




Article

Hybrid Multichannel-Based Deep Models Using Deep Features for Feature-Oriented Sentiment Analysis

Waqas Ahmad ¹, Hikmat Ullah Khan ^{1,*} , Tasswar Iqbal ¹, Muhammad Attique Khan ^{2,*} ,
Usman Tariq ³  and Jae-hyuk Cha ⁴

¹ Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah 47040, Pakistan

² Department of Computer Science, HITEC University, Taxila 47080, Pakistan

³ Management Information System Department, College of Business Administration, Prince Sattam Bin Abdulaziz University, Al-Kharj 16278, Saudi Arabia

⁴ Department of Computer Science, Hanyang University, Seoul 04763, Republic of Korea

* Correspondence: hikmat.ullah@ciitwah.edu.pk (H.U.K.); attique.khan@hitecuni.edu.pk (M.A.K.)

Abstract: With the rapid growth of user-generated content on social media, several new research domains have emerged, and sentiment analysis (SA) is one of the active research areas due to its significance. In the field of feature-oriented sentiment analysis, both convolutional neural network (CNN) and gated recurrent unit (GRU) performed well. The former is widely used for local feature extraction, whereas the latter is suitable for extracting global contextual information or long-term dependencies. In existing studies, the focus has been to combine them as a single framework; however, these approaches fail to fairly distribute the features as inputs, such as word embedding, part-of-speech (PoS) tags, dependency relations, and contextual position information. To solve this issue, in this manuscript, we propose a technique that combines variant algorithms in a parallel manner and treats them equally to extract advantageous informative features, usually known as aspects, and then performs sentiment classification. Thus, the proposed methodology combines a multichannel convolutional neural network (MC-CNN) with a multichannel bidirectional gated recurrent unit (MC-Bi-GRU) and provides them with equal input parameters. In addition, sharing the information of hidden layers between parallelly combined algorithms becomes another cause of achieving the benefits of their combined abilities. These abilities make this approach distinctive and novel compared to the existing methodologies. An extensive empirical analysis carried out on several standard datasets confirms that the proposed technique outperforms the latest existing models.

Keywords: sentiment analysis; aspect extraction; word embedding; attention mechanism; contextual positional information; multichannel convolutional neural network



Citation: Ahmad, W.; Khan, H.U.; Iqbal, T.; Khan, M.A.; Tariq, U.; Cha, J.-h. Hybrid Multichannel-Based Deep Models Using Deep Features for Feature-Oriented Sentiment Analysis. *Sustainability* **2023**, *15*, 7213. <https://doi.org/10.3390/su15097213>

Academic Editors: Ming Hour Yang, Vijayalakshmi Murugesan and Mercy Shalinie Selvaraj

Received: 19 February 2023

Revised: 10 April 2023

Accepted: 20 April 2023

Published: 26 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the age of social media, diverse industries and firms' advisories broadly depend upon user-generated opinions for forecasting their future earnings. These opinionated unstructured data are available as reviews, blog discussions, graphics, audio, videos, and other types of media that do not support any structure. That made this field challenging due to ambiguities of natural language, exponential increases in social media web content, and indirect sentiments expressed in user-generated context [1]. In that situation, data analysts widely considered ABSA to understand the users' or consumers' requirements, filtration of unrequired data, and obtain relevant suggestions that make their organizational and industrial decision appropriate. Generally, two types of online attitudes, reviews, or opinions are observed, i.e., product reviews and experience sharing regarding these products or services. The first type discusses features of a particular entity, such as a product or service. Moreover, the second type compares various features of entities to identify their pros and cons [2].

The extraction of accurate features of a targeted entity became a critical issue in the NLP field due to the complex nature of contextual information. The contextual information around these advantageous features of a targeted entity considers highly important in these circumstances because they provide valuable clues for accurate identification and extraction [3,4]. Moreover, the precise identification and extraction of these features still demand attention from the research community. Traditionally, these extractions and identifications of features accomplish through various existing methodologies, such as machine learning [5–7]; topic modeling [8–10]; and lexicon-based [11–13], rule-based, and syntactic relation-based [14–18] methods. Syntactic pattern techniques perform well while extracting features and classifying their sentiments. However, these methodologies are discouraged due to their time consumption and specialists' demand for creating rules and lexicons, which restrict them to a specific domain and language [19,20]. Additionally, the supervised methodologies of machine learning highly rely upon a large volume of labeled datasets that expresses a bottleneck of these methods [21]. On the other side, the semi-supervised methodology demands less labeled data for training, whereas their methods' complexity for feature selection becomes the prominent cause of obstinate such approaches. While unsupervised methods highly relied upon the manual feature engineering mechanism. The quality of such extracted features depends on the manual feature engineering process, which affects the scalability and adaptability of these approaches to various fields' applications in daily life [22,23].

One of the most highly recommended approaches in the machine learning field is known as deep learning (DL). It addresses diverse NLP challenges such as machine translation, named entity reorganization, and sentiment analysis (SA) [24]. Recent advancements in the field of NLP highly relied upon DL architectures for the extraction of such valuable features along with their sentiment classification. Recurrent and convolutional neural networks are the two well-leading architectures of DL methodologies. CNN has achieved that position due to the utilization of convolution kernels, which make it distinctive while extracting targeted features. On the other hand, recurrent neural networks (RNN) and their variations, such as long short-term memory (LSTM) or GRU, do not resemble during the extraction of contextual information, which can be versatile under varying circumstances [25].

RNN analyzes the whole sentence, word by word, to capture the semantic information of the sentence in the form of hidden layers. It also captures long semantical dependencies of long-contextual facts but in a biased manner. It means each upcoming word is more dominant compared to the previous work. However, worthwhile terms/words can occur in any position of a sentence, which becomes the prominent cause of the model's effectiveness reduction. Generally, the RNN-based model captures the sequential patterns through temporal features and long-term semantical dependencies among the pairs of words/terms while learning. In addition, these methods equally drew their attention to each word of the targeted sentence. Due to this, they did not distinguish between ordinary and prominent words, which are (compared to common words) more dominant and influential in contextual knowledge. It becomes the cause of performance degradation of RNN-based approaches [25–27].

On the other side, CNN has presented itself as a non-biased model. It comprises kernels and max-pooling layers to extract the prominent features of a targeted entity within a sentence. As a result, CNN captures the semantics of a sentence more effectively when compared to RNN. However, CNN is facing an issue while determining the optimistic size of the kernel. It means that their small size may lose some critical information. While on the other hand, the larger size may consider all non-required terms, which could cause erroneous training of the model. In addition, their utilization of filters effectively captures the local features that proved beneficial during the extraction of semantically highly effective terms. Moreover, such a model never demands order-sensitive long-term semantic dependencies of the sentence during training. Instead, they can only be trained based on the local feature information [25,26,28].

Recently, the researchers presented various developments in this trend with the involvement of the attention mechanism, which improves sentiment classification with a credible exposition of opinion targets [29]. Thus, we can conclude that both models (CNN and RNN) separately cannot deliver state-of-the-art performance while extracting feature terms and classifying their sentiments. Moreover, their combination can improve sentiment classification through the accurate extraction of features. Consequently, the existing approaches regarding feature terms identification and extraction found in the literature have considered such methodology. However, they combined them in a sequential/serial and parallel manner. In the case of the serial method, one of the models has obtained the actual text while the other acquires the output of the first model, which becomes the cause of information loss. In the case of the parallel technique, both models have never been treated equally in terms of inputs. It means that one of the models uses a whole set of inputs while the other obtains its subset. This situation underutilizes the abilities of one model compared to the other, which obstinate the benefits of parallel or simultaneous algorithms adhesion approach. This scenario demands the research community to develop such techniques where both models directly interact with the textual context in the form of an equal number of input parameters and then combine their learnings to exploit these models.

