

ISSN 1840-4855

e-ISSN 2233-0046

Original scientific article

<http://dx.doi.org/10.70102/afts.2026.1836.575>

CITATION IMPACT PREDICTION USING MULTI CHANNEL FUSION FRAMEWORK

Zahraa Yahya Mahdi Al-Mayali^{1*}, Salam Al-augby²

^{1*}Assistant Lecturer, College of Computer Science and Information Technology, Wasit University, Kut, Wasit Governorate, Iraq. e-mail: zmahadi@uowasit.edu.iq,
orcid: <https://orcid.org/0009-0007-0595-5948>

²Assistant Professor, Faculty of Computer Science and Mathematics, Kufa University, Kufa, Najaf Governorate, Iraq. e-mail: salam.alaugby@uokufa.edu.iq,
orcid: <https://orcid.org/0000-0001-8247-9497>

Received: May 25, 2026; Revised: July 06, 2026; Accepted: August 17, 2026; Published: September 29, 2026

SUMMARY

Citation impact prediction is complicated by the limitations of previous methods of prediction that are constrained to individual information channels such as textual features, citation networks, or bibliometrics. In this study, proposes a Multi-Channel Fusion Framework that leverages semantic, structural, and contextual information to predict citation impact. The proposed multi-channel model uses SciBERT-based text encodings obtained from titles, abstracts, and keywords, citation network-based graph embeddings for modeling relationships between publications, and features that describe author influence, journal quality, and source. An adaptive attention-based fusion architecture is used to learn the importance of each information channel prior to the classification of citation impact. Validate the framework using a Scopus dataset that consists of 11,543 Computer Science papers published from 2022 -2024, 145,329 reference relations, 48,957 authors, and 92,145 journal-level indicators. The experiment evaluation conducted with varying data split methods revealed that the proposed model scored an 84.33% accuracy and an 83.97% F1-score in comparison to the state-of-the-art methods for citation prediction. It is clear that the combination of textual semantics, citation graph, and bibliometric features gives a more complete representation of the research impact. The framework can help in identifying influential papers, evaluating research, and making decisions in academic settings. The future directions of research will consider multilingual data sets, citation graph development over time, explainable predictions, and interdisciplinary research.

Key words: *citation impact prediction, multi-channel fusion, SciBERT, graph neural networks, bibliometric analysis, attention-based learning, research impact assessment.*

INTRODUCTION

With the rapid rise in the publication rate of scientific papers, there have been significant changes in the areas of scholarly communication and research evaluation, thus bringing about problems in terms of determining the most influential scholarly outputs [2]. Even though the citation count of scientific documents is a commonly used measure to assess the influence of scholarly works, its use poses a problem since citations tend to accrue over time, making it impossible to determine the influence of newly published documents early enough [1][5]. This is an issue that may affect decisions related to research funding and institutional evaluation, among others.

Due to the rapid development of machine learning, natural language processing, and data mining technology, it has become possible for researchers to mine large-scale data sets of scholars and recognize some complicated patterns related to citation behavior [15]. So far, there are mainly three types of information that are considered in existing research. First of all, text-based models are employed to extract semantics through the title, abstract, and keywords in order to find out themes of research and new topics. However, only by using text information, it is difficult to find out social influence, collaboration patterns, and structural connections between science papers [8]. The typical prediction of citations relies on textual, bibliometric, or citation network attributes individually to predict the future impact of citations. The framework presented is different from the traditional method by integrating the three aspects together.

In addition to author reputation, journal prestige, institutional reputation, and other metadata variables, bibliographic metadata offers further relevant data that becomes available shortly after the paper's publication date. Yet, many current methods do not consider interactions between the various metadata features and use them independently of one another. Consequently, an effective framework that would predict citations through the combination of semantic, structural, and contextual information is required. Multi-channel fusion allows obtaining a holistic view of citations by taking into account text-based, graph-based, and metadata-based features [9].

The key contribution of the paper is the creation of a multi-modal fusion framework that simultaneously takes into account the semantics, structure, and context involved in citation impact. Unlike other models that use single features such as textual data, citation networks, or bibliometric measures, this framework utilizes representations of text through SciBERT, reference graphs with graph embeddings, and quality scores based on metadata. Additionally, an adaptive attention-based fusion approach has been employed to assign weights to different information modalities so as to make better use of diverse citations. Contributions of this paper include the following: (1) creation of an integrated citation prediction framework using textual, graph, and metadata representations; (2) creation of an adaptive fusion approach to learn the importance of modalities; and (3) evaluation under different data splitting schemes and comparison with existing citation prediction frameworks.

The outline of this paper is as follows: Section 1 presents background on the prediction of citation impact task and its research contributions. Section 2 gives an overview of existing methods for citation predictions. Section 3 describes the proposed multi-channel fusion methodology. Section 4 explains the experimental setup and evaluation. Section 5 discusses results, limitations, implications, and future work, whereas Section 6 summarizes the work.

