Journal of King Saud University – Computer and Information Sciences 35 (2023) 101578

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Integrated deep learning paradigm for document-based sentiment analysis

Peter Atandoh [a,*], Fengli Zhang [a], Daniel Adu-Gyamfi [b], Paul H. Atandoh [c], Raphael Elimeli Nuhoho [d]

[a] School of Information and Software Engineering, University of Electronic Science and Technology of China, North Jianshe Road, Chengdu 610054, Sichuan, China
[b] Department of Cyber Security and Computer Engineering Technology, C. K. Tedam University of Technology and Applied Sciences, Navrongo, UE/R, Ghana
[c] Department of Mathematics, Mercer University, 1501 Mercer University Drive, Macon, GA 31207, USA
[d] School of Computer Science and Engineering, University of Electronic Science and Technology of China, No. 2006, Xiyuan Avenue, Chengdu 610054, Sichuan, China

ARTICLE INFO

Article history:
Received 27 March 2023
Revised 12 May 2023
Accepted 14 May 2023
Available online 25 May 2023

Keywords:
Sentiment Analysis
Deep Learning
Bidirectional Encoder from Transformers
Convolutional Neural Network

ABSTRACT

An integrated deep learning paradigm for the analysis of document-based sentiments is presented in this article. Generally, sentiment analysis has enormous applications in the real world, particularly in e-commerce and/or cloud computing-oriented businesses. Integrated deep learning paradigms for document-based sentiment analysis seek to efficiently categorize the polarity of contextual sentiments into positive, negative, and neutral to aid organizations in making informed decisions. Nonetheless, the sparsity of text and disambiguation of natural languages make it relatively difficult for existing methods to provide precise identification and extraction when subjected to document-based data. As a result, this study introduces BERT-MultiLayered Convolutional Neural Network (B-MLCNN) as a computationally viable integrated deep learning paradigm. The B-MLCNN considers the overall textual review as a single document and classifies the available sentiments. First, the BERT pre-trained language model handles the feature vector representation and captures any global features. Further, the multi-layered convolutional neural network (MLCNN) with different kernel dimensions handles feature extraction. A softmax function produces classification results. The experimental setup with B-MLCNN based on IMDB movie reviews, 2002 movie reviews, 2004 movie reviews, and Amazon review datasets achieved accuracies of 95%, 88%, 95%, and 95% respectively, which promises to be efficient to deploy in practical applications.

1. Introduction

E-commerce technology is rapidly evolving and is being used more regularly, which is making online shopping a more appealing option for customers.

In contrast to brick-and-mortar stores, consumers can make purchases when- ever and from wherever they desire. The fact that they can go shopping throughout the week rather than on the weekend is also convenient. In addition, e-commerce sites provide customers with a plethora of products to choose from, allowing them to do their shopping in the comfort of their own homes regardless of the weather (Liang and Wang, 2019). However, while on– line shopping provides convenience to consumers, the virtuality of products sold on e-commerce sites includes mismatches in product descriptions and what customers receive, low-quality items, and a lack of follow-up support, to name a few (Ji et al., 2019). The World Wide Web contains a massive volume of opinion writing on items and services, and sorting through all of it to learn about the product is a challenging effort. As a result, it is critical for the discipline of sentiment analysis to automatically extract and structure positive and negative opinions from the web (Atandoh et al., 2021). A customer's emotional disposition can be automatically determined through the use of sentiment analysis for product reviews. This is accomplished by matching the emotive tone and the subjective comment language (Zeng et al., 2019). As of late, the deep learning-based categorization method has become very popular due to its superior capacity to extract text feature data (Zhao et al., 2020). Improving the effectiveness of classification and consuming the large quantity of user review texts published

* Corresponding author.
E-mail addresses: peteratandoh@gmail.com (P. Atandoh), fzhang@uestc.edu.cn (F. Zhang), dadugyamfi@cktutas.edu.gh (D. Adu-Gyamfi), atandoh_ph@mercer.edu (P.H. Atandoh), Raphmike46@yahoo.com (R.E. Nuhoho).

Peer review under responsibility of King Saud University. Production and hosting by Elsevier.

ELSEVIER Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2023.101578
1319-1578/

on the Internet may prove to be a difficult task in the present socioeconomic times when the sheer quantity of information and the quantity of review texts are both rapidly expanding. This is because, in the modern age, there has been a corresponding rise in the number of review texts. They add weighted sentiment features to an existing dependency graph-based position encoding method as part of the feature representation process. As input to a re- worked deep CNN model for email sentiment classification, they combined encoded sentiment sequence features with standard word embedding features (Liu and Lee, 2021). Convolutional neural networks (CNNs) are useful for recognizing local and position-invariant patterns. The use of convolutional neural networks was extensively revealed in this study in text classification across a wide variety of domains (Amin and Nadeem, 2018). However, this model has the pitfall of not capturing global features. As a result, (Hameed and Garcia-Zapirain, 2020) introduced a single global pooling mechanism and a single bidirectional long-short-term memory (BiLSTM) layer. Based on the attention mechanism and the bidirectional long-short-term memory network, the authors propose a sentiment classification method for massive amounts of microblog text (Wu et al., 2021). A deficiency of sensitivity to random weight initialization and its vulnerability to over-fitting are also observed in long-short-term memory while executing text classification. Due to this, (Zulqarnain et al., 2021) proposed an enhanced novel GRU for text classification. The Gated Recurrent Unit (GRU), a deep learning system incorporating revise and restart gates, has been a popular text categorization approach in recent years, particularly for sequential datasets. The GRU is less susceptible to overfitting and is capable of adjusting to longer sequence datasets. To create a sentence-level feature representation, the convolutional neural network (CNN) extracts n-gram data from each sentence at various granularities. The phrases are sequentially integrated using a bidirectional gated recurrent unit to retrieve the text's contextual semantic information (BiGRU). The CNN-BiGRU model has been updated to include an attention mechanism (Yan et al., 2021). For the purpose of citation filtering, the study detailed a feed-forward neural network that was trained using parallel CNN architectures of diverse kernel dimensions (van Dinter et al., 2021). In order to learn more about the meaning of each word, (van Dinter et al., 2021) employed glove embedding. The proposed method used bidirectional encoder representation from transformers as embedding, which solves the problem of polysemous words since words and vectors have a one-to-one relationship, a major problem associated with glove embedding and Word2veC. The model is then fed into a multi-layer convolutional neural network (CNN) with various kernel dimensions to categorize e-commerce textual reviews from end users. The Bidirectional Encoder Representations from Transformers is a novel language representation paradigm (Devlin et al., 2019) introduced in 2019. Advanced methods for discovering word representations from context have greatly benefited from this idea. The following are the paper's main contributions:

- BERT is used to implement word vectorization of text to overcome the disadvantages of word2vec, glove embedding, term frequency-inverse document frequency, and any additional text representation techniques that lack contextual information.
- In order to categorize online reviews, we further input the model into multi-layer convolutional neural networks with variable kernel sizes. Afterward, a softmax layer for classifying text follows.
- To improve the accuracy of the proposed DBSA, we extensively experimented with how the maximum length of sentences, learning rate, batch size, and epoch size affected the B-MLCNN model.
- We designed a system for classifying text using BERT multilayered CNN and extensively simulated it on IMDB, movie reviews, and Amazon review datasets.
- A statistical analysis (Friedman test) was used to scientifically test whether or not there were any significant variations in performance.
- We went a step further and contrasted our findings with those from other cutting-edge models.

The remaining sections of this work are structured as follows: The most recent research on sentiment classification is discussed in Section 2. We presented our proposed model, which drew inspiration from prior research, in the third and fourth segments. Finally, in Section 5, we conducted experiments to evaluate our unique framework proposal. Section 6 concludes with a summary of our findings and recommendations for where to take our research next.

2. Literature review

The process of grouping text documents into two or more categories is known as text classification. Document classification is the end goal of text categorization, not information extraction. The most recent breakthroughs in deep learning for text classification often center on developing increasingly complex models based on neural networks in order to handle enormous datasets. Here we look at how lexicons, machine learning, and deep learning may all be used to better categorize texts. The goal of this exercise is to establish the polarity of words. Early studies on classifying the text of product reviews mostly relied on rules or lexicons. The author uses the PMI-IR method of determining the review's semantic orientation by looking at word and phrase polarity (Turney, 2002). A major challenge associated with the lexicon method is that it considers the semantic orientation of adjectives and adverbs in reviews, which may not capture the full meaning or sentiment expressed in a review. There is also overreliance on mutual information between specific words ("happy" and"sad") to calculate semantic orientation, which may not be applicable to all domains or languages.

In this work, the authors presented a synset of WordNet that is tagged with three labels: objective, positive, and negative in the SentiWordNet lexical resource (Baccianella, 2010). The authors employed SENTIWORD- NET 3.0, which was an improved version of SENTIWORDNET 1.0. For the task of classifying review documents, they describe multiple classifiers. Three classifiers based on support vector machine (SVM), maximum entropy (SVM), and score calculation make up the methodology. When integrating a single classifier, they use SVMs and two voting techniques (Tsutsumi et al., 2007). Experimental results showed that two voting techniques and SVMs integrated with these classifiers increased classification accuracy. The traditional approach to resolving the challenge of text classification is based on machine learning. However, because these models ignore the semantics of the text, they are unable to comprehend the document's entire richness, which derives from a variety of social media platforms, individual web pages, blogs, and other online publications (Williams et al., 2015). (Mohd Nafis and Awang, 2021) research showed an improved hybrid feature selection method to boost sentiment classification using machine learning techniques. The authors trained using support vector machines and term frequency-inverse document frequency on the IMDB dataset. Concerns with document-level text sentiment classification were addressed. A major drawback of conventional machine learning models is that a text classifier that has only been trained on data from one domain often performs poorly when applied to data from another domain. In this paper, they offer a recurrent naive Bayes learning system for identifying the sentiment of user-generated material in online marketplaces. They took the naive Bayes parameter estimation method and applied it to a more thorough training style without compromising the excel- lent computational efficiency of the original naive Bayes model. (Xu et al., 2020). A prospective replacement for conventional machine learning techniques is deep learning. For larger datasets, it has demonstrated great performance for a wide range of text classification and other NLP tasks. The authors (Ouyang et al., 2015) used CNN and word2-vec to analyze sentiment. Word2vec's con– version of text into vectors provides input for CNN. This study demonstrates that, when compared to the RNN approach, CNN Plus word2vec can offer good results. Three convolutional layers are utilized in their model architecture to derive features from the input, and three pooling layers are utilized to get some instances of the convolutional structure. However, the pitfall associated with using word2vec as a text representation is the inability to handle words.

