IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 3 August 2023, accepted 18 September 2023, date of publication 27 September 2023, date of current version 19 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3320038

RESEARCH ARTICLE

Exploiting Contextual Word Embedding for Identification of Important Citations: Incorporating Section-Wise Citation Counts and Metadata Features

ARSHAD IQBAL[1], ABDUL SHAHID[1], MUHAMMAD ROMAN[1], MUHAMMAD TANVIR AFZAL[2], AND MUHAMMAD YAHYA[3]

[1]Institute of Computing, Kohat University of Science and Technology, Kohat 26000, Pakistan
[2]Department of Computing, Faculty of Computing, Shifa Tameer-e-Millat University Islamabad 44000, Pakistan
[3]Data Science Institute, University of Galway, Galway, H91 TK33 Ireland

Corresponding author: Muhammad Yahya (M.Yahya1@nuigalway.ie)

ABSTRACT Finding relevant research papers can be a challenging task due to the enormous number of scientific publications released each year. Recently, the scientific community has been diving deep into citation analysis, specifically examining the content of papers to identify more crucial documents. Citations serve as potential parameters for establishing connections between research articles. They have been widely utilized for various academic purposes, including calculating journal impact factors, determining researchers' h-index, allocating research grants, and pinpointing the latest research trends. However, researchers have argued that not all citations carry equal weight in terms of influence. Consequently, alternative techniques have been proposed to identify significant citations based on content, metadata, and bibliographic information. Nonetheless, the current state-of-the-art approaches still require further refinement. Additionally, the application of deep learning models and word embedding techniques in this context has not been extensively studied.In this research work, we present an approach consisting of two primary modules: 1) Section-wise citation count, and 2) metadata-based analysis of citation intent. Our study involves conducting several experiments using deep learning models in conjunction with FastText, word2vec, and BERT-based word embeddings to perform citation analysis. These experiments were carried out using two benchmark datasets, and the results were compared with a contemporary study that employed a rich set of content-based features for classification. Our findings reveal that the deep learning CNN model, coupled with FastText word embeddings, achieves the best results in terms of accuracy, precision, and recall. It outperforms the existing state-of-the-art model, achieving a precision score of 0.97

INDEX TERMS Citation context analysis, metadata Information retrieval, convolutional neural networks (CNNs) deep learning, convolutional neural network, gated recurrent units (GRUs), word embedding, scholarly dataset, importance citation identification, binary classification.

I. INTRODUCTION

Finding relevant research papers poses a significant challenge due to the vast number of scientific publications released annually. In the realm of computer science alone, approximately 100,000 articles are published each year [1]. Similar trends are observed across other disciplines. Notably, PubMed, a medical database, reported that the volume of research papers published in 2014 (514,395) was more than double that of 1990 (136,545) and over 100 times greater than that of 1950 (4,432) [2]. Consequently, the phenomenon of information overload, often referred to as a "tsunami of papers," has made it increasingly arduous to identify pertinent publications. Throughout recent decades,

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Li.

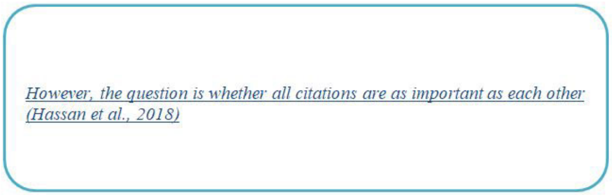
However, the question is whether all citations are as important as each other (Hassan et al., 2018)

FIGURE 1. Example of citation in a research article.

diverse methodologies have been employed to locate significant papers, encompassing content-based [3], [4], [5], [6], collaborative filtering [7], [8], co-citation [9], graph-based [10], [11], bibliographic analysis [12], and hybrid algorithms [13], [14]. Each approach evaluates the importance of papers from a distinct standpoint. Co-citation, for instance, posits that two research publications are noteworthy if they have been cited by one or more common citing papers, without explicitly addressing the citation rationale and purpose [9]. Conversely, bibliographic coupling determines the significance of two works based on their shared references [12]. However, traditional bibliographic coupling disregards the citation patterns of comparable references across logical sections of the citing publications. Collaborative filtering, even in the absence of content-based features, relies on information about users' interests, which proves valuable in recommending comparable articles but is susceptible to the "cold star problem" [4]. Content-based techniques utilize the article content to compute similarities between papers; however, the accessibility of such content is not always freely available [13]. The term "hybrid" denotes a technique that combines collaborative filtering with content-based strategies.

Citation-based strategies represent a significant approach for identifying relevant documents in academic research. Citations hold immense value in the scientific community and are increasingly utilized as indicators of research performance [14]. They play a unique role in scientific discoveries, understanding scholarly work [14], and calculating various academic elements such as impact, grant allocation, institutional listing/ranking, peer evaluations, and impact factor [8], [15], [16]. Previous studies have explored the multitude of reasons behind citing a study, including the utilization of past information, result enhancement, result comparison, and more [17], [18]. However, most citation analysis techniques assign equal weight to all citations without considering the diverse reasons behind their usage [19]. We argue that the quality of citations should not solely rely on their quantity, as negative citations, self-citations, and citations from colleagues can skew the evaluation process. Thus, it is crucial to consider citation intent throughout the evaluation process.

Researchers have proposed various models and strategies for categorizing citations based on their intent. Initially, manual classification through interviews with authors was employed [20]. However, due to limitations, this approach was replaced by automated classification methods. These automated techniques categorized citations into different classes based on their reasons, including citation context [21], citation count [22], citation sentiment [23], and in-text citation frequency [24]. Jurgens et al. [25] argued for six distinct categories of citation reasons, each with varying degrees of relevance. Over time, these multiple categories were gradually reduced to binary classes.

Metadata plays a significant role in establishing the relationship between cited and citing papers. Research article metadata, including title, abstract, keywords, and author list, provides a concise and descriptive summary of the entire research paper [17]. In situations where journal publishers restrict open access to their papers, leveraging openly available resources such as metadata becomes an alternative [18].

Word embedding has emerged as a crucial technique in the field of natural language processing (NLP) and text analysis. It provides a powerful representation of words in a continuous vector space, capturing semantic relationships and contextual information. Notably, word embedding models like Word2Vec, FastText, and BERT have significantly enhanced various NLP tasks, including citation intent classification and information retrieval [26]. The effectiveness of word embedding techniques has been extensively documented in the literature [26], [27], [28].In recent years, deep learning has significantly advanced research in the domain of natural language processing. Among the multitude of deep learning methods explored in this field, Convolutional Neural Networks (CNNs), Gated Recurrent Units (GRUs), and Long Short-Term Memory networks (LSTMs) have garnered substantial attention [29]. However, it is important to note that despite the extensive exploration of deep learning models for text classification, their effectiveness remains a subject of concern.In the specific domain of citation analysis and identification of important citations, the utilization of word embedding techniques in conjunction with deep learning models has remained relatively limited. This research aims to bridge this gap by exploring the potential of contextual word embedding models, such as BERT, fasttext, and Word2Vec, for capturing the significance of citations in scholarly documents. This study leverages the deep learning approach and existing state-of-the-art models for the automatic citation classification and makes the following key contributions

- A framework is introduced for citation analysis utilizing deep learning models (CNN, GRU, LSTM) combined with a word embedding approach. Additionally, the efficacy of FastText, Word2Vec, and BERT for word representation is individually analyzed to identify "important" citations.
- The proposed framework is evaluated by comparing its results with state-of-the-art models commonly used for citation classification.
- Furthermore, this study incorporates freely available metadata such as the title and abstract of the research paper, along with section-wise citation count features,

to enhance the accuracy and precision of identifying important citations
- This research provides unique insights into the application of contextual word embedding for identifying important citations in academic literature

The evaluation results of the proposed approach demonstrate a significant improvement of in precision for binary classification. The proposed technique outperforms state-of-the-art methods, highlighting its effectiveness in accurately identifying important and non-important citations. Based on these results, This paper establishes the claim that the proposed technique achieves a higher level of accuracy and precision in the task of important citation identification compared to existing approaches.

II. RELATED WORK

Several studies have been conducted in the field of citation analysis and identification of important citations. These works have explored various approaches and techniques to tackle the challenges associated with this area of research.

Researchers have expressed the viewpoint that each citation made by a researcher serves distinct purposes [30]. Therefore, treating all citations as equal is deemed ineffective [31]. Garfield et al. [32] was the pioneer in differentiating citations by studying researchers' motivations behind citing. He identified 15 categories for citation motivations. Following this, Moravcsik and Murugesan [33] proposed a technique to classify citations into four categories based on their functions. Using an automatic classification technique, Garzone and Mercer [34] expanded the classification to 35 different types. Teufel et al. [35] introduced a supervised machine-learning approach to categorize citations into four categories and 11 subcategories. Agarwal et al. [36] developed classifiers using support vector machine (SVM) and Multinomial Naïve Bayes (MNB) models for citation classification. Jürgen et al. [25] devised a machine-learning scheme to annotate citations into seven categories. Hamedani et al. [37] classified citations into six classes by analyzing keywords within citations. Kulkarni [38] proposed an automated citation extraction and analysis system that utilizes machine learning algorithms to identify and classify citation contexts based on their functions and intents. Bakhti et al. [39] proposed a classification model that combines ontology and convolutional neural network (CNN) for categorizing citations into six categories. Researchers have also delved into the use of citation patterns and co-citation analysis to uncover connections and relationships among scholarly works. Co-citation analysis, as proposed by Small [9] involves examining the frequency with which two works are cited together by other articles, revealing the intellectual associations between them. This technique helps identify influential works and provides insights into the intellectual structure of a research field. Smith et al. [40] investigated the impact of citation characteristics on the credibility and reliability of scientific papers. They analyzed factors such as citation age, self-citations, and citation patterns to determine their influence on the perceived quality of research. Their findings highlighted the importance of considering not only the quantity but also the quality and context of citations in evaluating the credibility of scholarly articles.

