Article

Classification of Research Papers on Radio Frequency Electromagnetic Field (RF-EMF) Using Graph Neural Networks (GNN)

Youngsun Jang [1], Kwanghee Won [1,*], Hyung-do Choi [2] and Sung Y. Shin [1]

[1] Electrical Engineering and Computer Science Department, South Dakota State University, Brookings, SD 57007, USA; youngsun.jang@sdstate.edu (Y.J.); sung.shin@sdstate.edu (S.Y.S.)
[2] Radio Research Division, Electronics and Telecom Research Institute, Daejeon 34129, Republic of Korea; choihd@etri.re.kr
* Correspondence: kwanghee.won@sdstate.edu

Abstract: This study compares the performance of graph convolutional neural network (GCN) models with conventional natural language processing (NLP) models for classifying scientific literature related to radio frequency electromagnetic field (RF-EMF). Specifically, the study examines two GCN models: BertGCN and the citation-based GCN. The study concludes that the model achieves consistently good performance when the input text is long enough, based on the attention mechanism of BERT. When the input sequence is short, the composition parameter λ, which combines output values of the two subnetworks of BertGCN, plays a crucial role in achieving high classification accuracy. As the value of λ increases, the classification accuracy also increases. The study also proposes and tests a simplified variant of BertGCN, revealing performance differences among the models under two different data conditions by the existence of keywords. This study has two main contributions: (1) the implementation and testing of a variant of BertGCN and citation-based GCN for document classification tasks related to radio frequency electromagnetic fields publications, and (2) the confirmation of the impact of model conditions, such as the existence of keywords and input sequence length, in the original BertGCN. Although this study focused on a specific domain, our approaches have broader implications that extend beyond scientific publications to general text classification.

Keywords: document classification; BERT; graph convolution neural network; electromagnetic fields

1. Introduction

Recent efforts have been made to systematically classify the scientific literature on the radio frequency electromagnetic field (RF-EMF) and extract key information on its potential health effects [1]. Despite the urgency of these efforts, there was no practical method available to enhance the extensive manual document classification task carried out by experts. The labor-intensive process of manual classification by experts is limited by biased evaluation and time inefficiency when compared to an automated classification system. To alleviate these issues, this study explores deep learning-based approaches to document classification in the field of RF-EMF. This work can contribute to the improvement of decision-making systems for experts or authorities on public health related to RF-EMF.

Our previous studies have demonstrated the effectiveness of natural language processing (NLP) models [2,3] in classifying scientific papers based on their experimental conditions, such as in vivo and in vitro, and extracting information from them. The classification of publications on RF-EMF into in vivo and in vitro categories is valuable because it plays a crucial role in the early stages of the systematic review process of RF-EMF literature [1]. The term "in vivo" pertains to research conducted on living organisms, often involving animal experimentation. The term "in vitro" refers to studies performed outside of a living organism, typically in a cellular context. This classification aids in the extraction

of relevant information for subsequent analysis. For instance, if a study is in vivo, it is possible to glean details such as specific animal species and the number of animals used.

However, these models are limited to analyzing individual documents and do not take into account the relationships between papers. To address this limitation, we propose a novel approach that utilizes a graph structure to explore the inter-document relationships among papers. Specifically, we apply this approach to the corpus of RF-EMF-related literature, which is relatively small compared to general document corpora and technical vocabulary.

This study aims to improve existing document classification approaches by utilizing graph convolutional neural networks (GCN) [4]. In this approach, nodes in the graph network represent articles, and weighted edges capture relationships between them. We hypothesize that citation practices play a significant role in document classification [5–7]. In addition, we compare the performance of GCN with Transformer-based BERT, which is a standard model used in NLP classification tasks. [8,9].

For the experiment, we constructed two types of datasets: (a) abstracts of papers on RF-EMF and their corresponding in vivo/in vitro labels, and (b) citation data for the papers. While the citation information can be utilized to build a citation graph, a statistically connected graph can be created by harnessing the frequency and co-occurrence property between documents and words [10]. Our study evaluates the efficiency of both these approaches for document classification. Furthermore, we draw upon the PECO framework, a decision tree used for evaluating RF-EMF studies [1,3], for the development of our work.