Both above-mentioned algorithms have their distinctive pros and cons. In the state of these motivations, the attention-based joint model (Att-JM) jointly utilizes them and applies the attention mechanisms to these for accurately identifying informative features and classifying their expressed sentiments. Consequently, while integrating these algorithms parallelly, the Att-JM shares information of hidden layers between them to achieve the benefits of their combined learnings. Additionally, it distributes an equal number of inputs between them to accomplish the main tasks of aspect-based sentiment analysis (ABSA). Thus, the main contribution of this paper has given as under:

1. The proposed approach proposes the parallel fusion of the multichannel convolutional neural network (MC-CNN) and multichannel gated recurrent unit (MC-GRU) with various deep features while accomplishing the main tasks of ABSA.
2. The approach proposes and explores the collective utilization and the uniform distribution effect of word2vec embedding and contextual position information on the performance of the merged deep learning-based algorithm during the aspect's identification and sentiment classification.
3. The proposed approach shares the information of hidden layers between merged models; thus, they attain the advantages of their combined abilities and learnings while predicting aspects and classifying their sentiments.
4. The proposed approach outperforms when assessed through evaluation metrics, e.g., precision, recall, and F1 measure on standardized datasets comprising SemEval and Twitter. Therefore, the F1 measure depicts 95% achievement of the proposed approach in the aspect term extraction (ATE) task and a performance of 92% in the sentiment classification (SC) task.

The rest of the paper is organized as follows: Existing work related to aspect extraction and sentiment classification is presented as "Related Work" in Section 2. Section 3 expresses the overall methodology and detail of the proposed approach as the "Proposed Research Methodology". The experimental environment considered for the development of this model is described in Section 4 as "Experimental Arrangements". The performance comparison of this model is presented in Section 5 as "Results and Discussion". The last section, Section 6, concludes the whole approach with future work as "Conclusion".

2. Related Work

The extraction of sentiment from a given piece of text is known as SA. Estimating the feelings, thoughts, and attitudes of users towards groups, individuals, and products or brands is the leading initiative of SA. It can automatically recognize and extract aspects and their corresponding opinions and then classifies their polarities related to these opinions from online textual reviews [30]. In addition, document-level and sentence-level SA cannot

explain users' likes or dislikes relating to a specific feature of an entity. They only focus on the entire document or sentence-oriented sentiments, which cannot be beneficial in all daily life scenarios. Sometimes, users are interested in specific aspects of products or services (that are out of the scope of document and sentence level SA), which demands ABSA for handling such a scenario. It is a unique genre of text-mining and a fine-grained SA, which can extract aspects and their corresponding sentiment polarities from a sentence. Moreover, it can summarize these sentiments related to aspects of user-generated reviews, which is a general task of ABSA [31,32].

Additionally, diverse challenges regarding ABSA have still become the cause of their performance degradation, such as identification of those textual parts of context that depict identical aspects, determining the relationship between features and the text, and handling comparative sentences [33]. In addition, the task of ABSA is accomplished in two phases. The first phase specifies all the conceivable aspects (either implicit or explicit) related to a specific topic or product. As a result, all the potential features are known and available to nominate their corresponding polarities. Therefore, the task of the second phase is to determine the sentiment polarities and assign them to their interrelated aspects. Moreover, explicit aspect extraction/identification has been considered the main subtask of ABSA. It is related to the extraction of those aspects or features of a specific entity that are explicitly mentioned in a review and discussed among users upon which they express their opinions and comments [34].

In the early days of ABSA, traditional approaches regarding aspect identification/extraction have highly relied upon machine learning methods (e.g., Nearest Neighbor, Support Vector Machine, Naïve Bayes, Decision Tree), nouns frequency-based, lexicons (e.g., WordNet, SenticNet), Topic modeling (e.g., LDA), n-grams combination, and rule-based approaches. These feature engineering-based procedures depend on manual annotations, rule creations, and handcraft features, which are laborious, time-consuming, domain-dependent, and cause performance bottlenecks [35]. The remarkable achievements of DL methodologies in NLP inspired researchers to involve these methodologies while accomplishing the main tasks of SA. At present, DL methods are famous for ABSA tasks, but the inclusion of reasoning, such as the human brain, is still an open research area for future contributions of researchers [36,37]. The success of DL methodologies in the NLP task made them commendable and created space for applying these methods within the task of aspect extraction and classifying their corresponding sentiments [38].

According to the literature, these aspects are extracted based on handcraft features. These are laborious and complicated or time-consuming methods that demand much effort from analysts. This motivated Xu et al. [39] to present a supervised approach that identifies the potential features using the DL method. Therefore, their methodology uses CNN with two embedding layers. One of the embeddings has trained generally, whereas the other acquired training according to a specific domain for extracting aspects. Additionally, Shu et al. [40] proposed a modified form of CNN named controlled CNN (Ctrl). It consists of two control modules: the embedding control and the CNN control module. Asynchronous updating of CNN's parameters prevents it from over-fitting while boosting the model performance significantly. Furthermore, A. Da'u and N. Salim [41] presented a model based on a multiple-channel CNN that uses word embedding and PoS tags as textual features. The introduced model was presented along two input channels. Among those, the first channel takes the word embedding as input, while the other takes PoS tag sequential information as input for aspects identification.

RNN-based approaches provide state-of-the-art performance due to their dependence on long-term dependencies with temporal features that enhance their learning procedure of textual representation and sequential information. Due to this, Li et al. [42] proposed a framework comprised of two LSTMs that perform ATE based on the summary of previously identified opinions and aspects. At each step based on historical information, the historic attention truncates the unrequired terms from the recently predicted representations and identifies valuable features. On the other hand, Saraiva et al. [43] proposed an approach

known as POS-AttWD-BLSTM-CRF, which utilizes the attention mechanism as an encoder that determines the grammatical dependencies found among the targeted words. Instead of electing a subsection of PoS-tagged features, their approach selects the most relevant among them. These features collectively provided to the Bi-LSTM-CRF classifier that accomplishes the task of ABSA.

Exploration of the relevant literature represents that the combination of Bi-GRU and conditional random field (CRF) is the most widely utilized method to accomplish the most challenging tasks of ABSA, i.e., aspect terms identification/extraction and their sentiment classification. These models become trained through SemEval 2014 labeled dataset, which uses either pre-trained GloVe or word2vec embedding [44]. In addition, the literature depicts that supervised methods outperformed as compared to rule-based methods. In order to achieve such high performance, these methodologies have to pay in the form of a large volume of annotated or labeled samples to train their models, which is time-consuming and expensive. This situation motivated Wu et al. [45] to propose a hybrid unsupervised approach that performs the identification of aspects from the targeted context. Hence, they combined linguistic-based rules and GRU algorithms to classify the targeted terms as aspect and non-aspect. Moreover, the accurate aspect terms extraction and identification phenomena depend upon long-term dependencies of the textual sentence or noun phrases, which makes them inadequate in the case of usability (such as across-domain scenarios) and accuracy. With these motivations, Chauhan et al. [46] proposed a hybrid two-step unsupervised model, which integrates linguistic patterns and an attention-based Bi-LSTM to perform the task of ATE. The first step comprises the linguistic rules and is responsible for extracting potential aspects composed of single or multi-words. Domain correlation then evolves to filter these terms, which relate to a specific domain. These filtered terms transform into a fine-tuned word embedding. Furthermore, these determined aspects from the first step utilize as labelled data during the second step. Based upon these identified terms, the training of the attention-based Bi-LSTM model is accomplished.