RELATED WORK

Prediction of citation impact has grown to be a significant field of study in assessing scholarly influence and predicting the future prominence of scientific documents. Early research was concerned with citation metrics and bibliometric factors in order to understand the patterns of research impact, emphasizing the need for citation-based evaluation metrics in scientific assessment [1]. There have been several quantitative approaches used to investigate the patterns of citation growth and scientific prominence over time [2]. Factors influencing citation growth have also been explored by some researchers [3].

As artificial intelligence progresses, the application of text mining techniques has become popular for the extraction of semantically rich information from scientific texts [4]. The use of transformer-based language models for the improvement of scientific text representation through the capture of contextual relations has also been proposed [5]. Information retrieval approaches have also been used to find research patterns and document similarities [6]. Techniques for citation network analysis have highlighted the role of structural relations between scientific papers for the prediction of research influence [7]. Graph-based approaches have contributed to the area of citation modeling through relation learning [8].

Hybrid methods for early citation prediction by considering both semantic and metadata information have recently been proposed. Metadata-based approaches used author reputation, quality of journals,

and other features of publications to predict the citation potential [9]. Sentence-level embeddings were developed to measure the semantic similarity of research papers [10]. Graph attention mechanism was developed to discover heterogeneous relations within scholarly networks [11]. Also, ensemble approaches for learning from different prediction models have been studied [12]. Machine learning-based citation prediction models using boosting methods obtained high classification accuracy [13].

Multi-modal models with contextual features, adaptive attention, and fusion mechanisms have been developed in recent times to enhance the accuracy of predictions [14]. Recommendation and prediction models for citations have adopted transformers and graph-based representation schemes for better scientific investigation [15]. The effectiveness of multi-modal fusion methods has been demonstrated in utilizing heterogeneous sources for predictive tasks [16]. However, current solutions are challenged in the aspect of incorporating semantic, structural, and contextual citation factors effectively [17]. In this regard, this paper proposes an adaptive multi-channel fusion method that leverages text, graph, and metadata representations [20].

The prior work on citation prediction primarily concentrates on one of the above-mentioned aspects, such as semantic similarity, citation graph, or metadata. The metadata-based approach makes use of author, journal, affiliation, and source features, whereas the text-based approach uses semantic representation through the usage of SBERT, BM25, and similarity, which helps in capturing the content of the research. The graph-based approach considers the structural dependencies between publications. The classification approaches consider all the aforementioned heterogeneous features using different models such as Logistic Regression, Random Forest, and XGBoost; SHAP-based interpretation can be done to make the approach more interpretable.

Limitations of current citation prediction models include reliance on single data sources, incomplete use of textual, graphical, and metadata features, and inability to provide interpretability. They also fail to consider aspects like source quality, the reputation of authors, and trends. Hence, a need for a multi-modal approach for citation impact prediction exists.