In order to classify Arabic text, the authors used a number of deep learning models. The researchers applied CNN, GRU, LSTM, and hierarchical attention networks on the SANAD and NADIA datasets (Elnagar et al., 2020). They looked into how effectively employing word2vec embedding models boosted classification results as well. In order to classify text data, this study creates an LSTM model for sentiment classification. LSTM combines its region embedding technology with the capacity to extract words and phrases with different contributions to categorize text (Zhang, 2021). The model struggles with sensitivity to random weight initialization of text. After carefully studying the merits of convolutional neural networks and CapsNet, this study provides an EEG emotion recognition model based on the attention mechanism and the pre-trained convolutional CapsNet network (Liu et al., 2023). In a recent CNN study, researchers used a self-supervised method to teach a deep multi-task convolutional neural network (CNN) to classify emotions from electrical brain activity. There are two distinct approaches to emotion recognition: those that rely on the speaker's facial or vocal cues and those that rely on textual cues. The true emotional state may be more precisely reflected by physiological signals because they are subjective and spontaneous (Liu et al., 2022; Wang et al., 2023). (Wu et al., 2022) designed a model specifically for building sentiment classification using a two-level long-short-term memory network, attempting to resolve the complexity of subjectivity annotation and sentiment representation. Delvin et al. (Devlin et al., 2019) presented the linguistic representation using the BERT model, which represents a breakthrough in natural language processing tasks by pre-training word embedding on a large network that is fine-tuned on enormous data. The BERT-BiGRU model's structure is proposed in this study. First, they used the BERT model instead. Word representations can be adjusted to better fit the intended meaning of the term after contextual details have been taken into account. Bidirectional Gated Recurrent Unit (BiGRU) is the second add-on, and it is in charge of extracting text information characteristics in both directions at the same time (Yu et al., 2021). Another hybrid model was proposed in this study; the authors combined BERT and CNN to classify text using Chinese reviews (Cui and Huang, 2021). (Li et al., 2021) enhanced the sentiment features of the vectors by processing the processed sentiment dictionary and first utilizing BERT to make a recreatable word-vector representation of the text. The topic and sentiment of online reviews are analyzed synchronously in this study. Researchers used a Joint Sentiment Table 1.

Topic Model with Many Granules, which combined sentiment and topic detection using a probabilistic model on the document level (Huang et al., 2022b). (Kit and Mokji, 2022) proposed a feature-reduction-based BERT model that doesn't require fine-tuning to reduce phrase vector representation and requires less training time. (Bello et al., 2023) presented BERT, in addition to other methods, as a text classification for use in natural language processing. The experimental results show that, compared to Word2vec and no variation, the model performance is much higher when BERT is used as text representation. In this paper, they propose a convolutional neural network (CNN) as an improvement over the BERT-GCN model for aspect-level sentiment analysis (Phan et al., 2022). In order to extract contextual semantic information, syntactic structures, and position information from Chinese microblogs, the BERT-CNN-ATT sentiment classification model is proposed in this paper. This model combines the benefits of BERT, a CNN, and an attention mechanism (Jia, 2022).

Table 1
Dataset.

Dataset name	Positive	negative	Total
IMDB	25,000	25,000	50,000
Movie Review (2002)	700	700	1400
Movie Review (2004)	1000	1000	2000
Amazon Review	7287	736	8023

In conclusion, when it comes to extracting text semantic and syntactic features, the BERT model is the most effective vectorization model currently available. Its text-representation skills are very useful for other computational tasks. To better represent text, the convolutional neural network can mine its local properties using convolution kernels of varying sizes.

2.1. Research objectives

- To study and analyze the problem of document-based sentiment classification from online reviews.
- Design a B-MLCNN architecture for document-based sentiment classification.
- Experimentally study and provide empirical analysis of the proposed model results with some existing state-of-the-art models.

2.2. Research outcome

The outcome of this research is to produce the BERT Multilayered Convolutional Neural Network (B-MLCNN) as an integrated deep learning paradigm to improve on the efficiency of document-based sentiment analysis.

3. Preliminary

3.1. Bert

The BERT model, which stands for "Bidirectional Encoder Representations from Transformers," is an implementation of the Transformers deep learning framework. In this model, every output element is related to every input element, and the weightings between the input and output elements are dynamically selected based on their relationship to one another. A sequence encoder and decoder are at the heart of the original Transformer design. The input is first transformed into an embedding by the encoder, and then, by the decoder, the text representation of that embedding is reconstructed.

3.2. Multi-layer convolutional neural network

A multilayer convolutional network has several pooling layers and a range of kernel sizes. Yet, all of the convolutional layers have the same number of filters and the same size. The form of the convolved feature map determines the size of the pooling layer.

Aftermath, the retrieved features are mapped into the output using a fully connected layer.

- **Convolutional layer**: A key component of any CNN is the convolutional layer. Understanding that CNN's parameters and channels are built from different types of trainable channels, or neurons, is crucial. These channels have a limited perceptron. Each feed-forward channel considers the input's dimensions and computes the dot product of the filter and input pixels. When a certain feature type is spotted in a certain location on the feature map, the system is able to learn the channels that are subsequently created.
- **Pooling layer**: The pooling layer creates a single neuron from the sum of all the outputs from the convolutional layer's sublayers. Maximum pooling and average pooling are both viable options for the pooling operation. One takes the mean of its inputs, while the other looks at the largest one.
- **Fully connected layer**: The neurons in the fully connected layers have links to a wide range of neurons in the layers above them. It offers comprehensive links. Generally speaking, fully connected layers work by decreasing the number of dimensions from the original feature map to a single-dimensional (1D) feature vector. Hence, the 1D feature vector can be used as a classifier in this application.

4. Proposed integrated BERT-MultiLayed convolutional neural network paradigm

Four review data sets, feature text representation creation, and sentiment classification are the components of the proposed model. Fig. 3 shows the system architecture of the proposed model. Subsequent subsections provide a more in-depth overview and accompanying algorithms for each stage of the proposed paradigm. There are essentially four parts to the BERT-MLCNN. The BERT model is trained to first obtain a vector representation of each word and then a semantic representation of the text. After that, a multilayered convolutional neural network receives the word's vector representation as its input, using a variety of kernel sizes to facilitate additional analysis and the extraction of features and meaning. The softmax layer then classifies the output as positive or negative.

4.1. Problem Definition

The Document-Based Sentiment Classification is an example of a sentiment analysis task that considers documents as a whole rather than analyzing individual entities or sections of information to determine sentiment. The goal is to label a piece of writing with a positive or negative orientation, or polarity, based on the tone it conveys. It is generally agreed in the current literature on document sentiment classification that the following will serve as a solid foundation for practical outcomes (Liu, 2015). The sentiment classification of documents relies on the premise that a document with an opinion, such a product review, will express opinions on a specific topic and come from a single opinion source. In the following, we present a mathematical description of DBSA.

Definition 1. *A corpus containing many sentences in a document with a total word count of n, shown as* $s_{sent} = \{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9 \cdots w_n\}$ *where* $d_{document} = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9 \cdots d_t\}$ *is a vocabulary set* of t words used in a particular context. $p = \{p_1, p_2, p_3, p_4, \; p_5, p_6, p_7, p_8, p_9 \cdots, p_j\}$ of the s_{sent} in $d_{document}$ represents the sentiment polarity, wherein l refers to the total number of sentiment polarity labels.

Definition 2. *For the purpose of predicting whether a sentiment is positive or negative of"*s_{sent}*" in the"*$d_{document}$*" problem is the basic purpose of DBSA, which can be represented mathematically as follows.* θ_{max} *is a quantificational function that identifies the highest degree of matching in"*$d_{document}$*" for a given sentence word* s_{sent} *and sentiment polarity* p_i*. The input is represented as*:

$$d_{document} = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9 \cdots d_t\} \tag{1}$$

$$s_{sent} = \{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9 \cdots w_n \forall w :\in s_{sent} \subset d_{document}\} \tag{2}$$

the output is represented as:

$$P_k = \theta_{max}(d_1, p_1) s_{sent} \forall i \in [1, l] \tag{3}$$

4.2. The BERT structure

One of the most well-known and widely used pre-trained language models that includes a transformer is the BERT model. It pretrains a large corpus.

with the use of bidirectional transformers, and then adjusts it to do particular tasks. (Devlin et al., 2019). BERT is a preprocessing approach for natural language processing that uses neural networks. BERT can be used in place of word2vec due to its superior benefits and adaptability. Word2vec does not adjust how it represents context-dependent words. Because of this, the ac- curacy of the results achieved in the following tasks would suffer if word2vec was used as a word vector representation. This is due to the fact that the se- mantic features of natural language make it feasible for a single word to have several meanings. As a result of its intricate construction, BERT necessitates a large commitment of both time and money to learn. Google, on the other hand, has made its source code publicly available. Users only need to make minor adjustments to the pre-trained BERT model based on the specifics of the current work in order to save both time and money. Using the trained BERT model, it is possible to apply it to a wide range of downstream tasks in natural language processing. WordPiece is the tokenizer used by BERT. The top hidden layer of the transformer network represents the word vector matrix S_D of the sentence S.

$$S_D = [X_0, X_1 \cdots X_n, X_{n+1}] \tag{4}$$

A sub matrix of is used to represent the desired words D_r.

$$D_r = [X_1, X_1 + 1, \cdots X_{1+m-1}] \tag{5}$$

$D_r \in R^{ml}$, where m is the desired length. The target vectors are applied to the max-pooling procedure, which selects the most essential features from all words in the target at each dimension.

$$V = \max\{D_r, L = 0\}, V \in R^{1*L} \tag{6}$$

4.3. The MultiLayered convolutional neural network

This section offers the final classification using MLCNN, following the successful creation of the feature set. We employ a multithreaded convolutional neural network that can read text using various kernel sizes. Consequently, a multichannel convolutional neural network is created. The convolutional neural network is competitive for document modeling given its convolutional layer as well as the maximum pooling layer. This is due to the fact that it can detect even the most minute grammatical tendencies throughout the full corpus of data used for training. Because of their unique nomenclature and conceptual framework, convolutional neural networks stand out from other varieties of neural network architecture. The embedding dimensions are a representation of corpus words. An illustration of a sentence's n-word composition is shown sequentially. The matrix results are used as input to the convolutional network. One of the reasons we employed CNN was to learn the categorization of encoding our texts as a sequence of embed- dings. Extracting the most essential local properties of the input matrix from the convolutional layer. The input layer accepts the review text vector $x_i^0 = x_1, x_2, x_3, x_4, x_5 \cdots .x_n$ produced the input representation layer and performs a convolution operation using the output of the convolutional layers to accurately extract the local sentiment data included in the review text, as indicated in Eq. (4). The term b_j^l represents the bias mapped by feature j; w represents the convolutional kernel weight. m for filter index; and σ for ReLU activation function. R and T represent the pooling stride respectively.

$$y_{ij}^l = \sigma\left(\sum_{m=1}^{M} w_{m,j}^l x_{i+m-1,j}^{l-1} + b_j^l\right) \tag{7}$$

In order to further reduce the number of parameters, data dimensionality, and local sentiment information from the review text, it is passed to the pooling layer after the convolution procedure.

$$x_{hidden} = \max_{r \epsilon R}(y_{i \times T+R,J}^{l-1}) \tag{8}$$

4.4. The output layer

After concatenation, the document feature obtained is passed to the softmax layer. Let z denote the document feature, we have:

$$z = concat(\lambda z_c, (1-\lambda) z_w), \lambda \in [0,1] \tag{9}$$

Classifying the input feature matrix is the primary goal of this layer. An integer between 0 and 1 is assigned to each feature in this layer. As the value gets closer to zero, the input text becomes more adversely emotional (negative). A value closer to one is interpreted as representing the input text, and it implies the polarity of the sentiment leans more toward the positive.