The primary contributions of this study are as follows:

- First, we implemented and tested a variant of BertGCN, called modified BertGCN, and a citation-based GCN for document classification tasks on RF-EMF-related publications. The modified BertGCN has a simplified architecture by removing the composition parameter of the original BertGCN. Additionally, we demonstrated how the citation relation of scientific publications can be utilized for document classification in the context of citation-based GCN.
- Second, we designed experiments to investigate whether graph-based deep learning helps classify paper abstracts in the absence of direct keywords in the text, i.e., without clues and with a limited length of information available (i.e., short sentence lengths as input to the models). These experiments validated the impacts of various conditions on the performance of the original BertGCN.

2. Related Works

We hypothesized that certain types of relationships between scientific publications could influence classification performance, as previously discussed in our work [2]. Since citations are a well-known form of interrelations between papers, we explored the application of citation graphs and citation analysis to NLP, specifically in document classification [11,12]. Since the early and mid-2000s, several studies [5,6,13] have shown that analyzing relationships between publications, such as citation information, can significantly improve the classification performance of scientific literature, rather than solely focusing on each isolated literature. With the study of Egghe and Rousseau (1990), Infometrics defined the citation graph as a representation of a document d_i that cites a document d_j through an arrow from node d_i to d_j [14]. By extending this notation to all documents in a collection, a directed citation graph can be constructed.

Several studies have employed machine learning techniques for document classification using citation graph structures. For instance, Nguyen and Do (2017) applied Latent Dirichlet allocation (LDA) on a large-scale citation network [7]. Li et al., in one of the earliest studies on document classification that leveraged the relations among nodes in a citation graph, proposed an efficient kernel method based on the Support Vector Machine (SVM) [6]. The kernel method is a learning algorithm that relies on a similarity value, which is represented as the kernel values. To improve the representation of the citation network, they utilized graph kernels, which outperformed the contemporary model of simply directed citations by capturing the richer features of the citation network context.

Co-citations in citation networks have also been studied by several researchers [5,15]. For example, Kolchinsky et al. (2010) developed a Naïve Bayes classifier using features extracted from the network [5]. Their study is noteworthy as it achieved encouraging performance and focused on a research topic similar to our own. The study also used PubMed publication data for classifying biomedical literature. However, applying their method to our research is challenging as their model is limited to binary classification, whereas our work aims to classify at least three groups.

A number of studies have employed deep neural networks (DNN) and reported impressive performance. Notably, some approaches involve developing DNN models with a graph structure. This concept has its roots in the pioneering work by Kipf and Welling (2017) [4], who applied the classical image convolution concept to datasets with graph structures such as citation relationships. They developed an improved Graph Convolution Neural Network (GCN) model, which has paved the way for various follow-up studies by overcoming the limitations of classical spatial convolution. Spatial convolution in graph networks faced two critical issues: exploding similarity of nodes and edge information loss. To address the issues primarily caused by the Laplacian regularizer in spatial convolution, they employed Fourier transformation to extract one of the candidate graphs with the lowest node similarities. This novel convolution, based on such a spectral perspective, achieved excellent performance in node classification tasks.

After Kipf and Welling introduced the GCN model, researchers made various attempts to apply it to NLP tasks. Among these, TextGCN has been the most successful [10,16]. The authors of TextGCN, Yao et al. (2018), proposed a method to improve document classification in NLP using GCN. Their study specifies how to construct node features consisting of all the words and documents and how to relate those nodes to build edges and their corresponding weight values. To calculate weight values, they applied several NLP techniques such as (a) positive pointwise mutual information (PPMI) between two different words and (b) term frequency-inverse document frequency (TF-IDF) between a word and a document.

Several studies have been conducted since the release of TextGCN to further improve the classification performance by extracting features from better language models. The main objective of these studies is to identify optimal conditions by combining various models from the NLP, CNN, and GCN fields. Dong et al. (2022) combined GCN with the bidirectional gated recurrent unit (Bi-GRU) model to enhance classification accuracy [17]. Yang et al. (2022) employed Bi-LSTM to improve classification performance [18]. They used GCN to capture co-occurrence information between words, words, and documents and Bi-LSTM to capture local context information. The attention mechanism was employed to capture long-distance semantic information.

In recent years, researchers have explored the integration of the Bidirectional Encoder Representations from Transformers (BERT) model into graph neural networks. Lin et al. (2022) proposed a mixing model of BERT and TextGCN, known as BertGCN. The model incorporated BERT embeddings as features into the transductive graph neural network, which is considered state-of-the-art in recent research [19–21]. The authors of BertGCN observed a slight improvement in document classification by replacing the document features of TextGCN with BERT embeddings. Furthermore, they demonstrated the optimal value of the composition parameter λ for achieving the best model performance, which has significant implications for our work.

