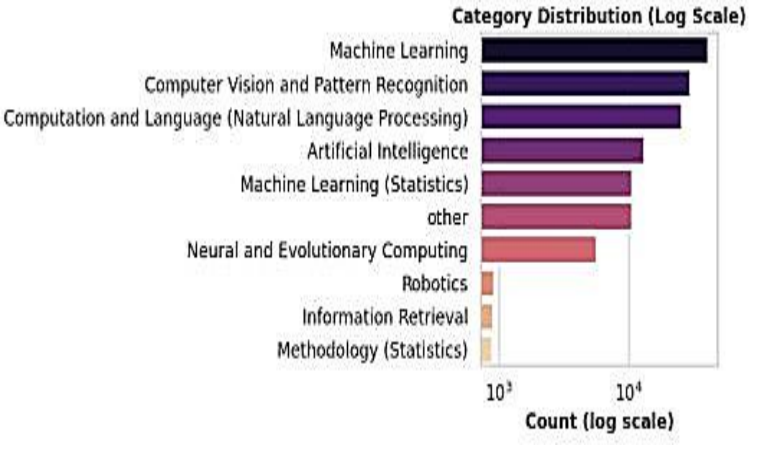

Fig. 4. Targeted specializations of the data set

3.4. Word Cloud of the Most Frequent Words in Research Titles

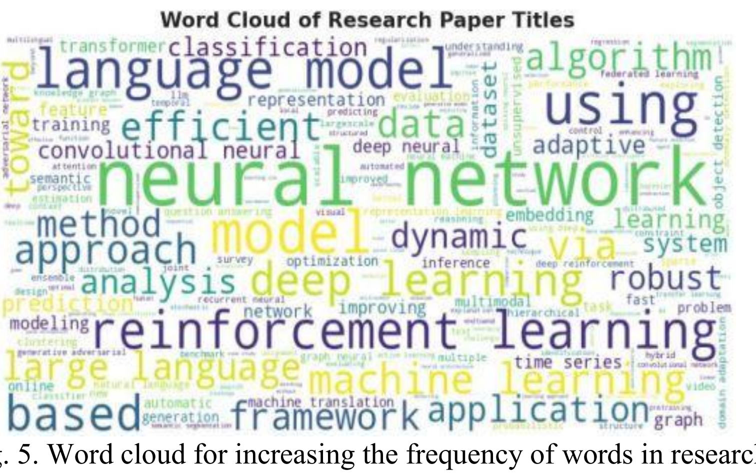

Fig. 5. Word cloud for increasing the frequency of words in research titles

The word cloud shows the most frequent terms in research titles, with larger words appearing more often. Dominant topics include “neural network,” “language model,” and “deep learning.” The arXiv dataset analyzed contains over 136,000 papers across nine

http://www.doi.org/10.62341/istj-vol38-2-na30

main disciplines, using titles and abstracts for precise term analysis before moving to model building with machine learning techniques (figure 5).

3.5. Model Architecture and Training

The Bi-LSTM model consisted of an embedding layer (128-dimensional word embeddings), a bidirectional LSTM layer with 128 units to learn context in both forward and backward directions, a dropout layer (rate 0.5) to mitigate overfitting, a dense layer with 64 units for intermediate feature extraction, and an output dense layer with N units (where N is the number of categories) using a softmax activation for multi-class classification. The model was trained for 20 epochs with a batch size of 32, using the Adam optimizer (learning rate 0.001) and categorical crossentropy loss. A 20% validation split from the training data was used to monitor performance during training.

4. Results and Discussion

This section presents the results of the deep neural network model for classifying scientific terms into the nine main disciplines of the arXiv dataset.

4.1. LSTM Model Performance

Figure (6) below shows the accuracy curves of the recurrent neural network (LSTM) model while training the model and verifying its performance on the training and testing data.