A notable line of research focuses on citation intent classification, where machine learning and natural language processing techniques are employed to categorize citations based on their purpose for being included in research papers. These approaches consider features such as citation context, sentiment, and in-text citation frequency [17], [18], [19]. Citation frequency and author overlap are also important features in identifying important citations [19]. Incorporating citation context information has been shown to improve classification performance [41], [42].The role of similarity between titles and abstracts has been discussed in measuring the value of citations [17], [18], [31]. Additionally, factors such as clue words [18], [35], keywords [43], and the time characteristics of citations are considered. Moreover, incorporating citation context information along with section-wise citation count features has shown promising results [22]. Furthermore, the use of citation sentiment analysis has gained attention as a means to assess the emotional tone and attitude expressed towards cited works. For example, Yang and Chang [44] employed sentiment analysis techniques to classify citation contexts as positive, negative, or neutral, providing insights into the evaluative nature of citations and their impact on the perceived value of research. Aljuaid et al. [20] proposed a content-based approach for binary classification of citations by analyzing in-text citation sentiments. Nazir et al. [45] presented a hybrid approach that combines quantitative methods and citation sentiment analysis to identify credible citations.

In the realm of word embedding and deep learning, several studies have demonstrated the effectiveness of these approaches in various natural language processing tasks. Word embedding models such as Word2Vec, fasttext, and BERT have been widely used to capture semantic relationships and contextual information in text [26], [28] utilized Word2Vec, a popular word embedding model, to explore the semantic similarity between cited and citing papers. BERT (Bidirectional Encoder Representations from Transformers), a powerful language model, has been widely adopted in citation analysis tasks. Researchers have fine-tuned BERT models on citation-related datasets to extract features and capture the contextual information within citation contexts [46]. Roman et al. [26] applied contextualized word embedding techniques (BERT) to convert text into numerical representations, resulting in a better understanding of the text's purpose and the reasons for citation. Pre-trained word embeddings such as FastText have been utilized to analyze citation contexts and identify the key topics or concepts within a research field.From a machine learning perspective, citation analysis can be regarded as a classification problem with the goal of categorizing citations into either important or non-important classes. In this context, several representative supervised learning methods have been applied, including Support Vector Machine (SVM), Convolutional Neural Network (CNN), Recursive Neural Network (RNN), and Long Short-Term Memory (LSTM). These methodologies have been employed to tackle the task of citation classification and have shown promise in enhancing the efficiency and accuracy of this critical analytical process [29]. Huang et al. [47] proposed a deep learning-based approach for citation recommendation using a convolutional neural network (CNN) to capture contextual information and learn citation patterns. Zhang et al. [48] focused on citation intent classification using deep learning models, employing a combination of bidirectional long short-term memory (BiLSTM) and attention mechanisms. Wang et al. [49] proposed DeepCite, a deep neural network model, to predict the future citation count of a paper based on its textual features and citation history.

To address citation intent analysis, Cohan et al. [46] employed a bi-directional LSTM with an attention mechanism, complemented by ELMO vectors and structural scaffolds. Beltagy et al. [50] introduced SciBERT, a variant of BERT trained on a large corpus of scientific publications. Dominique et al. [51] utilized ImpactCite, an approach based on the XLNet model, for citation impact and sentiment analysis. These studies utilized publicly available datasets such as SciCite and CSC. Muppidi et al. [52] suggested that combining CNN and LSTM improves the capability to identify local patterns and textual order. Deep learning methods offer effective means to analyze the importance of citations in scientific papers, overcoming the limitations of traditional approaches [53] and Researchers have extensively employed these deep learning models like LSTM, GRU, and CNN in the context of text classification tasks [29].

Furthermore, researchers have explored the role of metadata in understanding citation relationships. Utilizing metadata has proven valuable in establishing links between cited and citing papers, thereby enriching our knowledge of citation networks.The Study [18] utilized metadata to identify key citations and compared their findings with a comprehensive content-based categorization. Incorporating the context of metadata and section-wise citation information can provide a more comprehensive understanding of the relationship among research articles.

While existing studies have made significant contributions to citation analysis and identification, challenges still exist. The rapidly evolving landscape of scientific literature and the emergence of new research topics pose challenges in accurately categorizing and classifying citations. The availability of large-scale datasets for training and evaluating citation analysis models remains a concern, as access to comprehensive citation databases can be restricted or limited. However, integrating contextual word embedding models, metadata, and section-wise citation count features shows promise in enhancing the accuracy and precision of identifying important citations. This current research aims to bridge this gap by proposing a novel approach that integrates these elements, ultimately enhancing the accuracy and precision of identifying important citations in academic literature. By building upon existing research on word embedding, citation analysis, and metadata utilization, this study strives to provide valuable insights and advancements in the field of citation analysis and identification.

In conclusion, citation analysis plays a crucial role in understanding the relationships between scholarly works, assessing their impact, and identifying important citations. Researchers have employed various methods and techniques, ranging from traditional machine learning to deep learning models, to analyze citation patterns, classify citation intent, and predict citation impact. Additionally, the utilization of word embedding models, such as Word2Vec and BERT, has proven effective in capturing semantic relationships and contextual information within citation contexts. Moreover, the incorporation of metadata and section-wise citation information has further improved the understanding of citation networks.

Despite the progress made, challenges such as the evolving research landscape and limited access to comprehensive datasets persist. However, by leveraging advancements in machine learning, natural language processing, and the integration of contextual information, researchers can further improve the accuracy and effectiveness of citation analysis.

The primary objective of this study is to make a significant contribution to the field by introducing an innovative approach that integrates word embedding models, metadata, and section-wise citation information. This combined methodology aims to identify crucial citations within academic literature. By addressing these challenges and leveraging the power of advanced techniques, the research endeavors to provide valuable insights into citation analysis. Through this effort, it seeks to illuminate the importance and impact of citations, ultimately enhancing our understanding of scholarly communication and research evaluation.

III. PROPOSED METHODOLOGY

The proposed methodology comprises several steps, including dataset selection, metadata and section information extraction, and data preprocessing, as illustrated in Figure 2 of the architecture diagram. The subsequent points provide a detailed explanation of each step.

A. DATA PREPARATION

The data preparation phase of this study involves extracting data from two freely available standard benchmark datasets, namely ACL-ARC and SciCite, which are widely used for citation classification tasks.

The first dataset used in this study is ACL-ARC, which was collected by Valenzuela et al. [19]. This dataset consists of approximately 465 records and provides valuable information for analysis. It includes details such as the context of the citation, the location of in-text citations, the ID of the citing and cited papers, the publication years, the paper titles, the author IDs, the title of the section, the section number, the phrase

TABLE 1. List of citation classification techniques.

S.No	Author Name	Classes	Accuracy	Content-Based Features	Metadata-Based Features
1	Finney et al. (1979)	1) Background knowledge 2) Tentative references 3) Conformations references 4) Methodological references 5) Navigational references 6) Interpretational references 7) Future research references	not mentioned	Yes	NO
2	Garzone et al. (2000)	1) Negational 2) Afrmational 3) Assumptive 4) Tentative 5) Methodological 6) Interpretational 7) Future research 8) Developmental 9) Use of conceptual 10) Contrastive 11) 25 more references	Good for Seen Article and Average for unseen	Yes	No
3	Teufel et al. (2006)	1) Neutral 2) Weakness 3) Comparisons 4) Compatibility	F.Score 0.71%	Yes	No
4	Jochim et al. (2012)	1) Negative 2) Positive	F.Score: 0.68%	Yes	NO
5	Li et al. (2013)	1) Negative 2) Positive 3) Neutral	F.Score: 0.67%	Yes	NO
6	Abu-Jbara (2013)	1) Negative 2) Positive	F.Score: 0.68%	Yes	NO
7	Zhu et al. (2015)	1) Influential 2) Incidental	Precision 0.35%	YES	YES
8	Valenzuela et al. (2015)	1) Important 2) Incidental	Precision 0.65%	YES	YES
9	Qayyum et al. (2019)	1) Important 2) Not important	Precision 0.72%	YES	YES
10	Nazir et al. (2020)	1) Important 2) Not important	Precision 0.84%	YES	YES
11	Nazir et al. (2022)	1) Important 2) Not important	Precision 0.94%	YES	YES

before the citation context, and most importantly, the citation purpose. The citation purpose is specified through class labels such as background, usage, comparison, inspiration, extension, and future work. The dataset is available at `https://allenai.org/data/data-all.html`. The annotators categorized the citations into four different