METHODOLOGY

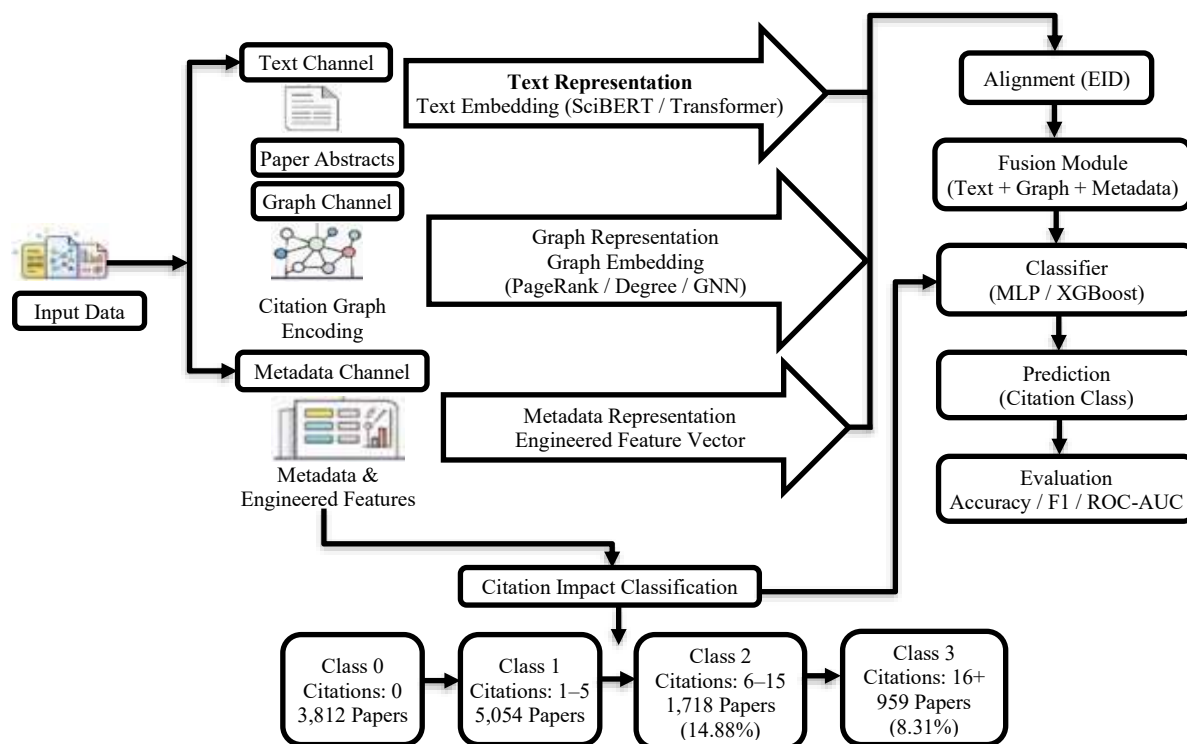


Figure 1. Proposed multi-channel fusion framework for citation impact prediction

This framework employs various academic databases as sources of data for building a complete citation prediction system, as shown in figure 1. The research papers are sourced from Scopus, quality indicators for journals from SCImago and CWTS, and the author's indicators like h-index from the OpenAlex database. All the above-mentioned databases are processed and transformed into a structured form. Features of the data can be derived in three different ways: through text features by analyzing the title, abstract, keywords, and their trends, graph features by analyzing the references between the documents, and metadata features like author reputation, journal quality, and source quality.

Dataset Collecting and Preparation

The dataset was created based on a systematic acquisition and preprocessing process for citation impact prediction. The bibliometric data were retrieved from the Scopus database for Computer Science articles between 2022 and 2024, with a resulting 11,543 scientific papers after excluding incomplete, duplicate, and off-topic documents. References were acquired to construct the source-level dataset, which contains 145,329 valid sources after filtering out the retracted sources. The author-level data were collected and enhanced by including the h-index from the Scopus and OpenAlex databases, whereas the SJR, CiteScore, and SNIP at the journal level were added from SCImago and CWTS. The final multi-layered dataset combines paper metadata, reference-level semantic information, author influence indicators, and journal quality measures, as summarized in table 1.

Table 1. Dataset component description

Dataset Component	Dataset Description	Dataset size
Paper-level dataset	Cleaned Computer Science papers from Scopus (2022–2024)	11,543
Reference-level dataset	Valid extracted references after excluding retracted references	145,329
Author-level dataset	Authors extracted as separate entries linked to papers	48,957
Journal metrics dataset	Journal-level indicators collected from SCImago, CWTS, and related metric files	92,145

Feature Engineering

Data wrangling involves feature engineering to structure the data into a representation that can be used by the proposed model pipeline. The proposed approach consists of four feature groups: text features, graph features, metadata features, and source equality features.

For each paper p_i , the final feature vector is defined as equation (1):

$$X_i = [T_i, G_i, M_i, Q_i] \quad (1)$$

Where:

- X_i : Final features vector
- T_i : Text features
- G_i : Graph features
- M_i : Metadata features
- Q_i : Source quality features

What's more, derived features are also added to improve the prediction. The features were derived from the underlying data and designed to capture corresponding features across the text, graph, and metadata representation of the data, leading to better generalization or prediction of the impact of a citation.

Proposed Approach

The proposed citation prediction system was designed as a multi-channel system with the following three complementary sources: the textual content of the paper, the graph structure derived from the references, and a metadata representation that includes author, journal, affiliation, open access, source quality, and topic-related signals. These channels are then treated separately and subsequently combined

using an integration model based on a transformer for the final prediction of the citation class. In the dataset, there are N papers, with each paper p_i being represented as $P_i = \{T_i, G_i, M_i, y_i\}$ where:

1. T_i - Text features
2. G_i - Graph features
3. M_i - Metadata features
4. $Y_i \in \{0, 1, \dots, C-1\}$ is the citation class label

The overall model end-to-end learns a mapping function $F: (T_i, G_i, M_i, y_i)$, which is realized by a late fusion transformer-based classifier. The following sections outline the proposed approach, methodology, and details of the core parts thereof.

Text Channel

The text channel extracts semantic information from research papers by combining titles, abstracts, and keywords into a unified representation. The processed text is encoded using SciBERT to generate contextual token embeddings. Instead of relying on the [CLS] token, mean pooling is applied to obtain a stable document-level representation (Equation 2):

$$t_i = \frac{\sum_{j=1}^L m_{ij} h_{ij}}{\sum_{j=1}^L m_{ij}} \quad (2)$$

Where h_{ij} represents token embeddings and m_{ij} denotes the attention mask, the generated embeddings are normalized using L2 normalization and optionally reduced through incremental PCA to decrease computational complexity. These semantic representations are further used for topic modeling and trend analysis to identify research patterns influencing citation impact prediction. The preprocessed text consists of the text after lowercasing, noise reduction, removal of repetitive keywords, and space correction to get an equal representation. Then, the tokenized text is created from the processed text using SciBERT with a maximum sequence length of 256 tokens, and the embeddings are calculated using mean-pooling and L2 normalization. These embeddings contain semantic connections, research topics, and new topic trends that can be used as signals for the classification of citation impacts. The use of text embeddings in the fusion process, together with graph and metadata information, increases the accuracy of citation impact prediction.