5. Experimental setup

Here, we first introduce the experimental setup that will be employed in the subsequent sections of the paper. The datasets used in the tests of that method are then presented. The pre-processing methods for filtering those datasets are then discussed in detail, and lastly, the baselines utilized for comparison to that technique are expressed. These are the metrics most frequently used in the associated literature on sentiment analysis. This study uses accuracy, precision, recall, and F1 measure metrics to evaluate the technique.

To create our proposed deep learning architecture, we resorted to a 12- core AMD Ryzen 9 5900X CPU on a single GPU, the GEFORCE GTX running a python-based jupyter IDE consisting of Tensorflow 2.0, and Keras 2.1.0.

5.1. Reproducibility of the proposed model

The pre-train BERT is employed in the role of word vector representation. The BERT layer is responsible for providing the model with a word vector representation, which enables the model to better comprehend the textual reviews and to learn variability in text patterns. By utilizing a variety of convolutional layers with various filter sizes, the multi-convolutional neural network was able to enhance the process of semantic feature extraction. Fig. 2 is a representation of the overall implementation of the model that has been proposed. We tested our model on four different datasets so that we could determine the full extent of the capabilities of the proposed model. Ablation research on BERT, CNN, BERT-CNN, and B-MLCNN was continued here. A comparative study with regard to the use of various baseline training methods will be carried out.

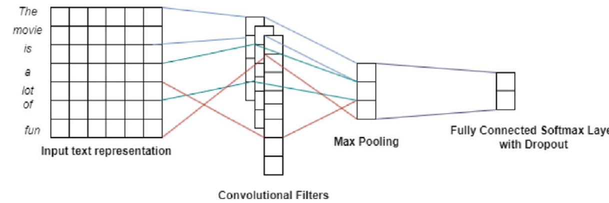

Fig. 1. CNN for sentiment analysis.

5.2. The dataset

We conducted experiments on four document-level sentiment classification datasets, including IMDB (Yenter and Verma, 2017), the 2002 and 2004 versions of Movie Review (Pang et al., 2002) (Pang and Lee, 2004), and Amazon reviews (Ni et al., 2019).

- The IMDB dataset is essentially a collection of reviews for movies that have been copied and pasted from the IMDB website. A long composition with a 1–10-star rating system is an example.
- The movie review dataset is comprised of sentences that have been labeled with one of two categories: either positive or negative. Amazon data sets are categorized according to the type of product. Our analysis used reviews for gift cards, fashion, software, magazines, and appliances. Reviews are given a rating from 1 to 5, with a score ≥ 3 denoting a positive review and a score of ≤ 2 denoting a negative review. The fact that we're looking at reviews that are so polarized, this study limited the number of sentiment classes to two (Lyu et al., 2020). The datasets are available for public access https://nijianmo.github.io/amazon/index.html.

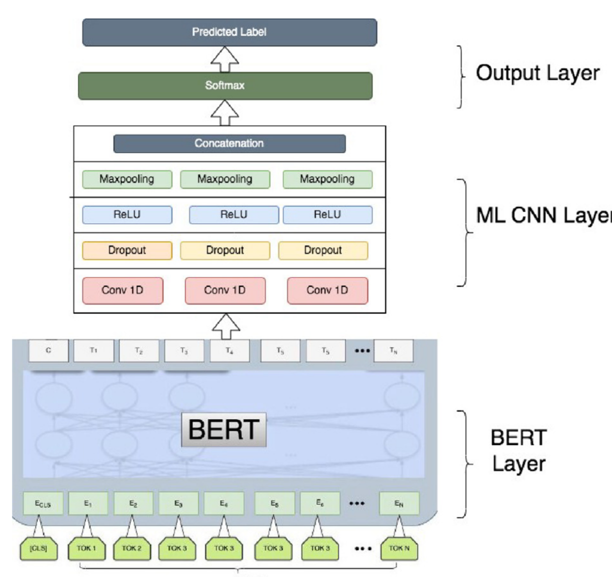

Fig. 2. Framework of BERT Multi-Layer Convolutional Neural Network.

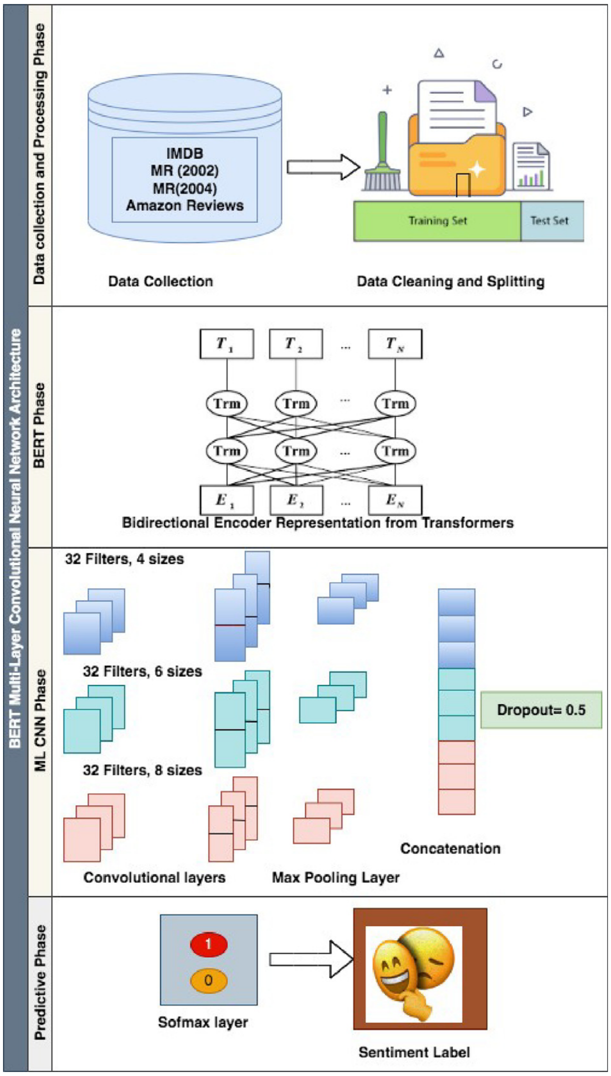

Fig. 3. Proposed Model of BERT Multi-Layer Convolutional Neural Network.

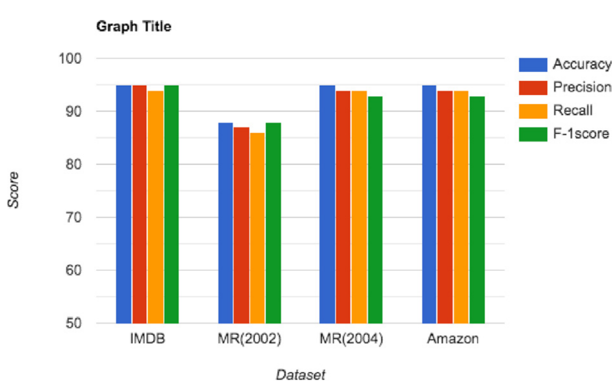

Fig. 4. Experiment Results of B-MLCNN on different dataset.

5.3. The data pre-processing

In this phase, we did a data cleaning on the train and test dataset. All non-verified and duplicate reviews from the entire collection were removed. In addition, very short reviews were removed after removing stop words as described. A full-stop marks the end of a sentence in the text reviews.

All punctuation is removed from the textual content sentences, and the white spaces between each sentence are divided into tokens. Removed all sentences that contained words made up of impure alphabetic (alphanumeric) characters. All stop-word-containing words in sentences have been eliminated, as well as all the words with special characters.

Algorithm 1 (*Proposed BERT Multi-Layer Convolutional Neural Network*). *Initialize*

- *Input*: $d_{document} = \{d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9 \cdots d_t\}$
- $s_{sent} = \{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9 \cdots w_n \forall w :\in s_{sent} \subset d_{document}\}$
- ***Output:*** $P_k = \theta_{max}(d_1, p_1) s_{sent} \forall i \in [1, l]$
- ***Data:*** *data*

$T_{model} \leftarrow Load(B)$ {#B represents BERT}
d, l = DataCleaning(d)

{#d represents data}.

$d_{input} TextRepresentation(B,\ d)$
$d_{train},\ d_{test} \leftarrow DataSplitting(d_{input})$
$^{B}model \leftarrow (^{T}model,\ ^{data}train)$
$F_{output}(d_{train},\ B)$ {#F represents Feature}
$TrainMLCNNwithF_{output}$
$Epochs \leftarrow epoch_{max}$
$ErrorRate_B \leftarrow best_{error}$
Epoch = 0
$Train(d_{train},\ B)$
$F \leftarrow Foutput(dtest,\ B)$
$Result \leftarrow Predict(F,\ MLCNN)$
$test_{error}$ = *ComputError*(*Results*)
if $test_{error} \leq B_{error}$ ***then***
$best_{error} \leftarrow test_{error}$
$Save(B)$
^{F} = $Foutput^{(d}train,\ ^{B)}$
$TrainMLCNNwithF_{output}$
$Save(MLCNN)$

5.4. Experimental results

A polarity prediction is presented in this paper on online reviews based on B-MLCNN-ATT. BERT is a robust text representation model with the ability to understand ambiguities in textual data by establishing context by referring to the surrounding text. The proposed model also utilized a multi-channel convolutional neural network with various kernel dimensions for feature extraction. As a result, these elements work together to enhance the proposed approach's performance in the experiment. When it comes to the highest level of performance and accuracy in NLP tasks, the pre-training word representation model known as BERT plays a crucial role. Furthermore, BERT pre-training is regarded as a significant performance component for text representation. As a result, the suggested strategy achieves its goal while leveraging BERT. Max pooling and convolutional layers of the model's multi-layer convolutional neural network improved feature extraction and found subtle grammatical trends in the training set. Because of the combined impact of various input parameters, this approach is the most reliable and thorough.

Through the use of four common datasets, including IMDB, Movie Review (2002 and 2004) versions, and Amazon Reviews datasets, the B-MLCNN is evaluated for its effectiveness in this study. As far as our knowledge and expertise allow, we hold that this method is one of the first to extract and categorize sentiment data at the document level using the B-MLCNN. A significant

Table 2
Baseline techniques.