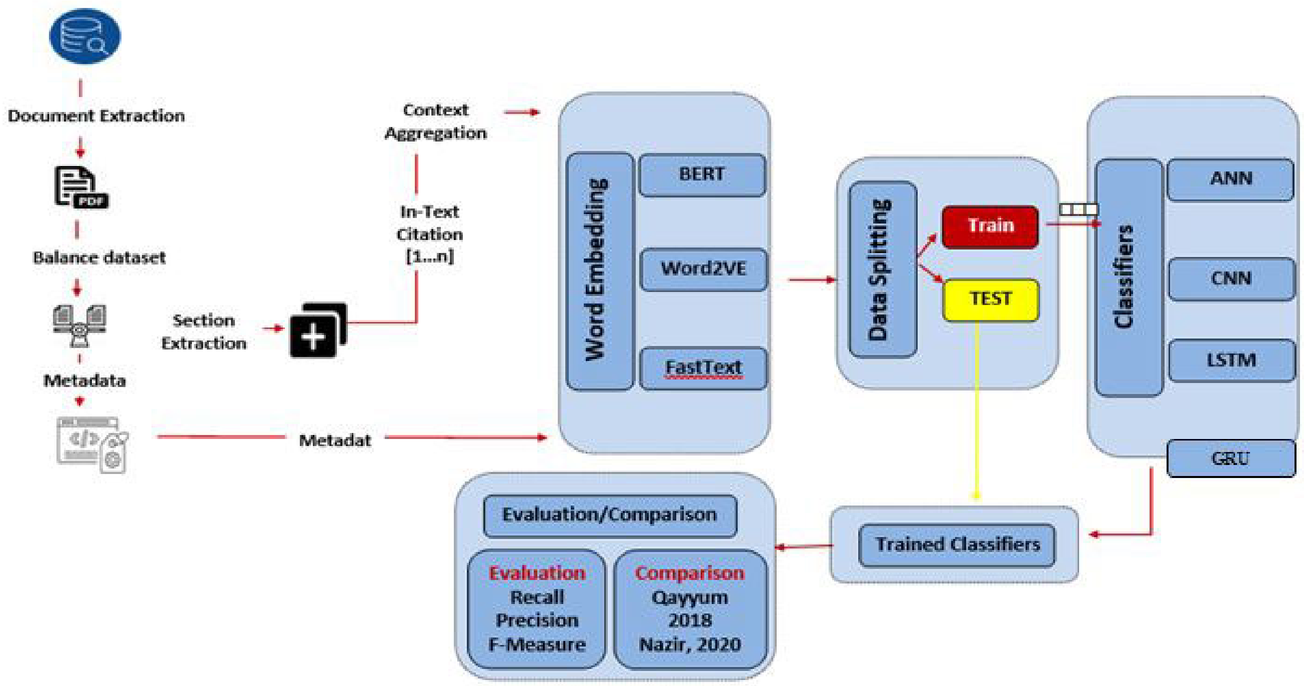

FIGURE 2. The architecture of the proposed methodology for important citation identification.

TABLE 2. Detail of dataset D1.

Attributes	*No of Records*
No of Citation	465
Not Important Citations	398
Important Citations	67

TABLE 3. Detail of dataset D2.

Attributes	*No of Records*
No of Citation	8,243
Methodology Citations	4,840
Background Citations	2,294
Result Citations	1,109

groups based on their importance. Group 0 represented relevant work, group 1 represented comparisons, group 2 indicated utilization of the work, and group 3 indicated extensions of the work. These four groups were then consolidated into two categories. The merged category of 0 and 1 was labeled as 0, representing non-important work, while the merged category of 2 and 3 was labeled as 1, indicating important citations.

The Second dataset used in this study is SciCite, which was created by Cohen et al. [46]. The SciCite dataset was chosen for several reasons:

- It is a well-known publicly accessible dataset that focuses on citations in the field of computer science.
- The dataset is of considerable size, providing a substantial amount of data for analysis.
- The SciCite dataset has been widely used in state-of-the-art approaches for citation classification tasks.

However, it is important to note that the SciCite dataset exhibits an unbalanced class distribution. Additionally, the dataset includes information such as the section name in which the in-text citation is inserted, the ID of the citing and cited papers, the context of the citation, and the class label for citation intent.

Based on Valenzuela's definition of important and not important citations, the dataset was further classified into binary categories. The category labeled as 0 represented not important citations, while the category labeled as 1 represented important citations.

Since both datasets display an unbalanced class distribution, it is essential to create balanced versions of them for comparison purposes. This balance will be achieved using the Synthetic Minority Oversampling Technique (SMOTE), which generates synthetic instances artificially, thus improving the accuracy and efficiency of the classification process [29]. The objective is to ensure that each class has an equal number of instances, facilitating fair and meaningful comparisons between different citation classes.

B. COMPLETING MISSING INFORMATION

Because the dataset D2 don't have metadata and abstract information. To overcome the problem of partial information in dataset D2, this study employed the semantic scholar API using python script to retrieve comprehensive details about the research articles, specifically focusing on acquiring title and abstract information. The semantic scholar API is a valuable resource that provides access to a wide range of

TABLE 4. Title of cited and citing papers.

Attributes	*No of Records*
Sentence Reduction for Automatic Text Summarization	Flexible Summarization of Spontaneous Dialogues in Unrestricted Domains3
Sentence Reduction for Automatic Text Summarization	A Noisy-Channel Model for Document Compression
A non-projective dependency parser	An Intrinsic Stopping Criterion for Committee-Based Active Learning
A non-projective dependency parser	A Classification Approach to Word Prediction

scholarly information, including metadata such as authors, publication date, abstracts, and other relevant details about research articles. By incorporating this API, this study enhances the quality and comprehensiveness of the dataset used.

C. BALANCING DATASET

One crucial aspect of this study is addressing algorithm bias by creating a balanced version of the dataset. By balancing the dataset, we aim to mitigate any potential biases and ensure fair and accurate evaluation of the deep learning models and word embedding techniques. SMOTE is a widely used sampling technique that generates synthetic samples from the minority class by interpolating between neighboring instances. It has been shown to effectively alleviate class imbalance problems and improve the performance of classification models in various domains [54]. Monard [55] conducted experiments to assess the performance of different methods for handling class imbalance. They concluded that SMOTE is one of the best techniques for addressing imbalanced datasets, providing significant improvements in classification accuracy [29]. Their findings further validate the suitability of SMOTE for balancing datasets in the context of this study for important citation identification. We have utilized the same techniques for creating equal number of instance of each class

D. FEATURES SELECTIONT

This study incorporates a combination of text-based and metadata-based features for the task of important citation identification. These features include the title, abstract, keywords, in-text citation count, and section-wise in-text citation information. Notably, a survey paper by Beel et al. [56] revealed that over 55% of 200 articles on research paper recommendation in the past two decades have utilized a content-based filtering approach. This emphasizes the significance of leveraging the content of research papers in recommendation systems.

Furthermore, the study draws support from previous research. In 2018, Hassan et al. [42] claimed that abstract and text similarity provide more informative signals compared to other features. This indicates the relevance of considering abstract and textual content in assessing citation importance.

TABLE 5. Selected features.

Feature	*Descriptions*
F1	Total Citation Count
F2	Total Citation in Mythology Section
F3	An Total Citation in Result Section
F4	Total Citation in Introduction Section
F5	Total Citation in Discussion Section
F6	Number of sections of citing paper in which a particular citation occurs
F7	Abstract Similarity
F8	Title Similarity

Similarly, in 2020, Saboor et al. [57] argued that the abstract alone is sufficient for making decisions regarding article similarity, further supporting the emphasis on abstract-based analysis.

Additionally, in 2019 Qayyum et al. [18] demonstrated the improved performance achieved by Valenzuela's approach by relying solely on freely available metadata database features. This highlights the usefulness of leveraging metadata in enhancing the accuracy and effectiveness of important citation identification.

By incorporating these text-based and metadata-based features, this research study aims to capture comprehensive information and exploit the potential of these various components to accurately identify important citations

E. TOKENIZATION

Tokenization is an essential step for text analysis, where paragraphs and sentences are segmented into individual words or tokens. This process enables more granular analysis of the text and facilitates subsequent text-processing tasks. By breaking down the text into tokens, the study can capture the fine-grained details and relationships between words, which are crucial for effective natural language processing

F. STOP WORD REMOVAL

Stop words are commonly occurring words that do not carry significant meaning in the context of the topic, and are eliminated. Examples of stop words include prepositions, conjunctions, and articles. To implement the removal of stop words in Python, the study utilizes the Natural Language Toolkit (NLTK), a widely used library for natural language processing tasks. The NLTK provides efficient methods and resources for stop word removal, enabling the elimination of these non-informative words from the citation text. various studies' results demonstrated that removing stop words effectively reduced noise and improved the performance of sentiment classification models [58].

G. CASE CONVERSION STEMMING

In the pre-processing step of the research study, two important techniques, namely case conversion and stemming, are applied to the text data. Firstly, all the upper-case terms in the text are converted to lower-case. This conversion is performed because, in the context of the study, upper-case and lower-case versions of words are considered to have the same meaning. By converting all terms to a consistent case, the study ensures that words with similar meanings are treated equally, regardless of their capitalization. Secondly, stemming is applied to the words in the text. Stemming involves reducing words to their base or root form by removing prefixes, suffixes, and other inflectional endings. This process helps to normalize the words and bring them to their basic form, which aids in capturing their core meaning and improving the efficiency of subsequent text analysis tasks. By applying case conversion and stemming, the study standardizes the text data, reduces unnecessary variations, and facilitates more accurate and meaningful analysis of the citation content.

H. EXTRACTING SECTION INFORMATION

Due to the presence of section information in both datasets, the research study extracts this information and represents it using the following equation.

$$LS = (LS_1, LS_2, LS_3, \ldots, LS_N) \tag{1}$$

Here, LS represents the logical sections of a research papers.

I. CITATION TAG AND CONTEXT

In this step the study generates a list of citation tags and their corresponding citation contexts, as expressed by Equation. 2. The citation context refers to a snippet of words that surround a citation tag within a document. the study keeps track of a vector representing the occurrence of citations within logical segments, as specified in Equation. 3. This is done to account for the possibility of a work being cited in multiple parts or sections of the document.

$$CD_i = (C_1, C_2, C_3, \ldots, C_N) \tag{2}$$

The function $f(C_i)$ is defined as:

$$f(C_i) = \begin{cases} 0 & \text{if } C_i \in LS_1 \\ 0 & \text{if } C_i \in LS_2 \\ \vdots & \\ \vdots & \\ 1 & \text{if } C_i \in LS_N \end{cases} \tag{3}$$

In the above equation, CD represents citations within the research document, while 'c' signifies individual citations. The equation serves the purpose of meticulously tracking all citations, discerning their respective categories as either '1' for important or '0' for not important.