Algorithm 1. Text Channel

Input:

- Paper-level dataset

Output:

- Text embedding matrix X_{text} , hot topic metrics.

Step 1: Text construction

- For each paper, I , the text embedding build

Step 2: Text cleaning

- Lowercase the text.
 - Remove noise.
-

- Extra spaces are removed.
- Remove repeated keywords.

Step 3: Tokenization

- For each constructed text embedding, T_i is tokenized using the SciBERT transformer with a maximum sequence length of 256 tokens.

Step 4: Mean pooling and normalization

- Mean pooling is applied.
- L2 normalization applied to text embeddings.

Step 5: Dimensionality reduction

- Incremental PCA was applied to reduce the dimension to 256

Step 6: Save output metrics

- Text embedding matrix X_{text} into a separate file

This text feature extraction algorithm 1 is applied to paper-level textual data in order to create semantic representations for citation prediction. At first, the title, abstract, and keywords are all used in order to make a unified representation of the text. This text is cleansed of noises, white spaces, and duplicates in keywords. Then, the text is tokenized by using SciBERT, where the max sequence length is set to 256. Contextual embeddings are created by using the mean pooling method, while L2 normalization is used in order to normalize those embeddings. Dimensionality of the embeddings is reduced to 256 by the incremental PCA method.

Graph Channel

The graph channel captures the structural linkages between research articles using a reference-induced similarity graph, which does not depend solely on citation relations. The references are captured using a combination of their titles and abstracts (Equation 3):

$$R_j = \text{Concat}(Title_j^{ref}, Abstract_j^{ref}) \quad (3)$$

These references are mapped into numerical values by making use of features of TF-IDF with unigrams and bigrams. Truncated SVD technique is used to make reference embeddings compact (Equation 4):

$$e_j^{ref} = \text{SVD}(r_j), e_j^{ref} \in R^{128} \quad (4)$$

Paper-level structural features are formed based on the embeddings of the references associated with the respective paper. A k-nearest neighbor (kNN) graph is built by computing the cosine similarity between the papers (Equation 5):

$$N_i = kNN(i, K) \quad (5)$$

This graph captures hidden similarities between publications even when direct citation information is unavailable. To learn high-level structural patterns, a Graph Transformer based on TransformerConv layers is applied (Equations 6 and 7):

$$H^{(1)} = \text{TransformerConv}(X, A) \quad (6)$$

$$H^{(2)} = \text{TransformerConv}(H^{(1)}, A) \quad (7)$$

The final graph representation of each paper is obtained as (Equation 8):

$$G_i = H_i^{(2)} \in R^{128} \quad (8)$$

These graph representations encode both local and global dependencies between the scientific papers and are given to the fusion block for predicting citation impact. A linear classifier with the Softmax function is employed as a predictor and cross-entropy as the objective function. The temporal split is performed for assessing realistic future citation prediction ability.

Metadata Channel

This particular channel contains the bibliometric, contextual, and quality information related to the research article using the author, journal, affiliation, reference quality, availability, and topic trends information. In contrast to the other two channels (text channel and graph channel), it considers the external variables affecting citation behavior. The author's prestige is measured using h-index values (mean and maximum) using log transformation (Equation 9):

$$\log(h_i^{\text{mean}} + 1), \log(h_i^{\text{max}} + 1) \quad (9)$$

where h_i represents the h-index information of authors associated with paper i . Journal influence is measured using quality indicators such as SJR, SNIP, CiteScore, and journal h-index, transformed as (Equation 10):

$$\log(\text{Metric}_i + 1) \quad (10)$$

Where Metric_i denotes journal-based indicators, the reputation of the affiliation is captured based on institutional productivity, author distribution, and collaboration. In order to capture the interactions between metadata attributes, interaction attributes have been used. For instance, the combination of author reputation and journal reputation is denoted as (Equation 11):

$$\text{Int}_1 = \log(h_i^{\text{mean}} + 1) \times \log(\text{SJR}_i + 1) \quad (11)$$

There are other interaction terms such as the interaction between topic popularity and journal quality, source quality and journal prestige, institution reputation and journal quality, and open access and topic trends. Interaction terms help the model understand that the impact of citation is affected not just by the variables individually but also through their interactions.