3. Model and Dataset

3.1. Model

This study presents a modified BertGCN that achieves comparable performance to the original model without the need for various parameter conditions. We simplify the hyperparameter search step by replacing the final layers of the BERT and GCN subnetworks with one fully-connected layer. Additionally, we adopt a citation-based GCN, which can capture inter-document relationships by constructing edges based on the citation information from papers.

Furthermore, we designed and tested experiments to investigate whether the original BertGCN validation performance is prone to be influenced by additional factors, such as the composition parameter setting (λ from 0.0 to 1.0), length of the input sequence (length of 128 and 280 tokens), and keywords (clued and un-clued abstracts).

3.1.1. BertGCN: Statistical Graph with BERT Features

The BertGCN consists of three graph convolution layers, namely the input layer, the hidden layer, and the output layers, along with a dropout layer. After initializing the graph network, it is subsequently trained using the 'RoBERTa-based' model specification, as previously reported [22]. The feature information of nodes is transferred from the BERT embedding, and the pre-trained 'RoBERTa-based' model is imported for this purpose. In the next step, the pre-trained BERT model is fine-tuned using our own dataset, which is labeled as in vivo, in vitro, and other. Once the fine-tuning process for BERT is complete, the feature is updated, and the extracted BERT features are transferred back to the BertGCN to construct node features (Figure 1a). The BertGCN follows the same structure as the TextGCN, comprising node features (i.e., words and documents) and weighted edges, as per a previous study [10].

The multiplier lambda adjusts the trade-off between the two objectives, where $\lambda = 1$ corresponds to the full BertGCN model and $\lambda = 0$ corresponds to using only the BERT module [19]:

$$Z \;=\; \lambda \, Z_{GCN} \;+\; (1 - \lambda) \, Z_{BERT} \tag{1}$$

The terms Z_{GCN} and Z_{BERT} represent the classification output of each model, with the aggregated Z reflecting the combined output that incorporates both models. In this case, the performance of the full BERT model with a composition parameter of 0 ($\lambda = 0$) serves as the baseline for comparison.

3.1.2. Modified BertGCN

The original BertGCN combines the outputs of the BERT-subnetwork and the GCN-subnetwork at the logit level using a composition parameter λ. However, the optimal value of λ depends on the data, network architecture, and other hyperparameters, and the parameter search process proposed by the authors to obtain it requires ten times more computational resources than the case without this parameter. To address this limitation, we made a minor modification to remove λ from the BertGCN. As shown in Figure 1b, the final layers of the BERT subnetwork and the GCN subnetwork are replaced with 9-dimensional linear layers that have a Rectified Linear Unit (ReLU) activation function and are concatenated. Then, an output layer with three classes is connected to the linear layer. This modification allows the additional layer to learn how to combine the intermediate outputs from the two subnetworks for optimal decision-making.