J. CITATION COUNT

During this phase, we focused on determining the frequency with which a particular publication was mentioned and cited in research articles, as well as the specific sections in which these citations occurred. The frequency of citations provided valuable insights into the reception and reputation of a paper within the scholarly community. By analyzing the dataset, we were able to identify how frequently a specific publication was referenced by other researchers. This information shed light on the level of acceptance and recognition that the paper had achieved within the academic community. A higher frequency of citations indicated that the paper was widely acknowledged and considered influential in the field [13]. Furthermore, we examined the distribution of these citations across different sections of the research articles. By identifying the parts in which the citations occurred, such as the introduction, methodology, Result, and Literature Review. we gained a deeper understanding of how the cited publication was being used and referenced by other researchers

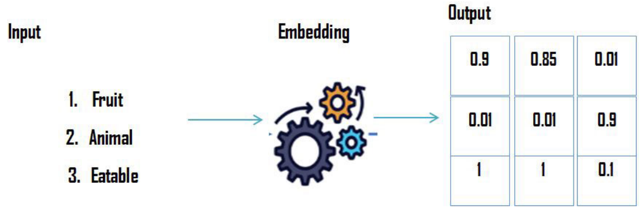

FIGURE 3. Example of text-to-numeric conversion.

K. WORD EMBEDDING REPRESENTATION

The process of converting textual representations to numerical forms, known as word embedding, is employed to capture contextual information and the relationships between individual words within citation contexts and among citation sentences. Three different word embedding methods, namely FastText, word2vec, and BERT are used and compared in the study. Let's define S as the complete set of citation contexts, which is represented as `S = {C_1, C_2, \ldots, C_n}`, extracted from the papers in the dataset D. Each paper `i` has a set of citation contexts, denoted as `C_i = {c_1, c_2, \ldots, c_n}`. To represent these citation contexts in a numerical form, word vectors Vi of a specific dimension are created. These word vectors serve as representations of the sentences, ensuring that semantically similar citation contexts are closer in the vector space and have similar representations.

One of the word embedding techniques, such as FastText, word2vec, or BERT, is applied to generate these word vectors. For example, the words "apple" and "mango" will have vectors that are closer to each other in the vector space compared to the word "elephant" since the former two words are used in similar contexts.

The Word2Vec word embedding generator, as mentioned in the study, aims to understand the implications and semantic relationships between words [33]. Both pre-trained versions are available at https://code.google.com/archive/p/word2vec/ and domain-trained versions of Word2Vec will be utilized for comparison purposes in the study.

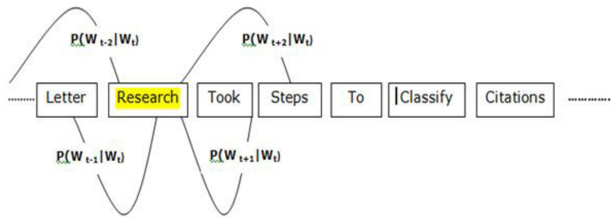

FIGURE 4. Representation of in-text Citation in the different context window.

In addition to Word2Vec, the study incorporates the use of Bidirectional Encoder Representations from Transformers (BERT) to refine word vector representations based on the semantic context of words in specific situations. BERT is a powerful language representation model that has revolutionized various natural language processing tasks. It leverages a deep bidirectional transformer architecture to capture contextual information and provide a rich representation of words [59]. BERT has been widely adopted in the field of NLP due to its ability to handle complex language phenomena such as word ambiguity and syntactic relationships. It learns contextualized word embeddings by considering the entire sentence rather than just the neighboring words. This allows BERT to capture intricate semantic nuances and produce highly informative word representations [59]. By utilizing BERT for word embedding, this study aims to further enhance the understanding of citation context and improve the accuracy of identifying important citations. The contextual word embeddings generated by BERT enable the model to capture the semantic significance of words within the specific citation analysis context.

Moreover, this study also incorporates FastText word embedding as part of its methodology. FastText is a word embedding technique that builds upon the Word2Vec model by considering subword information. It represents words as bags of character n-grams, which enables capturing morphological information and handling out-of-vocabulary words more effectively. The FastText model tries to capture the meaning of words based on the character-level composition of the word, allowing it to handle words with similar morphological structures. By utilizing FastText word embedding, the research article aims to enhance the identification of important citations in the academic literature by considering both semantic relationships and subword information. Previous research studies have effectively justified the utilization of FastText word embedding for the identification of important citations. Several studies in the field have highlighted the advantages of FastText in capturing morphological information and handling out-of-vocabulary words [27], [60]. These characteristics make FastText particularly suitable for the task of citation analysis, where the context and meaning of words play a crucial role.

The study extensively evaluated the performance of the three techniques: FastText, Word2Vec, and BERT, in the task of citation classification. The objective was to identify the technique or combination of techniques that yielded the best results in accurately classifying important citations. detailed analysis of the performance of each embedding technique in citation classification is given in the result section.

L. SIMILARITY SCORE

The similarity between citing and cited articles has been recognized as a valuable feature for determining the importance of a citation [42]. By leveraging metadata-based features, specifically the abstract and title of the articles, we aimed to quantify the degree of similarity between them. The abstract and title provide concise and descriptive information about the content and main findings of a research article. To calculate the similarity score, we employed established techniques such as cosine similarity. This method allowed us to measure the overlap and similarity of words, phrases, or concepts present in the abstract and title of both the citing and cited articles. A higher similarity score indicated a stronger alignment and shared thematic elements between the two articles.

1) ABSTRACT

The abstract similarity is determined by calculating the cosine similarity of the different embedding scheme scores, including Word2Vec, FastText, and BERT. The cosine similarity is computed using the equation:

$$\text{Similarity (Abstract)} = \cos(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \tag{4}$$

In this equation, x and y represent the embedding vectors of the abstracts being compared, while || x || and || y || denote the respective Euclidean norms of the vectors. The cosine similarity provides a measure of the similarity between the abstracts based on the direction and magnitude of their embedding vectors.

2) TITLE

By following a similar approach, the title similarity is calculated using the following equation

$$\text{Similarity (Title)} = \cos(\mathbf{xt}, \mathbf{yt}) = \frac{\mathbf{xt} \cdot \mathbf{yt}}{\|\mathbf{xt}\| \|\mathbf{yt}\|} \tag{5}$$

Here, `xt`and `yt` represent the embedding vectors of the titles being compared, while ||x|| and ||y|| indicate the Euclidean norms of the vectors. The cosine similarity formula allows us to quantify the similarity between titles based on the direction and magnitude of their embedding vectors. By incorporating the similarity score as a feature, we were able to assess the relevance and potential impact of a citation.

M. MODEL SELECTION

In this phase, we explored the use of deep learning neural networks, specifically the GRU (Gated Recurrent Unit)

TABLE 6. Abstract and title similarity.

Title	*Abstract*	*Class*
0.71	0.9	0
0.73	0.93	0
0.69	0.88	0
0.69	0.88	0
0.79	0.87	0
0.8	0.89	1

TABLE 7. Deep learning algorithms used in the study.

NO	*Classifier*
1	CNN
2	ANN
3	LSTM
4	GRU

classifier, for important citation classification. We compared the performance of different deep learning models including CNN, ANN, LSTM, and GRU. CNN is known for its ability to handle large data and learn complex features using convolutional, activation, pooling, and dropout layers [29]. The rectified linear unit (ReLU) is used in this work as an activation function.

$$y = \max(0, i) \tag{6}$$

where i stands for the input and y for the activation outcome. The cross-entropy error is employed as a loss function in binary classification; this has also been done in this work. It's calculated as:

$$\text{Cross-entropy} = -\,(i \cdot \log(p) + (1 - i) \cdot \log(2 - p)) \tag{7}$$

Once the input has been converted into a numeric representation, similar words are mapped close in the vector space. We are now ready to feed this data into the deep learning model for citation classification, aiming to discover the optimal deep learning model for predicting critical citation classes. The classification algorithms process the feature data. To define our problem, we focus on our training dataset.

$$D_n = \{R_1, R_2, R_3, \ldots, R_n\} \tag{8}$$

Each record Ri is assigned to one of two binary classes, denoted as C0 or C1, as depicted in Equation 9.

$$C_c = \{C_1, C_0\} \tag{9}$$

The task is to find the best deep-learning model, which correctly classify the new instance and assign a correct citation class to it

$$M_n(D) \rightarrow C \tag{10}$$

Several deep learning classifiers listed in Table 6 were utilized to assess the accuracy of deep learning for the task of identifying important citations. The input parameters from the ACL-ARC and SciCite datasets, as described in Table 4, were provided to the classification models.

TABLE 8. Experimental settings combining embedding and classification.

SetUp	*Embedding*	*Deep Learning Algorithm*	*Setting Name*
1	Word2vec	CNN	W2V+CNN
2	FastText	CNN	FastText+CNN
3	BERT	CNN	BERT+CNN
4	Word2vec	LSTM	W2V+LSTM
5	FastText	LSTM	FastText+LSTM
6	BERT	LSTM	BERT+LSTM
7	Word2vec	GRU	W2V+GRU
8	FastText	GRU	FastText+GRU
9	BERT	GRU	BERT+GRU
10	Word2vec	ANN	W2V+ANN
11	FastText	ANN	FastText+ANN
12	BERT	ANN	BERT+ANN

1) The dataset is divided with the ratio 80:20. 80% of the records were provided as training data and 20% were left for testing purposes.
2) The deep learning model was trained based on the input parameters, adjusting the input weights for the target class of citation intent.
3) The trained model was then used for predicting 20% of the remaining records that were left for testing purposes.
4) The predicted citation class was cross-checked with the actual class of the inputs.

IV. EXPERIMENTAL SETUP

We conducted a series of experiments using various combination of word embedding and deep learning algorithms to explore their performance in our task. The different settings of word embedding and deep learning algorithms are given in Table 8. To select the appropriate word embedding techniques, we followed the recommendations provided in [30]. Additionally, we carefully select the deep learning algorithms, that are well-suited for binary classification tasks. By comparing the results across all combinations of the selected algorithms, we aimed to identify the optimal setup for our specific environment.The experimental setup involved the following steps.