Fusion Block with Adaptive Gating

The fusion block combines the textual, graphical, and metadata features using a hierarchical attention-based learning scheme. Instead of basic concatenation schemes, it applies feature projection, normalization, weight adaptation, and cross-modality interaction to produce the final representation. The three synchronized feature representations are projected into a common dimensional space and augmented through non-linear functions. A gating scheme is used to assign dynamic importance weights to each modality based on the paper under consideration. This way, the model is able to focus more on reliable data and downweight less relevant features. The fused features are mapped to modality-specific tokens containing position information, whereas channel dropout increases robustness (Equation 12). The chosen feature engineering methods extract relevant citation features: textual semantic similarity, graph structural connections, and metadata-based influence. The chosen attention-based merging technique is better than the basic concatenation because it can automatically learn the importance of each modality.

$$\text{Attention}(X) = \text{Softmax}\left(\frac{(XW_Q)(XW_K)^T}{\sqrt{d_{\text{token}}}}\right)XW_V \quad (12)$$

Where X represents the input transformer representation, W_Q , W_k , and W_v represent internal attention weight matrices, and d_{token} is the size of each embedding after projection.

This allows the model to learn relationships such as text as citation structure and metadata for impact patterns.

EXPERIMENTAL RESULT

Experiment Environment and Data Initialization

The citation prediction system was constructed using Google Colab since it helped in dealing with computationally heavy transformer representations and feature extraction from graphs. All the data management was done using Google Drive, and the installation of dependencies was done without any change in core libraries. To make the system reproducible, random seed fixation and automatic CPU/GPU execution were enabled. The validation process for dataset access was done using file validation, encoding validation, and delimiter validation.

Dataset and Data Splitting

The dataset was split into three sets for training, validation, and testing in order to guarantee the correctness of the estimation process and avoid any data leakage. A stratified split with an 80:10:10 ratio was conducted in order to achieve the balance of citation classes in all splits. Also, temporal splitting was carried out, where the models were trained using older publications while being evaluated on newer ones in order to mimic the realistic environment of citation predictions. Another experiment excluded all 2024 publications during training in order to see how well the model could predict future citations (Table 2).

Table 2. Training-Testing size for each scenario

Split Type	Training Data	Validation Data	Testing Data	Description
Scenario 1	Random stratified without 2024	80% (5670) of records	10% (709) of records	10% (709) of records
Scenario 2	Random stratified with 2024	80% (8826) of records	10% (1104) of records	10% (1104) of records
Scenario 3	Temporal Split	80% (8826) of records	10% (1104) of records are only from 2024	10% (1104) of records are only from 2024

Training Scenarios

To make a fair and strong evaluation, many experimental scenarios were designed according to various data splitting approaches of the collected data. The following evaluation scenarios were considered:

1. Scenario 1 (Without 2024): Stratified random split including (2022,2023) with ensuring balanced class distribution, to simulate the real world for the prediction conditions.
2. Scenario 2 (With 2024): Stratified random split with balanced class distribution and all years (2022, 2023, and 2024).
3. Training on earlier years (2022, 2023) and testing on later years (2024) to assess temporal generalization (Scenario 3: Temporal Split).

Performance of the model under multiple scenarios reveals how well the model can cope with various cases of data splitting. Random stratified splitting without taking into account 2024 publications resulted in the best accuracy – 0.9676 because of equal temporal distribution in both train and test datasets. Adding 2024 publications led to decreased accuracy – 0.8432, meaning that the variance in data is larger. Temporal splitting showed decreased performance since the model was evaluated on the future publications in table 3.

Table 3. Metric values for each scenario

Scenario	Accuracy	Macro F1	Train Size	Val Size	Train Year	Val Year
Random Stratified (Without 2024)	0.9676	0.9676	5670	1418	2022, 2023	2022, 2023
Random Stratified (with 2024)	0.843226	0.848369	8826	2207	2022, 2023, 2024	2022, 2023, 2024
Mixed Temporal	0.799819	0.751718	8826	2207	2022, 2023, 2024	2024

Channel-Wise Performance Analysis

Evaluation experiments were run with each channel alone to assess the contribution of each channel. The results show a significant difference between the performance of various feature channels. The metadata channel performed best, with an Accuracy of 0.5359 and a Macro-F1 score of 0.3869, suggesting that indicators derived from the metadata of an article – such as author reputation, journal prestige, source quality – are good predictors of the impact of the article on citations. As can be seen in table 4, the text channel showed a moderate Macro-F1 performance of 0.2497 when compared with the other channels, which indicates that the text channel alone is not enough to capture the behavior of citations. Although the accuracy of the graph channel is lower, its Macro-F1 score is slightly higher than that of the text channel, demonstrating its effectiveness in dealing with class imbalance in the structural relationships. Lastly, the results here confirm that the most influential features are the metadata features, followed by graph features, which help to achieve balanced classification, and text features, which have limited predictive power on their own.