Baseline model	Innovation	Year
CNNrand (Kim, 2014)	CNN for sentence-level classification evaluations.	2014
CNN (Mchannels) (Kim, 2014)	CNN two-word vector model.	2014
LSTM (Tai et al., 2015)	Tree-LSTM sentiment classification.	2015
NCSL (Teng et al., 2016)	recurrent neural network context-sensitive lexicon.	2016
Doc2VecC (Chen, 2017)	Document Vector through Corruption	2017
LSTM + CBA + LA (Long et al., 2017)	Combination of two distinct attention	2017
LSTM with dynamic skip (Gui et al., 2019)	Long Short-Term Memory with Dynamic Skip Connections	2019
CNN + LSTM (Camacho-Collados and Pilehvar, 2018)	Text Preprocessing in Neural Network Architectures	2018
WALE-LSTM (Fu et al., 2018)	Combination of lexicon and attention with LSTM	2018
S-LSTM (Zhang et al., 2018)	Sentence-State LSTM for Text Representation	2018
XLNet (Yang et al., 2019)	generalized autoregressive pretraining approach	2019
VLAWE (Ionescu and Butnaru, 2019)	Novel Document-level Representation	2019
LSTM (Bodapati et al., 2019)	Long Short Term Memory for Sentiment Analysis	2019
DV-ngrams-cosine (Thongtan and Phienthrakul, 2019)	Document Embeddings Trained with Cosine Similarity	2019
BERT_{base} + *ITPT* (Sun et al., 2019)	How to Fine-Tune BERT for Text Classification	2019
BERT_{large} + *ITPT* (Sun et al., 2019)	How to Fine-Tune BERT for Text Classification	2019
T-Capsule (Chen et al., 2020)	capsule network-based text sentiment classification	2020
Single layered BILSTM (Hameed and Garcia-Zapirain, 2020)	A single layer BILSTM	2020
TextCNN (Chen et al., 2020)	Visual text analysis with CNN	2020
BERT (Chen et al., 2020)	BERT encodes, and softmax classifies it.	2020
SVM-RFE ()	TF-IDF-SVM hybridization	2021
Caps-BiLSTM (Dong et al., 2020)	BiLSTM capsule network	2021
MNB (Hassan et al., 2022)	MNB classifies distinct features	2022
SelfAtt (Tai et al., 2015)	self-attention sentence embedding	2022
SCL-NMA (Huang et al., 2022a)	Score SentiCNN	2022
MTL-ML_4 (Gui et al., 2022)	multi-tasking learning sentiment classifier.	2022
SentiCNNLBAM3 (Huang et al., 2022a)	Contextual and sentiment SentiCNN	2022

method for evaluating a model's efficacy with regard to unobserved data is cross-validation. For model optimization, we divided the training data for each dataset into ten folds using the cross-validation approach (Socher et al., 2012). These subsets are evenly distributed and mutually exclusive, with nine of them being utilized for training and one for validation. All of those sentences that were intended to be shorter than the allowed length have had zeros appended. Table 3 displays the proposed model's precise parameter settings. During training, we also looked at how the model's performance changed depending on max-sentence length, batch size, epoch size, and learning rate. The performance results are depicted in Tables 4–7. The B-MLCNN carefully examined all of the baseline studies' experimental designs. The suggested method determines which type of parameter setting leads to a performance increase or decrease during the analysis. The B-MLCNN, therefore, derives the best parameter settings from a thorough analysis of the baseline approaches depicted in Table 2 and the parametric optimization technique, relying on parameter optimization techniques and similar parametric studies. We also trained and compared our models in chronological order (BERT, BERT-CNN, and B-MLCNN). The performance of the four datasets used on the BERT, CNN, BERT-CNN, and B-MLCNN is noted here as well as in Tables 8, 9, 10, and 11. Classification performance comparisons between the proposed method and baseline approaches on the IMDB dataset are shown in Table 12. These tabular notations show the B-MLCNN's impressive accomplishments when measured against the acquisition of baseline approaches using standardized datasets, which denotes its empirical development across all domains of interest.

5.5. Evaluation metrics (EM)

The accuracy, loss, precision, recall, and F1 score were among the evaluation measures and benchmarks we used. The parameters utilized in this computation are described in detail below and are consistent with those found in prior research.

- True Positive (TP): This is a reference to a result that makes an accurate prediction of positive feedback.
- False Positive (FP): The term" false positive" refers to a result that shows inaccurately expected positive feedback.

Table 3
Optimal Configurations of the Proposed Model.

Parameters	IMDB dataset	MR (2002)	MR (2004)	Amazon dataset
Training Method	Cross val.	cross val.	cross val.	cross val.
$\text{Conv1D}_1(Filter, Kernelsize)$	32×4	32×4	32×4	32×4
$\text{Conv1D}_2(Filter, Kernelsize)$	32×6	32×6	32×6	32×6
$\text{Conv1D}_3(Filter, Kernelsize)$	32×8	32×8	32×8	32×8
Batch size	30	6	6	6
Learning rate	4e-3	4e-3	4e-3	4e-3
Max. length	400	100	100	100
Dropout	0.5	0.5	0.5	0.5
Activation Function	ReLU	ReLU	ReLU	ReLU
Predictive Function	Softmax	Softmax	Softmax	Softmax
Epochs	40	40	40	40

Table 4
The impact of max length on the B-MLCNN model.

Dataset	**Max-Length**	**Train Acc**	**Val.**	**Acc**	**Train Loss**	**Val.**	**Loss**
IMDB	32	0.89	0.87		0.48	0.42	
	64	0.90	0.89		0.49	0.41	
	100	0.96	0.89		0.39	0.22	
	128	0.90	0.88		0.36	0.40	
	256	0.88	0.87		0.42	0.25	
	400	0.97	0.95		0.24	0.22	
MR (2002)	32	0.89	0.87		0.48	0.42	
	64	0.89	0.87		0.49	0.41	
	100	0.91	0.88		0.35	0.21	
	128	0.90	0.87		0.36	0.40	
	192	0.88	0.87		0.62	0.45	
	256	0.88	0.87		0.76	0.40	
MR (2004)	32	0.89	0.87		0.48	0.42	
	64	0.89	0.87		0.49	0.41	
	100	0.97	0.95		0.39	0.20	
	128	0.90	0.87		0.36	0.40	
	192	0.88	0.87		0.52	0.42	
	256	0.88	0.87		0.78	0.57	
Amazon Reviews	32	0.89	0.87		0.48	0.42	
	64	0.89	0.87		0.49	0.41	
	100	0.98	0.95		0.39	0.22	
	128	0.90	0.87		0.36	0.40	
	192	0.88	0.82		0.38	0.30	
	256	0.87	0.80		0.72	0.40	

Table 5
The impact of learning rate on the B-MLCNN model.

Dataset	**Learning rate**	**Train Acc.**	**Val.**	**Acc.**	**Train Loss**	**Val.**	**Loss**
IMDB	1e-2	0.91	0.90		0.42	0.34	
	2e-5	0.91	0.90		0.42	0.34	
	3e-5	0.91	0.90		0.39	0.34	
	2e-4	0.95	0.92		0.42	0.31	
	4e-3	0.97	0.95		0.24	0.22	
	5e-2	0.93	0.91		0.35	0.26	
MR (2002)	1e-2	0.81	0.80		0.74	0.71	
	2e-5	0.82	0.80		0.56	043	
	3e-5	0.84	0.80		0.57	0.49	
	2e-4	0.89	0.80		0.57	0.49	
	4e-3	0.91	0.88		0.35	0.21	
	5e-2	0.76	0.74		0.28	0.24	
MR (2004)	1e-2	0.81	0.80		0.74	0.71	
	2e-5	0.88	0.87		0.59	0.47	
	3e-5	0.91	0.87		0.36	0.37	
	2e-4	0.85	0.83		0.47	0.40	
	4e-3	0.97	0.95		0.39	0.20	
	5e-2	0.75	0.74		0.22	0.27	
Amazon Reviews	1e-2	0.81	0.80		0.74	0.71	
	2e-5	0.88	0.84		0.55	044	
	3e-5	0.91	0.87		0.36	0.40	
	2e-4	0.89	0.86		0.49	0.41	
	4e-3	0.98	0.95		0.39	0.22	
	5e-2	0.78	0.74		0.22	0.23	

- True Negative (TN): If the results show that the positive reviews were correctly predicted, this is the case.
- False Negative (FN): This is significant in light of the result, which reveals that inaccurate projections of negative reviews were made.
- Accuracy: This is true for the outcome that shows the proportion of correctly anticipated reviews to all of the reviews.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

- Precision: This relates to the total number of projected positive reviews that were successfully forecasted.

$$\text{precision} = \frac{TP}{TP + FP} \quad (11)$$

- Recall: This is true for the outcome that shows all of the reviews in the dataset to have accurately predicted favorable reviews.

Table 6
The impact of Batch size on the B-MLCNN model.

Dataset	**Batch Size**	**Train Acc.**	**Val. Acc.**	**Train Loss**	**Val. Loss**
IMDB	5	0.93	0.90	0.45	0.34
	6	0.93	0.89	0.45	0.32
	7	0.93	0.89	0.45	0.34
	8	0.95	0.91	0.54	0.27
	30	0.97	0.95	0.24	0.22
	32	0.95	0.91	0.22	0.27
MR (2002)	5	0.90	0.89	0.73	0.67
	6	0.91	0.88	0.35	0.21
	7	0.91	0.87	0.36	0.37
	8	0.85	0.83	0.47	0.40
	16	0.86	0.80	0.39	0.20
	32	0.75	0.74	0.22	0.27
MR (2004)	5	0.93	0.81	0.74	0.71
	6	0.97	0.95	0.39	0.20
	7	0.91	0.87	0.36	0.37
	8	0.85	0.83	0.47	0.40
	16	0.85	0.89	0.39	0.20
	32	0.80	0.74	0.22	0.27
Amazon Reviews	5	0.95	0.86	0.40	0.23
	6	0.98	0.95	0.39	0.22
	7	0.95	0.91	0.36	0.40
	8	0.95	0.89	0.49	0.41
	16	0.93	0.86	0.41	0.24
	32	0.93	0.86	0.38	0.22

Table 7
The impact of epoch size on the B-MLCNN model.

Dataset	**Epochs**	**Train Acc.**	**Val. Acc.**	**Train Loss**	**Val. Loss**
IMDB	1	0.93	0.90	0.54	0.47
	2	0.93	0.90	0.52	0.43
	20	0.95	0.89	0.40	0.33
	25	0.95	0.92	0.41	0.34
	30	0.97	0.91	0.54	0.32
	40	0.97	0.95	0.24	0.22
MR (2002)	1	0.78	0.75	0.79	0.72
	2	0.78	0.74	0.79	0.65
	20	0.80	0.77	0.65	0.54
	25	0.80	0.78	0.58	0.36
	30	0.80	0.78	0.54	0.31
	40	0.91	0.88	0.35	0.21
MR (2004)	1	0.81	0.80	0.68	0.65
	2	0.82	0.80	0.72	0.67
	20	0.85	0.78	0.55	0.43
	25	0.85	0.80	0.51	0.32
	30	0.85	0.82	0.52	0.32
	40	0.97	0.95	0.39	0.20
Amazon Reviews	1	0.93	0.84	0.74	0.71
	2	0.95	0.87	0.54	0.52
	20	0.95	0.91	0.36	0.40
	25	0.92	0.86	0.42	0.27
	30	0.92	0.86	0.43	0.30
	40	0.98	0.95	0.39	0.22

Table 8
Experiment Results on IMDB dataset.