The analysis of the relevant literature exhibits that the field of ABSA highly utilizes RNN models, which perform superlatively. However, some inferiorities are identified within them regarding position invariance and local pattern sensitivity. As a result, these inferiorities raised the question of their performance; to prevail over this scenario, CNN presents its services, but long-term dependencies and sequence information modeling again raise difficulties in their success. Consequently, recent research trends changed towards the models' fusion approach that enhances their ability of feature identification and extraction [47,48], which encouraged Zhu et al. [49] to present an aspect-level attention-based recurrent convolutional neural network (AARCNN) model that extracts aspect-based sentiment from user-generated reviews and comments. Their approach combines the targeted information with the attention mechanism that enables the model to concentrate on the exact targeted aspects. Therefore, their models' Bi-LSTM provides entire sentence representation to CNN and from which it extracts the highly attentive parts of a sentence as potential features along with their sentiments. Under this influence, Akhtar et al. [50] proposed an approach for extracting aspects and classifying sentiment polarity using bi-directional long short-term memory (Bi-LSTM) with the CNN network for transfer learning perspective. Thus, their approach's Bi-LSTM learns the sequential pattern for predicting aspect terms of the provided review sentences. Additionally, CNN has been used to acquire local features related to identified aspect terms for SC. Both algorithms are used jointly in a serial manner to enhance the prediction rate of both before-mentioned tasks.

According to the literature, existing approaches have considered model fusion mechanisms, but they widely combined these models sequentially. It means that one of the models has received the actual input while the other acquires the output of the first model, which becomes the cause of information loss. This situation motivates researchers to consider such approaches, where both algorithms fetched the same input simultaneously and then combined their learnings to exploit their combined abilities. With this motivation, Guo et al. [25] proposed a hybrid parallel approach named CRAN, which comprises a

CNN and Bi-GRU. This approach is based upon the attention mechanism, in which the main objective is to combine both CNN and GRU output. In addition, their approach highlights those terms that acquire the focus of the whole sentence's contextual information. The sequential information gathered from GRU visualizes these valuable features and extracts them in light of contextual information learned from CNN. It also preserves the semantic information illustrated in the targeted sentence with the collaborative learning of both merged algorithms. However, the model performs well but lacks the knowledge of contextual positional information and PoS tags. It also excludes the information sharing of hidden layers among combined algorithms that can enhance the identification ability of valuable attributes.

Moreover, the analysis of the relevant literature on ABSA also observed those contributions that integrate both strategies (i.e., parallel and sequential). In this way, Yu et al. [3] proposed an approach (named ABLGCNN) that utilizes two parallel RNN architectures (comprising either LSTM or GRU) and separately joins each of them serially with CNN. However, according to the literature, RNNs can capture global contextual information but cannot capture local features efficiently, which becomes the cause of information loss in their approach. However, the model performs well, but the serial combination of RNN and CNN makes their model complex and becomes the cause of losing beneficial information. In addition, the attention mechanism has been applied only to the output of RNN while ignoring the outcome of CNN. The integrated algorithms lack the information transfer of contextual positions of context and the information concerning hidden layers. These issues can enhance the performance of the model during the classification process. Another hybrid approach named CNN_BiLSTM [51] also performs SC that relies on the parallel combination of both CNN and BiLSTM. Their CNN model extracts local features, whereas BiLSTM captures global contextual information. These extracted features have been combined and passed towards the softmax function for accomplishing the classification procedure. Both CNN and BiLSTM take only word2vec embedding as input. Whereas CNN comprises three input channels, word2vec embedding is the only parameter provided as input. In addition, Zhang et al. [52] proposed a parallel approach that combines multi-attention CNN and Bi-GRU. Three types of inputs, including attention-oriented word vector, PoS, and position information, have been provided to the multi-attention CNN. Paradoxically their Bi-GRU phase receives only word embedding (without an attention mechanism) for acquiring the contextual semantic facts from the targeted context, which determines sentence-oriented sentiment polarities. Their approach delivers inputs contrastively among variant algorithms such as attention-oriented vectors regarding PoS, and position information is only provided to CNN and does not deliver to the Bi-GRU, which is the prime deficiency of their approach. However, the involvement of these neglected features can further enhance their approach identification performance and their ability to predict their corresponding sentiment. In another contribution, Cheng et al. [53] proposed one more parallel procedure, which accomplishes text sentiment analysis and is composed of attention-based MC-CNN and attention-based Bi-GRU. The algorithms (MC-CNN and Bi-GRU) used within this framework utilize only attention-based word2vec embedding as the input parameter. However, they ignore the consideration of contextual positional information, dependency-based relations and even the sharing of hidden layers' information while accomplishing the classification. The reference of these features can enhance the performance and accuracy of the prescribed framework.

The traditional approaches are lacking in considering the influence of interrelated contextual words and the relationship between aspect terms and contextual phrases based on their distance. Therefore, Huang et al. [54] proposed an approach named CPA-SA, which accomplishes the task of ABSA while considering aspect-specific contextual location information. Their designed function has adjusted the weight of contextual words according to the position of potential terms alleviating the terms' inference on both sides of conceivable terms to determine their corresponding polarities. Additionally, this approach excluded the influence of syntactic and semantic relations during the accomplishment of this task.

The past deep learning approaches massively used either pre-trained language models or attention mechanisms, which apply similar attention weights to the whole context without any restriction on assigning the attention. Hence Feng et al. [55] proposed an approach that utilizes the attention procedure with a masked mechanism that imposes a threshold for attention weights and keeps only those scores above that threshold while removing the lower score terms. This approach only focuses on word2vec-based knowledge, whereas the importance of contextual position information is out of its scope. Moreover, Liao et al. [56] proposed an approach named FAPN, which is a phrase-aware CNN-based fine-grained attention mechanism that captures the word-level relations between the aspect and their corresponding context. Their methodology focuses only on the local contextual information, whereas the global contextual information neglects during this procedure.

Recapitulating the above discussion, we conclude that whenever either CNN or RNN implements as individuals, they can never perform extraordinarily. Moreover, their sequential combination losses valuable information. In addition, within a real-world scenario, a model's simple design always performs well and proves beneficial compared to a complicated model. In the relevant literature, the MC-BiGRU model conceivably never combines with other algorithms in a parallel perspective, which motivates the Att-JM to merge both algorithms (MC-BiGRU and MC-CNN) in one model. It enhances the performance and keeps the model architecture simple. According to the relevant literature on ABSA and best of our knowledge, the Att-JM is the first technique that integrates Att-MC-BiGRU with Att-MC-CNN, which shares their hidden layers for transfer learning. These discriminations improve the task of aspect extraction and sentiment prediction from the textual reviews and emerge as the main novelty of this methodology. Moreover, the utilization of attention mechanisms and contextual position information becomes the cause of accurately identifying aspects and classifying their sentiments. Furthermore, the proposed approach adequately distributes the input parameters, such as positional information and word2vec embedding, between algorithms. Due to this, they exploited their entire abilities while learning valuable features and then combined these learnings during the identification and extraction of targeted aspects with their corresponding sentiments, which became the cause of the distinction of the Att-JM compared to existing approaches.

3. Proposed Research Methodology

The proposed methodology aims to identify aspects accurately and classify their corresponding sentiments. Therefore, the proposed approach depends on the knowledge of contextual position, syntactic, and semantics facts of textual reviews that provide valuable clues to determine their appropriate polarities. In the relevant literature on ABSA, existing approaches have combined CNN and RNN in a sequential and parallel manner to obtain their combined benefits. However, their diverse distribution of input features has again appeared as the leading disadvantage to these methods. In the case of the serial approach, the former algorithm acquires the actual input, whereas the latter fetches the output of the former, which becomes the cause of information loss. In addition, in the case of the parallel approach, both merged algorithms were never moderately treated. It denotes that the former algorithm's input comprises various features, whereas the latter obtains a subset of them as input. Due to this, one of the algorithms acquires a more suitable acquaintance of contextual knowledge when compared to the other, which reduces the benefits of a parallel or simultaneous learning-based approach. Therefore, the proposed method fairly distributes the novel features among the combined algorithms to obtain the maximum advantages of parallel learning.