1) DATASET SELECTION

We selected two datasets for citation analysis after considering size, quality, and relevance to our study objectives. We have discussed the details of the dataset in the previous parts.

2) PREPROCESSING

For the selected dataset, preprocessing steps included text normalization, tokenization, and the deletion of extraneous letters and symbols. The steps are described in depth above.

3) WORD EMBEDDING INITIALIZATION

For each deep learning model (CNN, LSTM, GRU), we initialized the word embed dings using one of the following techniques:

- BERT: We utilized pre-trained BERT embeddings from the Hugging Face library1
- Word2Vec We used pre-trained Word2Vec embeddings for which we utilized the Gensim python library. The pretrained word2vec model is trained on the Google News dataset model, containing 300-dimensional embeddings for 3 million words and phrases. Available with the name 'GoogleNews-vectors-negative300.bin' (1.3 GB compressed) on https://code.google.com/archive/p/word2vec/.
- FastText: We employed pre-trained FastText embeddings obtained from the FastText library2

4) MODEL ARCHITECTURE DESIGN

We designed the architecture of each deep learning model (CNN, LSTM, GRU) by specifying the number and type of layers, activation functions, dropout rates, and other relevant parameters. The input to the models consisted of the word embeddings generated in the previous step.

5) MODEL TRAINING

We divided the preprocessed dataset into training and testing sets, typically using an 80:20 split. The training set was used to train the deep learning models using "Adam optimizer" optimization techniques and "binary cross-entropy" loss functions. The models were trained iteratively until convergence or for 50 epochs.

6) MODEL EVALUATION

After training, we evaluated the performance of each deep learning model on the testing set. We measured key evaluation metrics such as accuracy, precision, recall, and F1-score to assess the effectiveness of the models in predicting the importance of citations.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

$$\text{F-Measure} = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}} \tag{13}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{14}$$

7) COMPARISON AND ANALYSIS

Finally, we compared the performance of different deep learning models (CNN, LSTM, GRU) coupled with various word embeddings (BERT, Word2Vec, FastText). We analyzed the results to identify the best combination that yielded the highest accuracy, precision, and most reliable prediction of citation importance.

By conducting this experimental setup, we aimed to determine the most effective combination of deep learning models and word embeddings for citation analysis, facilitating accurate assessment of the importance of citations in research articles.

V. RESULTS

All experiments were conducted on a Lenovo ThinkCentre machine equipped with an Intel Core i5 processor and 8 GB of DDR4 RAM. The Kaggle notebook environment was utilized for performing the experiments, providing a convenient and scalable platform for running the code. The implementation of deep learning models and embedding techniques in Python relied on various libraries to facilitate the process. The sklearn library was used for data preprocessing and evaluation, while Keras and TensorFlow were employed for building and training the deep learning models. Additionally, the fasttext, hugging face, and Gensim libraries were utilized for integrating fastText, BERT, and Word2Vec embeddings.

A. COMPARISON OF PREDICTIVE PERFORMANCE OF MODELS USING DATASET-1

Extensive experiments were conducted to analyze citations and develop an efficient method for citation analysis. The experiments involved the utilization of deep learning models such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). Additionally, an Artificial Neural Network (ANN) classifier was also employed. These models were coupled with popular word embedding techniques including Word2Vec, FastText, and BERT. The main objective of these experiments was to identify important citations and examine the effectiveness of different deep-learning models in this task.

1) EXPERIMENTS USING FastText

Firstly, the models undergo training using fastText word embedding, and the outcomes are presented in Figure 5. All the models perform nearly identically, with only slight differences in precision when using fasttext embedding. Among the models, CNN attains the highest accuracy rate at 0.89%. Additionally, CNN also attains the highest precision score of 97%, surpassing the precision achieved by the other models. Both LSTM and GRU achieve a precision score of 96%. However, in terms of accuracy, LSTM performs the poorest with a accuracy rate of 83% and a recall rate of 75%. On the other hand, CNN delivers the most favorable outcome with the highest accuracy (89%), precision (97%), and recall (84%) for the important citation task.

2) EXPERIMENTS USING Word2vec

A separate series of experiments was conducted using deep learning models employing the word2vec word embedding. The performance comparison of these models for identifying important citations is presented in Figure 6. Upon examining the results, it is observed that both CNN and ANN exhibit a precision score of 0.97, while GRU and LSTM achieve a score of 0.95. Notably, the precision of ANN shows a slight improvement compared to the previous model, whereas GRU and LSTM experience a decrease. The overall performance remains relatively stable. Remarkably, CNN and ANN

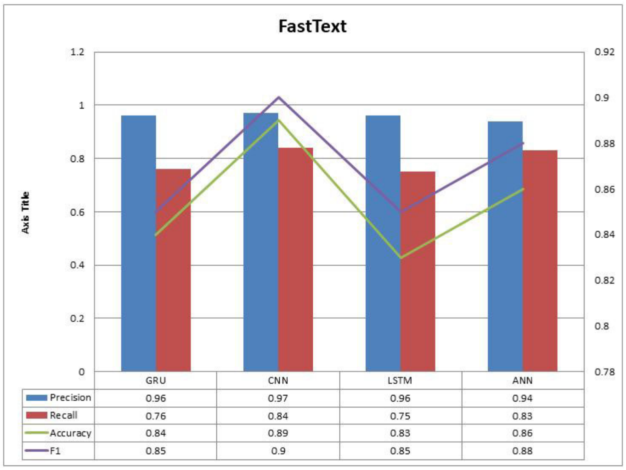

FIGURE 5. Experiment using FastText.

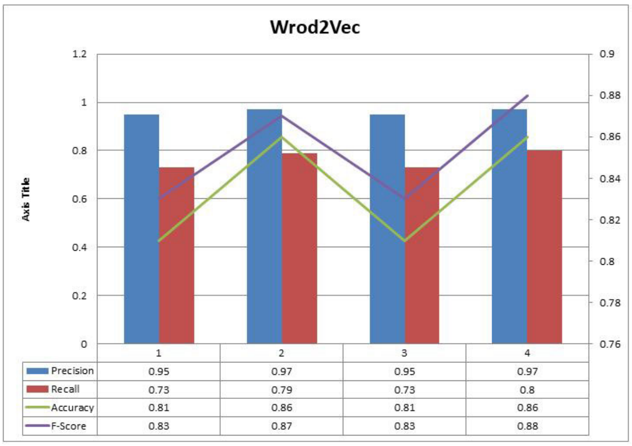

FIGURE 6. Experiments Using Word2Vec.

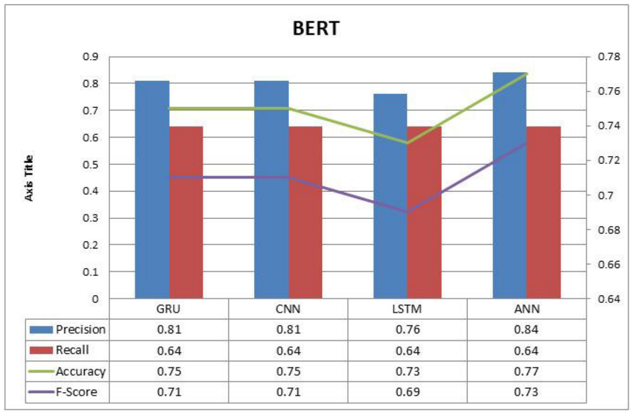

FIGURE 7. Experiment using BERT.

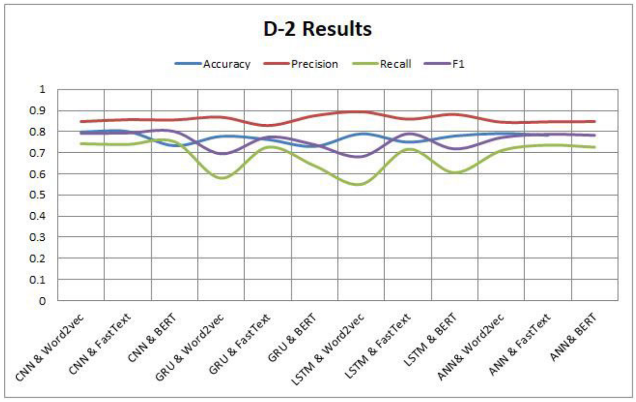

FIGURE 8. Dataset D-2 results.

achieve the highest precision of 0.97, while ANN attains the highest recall, and both CNN and ANN achieve the highest accuracy of 0.86. Both GRU and LSTM achieve identical recall scores of 0.73, along with the same accuracy rate of 0.81. The ANN achieved the highest F1 score of 0.88.

3) EXPERIMENTS USING BERT BASE

We conducted a separate series of experiments utilizing deep learning models that employed BERT-Base word embeddings. Figure 7 presents the performance comparison of these models in identifying important citations. Upon analyzing the results, we observed that all deep-learning models exhibited a similar recall of 0.64. However, the overall precision score decreased compared to the previous model's setup. Among the models, the ANN achieved the highest precision score of 0.84, while both GRU and CNN achieved a precision score of 0.81 and an accuracy of 0.71.

B. COMPARISON OF PREDICTIVE PERFORMANCE OF MODELS USING DATASET-2

Considering the limited size of D1, that consisting of only 465 paper-citation pairs, it is evident that this dataset may not offer sufficient information to conduct a comprehensive analysis of the outcomes. To ensure more reliable and conclusive findings, it becomes crucial to examine the performance trends across multiple datasets. Consequently, a series of experiments were conducted on dataset 2, using the same features, and the results are visually presented in Figure 8.