Model	Accuracy	Precision	Recall	F-1 score
BERT	0.94	0.93	0.95	0.93
CNN	0.91	0.91	0.93	0.92
BERT-CNN	0.93	0.90	0.93	0.93
B-MLCNN	0.95	0.95	0.94	0.95

Table 9
Experiment Results on MR (2002).

Model	Accuracy	Precision	Recall	F-1 score
BERT	0.88	0.85	0.86	0.87
CNN	0.84	0.85	0.86	0.85
BERT-CNN	0.87	0.84	0.87	0.86
B-MLCNN	0.88	0.87	0.86	0.88

Table 10
Experiment Results on MR (2004).

Model	Accuracy	Precision	Recall	F-1 score
BERT	0.89	0.89	0.88	0.88
CNN	0.89	0.89	0.90	0.89
BERT-CNN	0.90	0.88	0.91	0.90
B-MLCNN	0.95	0.94	0.94	0.93

Table 11
Experiment Results on Amazon Review.

Model	Accuracy	Precision	Recall	F-1 score
BERT	0.91	0.89	0.91	0.90
CNN	0.89	0.89	0.91	0.90
BERT-CNN	0.91	0.88	0.91	0.90
B-MLCNN	0.95	0.94	0.94	0.93

Table 12
Performance comparison of the proposed approach with the baselines.

Model	accuracy (IMDB)
S-LSTM	87.15
Doc2Vec	88.3
LSTM	88.6
CNN + LSTM	88.9
WALE-LSTM	89.50
LSTM with dynamic skip	90.1
LSTM + CBA + LA	90.1
Single layer BILSTM	90.58
DV-ngrams-cosine	93.13
B-MLCNN	95.01
BERT_{base} + *ITPT*	95.63
BERT_{large} + *ITPT*	95.79
XLNet	96.2

$$\text{recall} = \frac{TP}{TP + FN} \tag{12}$$

- F-1 score: When precision and recall are weighted together, the result is the weighted mean.

$$\text{F1} = \frac{2 * precision * recall}{recall + precision} \tag{13}$$

5.6. Parameters influence on B-MLCNN

To determine the optimal hyperparameters for the proposed model, we decided to fine-tune the BERT pre-train language model and adjust other parameters to improve its accuracy. A hyperparameter called" learning rate" defines the magnitude of model changes when adjusting model weights ac- cording to expected error. Selecting the learning rate is challenging since too little could lead to a lengthy training process that becomes stuck, while too much could lead to learning a suboptimal set of weights too rapidly or to an unstable training process. The learning rate controls weights based on the loss gradient in the model. We discovered that the model performs best when the learning rate value is 4e-3 after conducting experiments with different learning values. At this point, the loss is still decreasing. The paper recommends this as a good value for the model's learning rate. According to Table 5, the experimental results are shown. The BERT model was tested with different learning rates. Sentence length has a significant impact on our BERT method's performance value. From Table 4, the model performed well at a max length of 100; it achieved a validation accuracy of 96% and a validation loss of 0.22%. As the maximum sentence length of the model is increased beyond 128, the model overfits and eventually degrades performance. One of the most crucial hyperparameters to adjust in contemporary deep learning systems is batch size. Small batch sizes were more effective in our work on model generalization and achieved better accuracy. It has been demonstrated empirically that employing smaller batch sizes leads to faster convergence to good solutions. Smaller batch sizes enable the model to be- gin learning before needing to view all the data, which provides an intuitive explanation for this phenomenon. From Table 6, the validation accuracy of batch sizes 5, 6, 7, and 8 is 95% but with different validation losses. Batch size 6 achieves the best validation loss of 22% and batch size 8 has the least validation loss of 41%.

Model iteration affects experimental performance. From Table 7, The model's performance will initially increase and subsequently decline as the number of iterations increases. Results for the proposed model maximized around epoch 40. Aftermath, the performance of the model begins to de- cline.

5.7. Comparative study models

The original BERT, CNN, and BERT-CNN models are analyzed in comparison to the B-MLCNN model. The four models are run on the same training set and test set to check which of the models achieved higher accuracy. Table 8 additionally includes IMDb dataset performance numbers for BERT, CNN, BERT-CNN, and B-MLCNN. BERT has a 94% accuracy rate, a 93% precision rate, a 95% recall rate, and an F-1 score of 93%. Results for CNN are identical, coming in at 91%, 91%, 93%, and 92%. The same percentages were recorded by BERT-CNN: 93%, 90%, 93%, and 93%. A 95% accuracy, 95% precision, 94% recall, and 95% F1 score were achieved by our proposed framework (B-MLCNN). Table 9 also includes performance metrics for BERT, CNN, BERT-CNN, and B-MLCNN on the MR (2002) dataset. BERT has an F-1 score of 87%, an accuracy of 88%, a precision of 86%, and a recall of 86%. CNN also receives identical results:

84%, 85%, 86%, and 85%. Similarly, BERT-CNN found results of 87%, 84%, 87%, and 86%. Our proposed framework (B-MLCNN) achieved 88% accuracy, 87% precision, 86% recall, and an 88% F1 score in this instance. Table 10 also includes the results for the MR (2004) dataset for BERT, CNN, BERT-CNN, and B-MLCNN. BERT achieves an 89% accuracy, 89% precision, 88% recall, and an 88% F-1 score. Similar to the results, CNN has an 89, 89, 90, and 89% satisfaction rate. Similar results of 90%, 88%, 91%, and 90% were also reported by BERT-CNN. We recorded 95% accuracy, 94% precision, 94% recall, and a 93% F1 score with our proposed framework (B-MLCNN). Table 11 displays the metrics for the performance of BERT, CNN, BERT-CNN, and B-MLCNN on the Amazon dataset. BERT performs at 91% accuracy, 89% precision, 90% recall, and 90% F-1 score. CNN also receives similar results: 89%, 89%, 91%, and 90%. Similarly, BERT-CNN found scores of 91%, 88%, 91%, and a 90%. We measured 95% accuracy, 94% precision, 94% recall, and a 93% F1 score with our proposed framework (B-MLCNN).

The findings demonstrate the effectiveness of multi-channel convolutional neural networks in extracting information from reviews that are text-based. This combination made it possible to take advantage of each model's strengths.

5.8. Models performance comparison

Table 12 presents the accuracy of the competitive models over the IMDB datasets, from which we draw several conclusions. First, the XLNet model achieves overall better results than all the others. It is possible that the fact that the XLNet supports learning via bidirectional contexts in order to maximize the predicted likelihood of total permutations of the factorization order is one of the reasons why it performs so much better than other neural net- works. Among the competing models, BERTbase + ITPT and BERTlarge + ITPT placed second, indicating their expressiveness in word representation learning and the effectiveness of fine-tuning adaptation to the target domain task. On the IMDB dataset, the proposed model (B-MLCNN) demonstrated a notable performance improvement in accuracy measures.

As a results, it outperforms most baseline techniques with a respectable margin of success. 5.12% from Single layer BILSTM, 6.2% from WALE-LSTM, 5.6% from LSTM + CBA + LA, 2.57% from DV-ngrams-cosine, 8.55% from S-LSTM, 7.1% from LSTM, 7.4% from Doc2Vec, 6.8% from CNN + LSTM, and 5.6% from LSTM with dynamic skip. We derive numerous inferences based on the accuracy, precision, recall, and f-1 score displayed in Table 13 for the different models across the IMDB datasets. The proposed model achieved a higher level of accuracy than the baseline models by a margin of 10.6% for MNB, 5.14% for BERT, 9.23% for TextCNN, 3.9% for T-capsules, and 12.34% for SVM-RFE. MNB has a 10% precision margin, BERT 6.41%, TextCNN 10.63%, T-capsule 5.31%, and SVM-RFE 14.02, respectively. Recall percentages vary widely between different methods: MNB's is 7%, BERT's is 3.45%, TextCNN's is 7.98%, T-capsule's is 2.3%, and SVM-RFE's is 10.25. MNB's, BERT's, TextCNN's, T-capsule's, and SVM-RFE's f-1 score margins are all 9%, whereas SVM-RFE's is 12.66%. In Table 14, we compare the pro- posed model with baseline models on a variety of datasets in terms of their accuracy, precision, recall, and f-1 score. The proposed model achieved a 20% higher accuracy than CNNrand, 15.8% higher than CNN, 18.4% higher than LSTM, 13.3% higher than NSCL, 13.5 higher than SentiCNNLBAM3, 4.9% higher than SelfAtt, 4.8% higher than SentiCNNLBAM3, and 7.1% higher than SCL-NMA compared to the baseline models. CNNrand achieves a 20.9% precision margin, CNN achieves 15.5%, LSTM achieves 19.2%, NSCL achieves 12.9%, SentiCNNLBAM3 achieves 13.2%, SelfAtt achieves 5.3%, SentiCNNLBAM3 achieves 5.5%, and SCL-NMA achieves 1.42%. CNNrand has a 17.1% recall margin, CNN 15.3%, LSTM 15.3%, NSCL 13.6%, SentiC-NNLBAM3 12.9%, SelfAtt 3.5%, SentiCNNLBAM3 1.9%, and SCL-NMA 11.1%. For the f-1 score, CNNrand has a margin of 19.5%, CNN has a margin of 17.8%, LSTM has a margin of 13.5%, NSCL has a margin of 13.5%, Sen- tiCNNLBAM3 has a margin of 4.2%, and SCL-NMA has a margin of 7.1%. These results demonstrate the efficacy of B-MLCNN in the IMDB dataset for modeling interaction data and learning better representations for sentiment categorization.

Table 13
Accuracy, Precision, Recall, and the F-1 score are used in the Friedman Test (Test1) on the IMDB dataset. Friedman rank sum test; Friedman chi-squared = 13.688, df = 3, p-value = 0.003363.

Model	Accuracy	Precision	Recall	F-1 score
MNB	0.84	0.85	0.87	0.86
BERT	0.89	0.88	0.90	0.89
TextCNN	0.85	0.84	0.86	0.85
T-Capsule	0.91	0.89	0.91	0.90
SVM-RFE	0.82	0.80	0.83	0.82
BERT	0.94	0.93	0.95	0.93
CNN	0.91	0.91	0.93	0.92
BERT-CNN	0.93	0.90	0.93	0.93
B-MLCNN	0.95	0.95	0.94	0.95

Table 14
Accuracy, Precision, Recall, and the F-1 score are used in the Friedman Test (Test2) on different datasets. Friedman rank sum test; Friedman chi-squared = 0.78505, df = 3, p-value = 0.853.