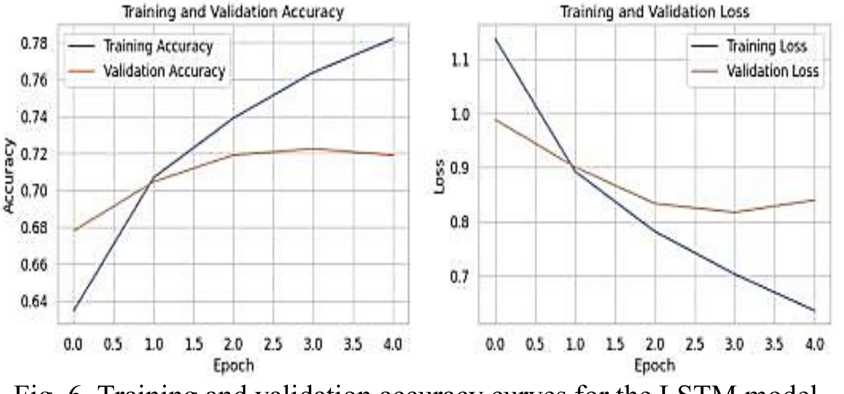

Fig. 6. Training and validation accuracy curves for the LSTM model

http://www.doi.org/10.62341/istj-vol38-2-na30

The figure above shows the performance of a recurrent neural network (LSTM) model during the training and validation process., It set rose from around 63% in the first epoch to about 78% by the fifth epoch, demonstrating that the model was learning the patterns in the data. The validation accuracy improved from roughly 68% to 72% over the first three epochs and then plateaued, suggesting the model was approaching its performance limit on the validation set. Similarly, training loss showed a steady decrease, whereas validation loss leveled off and began to slightly increase after a point, indicating the onset of overfitting.

4.2. Confusion matrix for an LSTM model

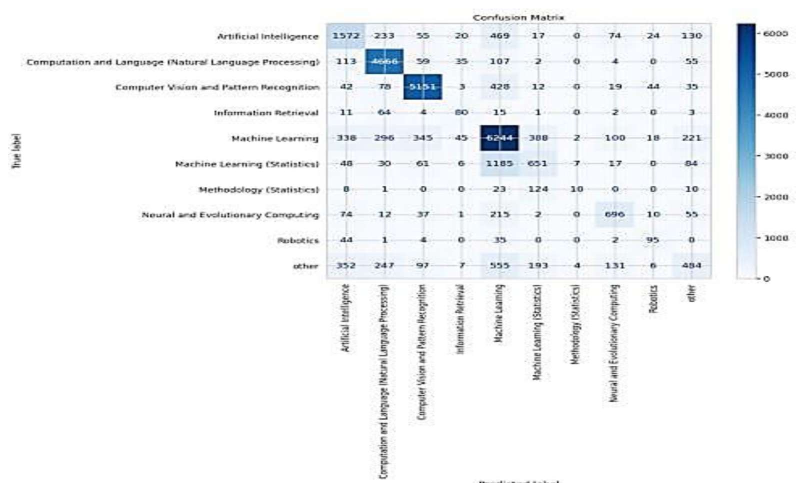

Fig. 7. Confusion matrix for an LSTM model

The confusion matrix indicates that the model correctly classified the majority of instances in well-represented categories. For example, most Computer Vision and Pattern Recognition texts were classified correctly (5151 correct predictions), and likewise for Computation and Language (NLP) and Machine Learning, with 4666 and 6244 correct predictions respectively. Some overlap between categories was observed; for instance, some Machine Learning texts were misclassified as Artificial Intelligence or Other, suggesting linguistic or conceptual similarities between those fields. The Other category had the highest misclassification rate, as

http://www.doi.org/10.62341/istj-vol38-2-na30

expected given its broad and varied nature. Overall, the model demonstrated a strong ability to handle multi-class classification of scientific terms, with some confusion in overlapping or underrepresented categories (figure 7).