3.1.3. Citation-Based Graph with BERT Features

In this study, a citation graph was used to represent academic papers and their citation relationships. Each node was assigned an index and a feature vector transferred from the BERT model. Edges were established between two papers when one cited the other. Specifically, the intermediate output of the BERT model before the last classification layer was employed as the node feature of the GCN (Figure 1c). The feature vector dimension was 768. The citation-based GCN consisted of a ReLU activation with 256 hidden features, followed by another convolution with 64 hidden features and a linear layer. The same pre-trained BERT model was used for the citation-based GCN in our experiment.
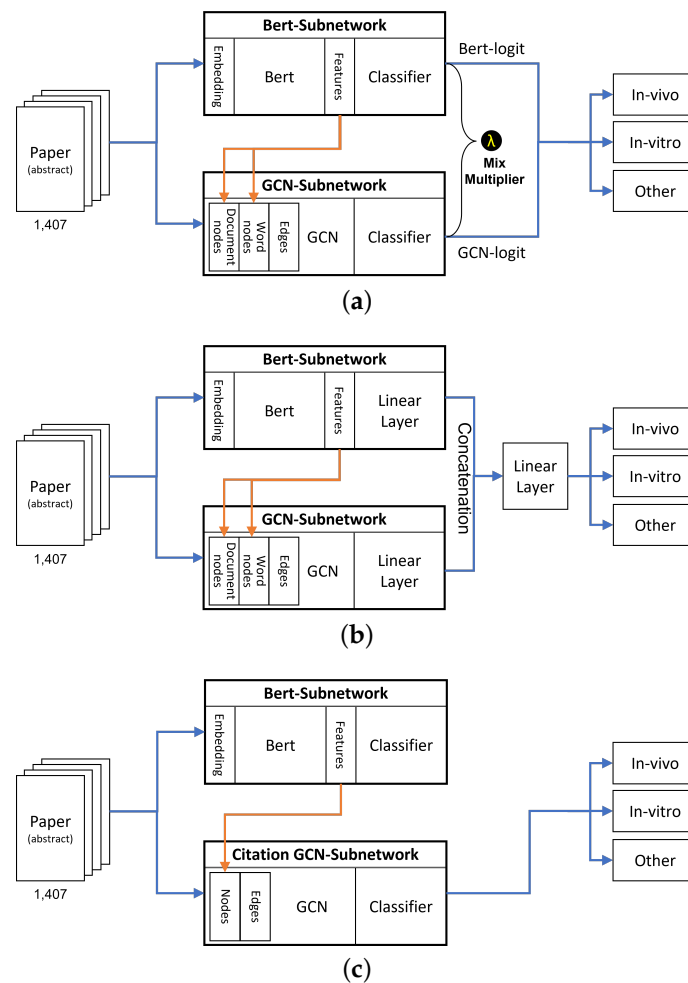
Figure 1. Three modes of GCN-based document classification. (**a**) Statistical GCN (BertGCN). The multiplier lambda adjusts the trade-off between the two objectives, where $\lambda = 1$ corresponds to the full BertGCN model and $\lambda = 0$ corresponds to using only the BERT module [19]; (**b**) Statistical GCN (BertGCN) without λ. Concatenating the output of both models replaces the mixture process with linear layers, which eliminates the need for the mixing parameter λ and simplifies the neural network architecture; (**c**) citation-based GCN. This model constructs a citation graph using the 396 papers and their corresponding citation papers.

3.2. Data

The publication data for this study were obtained from two sources: 347 RF-EMF publications from the open EMF portal database [2], and an additional 108 publications provided by the Electronics and Telecommunications Research Institute in Korea. These publications underwent expert review and labeling, resulting in an initial dataset of 455 papers. The EMF portal dataset was collected from four categories (cells, brain [cognitive function], brain metabolism, and DNA), and any unclassifiable data were excluded. The last access date for the EMF portal was on 22 September 2021. Subsequently, human annotators manually reviewed the abstracts and labeled them into three classes (in vivo, in vitro, and other).

To construct a citation graph, it was necessary to collect citation data between publications. PubMed's unique ID system was used to identify citation relationships between papers. The publications that were not registered in PubMed and any duplicate data were excluded from the dataset. Additionally, publications without citation information in PubMed were also removed. as a result, a total of 396 papers were identified as nodes in the citation graph.

The openly published PubMed database served as an essential resource for collecting citation data for our research. An additional set of 1011 papers was collected by tracing citations from the original 396 papers, and the entire dataset was used to create nodes in the graph neural network. Accessing abstracts of the cited papers was achieved by filtering PubMed IDs from the metadata of each paper. The citation graph (citation-based GCN) was constructed using a total of 1407 paper abstracts as nodes, while the statistical graph (statistical GCN) incorporated both paper abstracts and words as nodes. A total of 1407 paper abstracts were classified into three groups: (a) in vivo (n = 523), (b) in vitro (n = 332), and (c) Other (n = 552).

During the data selection process of the 1011 cited papers, two primary criteria were used. Firstly, this study focused specifically on RF-EMF generated from electronic devices, such as mobile phones and laptops, and excluded radiation sources that are distinct from RF-EMF, including gamma rays. Secondly, papers on ionizing radiation sources were also excluded, as RF-EMF from electronic devices is classified as a non-ionizing radiation source. The selection criteria included microwave radiation and short-wavelength visible light. Notably, the following types of light sources were excluded from the study:

- Gamma rays;
- Pulsed electric fields (either electric fields or pulse);
- Only magnetic field (MF);
- Ionizing radiation (e.g., brain CT scans leading to Alzheimer's and Parkinson's);
- X-rays.

Document nodes represent abstracts of the papers, and the word nodes represent all words used in the entire data set. Word nodes have statistical associations with other words and documents [10]. Both BertGCN and citation-based GCN utilized 1407 documents. However, BertGCN used 1407 document nodes and 33,115-word nodes together, whereas the citation-based GCN only used 1407 document nodes. BertGCN employed the transductive classification method (semi-supervised), which treated word nodes as unlabeled data [20,21].