Model	Accuracy	Precision	Recall	F-1 score	Dataset
CNNrand	0.75	0.74	0.76	0.75	MR
CNN(Mchannels)	0.79	0.79	0.78	0.77	MR
LSTM	0.76	0.75	0.78	0.77	MR
NCSL	0.81	0.82	0.80	0.81	MR
SentiCNNLBAM3	0.81	0.81	0.81	0.81	MR
SelfAtt	0.90	0.89	0.90	0.90	SLS
SentiCNNLBAM3	0.90	0.89	0.92	0.90	SLS
SCL-NMA	0.87	0.93	0.82	0.87	SLS
BERT	0.94	0.93	0.95	0.93	IMDB
CNN	0.91	0.91	0.93	0.92	IMDB
BERT-CNN	0.93	0.90	0.93	0.93	IMDB
B-MLCNN	0.95	0.95	0.94	0.95	IMDB

5.9. Friedman test

In order to empirically compare the performance of several approaches used for sentiment classification. This study uses a statistical significance assessment approach based on the influential work of (Kasihmuddin et al., 2022; Zamri et al., 2022). The Friedman test is used in this method. The goal of the Friedman test is to compare the efficacy of several deep learning approaches to sentiment analysis and determine whether or not there are any significant variations in performance. The following hypotheses are tested statistically using the Friedman test:

- H_o: Mean values for all four metrics are the same on the IMDB dataset.

  H_a: Mean values are different for four metrics on the IMDB dataset.

$$H_o : \mu_{Accuracy} = \mu_{Precision} = \mu_{Recall} = \mu_{F-1score} onIMDBdataset \tag{14}$$

$$H_a : \mu_{Accuracy} \neq \mu_{Precision} \neq \mu_{Recall} \neq \mu_{F-1score} onIMDBdataset \tag{15}$$

- H_o: Mean values for all four metrics are the same on different datasets.

  H_a: Mean values for all four metrics are different on different datasets.

$$H_o : \mu_{Accuracy} = \mu_{Precision} = \mu_{Recall} = \mu_{F-1score} ondifferentdatase \tag{16}$$

$$H_a : \mu_{Accuracy} \neq \mu_{Precision} \neq \mu_{Recall} \neq \mu_{F-1score} ondifferentdataset \tag{17}$$

Fig. 7 shows that the median scores for all measures are between 90 and 92, with accuracy and F1-score median values similar to precision and recall.

Fig. 8 shows that accuracy and F1-score are closer than precision and recall medians. The recall median is lower than the other three. From Table 13, At 5% level of significance, we reject the null hypothesis. From Table 14, we fail to reject the null hypothesis. According to the Friedman test in Table 13, there is statistical evidence that the means of the four evaluation measures are different.

6. Discussion

All the datasets' assessment metrics observed from the B-MLCNN studies are shown in Tables 4–7. The assessment procedure examines the B- MLCNN's testing performance in terms of accuracy, precision, recall, and f-1 score on standard datasets from Fig. 4. In this technique, the IMDB, MR (2004), and Amazon datasets reach the highest accuracy increase of the B-MLCNN task at 95%. The MR (2002) dataset shows a minimal accuracy of 88%. The IMDB dataset has a maximum precision of 95%, while the MR (2002) dataset has a minimum precision of 87%. The highest recall, 94%, is connected to the MR (2004), IMDB, and Amazon datasets. The MR (2002) dataset recorded a minimum recall of 86%. The IMDB dataset achieves the greatest f-1 score of 95%, and the least f-1 score recorded on the MR (2002) dataset of 88%. In Fig. 5a, the evaluation process analyzes BERT, CNN, BERT-CNN, and B-MLCNN's testing performance in terms of accuracy, precision, recall, and f-1 score on IMDB datasets. The B-MLCNN achieved the highest accuracy, precision, recall, and f-1 scores of 95%, 95%, 94%, and 95% respectively. CNN recorded the lowest score on all the evaluation metrics (EM). The B-MLCNN also achieved the greatest score on all EM, while CNN recorded the least on all EM on the MR (2002) dataset, as shown in Fig. 5b. From Fig. 6a and 6b, the B-MLCNN achieved the greatest results on all EM on the MR (2004) and Amazon reviews, respectively. The CNN model still recorded minimal results on all EM. Fig. 7 displays the receiver operating characteristic (ROC) curves for BERT, CNN, BERT-CNN, and the proposed model (B-MLLNN). The corresponding AUCs for BERT, CNN, BERT-CNN, and B-MLLN on the IMDB dataset are 95%, 93%, 96%, and 97%. With an AUC of 90%, BERT outperforms CNN (85%), BERT- CNN (89%), and B-MLCNN (90%) on MR (2002). On MR (2004), the AUCs for BERT is 91%, CNN is 90%, BERT-CNN is 93%, and B-MLCNN is 97%. Comparing the AUCs of BERT, CNN, BERT-CNN, and B-MLLNN on the Amazon reviews, we find that B-MLCNN achieves 96%, CNN 90%, and BERT-CNN 91%. These findings indicate the proposed approach is reliable and capable of accurately categorizing textual reviews.

Table 12 indicates that the BERT vectorization-equipped models outperformed the non-BERT vectorized models on the IMDB dataset. The findings of the experiments show that the BERT pre-train language model is useful in sentiment analysis because it captures the semantic and syntactic links present in textual reviews. In comparison to S-LSTM, Doc2Vec, LSTM, CNN + LSTM, WALE + LSTM, LSTM with dynamic skip,

LSTM + CBA + LA, single-layer BILSTM, and DV-ngrams-cosine, our model achieved better results by leveraging the combination of BERT and multilayered CNN, which demonstrates that B-MLCNN can identify complex features from textual re- views. On the IMDB dataset, $BERT_{base}$ + *ITPT and* $BERT_{large}$ + *ITPT* performed better than our model. Our model's use of multiple-feature ex- traction with additional parameters could be the reason for this, which might further contribute to the marginal effect on the accuracy at 95.01 as indicated in Table 12 for the proposed B-MLCNN. To experimentally examine the performance of a number of different methodologies that are used for sentiment categorization and to establish whether or not there are any significant variances in performance, the Friedman test was carried out. From Fig. 8, the median scores for all metrics are between 90 and 92, The accuracy and F1-score median values appear similar relative to the median values for precision and recall. From Fig. 9, accuracy and F1-score appear more similar relative to the median values of precision and recall. The median value for the recall is relatively lower compared to the other three.

The B-MLCNN illustrates a unique method for establishing document level sentiment classification by acquiring it implicitly as input. The pro- posed approach forecasts the sentiment of the documents based on these identified document reviews. So, using a variety of metrics viewpoints, that encompass accuracy, precision, recall, and F1 measures, we summarize the overall B-MLCNN performance in several domains of standardized datasets during the work of sentiment classification. In comparison to the majority of the baseline techniques, the framework performs much better on all datasets. As the illustration demonstrates, the B-MLCNN performs the task of document-level sentiment classification with higher accuracy. The more promising performance of the technique can be attributed to the utilization of multiple convolutional neural networks with various ambient kernel sizes in conjunction with BERT for the representation of text. This serves as the scientific foundation and promises to be efficient to deploy in practical

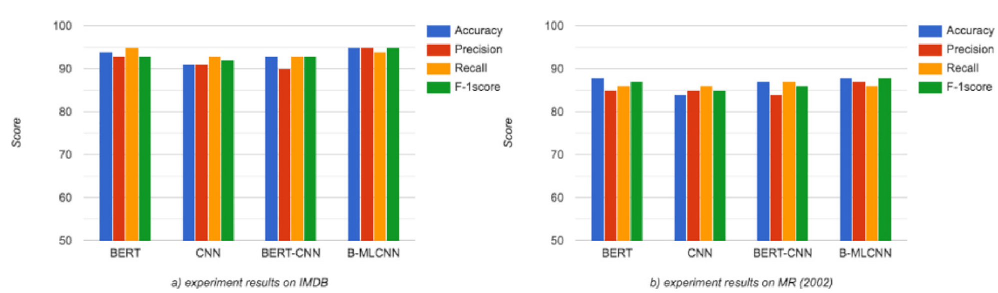

Fig. 5. Model Comparison on IMDB and MR (2002) Dataset.

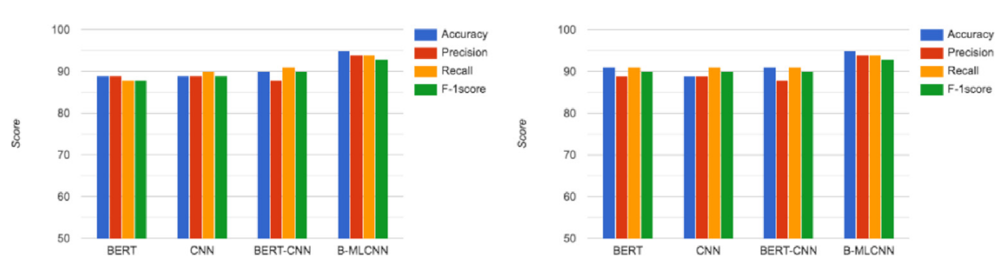

Fig. 6. Model Comparison on MR (2004) and Amazon Dat.

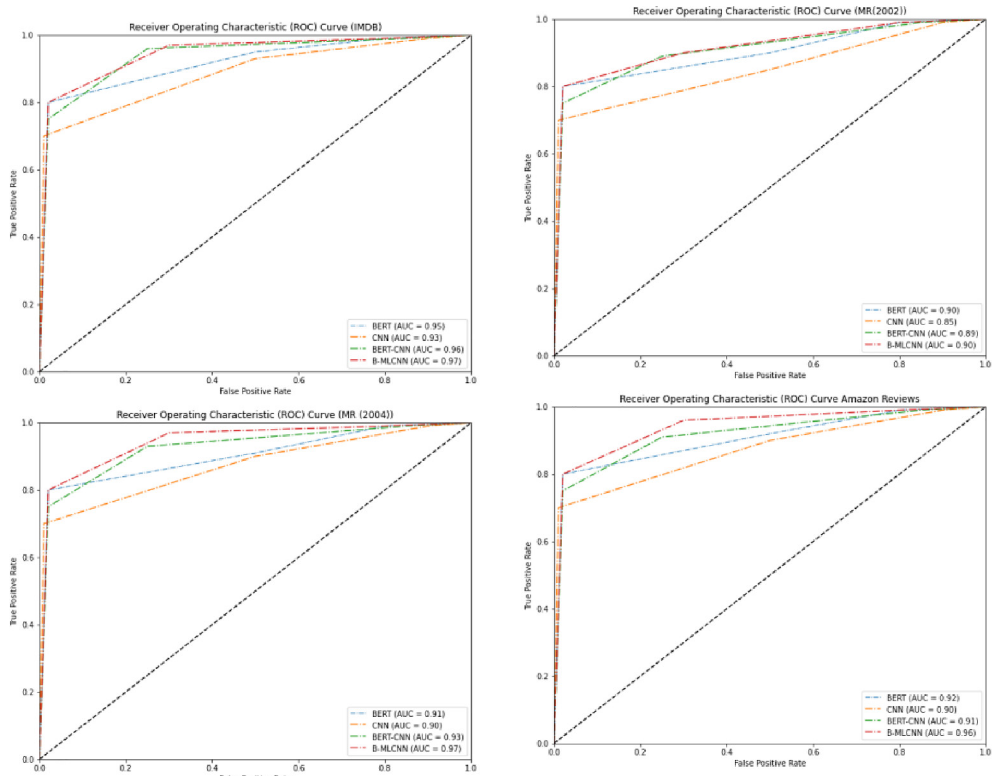

Fig. 7. ROC curves on IMDB, MR (2002), MR (2004) and Amazon Reviews Dataset.