Table 4. Channels performance

Channel	Accuracy	BAcc	Macro-F1	Micro-F1
Random Stratified (Without 2024)	0.9676	0.9676	5670	1418
Random Stratified (With 2024)	0.843226	0.848369	8826	2207
Mixed Temporal	0.799819	0.751718	8826	2207

Fusion Strategy Comparison

Full Fusion and Pairwise

1. A variety of fusion configurations were tested, including single-channel models, pairwise combinations, and full multi-channel fusion to assess the effectiveness of combining the different feature channels. To facilitate a fair comparison, all experiments were carried out with the same experimental set-up, data split, and evaluation protocol. The results of each configuration are summarized in table 5. Table 5 shows that there are some interesting points to note:
2. According to the results, the metadata-only model yields decent performance (accuracy: 0.6413 and Macro-F1: 0.6350), indicating that the signals provided by bibliometric features like citation statistics, journal metrics, and attributes related to the authors are very informative with respect to citation prediction.
3. The two models with only text or with only graph (a text-only model and a graph-only model) show less success, indicating that neither one nor the other was successful in capturing all the aspects of the problem.
4. Compared to single-channel models, the pairwise fusion models consistently improve performance, especially.

Text + metadata (Macro F1 = 0.6632)

Graph + metadata (Macro F1 = 0.8305)

This confirms that combining information sources enhances predictive capability.

1. The full version model achieves the best overall performance, particularly when using the proposed attention mechanism (Macro-F1= 0.8397). This demonstrates that integrating semantic, structural, and bibliometric features provides the most comprehensive representation of research impact.

Table 5. Comparison of fusion model performance

Model Configuration	Accuracy	Macro-F1
Text Only	0.3995	0.3599
Graph Only	0.5663	0.5456
Metadata Only	0.6413	0.6350
Text + Graph	0.6545	0.6348
Text + Metadata	0.6723	0.6632
Graph + Metadata	0.8369	0.8305
Full Fusion (Concatenation)	0.8297	0.8315
Full Fusion (Attention-Based)	0.8433	0.8397

Attention Fusion vs. Concatenation

For further analysis of the impact of the fusion mechanism, a comparison was conducted between simple concatenation and the proposed model attention fusion approach. The results in table 6 clearly demonstrate that the proposed attention fusion model significantly outperforms simple concatenation, which treats all the channels equally; the attention mechanism assigns importance weights to each channel. This allows the model to adaptively emphasize more informative features depending on the input instance. In this approach, the attention mechanism effectively leveraged the strong predictive power of metadata while still incorporating complementary information from text and graph channels.

Table 6. Attention fusion vs. concatenation

Fusion Method	Accuracy	Macro-F1
Simple Concatenation	0.8297	0.8315
Attention-Based Fusion (Proposed)	0.8433	0.8397

Features Engineering Comparisons

Full Model vs. No Smart Features Addition

The effectiveness of the proposed enriched feature set was investigated through the performance comparison of feature engineering approaches. The base model made use of raw Scopus metadata with text and graph features while ignoring other attributes such as the author's reputation, the journal's prestige, the source's quality, the topics' trends, and feature interaction. However, the proposed model included all the features engineered from text, graph, and metadata channels. Table 7 proved that there was a huge difference in performance, in which the proposed model had an accuracy of 0.9676 and a Macro F1 of 0.9647, while the base model had an accuracy of 0.3795 and a Macro F1 of 0.3609.

Table 7. Comparison of proposed framework and no smart features

Model Type	Accuracy	Macro-F1
Proposed Framework	0.9676	0.9647
No Smart Features (Scopus Raw Metadata)	0.3795	0.3609

Evaluation with the Traditional Machine Learning and Future Cite Baseline [17]

The proposed model was compared to other machine learning methods and Future Cite baseline models to determine its performance. FutureCite is the combination of classical machine learning methods like Naïve Bayes, Random Forest, and XGBoost with features based on text and graphs. While it represents a good traditional baseline method, its performance is limited, with the accuracy and Macro-F1 values of 0.7481 and 0.2764, respectively. The low Macro-F1 value shows that the method faces problems with unbalanced classes in the task and generating balanced predictions for different citation impact classes.

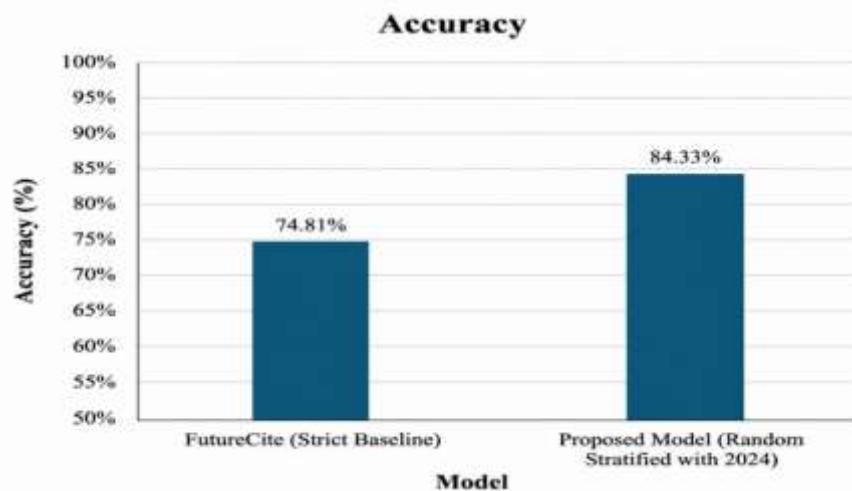


Figure 2. Accuracy comparison between futurecite baseline and proposed citation prediction framework