The utilization of contextual position information has proven itself to be a prominent factor that enhances the performance of various NLP tasks, which motivates the proposed method to include positional information vectors during the model learning phase. As a result, the Att-JM utilizes contextual positional embedding as position information vectors with word embedding. Therefore, the Att-JM uses the Google News dataset pre-trained word embedding that comprises three hundred dimensional vectors. In addition, instead of

merging similar algorithms, the proposed methodology combines two different algorithms (Bi-GRU and MC-CNN) that share their hidden layers. The Att-JM utilizes the attention layer to determine the highlighted and emphasized terms that achieve the entire focus of the whole context and then passes them toward the pooling layer. The finalized features from textual reviews deliver to softmax for categorizing these targeted terms as an aspect or non-aspect. The Att-JM utilizes these identified aspects and determines their related sentiments. Thereon, the context of the targeted sentence and their identified aspect terms are combined and given to the softmax to predict the relevant polarities as positive, negative, and neutral. Consequently, the Att-JM evaluates through a standard benchmark of datasets comprised of SemEval and Twitter. The detailed framework of the proposed methodology is shown in Figure 1.

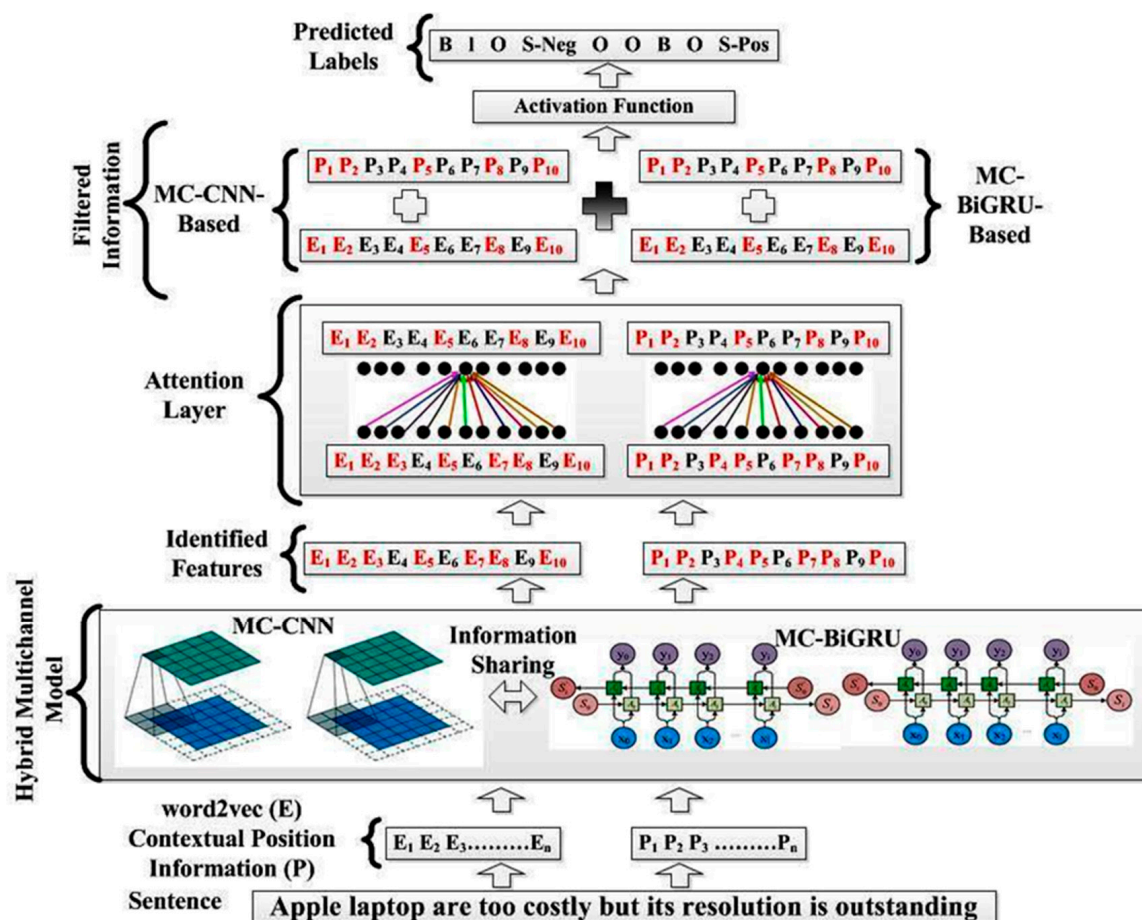


Figure 1. The framework of attention-based JM (MC-BiGRU and MC-CNN) for ATE and SC.

3.1. Proposed Model Description

The proposed method's objective is to determine aspects with the corresponding sentiment from the textual reviews. Wherefore the Att-JM combines MC-CNN and MC-Bi-GRU in a single framework where both algorithms comprise two channels for input. Therefore, the first obtains the word2vec embedding, whereas the other acquires the contextual position information vectors as input. Hence, the inclusion of an attention mechanism assists the proposed method during the accurate identification and extraction of these informative features and their relevant sentiments. In addition, the attention mechanisms utilize the precise location information from contextual position information during their recognition phase. Therefore, the precision of the Att-JM concerning the identification of potential terms is more authentic and reliable when compared with existing techniques.

The semantic information of word occurrence concerning a sentence is more significant to preserve their dependency knowledge. Thus, to accurately conserve this genre of facts, the Att-JM utilizes word2vec embeddings because these representations are well-known to achieve this objective. For this purpose, the Google News dataset pre-trained word2vec embedding is being used, which possesses three hundred dimensional vectors. These insights deliver towards the integrated algorithms that share their hidden layer's information to attain the benefits of their combined learnings. Consequently, the proposed methodology acquires the advantages of two different algorithms (MC-BiGRU and MC-CNN) and a parallel learning-based approach. According to the literature and to the best of our knowledge, the Att-JM is among the pioneers that perform the task of ATE and SC and merge different algorithms in a parallel manner that treats them equally in terms of input parameters. Table 1 depicts the utilized symbols in our procedure to elaborate it.

Table 1. List of abbreviations.

Symbol	Description	Symbol	Description
W	Weight matrix	f	Activation function
x	Input term	l	Length of sequence
b	Bias	j	Targeted word
p	Word position	u	Update vector
H	The combined output of Bi-GRUs	C	Convolutional vector
θ	Influence ratio	c	Length of sentence
i	Targeted sentence	Y	Resultant' representation
α	Attention score	S	Concatenation
CPI	A contextual position information vector	WE	Word embedding

3.1.1. Multichannel CNN

Therefore, MC-CNN exploits contextual position information vectors and word2vec embedding as input. The contextual position information depicts the effect of each targeted term within the context of a sentence through vector representation. These representations formulate with the help of mathematical notation, which has exhibited in the form of Equation (1) [57] given below:

$$CPI_i = \lim_{j \rightarrow n} \left(\theta_j + (1 - \theta_j) \left(\frac{p_j}{n} \right) \right), 0 \leq \theta \leq 1, \quad (1)$$

In Equation (1), CPI_i is the contextual position information vector of i th sentence of the dataset. In addition, θ_j is the term's locality-based presence ratio that expresses the range $[0, 1]$. It depicts the contributonal influence of the j th word presence concerning its position of occurrence within context. Additionally, p_j is the position of a word in the targeted sentence, while n expresses the length of the sentence. If the j th word lies in vocabulary, then Equation (1) calculates their corresponding position-oriented influence. Otherwise, zero adds in the case of the non-availability of the word. Hereafter, contextual positional information vectors deliver on the first channel of MC-CNN with their weight matrix W_{CPI} . Then features are extracted using Equation (2) [58] given below:

$$C_{CPI} = f(W_{CPI}x_{CPI} + b) \quad (2)$$

In the above equation, f is the activation function. Consequently, Leaky ReLU is utilized for this purpose, W_{CPI} indicates the weight of the targeting term x_{CPI} , and b is a bias term. The last term, C_{CPI} , denotes the identified features from the convolutional layer based on the contextual position information. In addition, as described previously, word2vec embedding uses to preserve the syntactic and semantic details of context. Therefore, these embeddings impart to the second channel of MC-CNN along with the weight matrix W_{WE} . Then features are extracted using Equation (3) given below:

$$C_{WE} = f(W_{WE}x_{WE} + b) \quad (3)$$

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn. The above Equation (3) comprises the same parameters as observed in Equation (2). The only difference is that these parameters in Equation (3) depict the information relates to word2vec embedding. Afterward, both C_{CPI} based and C_{WE} based features are combined using the mathematical notation expressed in Equation (4).

$$C = (C_{CPI}; C_{WE}) \quad (4)$$

After utilizing these convolution layers, a max-pooling layer deploys for feature mapping, and their maximum values use for corresponding feature identification. The reason is that the highest value depicts the prominent terms of interest.