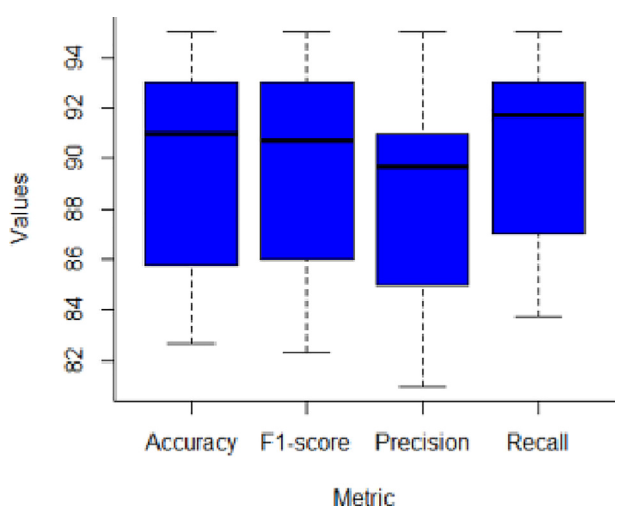

Fig. 8. A Comparison of the median scores of the evaluation metrics on IMDB data.

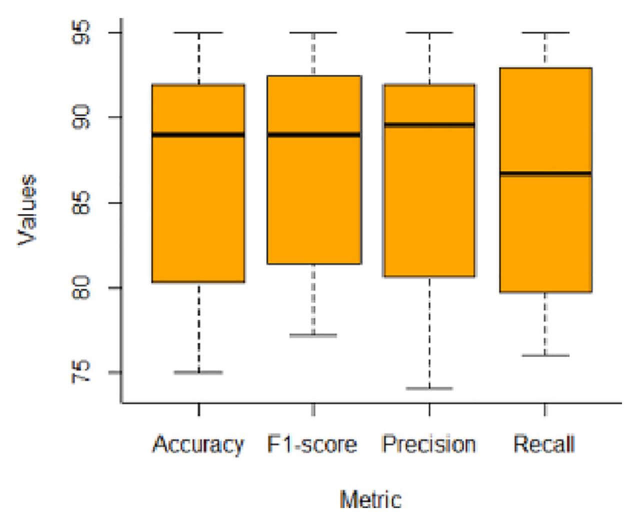

Fig. 9. A Comparison of median scores of four evaluation metrics on different datasets.

applications. As a result, the empirical analytical successes on multiple datasets demonstrate the value of deep feature collaboration and convolutional neural network multi-channels.

7. Conclusion

A BERT MultiLayered Convolutional Neural Network (B-MLCNN) as an integrated deep learning paradigm is presented to classify sentiments as present in the textual reviews of a document. With B-MLCNN, each review is treated as a separate document. The proposed B-MLCNN is computation- ally viable in terms of system performance as it derives advantages in terms of feature extraction and classification of document-based sentiments from both BERT and multi-layer CNN approaches. Furthermore, the effective- ness of the proposed B-MLCNN lends credence to the claim that merging BERT with a multi-layer convolutional neural network can enhance the solution to the problem of document-level sentiment categorization. Our findings demonstrated that the proposed model for document-based sentiment classification performed well on the majority of the datasets used. For all four metrics (accuracy, precision, recall, and f-1) the MR (2002) dataset had the lowest values. While the proposed model showed effectiveness in terms of learning and comprehending corpus complex semantic relations, it struggled in areas such as comprehending the contextual meaning of sentences in textual reviews. The large number of parameters also makes it difficult to determine the optimal combination via thorough parameter adjustment. Therefore, some of the parameters of the model are selected by studying the values that other related models use. A future study will focus on using B-MLCNN to identify the explicit polarity based on the contextual position of the text for sentiment classification when subjected to multiple datasets.

Declaration of Competing Interest

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

Acknowledgement

This work was supported by the Sichuan Province Science and Technology Program (Grant Nos. 2021YFS0391, 2022ZDZX0004, and 2023YFG002).

References

Amin, M.Z., Nadeem, N., 2018. Convolutional neural network: Text classification model for open domain question answering sys- tem. CoRR abs/1809.02479. URL: http://arxiv.org/abs/1809.02479, arXiv:1809.02479.

Atandoh, P., Fengli, Z., Adu-Gyamfi, D., Leka, H.L., Atandoh, P.H., 2021. A glove cnn-bilstm sentiment classification, in: 2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), pp. 245–249. doi:10.1109/ICCWAMTIP53232.2021.9674171.

Baccianella, S., Esuli, A., Sebastiani, F., 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in: Calzolari, N., Choukri, K., Mudguard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D. (Eds.), Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta, European Language Resources Association. URL: http://www.lrec-conf.org/proceedings/lrec2010/summaries/769.html.

Bello, A., Ng, S.C., Leung, M., 2023. A BERT framework to sentiment analysis of tweets. Sensors 23, 506. https://doi.org/10.3390/s23010506.

Bodapati, J.D., Veeranjaneyulu, N., Shareef, S.N., 2019. Sentiment analysis from movie reviews using lstms. Ingénierie des Syst'emes d Inf. 24, 125–129. https://doi.org/10.18280/isi.240119.

Camacho-Collados, J., Pilehvar, M.T., 2018. On the role of text preprocessing in neural network architectures: An evaluation study on text categorizat and sentiment analysis. In: Proceedings of the 2018 EMNLP Work- shop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Association for Computational Linguistics, Brussels, Belgium. pp. 40– 46. URL: https://aclanthology.org/W18-5406, doi:10.18653/v1/W18-5406.

Chen, B., Xu, Z., Wang, X., Xu, L., Zhang, W., 2020. Capsule network-based text sentiment classification. IFAC-PapersOnLine 53, 698–703. URL: https://www.sciencedirect.com/science/article/pii/S2405896321003001, doi:Doi: 10.1016/j.ifacol.2021.04.160. 3rd IFAC Workshop on Cyber-Physical Human Systems CPHS 2020.

Chen, M., 2017. Efficient vector representation for documents through corruption, in: 5th International Conference on Learn- ing Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings, OpenReview.net.URL: https://openreview.net/forum?id=B1Igu2ogg.

Cui, Y., Huang, C., 2021. A chinese text classification method based on bert and convolutional neural network, in: 2021 7th In- ternational Conference on Systems and Informatics (ICSAI), pp. 1–6. doi:10.1109/ICSAI53574.2021.9664066.

Devlin, J., Chang, M., Lee, K., Toutanova, K., 2019. BERT: pre-training of deep bidirectional transformers for language understanding, in: Burstein, J., Doran, C., Solorio, T. (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers), Association for Computational Linguistics. pp. 4171–4186. URL: Doi: 10.18653/v1/n19-1423, doi:10.18653/v1/n19-1423.

Dong, Y., Fu, Y., Wang, L., Chen, Y., Dong, Y., Li, J., 2020. A sentiment analysis method of capsule network based on bilstm. IEEE Access 8, 37014–37020. https://doi.org/10.1109/ACCESS.2020.2973711.

Elnagar, A., Al-Debsi, R., Einea, O., 2020. Arabic text classification using deep learning models. Inf. Process. Man- age. 57. URL: Doi: 10.1016/j.ipm.2019.102121, doi:10.1016/j.ipm.2019.102121.

Fu, X., Yang, J., Li, J., Fang, M., Wang, H., 2018. Lexicon-enhanced LSTM with attention for general sentiment analysis. IEEE. Access 6, 71884–71891. https://doi.org/10.1109/ACCESS.2018.2878425.

Gui, T., Zhang, Q., Zhao, L., Lin, Y., Peng, M., Gong, J., Huang, X., 2019. Long short-term memory with dynamic skip connections, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, AAAI Press. pp. 6481–6488. doi:10.1609/aaai.v33i01.33016481.

Gui, L., Leng, J., Zhou, J., Xu, R., He, Y., 2022. Multi task mutual learning for joint sentiment classification and topic detection. IEEE Transactions on Knowledge and Data Engineering 34, 1915–1927. https://doi.org/10.1109/TKDE.2020.2999489.

Hameed, Z., Garcia-Zapirain, B., 2020. Sentiment classification using a single-layered bilstm model. IEEE Access 8, 73992–74001. https://doi.org/10.1109/ACCESS.2020.2988550.

Hassan, S.U., Ahamed, J., Ahmad, K., 2022. Analytics of ma- chine learning-based algorithms for text classification. Sustainable Operations and Computers 3, 238–248. URL: https://www.sciencedirect.com/science/article/pii/S2666412722000101, doi:Doi: 10.1016/j.susoc.2022.03.001.

Huang, M., Xie, H., Rao, Y., Liu, Y., Poon, L.K.M., Wang, F.L., 2022a. Lexicon-based sentiment convolutional neural networks for online review analysis. In: IEEE Transactions on Affective Computing, vol. 13, no 3, pp. 1337–1348. https://doi.org/10.1109/TAFFC.2020.2997769.

Huang, F., Yuan, C., Bi, Y., Lu, J., Lu, L., Wang, X., 2022b. Multi-granular document-level sentiment topic analysis for online reviews. Appl. Intell. 52, 7723–7733. https://doi.org/10.1007/s10489-021-02817-1.

Ionescu, R.T., Butnaru, A., 2019. Vector of locally-aggregated word embeddings (VLAWE): A novel document-level representation. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota. pp. 363–369. URL: https://aclanthology.org/N19-1033, doi:10.18653/v1/N19-1033.

Ji, P., Zhang, H.Y., Wang, J.Q., 2019. A fuzzy decision support model with sentiment analysis for items comparison in e-commerce: The case study of http://pconline.com. IEEE Transactions on Systems, Man, and Cybernetics: Systems 49, 1993–2004. https://doi.org/10.1109/TSMC.2018.2875163.

Jia, K., 2022. Sentiment classification of microblog: A frame- work based on bert and cnn with attention mechanism. Computers and Electrical Engineering 101,108032. URL: https://www.sciencedirect.com/science/article/pii/S0045790622002956, doi:Doi: 10.1016/j.compeleceng.2022.108032.

Kasihmuddin, M.S.M., Jamaludin, S.Z.M., Mansor, M.A., Wahab, H.A., Ghadzi, S.M.S., 2022. Supervised learning perspective in logic mining. Mathematics 10 URL: https://www.mdpi.com/2227-7390/10/6/915. https://doi.org/10.3390/math10060915.