Figure 2 shows the baseline FutureCite by a margin of accuracy of 0.8432 and a Micro-F1 score of 0.8484 with the same evaluation settings. Improvement in the Macro-F1 score demonstrates that the new model performed better in the balance between the categories of citation impact. This was possible due to the use of enhanced features of text representation, graphs, and metadata to capture semantics, structure, and context. In addition, the 80/10/10 split performed the best with an accuracy of 96.76% and a Macro-F1 score of 96.47%. This was because of the greater availability of data for training.

Table 8 provides a comparative study between the suggested framework for multi-channel fusion and the existing citation prediction models, which make use of textual, graph, and metadata information. The differences in the features used, the learning methods adopted, and the prediction performances in terms of accuracy and F1-scores have been shown.

Table 8. Comparison with recent citation prediction methods

Method	Feature Representation	Learning Model	Accuracy (%)	F1-Score (%)
SChuBERT [18]	Scholarly text representation	BERT-based citation count prediction	72.40	68.50
Citation Trend Prediction using GNN [19]	Citation graph features	Graph Neural Network	81.30	77.90
Instant Citation Potential Prediction [20]	Metadata + semantic features	Hybrid prediction model	82.10	79.40
FutureCite [17]	Text + graph mining features	Machine learning ensemble model	74.81	27.64
Proposed Multi-Channel Fusion Framework	Text + Graph + Metadata	Adaptive Attention Fusion Transformer	84.33	83.97



Figure 3. Accuracy comparison under different data splitting strategies

As seen in figure 3, the accuracy for the split (70/15/15) was slightly lower at 0.9577, and the Macro F1 score was 0.9518. The performance measures are slightly lower, but this configuration gives a more balanced and statistically sound assessment because larger validation and testing sets give a more stable performance measure. A structured partitioning approach is also applied for those who wish to validate and test the implementation, so that proper representation of validation and test data is achieved.

DISCUSSION

The multi-channel fusion approach introduced in this paper proves that the use of multiple features, including semantic, structural, and contextual features, gives a better representation of research impact than the traditional single-channel citation prediction methods. Based on experimental results, the adaptive attention-based fusion enhances the performance of the prediction through dynamic assignment of weights to the text, graph, and metadata features. The framework shows an accuracy and Macro-F1 score of 84.33% and 83.97%, respectively, in the comparative experiment. This work, however, has some drawbacks. For instance, the data used only involves Computer Science articles from the Scopus database for the years 2022 to 2024. Furthermore, there are many external factors that affect citation processes, and they include societal impact, collaboration, and research policies, among others.

This method also has practical applications in evaluating research papers, allocating resources, and predicting potentially impactful papers. Through the creation of a holistic system of predictions based on text, citations, and bibliometrics, this method could help universities and scholars make strategic decisions. Further development of the framework could take into account multilingual papers, citation dynamics, open access indicators, and interdisciplinary data sources. Another possible improvement is the use of explainable AI to understand attention weights and determine what affects citation impact prediction.

CONCLUSION

The present study offered a Multi-Channel Fusion Framework for citation impact prediction using a combination of semantics, structure, and context from scientific papers. The main idea was to address the drawbacks of existing citation prediction methods, which make use of individual textual, graph, or bibliometrics-based features. The suggested framework used SciBERT for generating textual features, generated reference-based graph features, and used author reputation, journal quality, and source metrics as metadata-based features through an attention-based fusion framework. The experiment was carried out using a Scopus dataset of 11,543 Computer Science papers in the period of 2022-2024, which had 145,329 references, 48,957 authors, and 92,145 journal metrics.

The experimental results show the validity of the proposed method for citation impact estimation. The accuracy reached by the suggested model in the case of random stratified evaluation was 96.76%, and the value of the Macro-F1 score was 96.47%, whereas the comparison with recent works related to citation prediction resulted in values of accuracy and Macro-F1 score equal to 84.33% and 83.97%, respectively. In particular, the attention-based feature fusion mechanism proved to be better than simple concatenation of the features since the accuracy increased from 82.97% to 84.33% and Macro-F1 from 83.15% to 83.97%. Moreover, feature engineering analysis proved that joint use of text, graph, and metadata features is better than usage of metadata features only since the accuracy of the proposed framework reached 96.76% in contrast to 37.95% of the baseline model.

The suggested framework offers tremendous benefits to universities, grant-making organizations, and researchers by ensuring early detection of influential papers and facilitating evidence-based research evaluation. But the present study has used Computer Science papers from the Scopus database only and has not taken into account the external variables like collaboration, sociological, and policy impact. The future research will be centered on extending the framework by using multilingual and interdisciplinary databases, taking into account the citation dynamics over time, and using explainable artificial intelligence methods.