3.1.2. Multichannel BiGRU

Moreover, the second algorithm's, i.e., MC-BiGRU, first channel acquires each targeted sentence contextual position information vector with the help of Equation (1) and their corresponding weight matrix W_{CPI} . Features are then extracted using Equations (5)–(8) to (9)–(12) [59,60] for both the forward and backward directions, which are given below:

$$Z_{CPI} = f(W_{CPI}x_t + W_{z-1}h_{CPI-1} + b_z) \quad (5)$$

$$R_{CPI} = f(W_{CPI}x_t + W_{r-1}h_{CPI-1} + b_r) \quad (6)$$

$$\hat{h}_{CPI} = f(W_{CPI}x_t + W_{CPI-1}(R_{CPI} \odot h_{CPI-1}) + b_h) \quad (7)$$

$$\vec{h}_{CPI} = (1 - Z_{CPI}) \odot \vec{h}_{CPI-1} + Z_{CPI} \odot \hat{h}_{CPI} \quad (8)$$

$$Z_{CPI} = f(W_{CPI}x_t + W_{z-1}h_{CPI-1} + b_z) \quad (9)$$

$$R_{CPI} = f(W_{CPI}x_t + W_{r-1}h_{CPI-1} + b_r) \quad (10)$$

$$\hat{h}_{CPI} = f(W_{CPI}x_t + W_{CPI-1}(R_{CPI} \odot h_{CPI-1}) + b_h) \quad (11)$$

$$\overleftarrow{h}_{CPI} = (1 - Z_{CPI}) \odot \overleftarrow{h}_{CPI-1} + Z_{CPI} \odot \hat{h}_{CPI} \quad (12)$$

In the above Equations (5)–(8), \vec{h}_{CPI} represents the forward direction state's output vector about the contextual position information. Moreover, Z_{CPI} denotes the update vector of contextual position information, \vec{h}_{CPI-1} depicts the contextual position information's previous state vector, and \hat{h}_{CPI} shows the activation vector. In addition, the symbol \odot denotes the Hadamard product. However, the symbol f depicts the activation function. Therefore, Leaky ReLU is applied as an activation function in Equations (9) and (10). Additionally, Equation (11) uses the Tanh as an activation function. Furthermore, Equations (9)–(12) comprise the same elements described in Equations (5)–(8). However, they demonstrate the detail regarding the backward direction state. These (directions) states are then combined and make a single state H_{CPI} , as shown in Equation (13) given below:

$$H_{CPI} = \left(\vec{h}_{CPI}; \overleftarrow{h}_{CPI} \right) \quad (13)$$

On the other hand, the MC-BiGRU's second channel allocates for acceptance of the sentence's syntactic and semantic details as word2vec embeddings along with the weight

matrix W_{WE} as input. To analyze these details in both forward and backward directions, Equations (14)–(17) to (18)–(21) were then applied. These equations are equivalent to Equations (5)–(8) to (9)–(12) in terms of their comprised elements and functionality, which are given below:

$$Z_{WE} = f(W_{WE}x_t + W_{z-1}h_{WE-1} + b_z) \quad (14)$$

$$R_{WE} = f(W_{WE}x_t + W_{r-1}h_{WE-1} + b_r) \quad (15)$$

$$\hat{h}_{WE} = f(W_{WE}x_t + W_{WE-1}(R_{WE} \odot h_{WE-1}) + b_h) \quad (16)$$

$$\vec{h}_{WE} = (1 - Z_{WE}) \odot \vec{h}_{WE-1} + Z_{WE} \odot \hat{h}_{WE} \quad (17)$$

$$Z_{WE} = f(W_{WE}x_t + W_{z-1}h_{WE-1} + b_z) \quad (18)$$

$$R_{WE} = f(W_{WE}x_t + W_{r-1}h_{WE-1} + b_r) \quad (19)$$

$$\hat{h}_{WE} = f(W_{WE}x_t + W_{WE-1}(R_{WE} \odot h_{WE-1}) + b_h) \quad (20)$$

$$\leftarrow h_{WE} = (1 - Z_{WE}) \odot \leftarrow h_{WE-1} + Z_{WE} \odot \hat{h}_{WE} \quad (21)$$

The above Equations (14)–(21) identify the potential terms using word2vec embedding in both the backward and forward directions in light of syntactic and semantic information of the targeted context. The identified conceivable terms during the analysis of the bidirectional states are combined to make a single state H_{WE} , as shown in Equation (22) given below:

$$H_{WE} = \left(\vec{h}_{WE}; \leftarrow h_{WE} \right) \quad (22)$$

Eventually, the features indicated from both BiGRU layers, based on contextual position information as H_{CPI} and based on word2vec embedding as H_{WE} , are merged as H . It produces a combined form of feature representations, as depicted in Equation (23).

$$H = (H_{CPI}; H_{WE}) \quad (23)$$

After utilizing the BiGRU layers and acquiring their merged features representation as H , a max-pooling layer deploys for feature mapping because max values depict the prominent terms of interest from identified features. These pooling layers comprise the information of the hidden layers. Therefore, the pooling layers of both BiGRU and CNN are interchanged between them to achieve the benefit of their combined learning abilities.

3.1.3. Attention Layer

According to the literature, different words have different influences within the context of the targeted sentence; thus, the Att-JM considers an attention mechanism to identify the most effectual terms of the targeted sentence. Therefore, the contextual position information reinforces attention mechanisms to determine the accurate position of those potential terms. Generally, the attention mechanism consists of two modules, known as the encoder and decoder. The responsibility of the encoder is to transform the targeted sentence into a real-valued equal-length vector, which provides the semantic information related to each term. Moreover, the responsibility of the decoder module is to provide output after the

transformation of an encoded vector. Core equations regarding the attention mechanism are expressed in (24)–(26) [61] given below:

$$u_p = \tanh(W_p H + b_p) \quad (24)$$

$$\alpha_p = \frac{\exp(u_p^T u_w)}{\sum_p \exp(u_p^T u_w)} \quad (25)$$

$$H_T = \sum_p \alpha_p H \quad (26)$$

In the above equations, u_w expresses the updated context vector. Moreover, u_p represents the result of the hidden layer's vector H , and b_p and W_p are the bias and weight matrix of the attention mechanism, respectively. The last term, α_p , expresses the attention score regarding the single word of the sentence. The proposed methodology applies attention mechanisms separately for both algorithms (MC-BiGRU and MC-CNN). These models learn regarding potential terms in the light of word embedding with contextual position information, and then the attention mechanism filters out the most prominent features from each sentence.

$$A_{MC-BiGRU} = \text{Attention}(H_T) \quad (27)$$

$$A_{MC-CNN} = \text{Attention}(C_T) \quad (28)$$

3.1.4. Attention Layer

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn. The last step performs the concatenation of resultant output layers as Y , which comprises the information about the potential terms along with their contextual position and syntactic and semantic information of the targeting context. These outcomes have forwarded toward the softmax activation function for prediction and classification purposes. Thus all the targeted terms found in the targeted sentence categorizes as aspect or non-aspect terms. The detailed architecture of the proposed method is depicted in Figure 2.