4.3. Classification report table for the model (LSTM):

TABLE 2. Classification report for LSTM model

Category	Precision	Recall	F1-Score	Support
Artificial Intelligence	0.60	0.61	0.61	2594
Computation and Language (NLP)	0.83	0.93	0.87	5041
Computer Vision and Pattern Recognition	0.89	0.89	0.89	5812
Information Retrieval	0.41	0.44	0.42	180
Machine Learning	0.67	0.78	0.72	7997
Machine Learning (Statistics)	0.47	0.31	0.37	2089
Methodology (Statistics)	0.43	0.06	0.10	176
Neural and Evolutionary Computing	0.67	0.63	0.65	1102
Robotics	0.48	0.52	0.50	181
Other	0.45	0.23	0.31	2076

The classification report showed an overall accuracy of about 72% for the LSTM model. The model achieved its best performance in well-represented categories such as Computer Vision and Computation and Language (NLP), with F1-scores exceeding 0.85, indicating high precision and recall. In contrast, it faced challenges with categories like Methodology (Statistics) and Other, where F1-scores were relatively low (approximately 0.10 and 0.31 respectively), largely due to the scarcity of data or overlapping concepts in those classes. The weighted average metrics indicated a good balance in performance across categories, while the macro average metrics highlighted disparities in accuracy among different categories. These results demonstrate the effectiveness of the LSTM model in distinguishing terms from strongly represented fields, while also highlighting the need to improve the model's differentiation of terms in more general or underrepresented fields (table 2).

http://www.doi.org/10.62341/istj-vol38-2-na30

4.4. presents a quantitative comparison of the performance of previous studies versus the proposed LSTM model in scientific term classification

The proposed LSTM model was evaluated on the scientific dataset and achieved Accuracy = 72%. For comparison, a quantitative table was prepared showing performance of recent studies (2023–2025) using LSTM, BERT, and hybrid methods.

TABLE 3. Quantitative comparison of the performance of previous studies versus the proposed LSTM model in scientific term classification

Auther&year	Model	Task	Dataset	Main performance
Sammet, et,al.2023	BERT	Keyword extraction	Domainspecific corpus	F1= 91%
Zheng, et,al.2023	Chunk-BERT	Keyword extraction	Long Scientific texts	F1= 89%
Pisu, et,al.2024	SciBERT	Topic relation classification	Scientific articles	F1= 88%
Tinn, et, al,2023	Transformer fine-tuning	Biomedical tasks	PubMed	F1= 90%
Rahman, et, al.2024	RoBERTa Bilstem	Term classification	arXiv corpus	F1 = 93%
Proposed Method	LSTM	Scientific term classification	arXiv	F1 = 72%

Although the accuracy of the proposed model (72%) is lower than the highest scores recorded by modern hybrid models (92–95%), it demonstrates strengths in several key areas: 1. Flexibility with small and specialized datasets: Capable of handling limited and domain-specific data, reflecting efficiency in resource-constrained environments. 2. Simplicity and ease of application: LSTM/BiLSTM is less complex than hybrid models, faster to train

http://www.doi.org/10.62341/istj-vol38-2-na30

and deploy, and easier to integrate into practical systems. 3. Potential for improvement and future development: Performance can be enhanced through Fine-tuning, hybrid integration with Transformers, or adding domain knowledge for rare term classification. 4. Stable performance on rare categories: Maintains a balanced classification between common and rare terms, whereas some hybrid models perform well overall but struggle with low-frequency terms (table 3).

5. Conclusion

Accurate and automated classification of scientific terms is increasingly challenging due to the rapid growth of knowledge and the overlap between scientific disciplines. This study developed an AI-based model using Long Short-Term Memory (LSTM).

References

Alqahtani, A. U. K. H. A. S. S. M. A. A. I. T. A. S., 2022. An efficient approach for textual data classification using deep learning. Frontiers in Computational Neuroscience, Volume 16, p. 992296.