The obtained results strongly corroborate the effectiveness of the proposed CNN when utilized in conjunction with FastText. Moreover, it is observed that the performance of the CNN is consistently superior when combined with FastText, except in cases where BERT and Word2Vec features are employed. These findings validate the notion that for larger datasets, employing the LSTM model with word2vec and BERT embedding leads to improved precision, achieving an impressive 89% precision when utilized in combination with word2vec and BERT. Furthermore, the GRU model, when combined with BERT embeddings, also demonstrates exceptional results.

In terms of the F1 Score, the CNN model, when combined with word2vec and FastText embeddings, achieves high performance. Additionally, the overall accuracy of the CNN surpasses that of other models, further emphasizing its efficacy.

C. PERFORMANCE COMPARISON WITH STATE-OF-THE-ART STUDIES

A performance comparison is also made with other recent studies to demonstrate the significance of the proposed approach. In this regard, three recent studies were selected, namely Faiza et al. [18], [20], and [45]. The study employed various models for important citation identification, including support vector classifier (SVM), random forest, k-nearest neighbor, and LR. The comparison results provided in Table 8 indicate that the proposed framework, utilizing the CNN model coupled with fastText and blended features, outperforms the state-of-the-art models for important citation analysis. It is noteworthy that Faiza et al. [18] achieved a maximum precision of 0.72, an improvement over Valenzuela's work which obtained a precision of 0.65. the study [20] Further enhanced these results by attaining a precision of 0.85 using Random Forest on the same dataset employed by Valenzuela et al. [19] and Faiza et al. [18]. These achievements were accomplished by incorporating features such as (1) content similarity, (2) Citation Count, and (3) section-wise In-text Citation weights. Among these features, the novel addition in this research was the introduction of Section-wise In-text citation weights. In 2022, Nazir et al. [45] extended the results by incorporating sentiment analysis as additional features using the same classifier. However, In contrast to previous studies that relied on intricate and time-consuming calculations to determine citation weights, our approach excels by utilizing the same dataset and readily available features from research articles. We eliminate the need for complex computations and instead adopt a straightforward method, counting citations and harnessing the metadata features described in the preceding section. This pragmatic approach not only expedites the analysis process but also delivers outstanding results, as evidenced by our model achieving an impressive precision of up to 0.97%. By utilizing the CNN model coupled with fastText and blended features, the proposed framework surpasses the state-of-the-art models in the field of important citation analysis. These results underscore the effectiveness of incorporating features such as metadata base content similarity, Citation Count, and section-wise In-text Citation counts. As a result, our approach achieves a remarkable precision of up to 0.97, signifying a substantial improvement in the accuracy of important citation analysis. This simplicity, coupled with its remarkable precision, underscores the practicality and superiority of our model. The findings from the performance comparison highlight the significant advancements made by the proposed approach,offering a valuable contribution to the field of citation analysis in research articles.

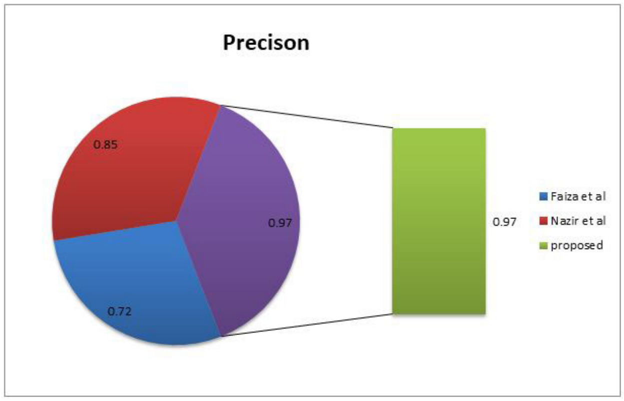

FIGURE 9. Dataset D-2 results.

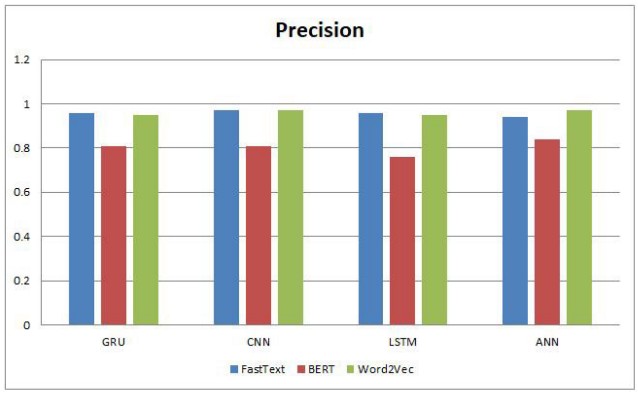

FIGURE 10. Precision-based comparison among fastText, BERT, and Word2Vec.

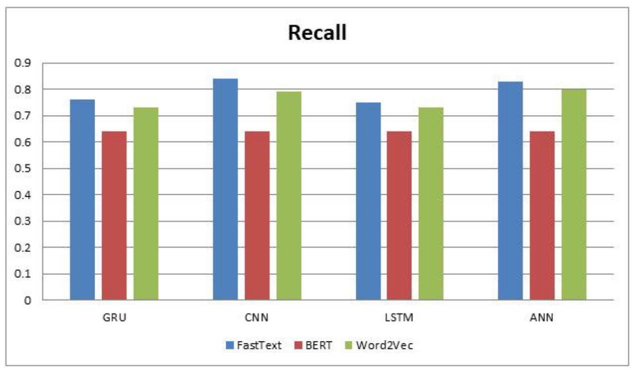

FIGURE 11. Recall-based comparison among fastText, BERT, and Word2Vec.

VI. DISCUSSIONS

To classify citations into important and not-important classes, the study explores state-of-the-art machine learning and deep learning models combined with word embedding techniques. Each word embedding technique undergoes evaluation using standard measures like accuracy, precision, recall, and F1 score. Figure 5-7 illustrates the comparative performance of classifiers for citation analysis. The results demonstrate that the deep learning model, CNN, achieves superior performance when trained on blended features.

Figure 10 provides a Precision comparison of models using fastText, BERT, and word2vec word embedding techniques.

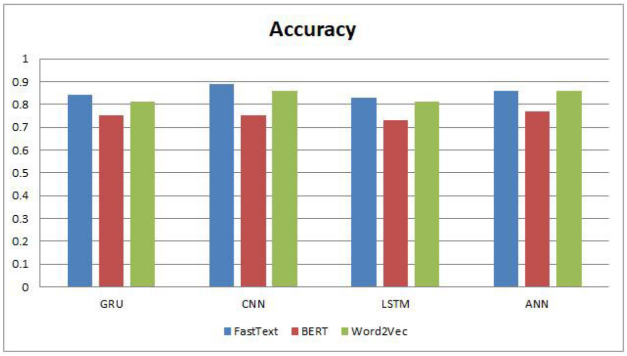

FIGURE 12. Accuracy-based comparison among fastText, BERT, and Word2Vec.

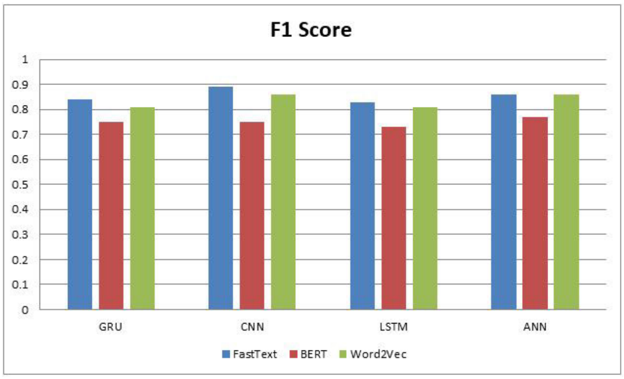

FIGURE 13. Performance comparison in term of F1-score.

The results indicate that CNN consistently achieved the best precision across all word embedding techniques, outperforming the other models. On the other hand, the ANN model demonstrated the highest precision when using combined features. Interestingly, the ANN model exhibited the lowest precision when using fastText embeddings, while achieving the highest precision with word2vec. It is worth noting that all models exhibited lower precision with BERT embedding compared to fastText. Moreover, the deep learning model CNN consistently showed superior precision when trained on fastText and word2vec embeddings. Additionally, it's important to mention that these findings suggest the effectiveness of deep learning models, particularly CNN when combined with specific word embedding techniques such as fast Text and word2vec. The superior precision achieved by CNN underscores its capability to capture intricate patterns and nuances within the citation data, thereby leading to more accurate classification of "important" citations. These findings contribute to the advancement of our understanding and the implementation of techniques for important citation analysis. They also underscore the significance of selecting appropriate word embedding techniques to enhance model performance.

In Figure 11, we present a recall comparison of models using fastText, BERT, and word2vec word embeddings. Among the models, the voting classifier demonstrated the highest recall of 0.84 when utilizing fastText word embedding. Notably, when trained on fastText, the CNN model achieved the highest recall rate of 84%. Additionally, the remaining models consistently exhibited better recall rates with fastText word embedding, while showing similar recall rates with BERT embedding. Furthermore, all models demonstrated improved recall rates with word2vec embeddings as compared to BERT. Both CNN and ANN models showcased nearly identical recall rates, with only a slight difference, while LSTM and GRU models also displayed identical recall rates with word2vec embedding.

Figure 12 illustrates the accuracy comparison of models using fastText, BERT, and Word2Vec word embedding techniques. The results indicate that CNN achieved the highest accuracy when trained on FastText embeddings. On the other hand, CNN and ANN demonstrated the same level of accuracy when using Word2Vec embeddings, in term of accuracy ANN achieving better results with BERT compared to other models. LSTM showed the lowest accuracy when utilizing BERT embedding. Both GRU and CNN models exhibited the same level of accuracy with BERT embedding. Furthermore, ANN and CNN achieved identical accuracies when using Word2Vec embedding.

These findings suggest that CNN consistently outperformed other models in terms of accuracy when trained on FastText and Word2Vec word embeddings. Additionally, the results highlight the variability in accuracy among different models when utilizing BERT embedding, with ANN achieving relatively better accuracy compared to other models.