In summary, the citation graph comprises 1407 nodes, consisting of 396 primary papers and their 1011 cited papers. In contrast, the statistical model includes the 1407 document nodes and their constituent 33,115 words, resulting in 34,522 nodes. The word nodes mentioned in this study refer to tokens, which are the basic units of a sentence in the field of NLP. In the BERT model, a single word is composed of smaller units called tokens, which are used as input for training the model. Additionally, each node has 768-dimensional BERT embeddings as features. The statistical model includes approximately 4 million edges connecting these nodes. Edges are weighted according to the term frequency and co-occurrence between words, and between words and documents as the model training advances [19].

Another critical parameter of the GCNs is the maximum length of the input sentences. This parameter determines the memory footprint of the model limited by the system capability and can also significantly impact the model performance. In this study, we evaluated the model performance with both short and long versions of the input data; the results are in the following section.

4. Experiment and Results

The statistical graph experiment involved varying the sentence lengths in the BERT model, specifically using shorter (128 tokens) and longer (280 tokens) lengths. Although the default length for BertGCN is 128 tokens [19], we studied whether there would be a performance change with increased sentence length. The subsequent section will discuss the effect in detail.

The presence of direct keywords within a document is a significant factor for classification performance. As such, the experiment was performed under two different conditions: an abstract containing critical keywords, such as 'in vivo' or 'in vitro' as clues (clued), and an abstract with no such keywords (un-clued). Out of the 1407 papers, 161 papers contained the critical keywords

(11.4%), while the un-clued experiment involved removing these keywords from the dataset. Of the 161 un-clued papers, 28 were from the in vivo class, 93 were in vitro, and the remaining 40 were 'other'. The performance was measured under various λ conditions according to the model configuration in Section 3 (λ = 0.0 to 1.0 with 0.1 intervals) in both clued and un-clued conditions. Notably, the λ value of zero represents a BERT-only model and served as the baseline.

The study involved multiple trials to evaluate the expected performance. A trial represents one round of training with validation accuracy measurements. The entire dataset was partitioned into disjoint training and testing sets with ratios of 0.75 and 0.25, respectively. Subsequently, in accordance with the experimental design of the transductive BertGCN model [19], a validation set was randomly selected from the training dataset with a ratio of 0.10. For each experimental condition (length, clued/un-clued, and λ condition), there were 40 trials for the statistical GCN and 20 trials for the citation-based GCN. The average validation accuracy was calculated across all trials to select the best model. Subsequently, the performance of the best model was evaluated on the test dataset.

The initial graph network was constructed by setting the word embedding dimension to 768, corresponding to the BERT embedding. Word co-occurrence was calculated during edge construction for the statistical graph using a sliding window of 20. The model configuration of the BertGCN consisted of a batch size of 25 for a length of 128 and a batch size of 40 for a length of 280, with early stopping applied during 50 epochs. The model had one hidden layer with a default of 200 dimensions, a dropout of 0.5, a learning rate of 1×10^{-5}, and employed an Adam optimizer. We chose these values empirically while keeping within the memory limit of the hardware. For fine-tuning the BERT model, the 'RoBERTa-base' was imported as a pre-trained model with a maximum input length of 128 tokens and 280 tokens, a learning rate of 1×10^{-5}, and an Adam optimizer. The experiment was conducted using CUDA version 10.1 and Nvidia Titan RTX (24 GB).

4.1. Statistical BertGCN with a Shorter Input Sequence

Table 1 shows that the baseline model attained a validation accuracy of 85.53% under the clued condition, which is lower than all other λ cases. In contrast, the BertGCN model exhibited higher performance than the baseline across all λ conditions ranging from 0.1 to 1.0. The highest validation accuracies for the clued cases were achieved at λ conditions of 0.6 and 0.8, with values of 87.11% and 87.20%, respectively. The test accuracy of the best model was 81.77% for the clued case.

The results also revealed that the baseline model achieved a validation accuracy of 86.13% for the un-clued dataset, where no keyword is given to some documents. While this performance is respectable, it falls short of most other λ cases, except for an outlier of 0.1. The BertGCN model achieved better performance than the baseline across all λ conditions ranging from 0.2 to 1.0, and the highest validation accuracy was achieved at a λ condition of 0.4 with a value of 87.11%. The test accuracy of the best model was 82.34% for the un-clued case.