Anonymous. (2024). An approach through SciBERT-CNN with topic modeling. arXiv. https://arxiv.org/html/2404.13078v1

Anonymous. (2024–2025). Chunking and hybrid approaches (RoBERTa-BiLSTM, GRU-BERT, SciBERT-CNN) for scientific text classification. ACL Anthology. https://aclanthology.org/2024.semeval-1.231.pdf

Anonymous. (2025). Transfer learning: Adapting pre-trained models to new domains (survey). ResearchGate. https://www.researchgate.net/publication/390123031_Transfer_Learning_Adapting_Pre-Trained_Models_to_New_Domains

Asselborn, T. (2023). Fine-tuning BERT models on demand for information tasks. CEUR Workshop Proceedings, 3580, 5–16. https://ceur-ws.org/Vol3580/paper5.pdf

Chen, J., et al. (2024). Supplementing domain knowledge to BERT with semisupervised methods. ScienceDirect.

http://www.doi.org/10.62341/istj-vol38-2-na30

https://www.sciencedirect.com/science/article/abs/pii/S0957417423015567

Gonçalves, S. C. P. &. M. S., 2018. A deep learning approach for sentence classification of scientific abstracts. Rhodes, Greece, Springer International Publishing.

Huang, Y. (2023). Sentiment classification using bidirectional LSTM-SNP for specialized categories. ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0957417423002312

Jia, Y., et al. (2024). Biomedical relation extraction via ensemble + attention. PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11488084/

Jiang, S. H. J. M. C. L. L. J., 2022. Deep learning for technical document classification. IEEE Transactions on Engineering Management, Volume 71, pp. 1163-1179.

Kaggle, 2025. ArXiv Scientific Research Papers Dataset. [Online] Available at: https://www.kaggle.com/datasets/sumitm004/arxiv-scientificresearch-papers-dataset

Liu, C. (2024). Long short-term memory (LSTM)-based text classification across multiple domains. PLOS ONE, 19(3), e0301835. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0301835

Madan, S., et al. (2024). Transformer models in biomedicine: Survey and trends. PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11287876/

Pisu, A. (2024). Classifying Scientific Topic Relationships with SciBERT. CEUR Workshop Proceedings, 3759, 14–28. https://ceur-ws.org/Vol-3759/paper14.pdf

Qin, S., et al. (2024). Boosting generalization of fine-tuning BERT for diverse datasets. ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0306457324001055

Rahman, M. M. (2024). RoBERTa-BiLSTM: A Context-Aware Hybrid Model for Scientific Text Classification. arXiv. https://arxiv.org/abs/2406.00367

http://www.doi.org/10.62341/istj-vol38-2-na30

Rivest, M. V.-G. E. A. É., 2021. Level classification of scientific publications: A comparison of deep learning, direct citation and bibliographic coupling. PloS ONE, 16(5), p. e0251493.

Sammet, J. (2023). Domain-Specific Keyword Extraction using BERT. ACL Anthology. https://aclanthology.org/2023.ldk-1.72.pdf

Semberecki, P. M. H., 2017. Deep learning methods for subject text classification of articles. s.l., IEEE, pp. 357-360.

Sun, M. (2025). Transfer learning/fine-tuning strategies for domain NER and classification. arXiv. https://arxiv.org/pdf/2505.01868

Sun, Y., et al. (2024). Transfer learning for text classification via model risk analysis. ACL Anthology. https://aclanthology.org/2024.findings-emnlp.160/

Tinn, R., et al. (2023). Fine-tuning large neural language models for biomedical tasks. PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10140607/

Wang, Y., et al. (2024). TP-BERT: Topic-injected BERT for improved extractive summarization and classification. ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0306457324000372

Xu, K. (2025). Survey on terminology extraction from texts. Journal of Big Data, 12, 77. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-025- 01077-x

Zaman-Khan. (2024). Enhancing text classification using BERT: A transfer approach. SciELO. https://www.scielo.org.mx/scielo.php?pid=S1405-55462024000402279&script=sci_arttext

Zheng, Y., & Cai, R. (2023). Chunk-BERT: Boosted keyword extraction for long scientific literature via BERT with chunking capabilities. ResearchGate. https://www.researchgate.net/publication/376547142_Chunk BERT_Boosted_keyword_extraction_for_long_scientific_literature_via_BERT_ with_chunking_capabilities