$$Y = (A_{BiGRU} + A_{MC-CNN}) \quad (29)$$

$$\text{Predicted aspects} = \text{Softmax}(W.Y + b) \quad (30)$$

Manipulating Equation (30) emphasizes the predicted aspects from the targeting sentences and categorizes them as aspect or non-aspect terms. Therefore, the identified aspect terms are admitted and neglect the non-aspect terms. Afterward, ATT-JM determines the remaining terms' polarity as positive, negative, and neutral. Therefore, these aspects are combined with their context to determine their corresponding sentiments. This process depicts in Equation (31). Thus, the softmax activation function acquires these predicted filtered terms integrated with the outcome of Equation (29) that classifies their sentiment as positive, negative, and neutral. This procedure is observable from the mathematical notation of Equation (32).

$$S = (\text{Predicted aspects}, Y) \quad (31)$$

$$\text{Predicted sentiments} = \text{Softmax}(S) \quad (32)$$

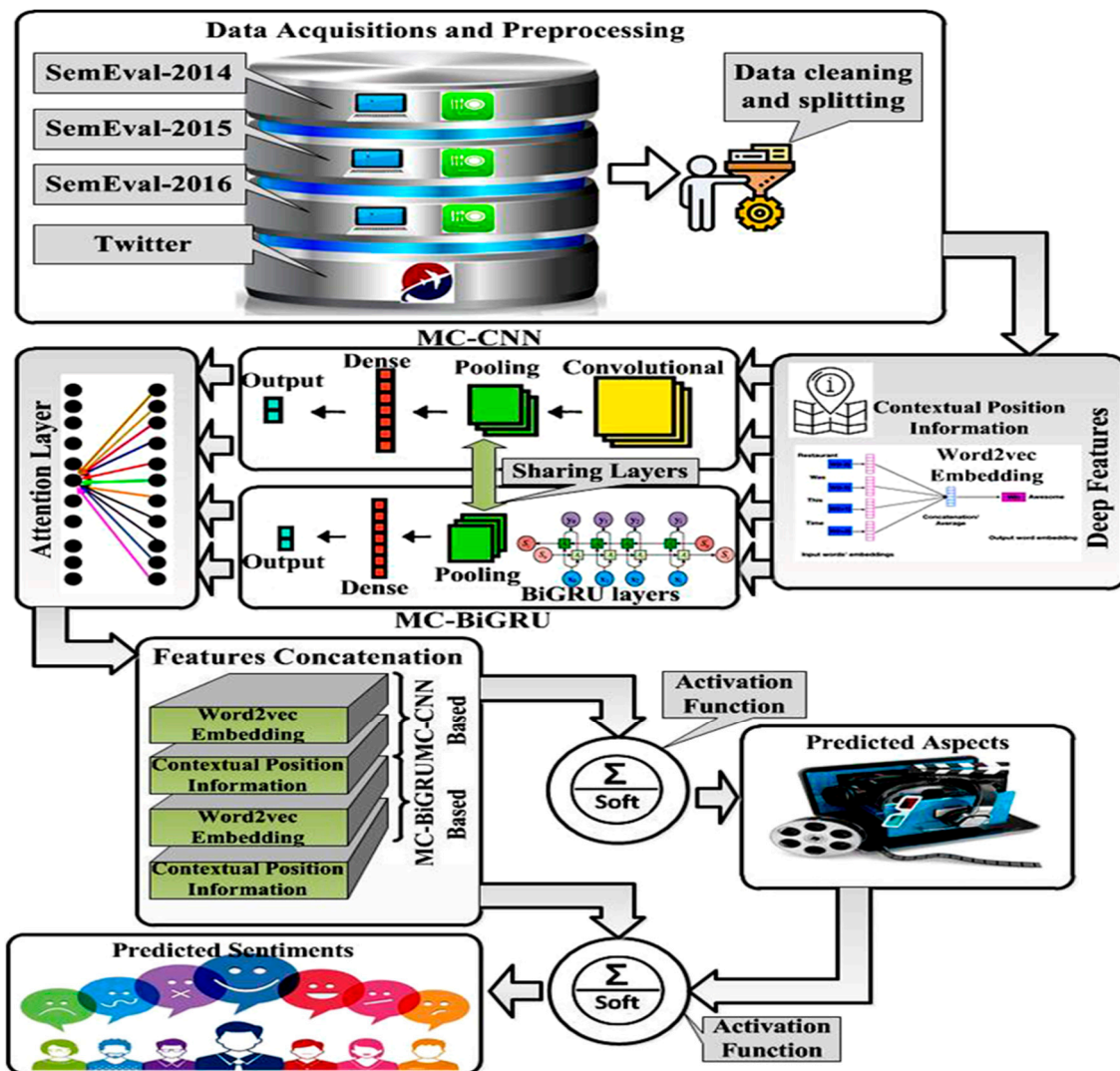


Figure 2. Attention-based (MC-BiGRU and MC-CNN) for aspect extraction and sentiment classification.

4. Experimental Arrangements

In this section, this study first illustrates the environmental arrangements for conducting these experiments. The datasets used during these experiments to evaluate that approach are then represented. Afterward, the preprocessing steps carried out for filtering these datasets are explained and finally express those baselines used for comparison to that approach. This study utilizes precision, recall, and F1 measure score for evaluating that approach because these are the most widely used metrics in the related literature on ABSA.

4.1. Environment for Experiments

The organized environment for conducting experiments regarding the Att-JM and evaluating their performance was based on the Windows 07 operating system. The hardware comprised a CPU Intel Xeon W3530 2.8 GHz, a GPU GeForce GTX 1060, and DDR3 RAM of size 16 GB. In addition, the implementation software included Python-oriented GPU Tensorflow 2.0, Keras 2.1.0, and the tool utilized is PyCharm IDE for Python 3.7.

4.2. Dataset Selection

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn. The performance of the proposed model has been extensively evaluated based on seven standard datasets. These comprise the test and training datasets and are made publically available by the SemEval organizers. The first two datasets belong to SemEval2014 [35,45,62], which contains the reviews regarding restaurant and laptop domains, respectively. The subsequent two datasets also belong to the fields of restaurant and laptop while belonging to the semeval2015 [35,45,62] dataset.

The fifth and sixth also belong to the restaurant and laptop domain, which relate to SemEval2016 [35,45,62]. The last one is the Airline Twitter Sentiment dataset, which concerns U.S leading airlines. This Twitter data comprises positive, negative, and neutral tweets [63,64]. The dataset developed by CrowdFlower that publically available since 2020 and updated in 2021 and 2022. The experimental study prefers this dataset to analyze the performance of their proposed method through the leading social media platform Twitter because this emerges as a primary source to express public opinions. Table 2 shows the statistics of these underlying datasets.

Table 2. Statistics of the datasets.

Datasets	Train					Test				
	Sentences	Aspects	Pos	Neg	Neu	Sentences	Aspects	Pos	Neg	Neu
SemEval-14L	3041	2358	1007	886	465	800	654	348	138	168
SemEval-14R	3045	3693	2195	835	663	800	1134	742	196	196
SemEval-15L	1739	1697	819	749	129	761	949	527	306	116
SemEval-15R	1315	1192	536	503	153	685	678	354	246	78
SemEval-16L	2500	2357	1150	749	458	808	914	661	209	44
SemEval-16R	2000	1743	929	739	75	676	622	311	257	54
Twitter Data-US Airline	2243	2389	1239	1049	101	764	923	611	264	48

4.3. Preprocessing

The proposed method performed preprocessing upon the textual reviews to obtain a structured and clean textual dataset. This preprocessing is defined as given below:

- A lowercase conversion occurs regarding all words of sentences that belong to the English language.
- The textual reviews' paragraphs are split into sentences from the symbol of full-stop.
- All sentences are split into tokens from the white spaces.
- All punctuation words are terminated from the sentences of textual content.
- All the words in sentences that comprise an impure form of alphabetical (alphanumeric) characters are removed.
- All the words in sentences that incorporate stop words are removed.
- All the words in sentences that contain special characters are removed.
- All the words whose length is less than or equal to one character are terminated.