In summary, Table 1 shows that when training a model with a fixed input length of 128 tokens, increasing the composition parameter λ led to better overall model performance. This result is consistent with BertGCN's findings that the gap between the highest and lowest λ values was about 2%. However, no significant gap was revealed between the clued and un-clued datasets.

Table 1. Validation accuracy by λ (length = 128 tokens).

	λ = 0.0	λ = 0.1	λ = 0.2	λ = 0.3	λ = 0.4	λ = 0.5	λ = 0.6	λ = 0.7	λ = 0.8	λ = 0.9	λ = 1.0
Clued	0.8553	0.8569	0.8585	0.8661	0.8663	0.8679	0.8711	0.8665	0.8720	0.8644	0.8663
Un-clued	0.8613	0.8535	0.8635	0.8606	0.8711	0.8645	0.8629	0.8668	0.8665	0.8663	0.8674

4.2. Statistical BertGCN with a Longer Input Sequence

The longer model with 280 tokens showed a relatively consistent performance pattern (Figure 2), with no significant difference observed in the performance when varying the

composition parameter λ. The baseline model achieved a validation accuracy of approximately 89.50% for both the clued and un-clued datasets, which was close to the highest accuracy. The clued model performed best at a λ condition of 0.8, reaching a validation accuracy of 89.63%, while the un-clued model achieved its highest accuracy at a λ condition of 0.3, with a value of 89.79%. The test accuracy of the best model was 86.61% for the clued case and 85.75% for the un-clued case, respectively. Conversely, the lowest accuracy for both datasets was observed at the λ condition of 1.0, with the clued and un-clued models recording values of 88.16% and 88.67%, respectively (Table 2).

Table 2. Validation accuracy by λ (length = 280 tokens).

	$\lambda = 0.0$	$\lambda = 0.1$	$\lambda = 0.2$	$\lambda = 0.3$	$\lambda = 0.4$	$\lambda = 0.5$	$\lambda = 0.6$	$\lambda = 0.7$	$\lambda = 0.8$	$\lambda = 0.9$	$\lambda = 1.0$
Clued	0.8949	0.8874	0.8879	0.8839	0.8879	0.8897	0.8950	0.8847	0.8963	0.8913	0.8816
Un-clued	0.8959	0.8892	0.8931	0.8979	0.8885	0.8922	0.8885	0.8870	0.8917	0.8929	0.8867

In summary, Table 2 shows that the performance of a model trained with a fixed 280 tokens does not improve by varying λ. On the contrary, it decreases as λ approaches 1.0. These findings suggest that λ does not play a significant role when Bert embeddings can be fully used.
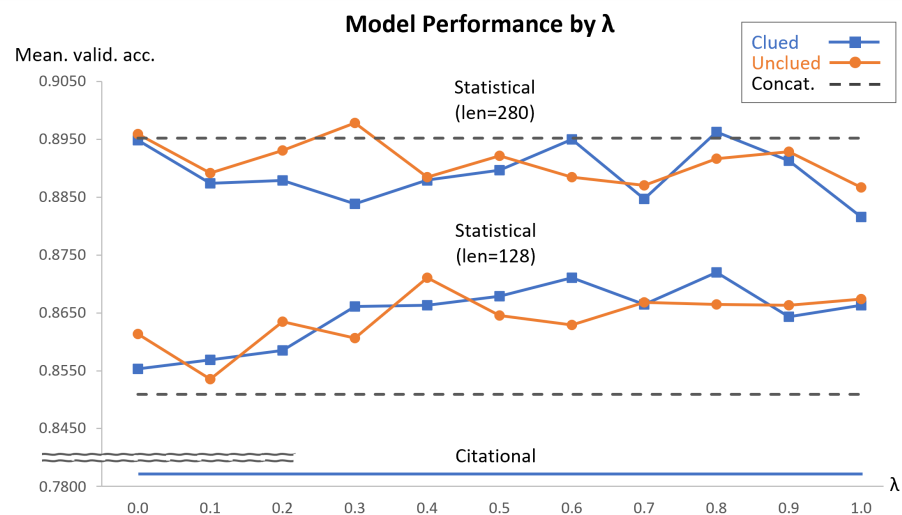

Figure 2. Document classification performance of the three modes. This shows the document classification performances of the three modes, including the modified BertGCN, the original BertGCN, and the citation-based GCN models.