Kim, Y., 2014. Convolutional neural networks for sentence classification CoRR abs/1408.5882. URL: http://arxiv.org/abs/1408.5882, arXiv:1408.5882.

Kit, Y., Mokji, M.M., 2022. Sentiment analysis using pre-trained language model with no fine-tuning and less resource. IEEE Access 10, 107056–107065. https://doi.org/10.1109/ACCESS.2022.3212367.

Li, H., Ma, Y., Ma, Z., Zhu, H., 2021. Weibo text sentiment analysis based on bert and deep learning. Applied Sciences 11. https://doi.org/10.3390/app112210774.

Liang, R., Wang, J.K., 2019. A Linguistic Intuitionistic Cloud Decision Support Model with Sentiment Analysis for Product Se- lection in E-commerce. International Journal of Fuzzy Systems 21, 963–977. https://doi.org/10.1007/s40815-019-00606-0.

Liu, S., Lee, I., 2021. Sequence encoding incorporated cnn model for email document sentiment classification. Applied Soft Computing 102,107104. URL: https://www.sciencedirect.com/science/article/pii/S1568494621000272, doi:Doi: 10.1016/j.asoc.2021.10710.

Liu, S., Wang, X., Zhao, L., Li, B., Hu, W., Yu, J., Zhang, Y., 2022. 3dcann: A spatio-temporal convolution attention neural network for EEG emotion recognition. IEEE J. Biomed. Health Informatics 26, 5321–5331. https://doi.org/10.1109/JBHI.2021.3083525.

Liu, S., Wang, Z., An, Y., Zhao, J., Zhao, Y., Zhang, Y., 2023. EEG emotion recognition based on the attention mechanism and pretrained convolution capsule network. Knowl. Based Syst. 265,. https://doi.org/10.1016/j.knosys.2023.110372 110372.

Liu, B., 2015. Sentiment Analysis Mining Opinions, Sentiments, and Emotions. Cambridge University Press. URL: http://www.cambridge.org/us/academic/subjects/computer-science/knowledge-management-databases-and-data-mining/sentiment-analysis-mining-opinions-sentiments-and-emotions.

Long, Y., Lu, Q., Xiang, R., Li, M., Huang, C.R., 2017. A cognition based attention model for sentiment analysis. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Copenhagen, Denmark. pp. 462–471. URL: https://aclanthology.org/D17-1048, doi:10.18653/v1/D17-1048.

Lyu, C., Foster, J., Graham, Y., 2020. Improving document-level sentiment analysis with user and product context. In: Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online). pp. 6724–6729. URL: https://aclanthology.org/2020.coling-main.590, doi: 10.18653/v1/2020.coling-main.590.

Mohd Nafis, N.S., Awang, S., 2021. An enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification. IEEE Access 9, 52177–52192. https://doi.org/10.1109/ACCESS.2021.3069001.

Ni, J., Li, J., McAuley, J., 2019. Justifying recommendations using distantly- labeled reviews and fine-grained aspects. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China. pp. 188–197. URL: https://aclanthology.org/D19-1018, doi:10.18653/v1/D19-1018.

Ouyang, X., Zhou, P., Li, C.H., Liu, L., 2015. Sentiment analysis using convolutional neural network, in: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Se- cure Computing; Pervasive Intelligence and Computing, pp. 2359–2364. doi:10.1109/CIT/IUCC/DASC/PICOM.2015.349.

Pang, B., Lee, L., 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, in: Scott, D., Daelemans, W., Walker, M.A. (Eds.), Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, 21-26 July, 2004, Barcelona, Spain, ACL. pp. 271–278. URL: https://aclanthology.org/P04-1035/, doi:10.3115/1218955.1218990.

Pang, B., Lee, L., Vaithyanathan, S., 2002. Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, EMNLP 2002, Philadelphia, PA, USA, July 6-7, 2002, pp. 79–86. URL: https://aclanthology.org/W02-1011/, doi:10.3115/1118693.1118704.

Phan, H.T., Nguyen, N.T., Hwang, D., 2022. Aspect-level sentiment analysis using cnn over bert-gcn. IEEE Access 10, 110402–110409. https://doi.org/10.1109/ACCESS.2022.3214233.

Socher, R., Huval, B., Manning, C.D., Ng, A.Y., 2012. Semantic compositionality through recursive matrix-vector spaces, in: Tsujii, J., Henderson, J., Pasca, M. (Eds.), Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL 2012, July 12-14, 2012, Jeju Island, Korea, ACL. pp. 1201–1211. URL: https://aclanthology.org/D12-1110/.

Sun, C., Qiu, X., Xu, Y., Huang, X., 2019. How to fine-tune BERT for text classification? in: Sun, M., Huang, X., Ji, H., Liu, Z., Liu, Y. (Eds.), Chinese Computational Linguistics - 18th China National Conference, CCL 2019, Kunming, China, October 18-20, 2019, Proceedings, Springer. pp. 194–206. URL: Doi: 10.1007/978-3-030-32381-3 16, doi:10.1007/978-3-030-32381-3_16.

Tai, K.S., Socher, R., Manning, C.D., 2015. Improved semantic representations from tree-structured long short-term memory networks. CoRR abs/1503.00075. URL: http://arxiv.org/abs/1503.00075, arXiv:1503.00075.

Teng, Z., Vo, D., Zhang, Y., 2016. Context-sensitive lexicon features for neural sentiment analysis, in: Su, J., Carreras, X., Duh, K. (Eds.), Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016, The Association for Computational Linguistics. pp. 1629–1638. URL: Doi: 10.18653/v1/d16-1169, doi:10.18653/v1/d16-1169.

Thongtan, T., Phienthrakul, T., 2019. Sentiment classification using document embeddings trained with cosine similarity. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Stu- dent Research Workshop, Association for Computational Linguistics, Florence, Italy. pp. 407–414. URL: https://aclanthology.org/P19-2057, doi:10.18653/v1/P19-2057.

Tsutsumi, K., Shimada, K., Endo, T., 2007. Movie review classification based on a multiple classifier. In: Proceedings of the 21st Pacific Asia Conference on Language, Information and Computation, The Korean Society for Language and Information (KSLI), Seoul National University, Seoul, Korea. pp. 481–488. URL: https://aclanthology.org/Y07-1050, doi:http://hdl.handle.net/2065/29106.

Turney, P.D., 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, USA. p. 417–424. URL: Doi: 10.3115/1073083.1073153, doi:10.3115/1073083.1073153.

van Dinter, R., Catal, C., Tekinerdogan, B., 2021. A multi- channel convolutional neural network approach to automate the citation screening process. Appl. Soft Comput. 112, 107765. https://doi.org/10.1016/j.asoc.2021.107765.

Wang, X., Ma, Y., Cammon, J., Fang, F., Gao, Y., Zhang, Y., 2023. Self- supervised eeg emotion recognition models based on cnn. IEEE Transactions on Neural Systems and Rehabilitation Engineering 31, 1952–1962. https://doi.org/10.1109/TNSRE.2023.3263570.

Williams, L., Bannister, C., Arribas-Ayllon, M., Preece, A., Spasić, I., 2015. The role of idioms in sentiment analysis. Expert Syst. Appl. 42, 7375–7385. https://doi.org/10.1016/j.eswa.2015.05.039.

Wu, P., Li, X., Ling, C., Ding, S., Shen, S., 2021. Sentiment classification using attention mechanism and bidirectional long short- term memory network. Applied Soft Computing 112, 107792. URL: https://www.sciencedirect.com/science/article/pii/S1568494621007134, doi:Doi: 10.1016/j.asoc.2021.107792.

Wu, O., Yang, T., Li, M., Li, M., 2022. Two-level lstm for sentiment analysis with lexicon embedding and polar flipping. IEEE Transactions on Cybernetics 52, 3867–3879. https://doi.org/10.1109/TCYB.2020.3017378.

Xu, F., Pan, Z., Xia, R., 2020. E-commerce product review sentiment classification based on a naive Bayes continuous learning framework. Inf. Process. Manag. 57,. https://doi.org/10.1016/j.ipm.2020.102221 102221.

Yan, W., Zhou, L., Qian, Z., Xiao, L., Zhu, H., 2021. Sentiment analysis of student texts using the cnn-bigru-at model. Sci. Program. 2021, 8405623:1–8405623:9. URL: Doi: 10.1155/2021/8405623, doi:10.1155/2021/8405623.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J.G., Salakhutdinov, R., Le, Q.V., 2019. Xlnet: Generalized autoregressive pretraining for language understanding, in: Wallach, H.M., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E.B., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, pp. 5754–5764. URL: https://proceedings.neurips.cc/paper/2019/hash/dc6a7e655d7e5840e66733e9ee67cc69-.

Yenter, A., Verma, A., 2017. Deep cnn-lstm with combined kernels from multiple branches for imdb review sentiment analysis. In: 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), p. 540546. https://doi.org/10.1109/UEMCON.2017.8249013.

Yu, Q., Wang, Z., Jiang, K., 2021. Research on text classification based on bert-bigru model. Journal of Physics: Conference Series 1746, 012019. https://doi.org/10.1088/1742-6596/1746/1/012019.

Zamri, N.E., Azhar, S.A., Mansor, M.A., Alway, A., Kasihmuddin, M.S.M., 2022. Weighted random k satisfiability for k=1,2 (r2sat) in discrete hopfield neural network. Applied Soft Computing 126, 109312. URL: https://www.sciencedirect.com/science/article/pii/S1568494622004902, doi:Doi: 10.1016/j.asoc.2022.109312.

Zeng, D., Dai, Y., Li, F., Wang, J., Sangaiah, A.K., 2019. As- pect based sentiment analysis by a linguistically regularized CNN with gated mechanism. J. Intell. Fuzzy Syst. 36, 3971–3980. https://doi.org/10.3233/JIFS-169958.

Zhang, Y., Liu, Q., Song, L., 2018. Sentence-state LSTM for text repre- sentation. In: Proceedings of the 56th Annual Meeting of the Associa- tion for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia. pp. 317–327. URL: https://aclanthology.org/P18-1030, doi:10.18653/v1/P18-1030.

Zhang, Y., 2021. Research on text classification method based on lstm neural network model. In: 2021 IEEE Asia-Pacific Conference on Im- age Processing, Electronics and Computers (IPEC), pp. 1019–1022. doi:10.1109/IPEC51340.2021.9421225.

Zhao, W., Zhang, G., Yuan, G., Liu, J., Shan, H., Zhang, S., 2020. The study on the text classification for financial news based on partial information. IEEE Access 8, 100426–100437. https://doi.org/10.1109/ACCESS.2020.2997969.

Zulqarnain, M., Ghazali, R., Hassim, Y.M.M., Aamir, M., 2021. An enhanced gated recurrent unit with auto-encoder for solving text clas- sification problems. Arabian Journal for Science and Engineering 46, 8953–8967. https://doi.org/10.1007/s13369-021-05691-8.