4.4. Baselines

The proposed methodology's performance compares with the performance of the following baseline models. The details regarding the baseline models have represented in a tabular notation in the form of Table 3, given below:

Table 3. Summary of the baseline methods considered for the proposed approach.

Baseline Models/Reference	Novelty	ATE/SC	Year	Dataset
Bi-GRU + WE + POS [45]	Considering linguistic rules and domain correlation to filter irrelevant terms, then train the GRU network for ATE and OTE.	ATE	2018	SemEval-14L, 14R, SemEval-15Hotel, 15R, and SemEval-16R
DE + CNN [39]	Both pre-trained domain-specific and general-purpose word embeddings utilize in the CNN model for aspect extraction.	ATE	2018	SemEval-14L and SemEval-16R
SRNN + GRU [62]	The joint model extracts the aspects for a specific domain of interest.	ATE	2019	SemEval-14L, 14R, and SemEval-15R
MCNN + WV + POS [41]	Using general embedding. Encoding contextual Information using word and POS Tag embedding for aspect extraction	ATE	2019	SemEval-14L, 14R, SemEval-15R, and SemEval-16R
MCNN + WV2 + POS [41]	Using domain-specific embedding. Encoding contextual Information using word and POS Tag embedding for aspect extraction	ATE	2019	SemEval-14L, 14R, SemEval-15R, and SemEval-16R
POS-AttWD-BLSTM-CRF [43]	The model considers grammatical relations and part-of-speech tag features, along with Bi-LSTM and attention mechanism for the extraction of aspect terms.	ATE	2020	SemEval-14L, 14R, and SemEval-16R
FAPN [56]	The FAPN is a phrase-aware CNN-based fine-grained attention mechanism that captures the word-level relations between the aspect and their corresponding context.	SC	2021	SemEval-14L, 14R, SemEval-15R, and SemEval-16R
HPNet-M [65]	This model proposes a hierarchical framework that uses the pre-trained languages hierarchical models to infer the potential terms.	SC	2021	SemEval-14L, 14R, SemEval-15R, and SemEval-16R
JTSG [66]	The JTSG utilizes a generative model comprising an encoder to encode the sentences and a decoder that indicates the start and end position of potential aspects.	SC	2021	SemEval-14L, 14R, SemEval-15R, and SemEval-16R
CPA-SAA [54]	Adjusts the weight of contextual words according to the position of potential terms and alleviates the inference of terms among both sides of conceivable terms to determine their corresponding polarities.	SC	2022	SemEval-14L, 14R, SemEval-15R, and SemEval-16R
GCNs [67]	The GCNs utilizes a graph-based CNN to infer both the general and structural semantics of words.	SC	2022	SemEval-14L, 14R, SemEval-15R, and SemEval-16R

5. Results and Discussion

In the empirical analysis, we noticed during the experiments that accurate performance regarding extraction/identification of aspects and their related sentiments highly relies upon word embedding. Moreover, the utilization of contextual position information endorses such genres of achievements. In addition, pre-training is another well-known significant factor concerns to the performance of word embedding. However, the best pre-training achieves when word embedding trains through a large volume of datasets. Due to this, the Att-JM utilizes the Google News dataset-based pre-train word2vec embeddings for performing tasks of ABSA. On the other side, Att-JM's performance evaluates through the English-based standard datasets comprising SemEval 2014, 2015, and 2016, and Twitter datasets, which belong to the laptop, restaurant, and US-Airlines domains, respectively. During the training phase, the proposed framework determines the length of the maximum sentence of the textual review, which is assumed to be the highest attainable length of a

single sentence within these reviews. All the sentences that conceived a size less than the maximum length were padded with zeros. The Att-JM considers those procedures for accurately identifying aspects and their sentiments. The detailed parameter setting of the proposed model is depicted in Table 4.

Table 4. Proposed approach parametric setting.

Proposed Model Parameters	Value
Dimensions of Word Embedding Vector	300
Size of Convolutional Layer Kernels	(6, 8)
Size of Bi-GRU Positional Vector Layer	128
Size of Bi-GRU Word2Vec Layer	128
Activation Function	Leaky ReLU
Batch Size	50
Number of Epochs	50
Optimizer	Adam
Learning Rate of Optimizer	0.001
Loss	Categorical Crossentropy
Prediction Function	Softmax
Dropout Rate	0.5

The proposed approach deeply analyzes the experimental arrangements of all the baseline studies. During the analysis, the proposed method determines which kind of parameter setting becomes the actual cause of performance increase and decrease. Depending upon such parametric analysis and utilizing the parameter optimization approach, the proposed method conceives the best parametric settings. These parameter settings are described in Table 4. Afterward, Table 5 represents the Att-JM's comparison with baseline approaches based on the F1 measure regarding ATE, whereas their SC-based performance demonstrates in Table 6. These tabular notations describe Att-JM's remarkable achievements when assessing its abilities through standardized datasets against the acquisitions of baseline approaches, which signifies its quantitative improvement in all domains of interest.

Table 5. Performance comparison of proposed approach based upon F1 measure score to baseline approaches regarding ATE.

Model	SemEval-14L	SemEval-14R	SemEval-16R
POS-AttWD-BLSTM-CRF	81.47	87.86	73.04
Bi-GRU + WE + POS	68.81	72.8	76.15
DE + CNN	81.59	73.62	74.37
SRNN + GRU	73.62	75.5	71.7
MCNN + WV + POS	79.84	84.69	72.62
MCNN + WV2 + POS	80.63	86.89	75.71
Att-JM (Proposed Model)	95	91	94

Table 6. Performance comparison of proposed approach based upon F1 measure score to baseline approaches regarding the SC.

Model	SemEval-14L	SemEval-14R	SemEval-15R	SemEval-16R
FAPN	72.91	73.07	63.07	73.03
GCNs	74.63	77.35	66.39	75.43
JTSG	64.28	74.76	74.66	74.56
HPNet-S	68.94	78.87	78.86	78.85
HPNet-M	69.11	78.95	78.94	78.93
CPA-SAF	70.87	72.81	60.26	71.47
CPA-SAA	71.5	73.38	60.15	72.43
Att-JM (Proposed Model)	89	88	90	92

5.1. Aspect Term Extraction Performance Comparison

The Att-JM acquires a significant performance improvement in the SemEval-14L domain dataset in the F1 measure score. It achieves an appropriate margin success from all baselined approaches in F1 measure performance. For example, it gained 14.37% from MCNN + WV2 + POS, 15.16% from MCNN + WV + POS, 13.41% from DE-CNN, 21.38% from SRNN-GRU, 26.19% from BiGRU-WE-POS, and 13.53% from POS-AttWD-BLSTM-CRF. This performance gain can observe graphically in Figure 3.