Figure 13 presents the comparison of F1 scores among models for important citation analysis. The results demonstrate that the highest F1 score of 90% is achieved by the CNN model coupled with FastText when using the combined feature set. Additionally, ANN has attained the second-highest F1 score when utilizing FastText word embedding. Both GRU and LSTM models exhibit the same F1 score when employing word2vec embedding. Similarly, CNN and GRU models achieve identical F1 scores when trained on BERT embedding. Notably, ANN achieves better F1 scores with Word2Vec embedding as compared to other models. On the other hand, LSTM exhibits the lowest F1 score among the models when utilizing BERT embedding.

These findings highlight the effectiveness of CNN coupled with FastText in achieving the highest F1 score for important citation analysis. Furthermore, the results emphasize the variations in F1 scores among different models when using different word embedding techniques. ANN stands out with better F1 scores when utilizing word2vec embedding, while LSTM shows comparatively lower F1 scores with word2vec embedding. These insights contribute to the understanding of model performance in the context of important citation analysis and can assist researchers in selecting appropriate models and word embedding techniques for achieving higher F1 scores.

VII. CONCLUSION

In recent years, the effective identification of important citations has garnered significant attention in the field of scientometrics. In this study, we introduced a novel approach for citation analysis that combines in-text citation counts and metadata features. We employed multiple word embedding techniques, including fastText, Word2Vec, BERT, and their combinations, in conjunction with deep learning models. The study utilized two benchmark datasets, namely D1 (comprising 465 paper-citation pairs), and D2, containing a significantly larger number of instances, totaling 8,243 paper-citation pairs. These datasets exhibit a high degree of imbalance. To address this data imbalance issue and enhance the accuracy of classification, we employed SMOTE techniques, which artificially generated new samples. To balance the dataset. Additionally, we developed a Python script to extract data from Semantic Scholar to complete the missing information of the dataset.We explored the efficiency of four deep learning models (GRU, CNN, LSTM, and ANN) coupled with three different word embedding techniques for the task of important citation identification. The results indicate that the CNN model coupled with fastText embedding outperforms other models in identifying important citations. Furthermore, we compared our results with state-of-the-art models for important citation identification [18], [20]. Our model consistently achieved superior performance. On average, our proposed model CNN+fastText, when used with combined features, achieved 97% precision, 89% accuracy, and a 90% F1 score for important citation identification. Based on these results, we confidently claim that our approach provides a more accurate reflection of important citations.The current approach has been exclusively assessed using binary classes. Therefore, our future research endeavors are directed towards broadening the scope of our experiments to encompass multi-class evaluations. Additionally, we intend to expand our assessment by integrating datasets originating from diverse academic disciplines and sources. At present, our focus is primarily centered on metadata features. However, in forthcoming research, we intend to extend our analytical scope to encompass the entire body of text within research documents. Furthermore, we are planning to delve into optimizing the embedding window size to enhance the precision and effectiveness of our model.

REFERENCES

[1] O. Küçüktunç, E. Saule, K. Kaya, and Ü. V. Çatalyürek, "Diversifying citation recommendations," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 4, pp. 1–21, Jan. 2015.

[2] M. Färber and A. Jatowt, "Citation recommendation: Approaches and datasets," *Int. J. Digit. Libraries*, vol. 21, no. 4, pp. 375–405, Dec. 2020.

[3] N. Ratprasartporn and G. Ozsoyoglu, "Finding related papers in literature digital libraries," in *Proc. Int. Conf. Theory Pract. Digit. Library*. Cham, Switzerland: Springer, 2007, pp. 271–284.

[4] Y. Ding, G. Zhang, T. Chambers, M. Song, X. Wang, and C. Zhai, "Content-based citation analysis: The next generation of citation analysis," *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 9, pp. 1820–1833, Sep. 2014.

[5] F. Narin, "The use of publication and citation analysis in the evaluation of scientific activit," Computer Horizons, Cherry Hill, NJ, USA, Tech. Rep. 704R, 1976.

[6] K. Sugiyama and M.-Y. Kan, "A comprehensive evaluation of scholarly paper recommendation using potential citation papers," *Int. J. Digit. Libraries*, vol. 16, no. 2, pp. 91–109, Jun. 2015.

[7] N. Sakib, R. B. Ahmad, and K. Haruna, "A collaborative approach toward scientific paper recommendation using citation context," *IEEE Access*, vol. 8, pp. 17328–17338, 2020.

[8] L. Bornmann and H. Daniel, "What do citation counts measure? A review of studies on citing behavior," *J. Document.*, vol. 64, no. 1, pp. 45–80, Jan. 2008, doi: 10.1108/00220410810844150.

[9] H. Small, "Co-citation in the scientific literature: A new measure of the relationship between two documents," *J. Amer. Soc. Inf. Sci.*, vol. 24, no. 4, pp. 265–269, Jul. 1973.

[10] F. Meng, D. Gao, W. Li, X. Sun, and Y. Hou, "A unified graph model for personalized query-oriented reference paper recommendation," in *Proc. 22nd ACM Int. Conf. Conf. Inf. Knowl. Manage.*, 2013, pp. 1509–1512, doi: 10.1145/2505515.2507831.

[11] L. Guo, X. Cai, F. Hao, D. Mu, C. Fang, and L. Yang, "Exploiting fine-grained co-authorship for personalized citation recommendation," *IEEE Access*, vol. 5, pp. 12714–12725, 2017.

[12] R. Habib and M. T. Afzal, "Sections-based bibliographic coupling for research paper recommendation," *Scientometrics*, vol. 119, no. 2, pp. 643–656, May 2019.

[13] A. Shahid, M. T. Afzal, A. Alharbi, H. Aljuaid, and S. Al-Otaibi, "In-text citation's frequencies-based recommendations of relevant research papers," *PeerJ Comput. Sci.*, vol. 7, p. e524, Jun. 2021.

[14] T. Luukkonen, "Is scientists' publishing behaviour rewardseeking?" *Scientometrics*, vol. 24, no. 2, pp. 297–319, Jun. 1992.

[15] H. Inhaber and K. Przednowek, "Quality of research and the Nobel prizes," *Soc. Stud. Sci.*, vol. 6, no. 1, pp. 33–50, 1976.

[16] F. Narin, "Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity," Computer Horizons, Cherry Hill, NJ, USA, Tech. Rep., 1976.

[17] X. Zhu, P. Turney, D. Lemire, and A. Vellino, "Measuring academic influence: Not all citations are equal," *J. Assoc. Inf. Sci. Technol.*, vol. 66, no. 2, pp. 408–427, Feb. 2015.

[18] F. Qayyum and M. T. Afzal, "Identification of important citations by exploiting research articles' metadata and cue-terms from content," *Scientometrics*, vol. 118, no. 1, pp. 321–344, 2019.

[19] M. Valenzuela, V. Ha, and O. Etzioni, "Identifying meaningful citations," 2015.

[20] S. Nazir, M. Asif, S. Ahmad, F. Bukhari, M. T. Afzal, and H. Aljuaid, "Important citation identification by exploiting content and section-wise in-text citation count," *PLoS ONE*, vol. 15, no. 3, Mar. 2020, Art. no. e0228885.

[21] E. Garfield, "Is citation analysis a legitimate evaluation tool?" *Scientometrics*, vol. 1, no. 4, pp. 359–375, 1979, doi: 10.1007/BF02019306.

[22] S. Nazir, M. Asif, and S. Ahmad, "Important citation identification by exploiting the optimal in-text citation frequency," in *Proc. Int. Conf. Eng. Emerg. Technol. (ICEET)*, Feb. 2020, pp. 1–6, doi: 10.1109/ICEET48479.2020.9048224.

[23] A. Athar, "Sentiment analysis of citations using sentence structure-based features," in *Proc. ACL Student Session*, 2011, pp. 81–87.

[24] A. Shahid, M. Afzal, and M. Qadir, "Discovering semantic relatedness between scientific articles through citation frequency," *Austral. J. Basic Appl. Sci.*, vol. 5, no. 6, pp. 1599–1604, 2011.

[25] D. Jurgens, S. Kumar, R. Hoover, D. McFarland, and D. Jurafsky, "Measuring the evolution of a scientific field through citation frames," *Trans. Assoc. Comput. Linguistics*, vol. 6, pp. 391–406, Dec. 2018.

[26] M. Roman, A. Shahid, S. Khan, A. Koubaa, and L. Yu, "Citation intent classification using word embedding," *IEEE Access*, vol. 9, pp. 9982–9995, 2021.

[27] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Trans. Assoc. Comput. Linguistics*, vol. 5, pp. 135–146, Dec. 2017.

[28] M. Karim, M. M. S. Missen, M. Umer, S. Sadiq, A. Mohamed, and I. Ashraf, "Citation context analysis using combined feature embedding and deep convolutional neural network model," *Appl. Sci.*, vol. 12, no. 6, p. 3203, Mar. 2022.

[29] H. Sadr, M. M. Pedram, and M. Teshnehlab, "Multi-view deep network: A deep model based on learning features from heterogeneous neural networks for sentiment analysis," *IEEE Access*, vol. 8, pp. 86984–86997, 2020.

[30] S. E. Cozzens, "What do citations count? The rhetoric-first model," *Scientometrics*, vol. 15, nos. 5–6, pp. 437–447, May 1989.

[31] M. Wang, J. Ren, S. Li, and G. Chen, "Quantifying a paper's academic impact by distinguishing the unequal intensities and contributions of citations," *IEEE Access*, vol. 7, pp. 96198–96214, 2019.

[32] E. Garfeld, "Can citation indexing be automated? in *Proc. Stat. Assoc. Methods Mechanized Document., Symp.*, Washington, DC, USA: Nat. Bureau Standards, Miscellaneous Publication, 1965, p. 269.