4.3. Modified BertGCN and Citation Graph

The simplified BertGCN model, in contrast to the original BertGCN model, requires no λ variation setting and produces a single-valued performance. The longer case yielded an average validation accuracy of 89.52% (the test accuracy of the best model was 84.33%), outperforming most statistical conditions. Conversely, the shorter case achieved an average validation accuracy of only 85.09% (the test accuracy of the best model was 86.04%), which is the lowest compared to any statistical condition. These results demonstrate a clear contrast in the model performance, with the shorter case approaching the lowest validation accuracy observed in the original BertGCN experiment, and the longer case approaching the highest. On the other hand, the citation-based GCN model achieved a mean validation accuracy of 78.83% under the clued condition with 280-token input sentences in 20 trials.

5. Discussion

It is worth noting that the modified BertGCN model with longer input length achieved a performance comparable to the highest validation accuracy achieved by the original

BertGCN. This finding indicates that it can consistently achieve high validation accuracy without relying on the composition parameter, simply by connecting the last layers of the BERT and GCN models. This idea deserves consideration when aiming to surpass existing BERT document classification performance using GCN, to implement a practical system.

However, it was found that the citation graph exhibited a lower performance than the statistical GCN, achieving a validation accuracy of about 10% lower. The smaller size of the nodes and edges constituting the citation graph may explain this gap, but further research is necessary to determine the exact reason.

As for various factors influencing the original BertGCN, The length of input data fed to the model is a primary factor for classification performance in BertGCN. The validation performance disparity of up to approximately 4% between the two length conditions of the input sequence (128, 280 tokens) is rooted in the richer BERT embeddings (Figure 2). This is because as the input sequence length increases, BERT can capture more accurate and informative feature embeddings with the help of the multi-headed self-attention mechanism. The longer sentence condition is slightly less sensitive to the λ variation compared to the shorter condition. While the shorter case shows a gradual validation accuracy increase from λ of 0.0 to 0.8, the longer case exhibits a bumpy pattern that requires more trials to establish a stable pattern. The shorter-sequence model shows a deviation of approximately 2%, whereas there is no significant difference under the long-sentence condition except for the sudden drop near λ of 1.0. Therefore, it implies that the importance of the composition parameter λ increases as the length of input sequences decreases.

In the BertGCN model, the final output is a combination of the logit values from the BERT and GCN subnetworks. The observed trend of increasing validation accuracy with higher values of λ for shorter input sequences suggests that the GCN model has a greater impact on improving classification accuracy than the BERT model as the value of λ increases.

The experiments revealed no noticeable performance gap between clued and unclued conditions. The observed improvement in classification performance without critical keywords can also be attributed to the ability of Transformer-related models to capture contextual information during training using multi-headed attention mechanisms, enabling accurate classification even in the absence of explicit keyword cues, such as "in vivo" or "in vitro".

6. Conclusions

While the modified BertGCN demonstrated consistently high performance when an adequate length of input sequence was provided, the citation-based GCN's performance was modest due to its relatively small graph size. This issue could potentially be mitigated with an increase in data. Given the limited scientific literature datasets, exploring data augmentation techniques could be a promising area for future research. Furthermore, employing an inductive approach may offer the potential to improve classification accuracy [20,21], which will be explored in our follow-up study.

This study confirmed that the existing BertGCN model effectively classifies documents even in the absence of keywords by capturing document context based on a multi-headed attention mechanism. However, Transformer-based models have limited contextual understanding in cases of short input lengths. The findings suggest that GCN is a more effective contributor than existing BERT models for short-input contexts (i.e., length = 128 tokens). Additionally, BERT models have limitations in handling long sentences, which is a critical issue that needs addressing not only for abstract classification but also for entire scientific paper classification. This limitation is even more critical when classifying other long and general texts, partly related to hardware specifications such as GPUs. Addressing this issue will be part of our subsequent research.

Although this study focused on RF-EMF experiments in a specific domain of science, our models have broader implications for text classification beyond scientific publications. In subsequent studies, we plan to explore the applicability of our models to a wider range of ordinary documents.