Declaration Statement

Funding: This research received no external funding

Data Availability Statement: In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. You might choose to exclude this statement if the study did not report any data.

Conflicts of Interest: “The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.”

Acknowledgment: This research was not funded by any grant

REFERENCES

- [1] Waltman L. A review of the literature on citation impact indicators. *Journal of informetrics*. 2016 May 1;10(2):365-91. <https://doi.org/10.1016/j.joi.2016.02.007>
- [2] Wang D, Song C, Barabási AL. Quantifying long-term scientific impact. *Science*. 2013 Oct 4;342(6154):127-32. <https://doi.org/10.1126/science.1237825>
- [3] Tahamtan I, Safipour Afshar A, Ahamdzadeh K. Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics*. 2016 Jun;107(3):1195-225. <https://doi.org/10.1007/s11192-016-1889-2>
- [4] Aggarwal CC, Zhai C. An introduction to text mining. In *Mining text data* 2012 Jan 7 (pp. 1-10). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4614-3223-4_1
- [5] Beltagy I, Lo K, Cohan A. SciBERT: A pretrained language model for scientific text. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP) 2019 Nov* (pp. 3615-3620). <https://doi.org/10.18653/v1/D19-1371>
- [6] Chen P, Xie H, Maslov S, Redner S. Finding scientific gems with Google’s PageRank algorithm. *Journal of informetrics*. 2007 Jan 1;1(1):8-15. <https://doi.org/10.1016/j.joi.2006.06.001>
- [7] Castelino RM, Pradhan S, Hahn A. Track-to-Track Fusion for Cooperative Perception Using Collective Perception Messages. *Sensors*. 2026 Mar 23;26(6):2003. <https://doi.org/10.3390/s26062003>
- [8] Onan A. Hierarchical graph-based text classification framework with contextual node embedding and BERT-based dynamic fusion. *Journal of King Saud University Computer and Information Sciences*. 2023 Jun 13;35(7):101610. <https://doi.org/10.1016/j.jksuci.2023.101610>
- [9] Dinh TN, Pham P, Nguyen GL, Nguyen NT, Vo B. A hybrid citation recommendation model with SciBERT and GraphSAGE. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. 2024 Nov 21;55(2):852-63. <https://doi.org/10.1109/TSMC.2024.3490774>
- [10] Wang X, Ji H, Shi C, Wang B, Ye Y, Cui P, Yu PS. Heterogeneous graph attention network. In *The world wide web conference 2019 May 13* (pp. 2022-2032). <https://doi.org/10.1145/3308558.3313562>
- [11] Dietterich TG. Ensemble methods in machine learning. In *International workshop on multiple classifier systems 2000 Jun 21* (pp. 1-15). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-45014-9_1
- [12] Reimers N, Gurevych I. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP) 2019 Nov* (pp. 3982-3992). <https://doi.org/10.18653/v1/D19-1410>
- [13] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 2016 Aug 13* (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>
- [14] Tian G, Zhang C, Shi Y, Li X. MultiWaveNet: A long time series forecasting framework based on multi-scale analysis and multi-channel feature fusion. *Expert Systems with Applications*. 2024 Oct 1;251:124088. <https://doi.org/10.1016/j.eswa.2024.124088>
- [15] Wang Q, Li H, Ma M, Li Z. A context-aware enhanced local citation recommendation model integrating SciBERT and self-adaptive attention. *Scientometrics*. 2025 Aug;130(8):4495-517. <https://doi.org/10.1007/s11192-025-05382-3>
- [16] Xu G, Ren M, Wang Z, Li G. MEMF: Multi-entity multimodal fusion framework for sales prediction in live streaming commerce. *Decision Support Systems*. 2024 Sep 1;184:114277. <https://doi.org/10.1016/j.dss.2024.114277>

- [17] Thafar MA, Alsulami MM, Albaradei S. Futurecite: predicting research articles' impact using machine learning and text and graph mining techniques. *Mathematical and Computational Applications*. 2024 Jul 21;29(4):59. <https://doi.org/10.3390/mca29040059>
- [18] Van Dongen T, de Buy Wenniger GM, Schomaker L. SchuBERT: Scholarly Document Chunks with BERT-encoding boost Citation Count Prediction. In *Proceedings of the First Workshop on Scholarly Document Processing 2020 Nov* (pp. 148-157). <https://doi.org/10.18653/v1/2020.sdp-1.17>
- [19] Cummings D, Nassar M. Structured citation trend prediction using graph neural networks. In *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP) 2020 May 4* (pp. 3897-3901). IEEE. <https://doi.org/10.1109/ICASSP40776.2020.9054769>
- [20] Zhu H, Shuhuai L. Instant prediction of scientific paper cited potential based on semantic and metadata features: Taking artificial intelligence field as an example. *PLOS one*. 2024 Dec 2;19(12):e0312945. <https://doi.org/10.1371/journal.pone.0312945>