[33] M. J. Moravcsik and P. Murugesan, "Some results on the function and quality of citations," *Social Stud. Sci.*, vol. 5, no. 1, pp. 86–92, Feb. 1975, doi: 10.1177/030631277500500106.

[34] M. Garzone and R. E. Mercer, "Towards an automated citation classifier," in *Proc. 13th Biennial Conf. Can. Soc. Computational Stud. Intell.*, vol. 13. Berlin, Germany: Springer, 2000.

[35] S. Teufel, A. Siddharthan, and D. Tidhar, "Automatic classification of citation function," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2006, pp. 103–110.

[36] S. Agarwal, L. Choubey, and H. Yu, "Automatically classifying the role of citations in biomedical articles," in *Proc. AMIA Annu. Symp. Amer. Med. Informat. Assoc.*, 2010, p. 11.

[37] M. Reyhani Hamedani, S.-W. Kim, and D.-J. Kim, "SimCC: A novel method to consider both content and citations for computing similarity of scientific papers," *Inf. Sci.*, vols. 334–335, pp. 273–292, Mar. 2016.

[38] R. Kulkarni and E. Di Minin, "Automated retrieval of information on threatened species from online sources using machine learning," *Methods Ecol. Evol.*, vol. 12, no. 7, pp. 1226–1239, Jul. 2021.

[39] K. Bakhti, Z. Niu, A. Yousif, and A. S. Nyamawe, "Citation function classification based on ontologies and convolutional neural networks," in *Proc. Int. Workshop Learn. Technol. Educ. Cloud*, vol. 7. Cham, Switzerland: Springer, Aug. 2018, pp. 105–115.

[40] J. Smith, A. Johnson, and R. Davis, "The impact of citation characteristics on the credibility and reliability of scientific papers," *IEEE Trans. Sci.*, vol. 12, no. 3, pp. 123–138, Nov. 2020.

[41] D. Pride and P. Knoth, "Incidental or influential? A decade of using text-mining for citation function classification," in *Proc. 16th Int. Soc. Scientometrics Informat.*, Wuhan, China, 2017, pp. 16–20.

[42] S.-U. Hassan, A. Akram, and P. Haddawy, "Identifying important citations using contextual information from full text," in *Proc. ACM/IEEE Joint Conf. Digit. Libraries (JCDL)*, Jun. 2017, pp. 1–8.

[43] S.-U. Hassan, I. Safder, A. Akram, and F. Kamiran, "A novel machine-learning approach to measuring scientific knowledge flows using citation context analysis," *Scientometrics*, vol. 116, no. 2, pp. 973–996, Aug. 2018.

[44] J. Yang, D. She, Y.-K. Lai, P. L. Rosin, and M.-H. Yang, "Weakly supervised coupled networks for visual sentiment analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7584–7592.

[45] S. Nazir, M. Asif, S. Ahmad, H. Aljuaid, R. Iftikhar, Z. Nawaz, and Y. Y. Ghadi, "Important citation identification by exploding the sentiment analysis and section-wise in-text citation weights," *IEEE Access*, vol. 10, pp. 87990–88000, 2022.

[46] A. Cohan, W. Ammar, M. van Zuylen, and F. Cady, "Structural scaffolds for citation intent classification in scientific publications," 2019, *arXiv:1904.01608*.

[47] X. Huang, Y. Hu, Y. Wang, B. Liu, and S. Liu, "Deep learning-based quantitative steganalysis to detect motion vector embedding of HEVC videos," in *Proc. IEEE 5th Int. Conf. Data Sci. Cyberspace (DSC)*, Jul. 2020, pp. 150–155.

[48] Y. Zhang, Y. Wang, Q. Z. Sheng, A. Mahmood, W. E. Zhang, and R. Zhao, "TDM-CFC: Towards document-level multi-label citation function classification," in *Proc. Int. Conf. Web Inf. Syst. Eng.*, vol. 22. Cham, Switzerland: Springer, 2021.

[49] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous graph attention network," in *Proc. World Wide Web Conf.*, 2019, pp. 2022–2032.

[50] I. Beltagy, A. Cohan, and K. Lo, "SciBERT: Pretrained contextualized embed-dings for scientific text," 2019, *arXiv:1903.10676*.

[51] D. Mercier, S. T. R. Rizvi, V. Rajashekar, A. Dengel, and S. Ahmed, "ImpactCite: An XLNet-based method for citation impact analysis," 2020, *arXiv:2005.06611*.

[52] S. Muppidi, S. K. Gorripati, and B. Kishore, "An approach for bibliographic citation sentiment analysis using deep learning," *Int. J. Knowl.-based Intell. Eng. Syst.*, vol. 24, no. 4, pp. 353–362, Jan. 2021.

[53] A. Yousif, Z. Niu, J. Chambua, and Z. Y. Khan, "Multi-task learning model based on recurrent convolutional neural networks for citation sentiment and purpose classification," *Neurocomputing*, vol. 335, pp. 195–205, Mar. 2019.

[54] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[55] R. C. Prati, G. E. A. P. A. Batista, and D. F. Silva, "Class imbalance revisited: A new experimental setup to assess the performance of treatment methods," *Knowl. Inf. Syst.*, vol. 45, no. 1, pp. 247–270, Oct. 2015.

[56] J. Beel, B. Gipp, and S. Langer, "Research paper recommender systems: A literature survey," *Int. J. Digital Libraries*, vol. 17, no. 4, pp. 305–338, 2016.

[57] F. Saboor, M. I. Malik, and K. N. Qureshi, "Recommendation system for research articles: A survey," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 6, pp. 2533–2551, 2020.

[58] A. Kumar et al., "Stop words removal and stemming for sentiment analysis on social media text," *Proc. Comput. Sci.*, vol. 122, pp. 324–331, Jan. 2017.

[59] Y. Mu, H. R. Tizhoosh, R. M. Tayebi, C. Ross, M. Sur, B. Leber, and C. J. V. Campbell, "A BERT model generates diagnostically relevant semantic embeddings from pathology synopses with active learning," *Commun. Med.*, vol. 1, no. 1, p. 11, Jul. 2021.

[60] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin, "Advances in pre-training distributed word representations," 2017, *arXiv:1712.09405*.

ARSHAD IQBAL received the M.S. degree in computer science from the Arid Agriculture University Rawalpindi, in 2018. He is currently pursuing the Ph.D. degree in computer science with the Institute of Computing, Kohat University of Science and Technology. He is also a programmer in Government organizations for the last seven years. His research interests include citation analysis, recommending relevant documents, expert systems, information systems, deep learning, and natural language processing.

ABDUL SHAHID received the Ph.D. degree in computer science from the Capital University of Science and Technology, Islamabad, Pakistan. He is currently a Faculty Member with the Institute of Computing, Kohat University of Science and Technology, Pakistan. He is also a professional software engineer and a consultant with software companies for the last 13 years. His research interests include information systems, and digital libraries, and recommending relevant documents with the help of in-text citation frequencies and patterns. In this field, he has authored or coauthored a number of good quality articles in different international conferences and journals.

MUHAMMAD ROMAN received the Ph.D. degree in computer science from the Kohat University of Science and Technology (KUST), Pakistan. He is currently a Research Fellow with the School of EEECS, Queens University Belfast, U.K., and a Faculty Member with KUST. His research interests include large language models, recommending relevant documents, information systems, deep learning, and natural language processing. He has authored a number of publications in quality journals in this area. Besides his research activities, he has worked on various software development projects for the last 13 years.

MUHAMMAD TANVIR AFZAL received the M.Sc. degree (Hons.) in computer science from Quaid-i-Azam University, Islamabad, Pakistan, and the Ph.D. degree (Hons.) in computer science from the Graz University of Technology, Austria. Previously, he was with the Capital University of Science and Technology, Islamabad, as a Professor, an Associate Professor, and an Assistant Professor of computer science. Furthermore, he was with NESCOM, COMSATS University Islamabad, and JinTech Islamabad. He was a master trainer and the program director of a national level training for a public sector organization in Pakistan on human factors engineering and conducted the training of over 100 hours for experts from the industry. He has been associated with academia and industry at various levels for the last 20 years. He is currently a Professor and the Director QEC of the Department of Computer Science, Namal Institute Mianwali. He conducted more than 100 curricular, co-curricular, and extra-curricular activities in the last five years, including seminars, workshops, national competitions (ExcITeCup), and invited international and national speakers from Google, Oracle, IICM, IFIS, and SEGA Europe. Under his supervision, more than 60 postgraduate students (M.S. and Ph.D.) have defended their research theses successfully, and a number of Ph.D. and M.S. students are pursuing their research with him. He has authored more than 120 research articles, including 50 published articles in impact factor leading journals in the field of data science, information retrieval and visualization, semantics, digital libraries, and scientometrics. He has authored two books and has edited two books in computer science. His cumulative impact factor is more than 110, with citations over 770. He played a pivotal role in making collaborations between MAJU-JUCS, MAJU-IICM, and TUG-UNIMAS. He was a recipient of multiple international research funding's. He served as the Ph.D. symposium chair, the session chair, the finance chair, a committee member, and an editor for several IEEE, ACM, Springer, and Elsevier international conferences and journals. He is serving as the Editor-in-Chief for reputed impact factor journal, such as *Journal of Universal Computer Science*.

MUHAMMAD YAHYA received the B.S. degree in computer science from KP Agricultural University, Peshawar, in 2013, and the M.S. degree in electrical and electronic engineering from Universiti Kuala Lumpur, in 2019. He is currently pursuing the Ph.D. degree with the Confirm SFI Research Centre for Smart Manufacturing, University of Galway, Ireland, under the supervision of Prof. John Breslin and Dr. Intizar Ali. He has published in major venues, such as *Semantic Web* journal, ESWC, and ISWC.