References

1. Henschenmacher, B.; Bitsch, A.; de Las Heras Gala, T.; Forman, H.J.; Fragoulis, A.; Ghezzi, P.; Kellner, R.; Koch, W.; Kuhne, J.; Sachno, D.; et al. The effect of radiofrequency electromagnetic fields (RF-EMF) on biomarkers of oxidative stress in vivo and in vitro: A protocol for a systematic review. *Environ Int.* **2022**, *158*, 1–11.
2. Won, K.; Jang, Y.; Choi, H-d.; Shin, S. Design and implementation of information extraction system for scientific literature using fine-tuned deep learning models. *ACM SIGAPP Appl. Comput. Rev.* **2022**, *22*, 31–38.
3. Jang, Y.; Won, K.; Choi, H-d.; Shin, S. Deep Learning Models for Multiple Answers Extraction and Classification of Scientific Publications. In Proceedings of ACM 2022 International Conference on Research in Adaptive and Convergent Systems Conference, Online, 3–6 October 2022; pp. 185–190.
4. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the 5th International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
5. Kolchinsky, A.; Abi-Haidar, A.; Kaur, J.; Hamed, A.A.; Rocha L.M. Classification of protein-protein interaction full-text documents using text and citation network features. *TCBB* **2010**, *7*, 400–411.
6. Li, X.; Chen, H.; Zhang, Z.; Li, J. Automatic patent classification using citation network information: an experimental study in nanotechnology. In Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, New York, NY, USA, 7–11 June 2007; pp. 419–427.
7. Nguyen, T.; Do, P. Managing and Visualizing Citation Network Using Graph Database and LDA Model. In Proceedings of the Eighth International Symposium on Information and Communication Technology, Nha Trang, Vietnam, 7–8 December 2017; pp. 100–105.
8. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
9. Devlin, J.; Chang, M-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL-HLT 2019, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
10. Yao, L.; Mao, C.; Luo, Y. Graph Convolutional Networks for Text Classification. In Proceedings of the 33rd AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 7370-7377.
11. Ji, T.; Self, N.; Fu, K.; Chen, Z.; Ramakrishnan, N.; Lu, C.-T. Dynamic Multi-Context Attention Networks for Citation Forecasting of Scientific Publications. In Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, CA, USA, 2–9 February 2021; pp. 7953–7960.
12. Ji, T.; Chen, Z.; Self, N.; Fu, K.; Lu, C.-T.; Ramakrishnan, N. Patent Citation Dynamics Modeling via Multi-Attention Recurrent Networks. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 2621–2627.
13. Rashed, A.; Grabocka, J.; Schmidt-Thieme, L. Citation Multi-Relational Classification via Bayesian Ranked Non-Linear Embeddings. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 4–8 August 2019; pp. 1132–1140.
14. Egghe, L.; Rousseau, R. *Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science*, 1st ed.; Elsevier Science Publishers: Amsterdam, The Netherlands, 1990; pp. 228–230.
15. Kobayashi, Y.; Shimbo, M.; Matsumoto, Y. Citation Recommendation Using Distributed Representation of Discourse Facets in Scientific Articles. In Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries, Fort Worth, TX, USA, 3–7 June 2018; pp. 243–251.

16.   Liu, X.; You, X; Zhang, X.; Wu, J.; Lv, P. Tensor Graph Convolutional Networks for Text Classification. In Proceedings of the 34th AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 8409-8416.

17.   Dong, Y.; Yang, Z.; Cao, H. A Text Classification Model Based on GCN and BiGRU Fusion. In Proceedings of the 8th International Conference on Computing and Artificial Intelligence, Tianjin, China, 18–21 March 2022; pp. 318–322.

18.   Yang, Y.; Cui, Q.; Ji, L.; Cheng, Z. Graph Convolution Word Embedding and Attention for Text Classification. In Proceedings of the 6th International Conference on Machine Learning and Soft Computing, New York, NY, USA, 15–17 January 2022; pp. 160–166.

19.   Lin, Y.; Meng, Y.; Sun, X.; Han, Q.; Kuang, K.; Li, J.; Wu, F. BertGCN: Transductive Text Classification by Combining GNN and BERT. In Findings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021), Bangkok, Thailand, 1–6 August 2021; pp. 1456–1462.

20.   Huang, Y-H.; Chen, Y-H.; Chen, Y-S. ConTextING: Granting Document-Wise Contextual Embeddings to Graph Neural Networks for Inductive Text Classification. In Proceedings of the 29th International Conference on Computational Linguistics, Gyeongju, Republic of Korea, 12–17 October 2022; pp. 1163–1168.

21.   Wang, K.; Han, S.C.; Poon, J. InducT-GCN: Inductive Graph Convolutional Networks for Text Classification. In Proceedings of 26th International Conference on Pattern Recognition, Montréal, QC, Canada, 21–25 August 2022; pp. 1243–1249.

22.   Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. In Proceedings of the 8th International Conference on Learning Representations, Addis Ababa, Ethiopia, 30 April 2020.