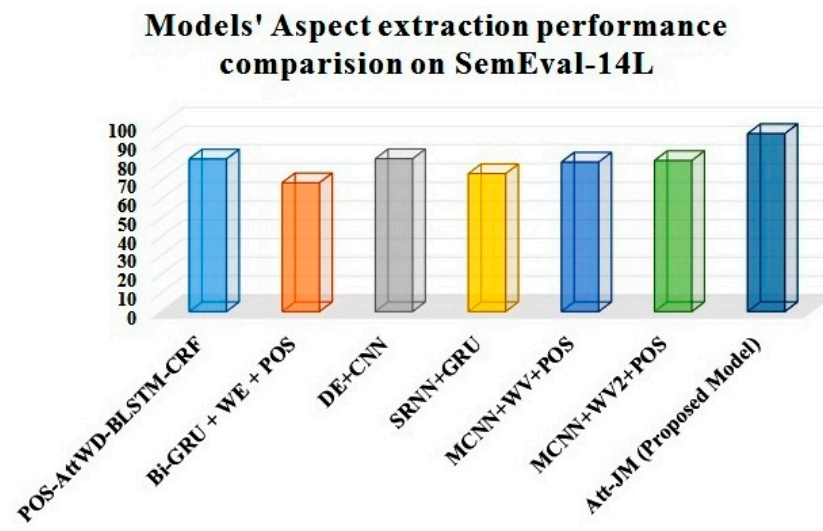


Figure 3. Model comparison with baseline approaches upon SemEval-14L domain.

The Att-JM abilities have not only been evaluated depending on a single domain. Therefore, in assessing the capabilities of the Att-JM, the SemEval-14R is also used. The collaborative consideration of different algorithms and sharing their learning information is the epicenter discrimination of this approach that provides a remarkable performance enhancement in terms of the F1-measure score compared to existing techniques, even when the domain varies. Therefore, it achieves 4.11% from MCNN + WV2 + POS, 6.31% from MCNN + WV + POS, 15.13% from SRNN-GRU, 17.38% from DE + CNN, 18.2% from BiGRU-WE-POS, and 3.14% from POS-AttWD-BLSTM-CRF. This performance improvement can observe in the form of a graphical representation in Figure 4.

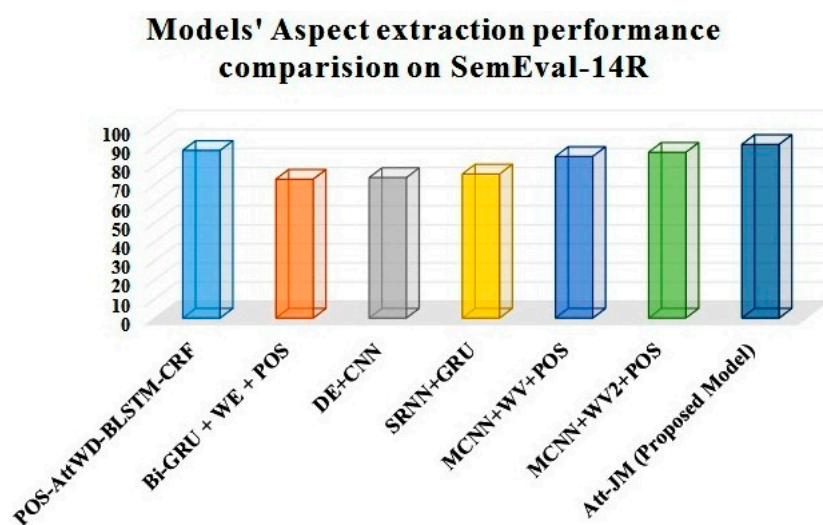


Figure 4. Model comparison with baseline approaches upon SemEval-14R domain.

These experiments also utilize the SemEval-16R domain during the Att-JM's performance evaluation. Consequently, a boost in performance concerns the proposed approach noticing in terms of the F1 measure score. Hence based on these improvements, the experiments analyze the outperforms of the proposed technique compared to baseline models. Therefore, it gains 18.29% improvement from MCNN + WV2 + POS, 21.38% improvement from MCNN + WV + POS, 19.63% improvement from DE-CNN, 22.3% improvement from SRNN-GRU, 17.85% improvement from BiGRU-WE-POS, and 20.96% improvement from POS-AttWD-BLSTM-CRF in the performance. This performance refinement can observe in the form of a graphical representation in Figure 5.

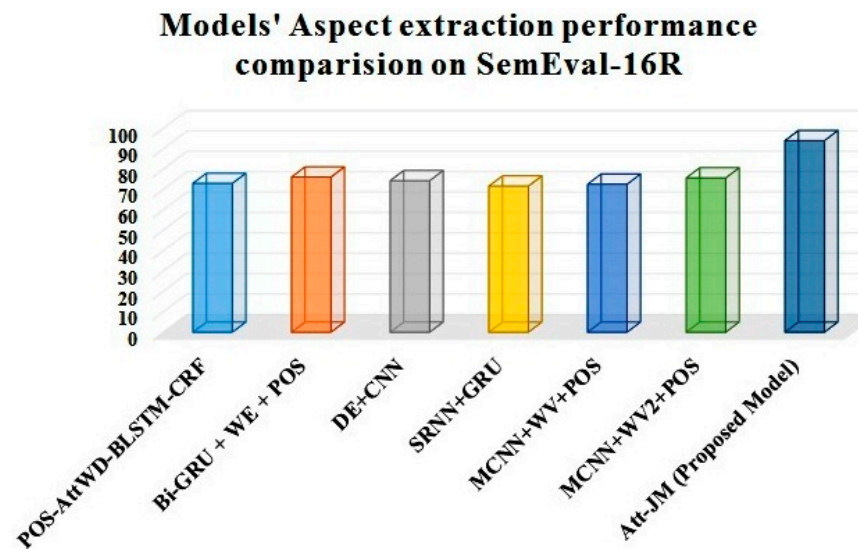


Figure 5. Model comparison with baseline approaches upon SemEval-16R domain.

5.2. Sentiment Classification Performance Comparison

The Att-JM's acquires a significant performance improvement during the classification process of identified aspects in the SemEval-14L domain dataset, which is 89% in the F1 measure score. It achieves an appropriate margin success from all baselined approaches in F1 measure performance. For example, their gained performance enhancement is 17.5%, 18.13%, 19.89%, 20.06%, 24.72%, 14.37%, and 16.09% compared to CPA-SAA, CPA-SAF, HPNet-M, HPNet-S, JTSG, GCNs, and FAPN, respectively. These performance improvements are depicted graphically in Figure 6.

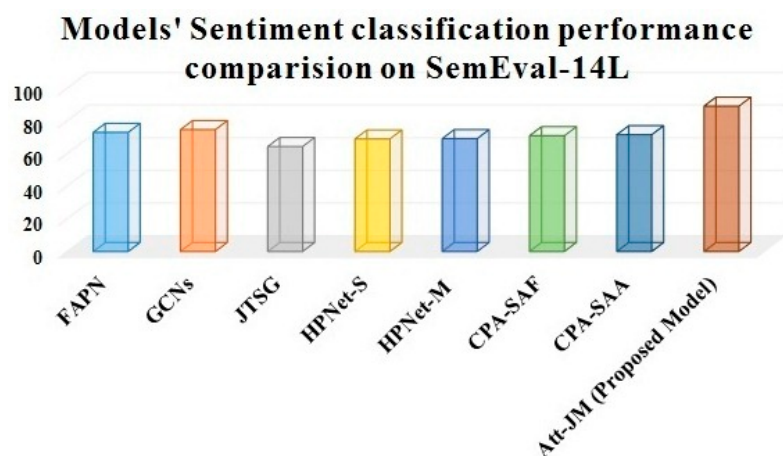


Figure 6. Model comparison with baseline approaches on SemEval-14L's domain.

Moreover, while classifying the identified aspects in the SemEval-14R domain dataset, the Att-JM obtained an enhancement, which is 88%. Therefore, its achievement improve-

ment is 14.62%, 15.19%, 9.05%, 9.13%, 13.24%, 10.65%, and 14.93% compared to CPA-SAA, CPA-SAF, HPNet-M, HPNet-S, JTSG, GCNs, and FAPN, respectively. Figure 7 depicts these performance enhancements against the baseline approaches in a pictorial configuration.

Models' Sentiment classification performance comparison on SemEval-14R

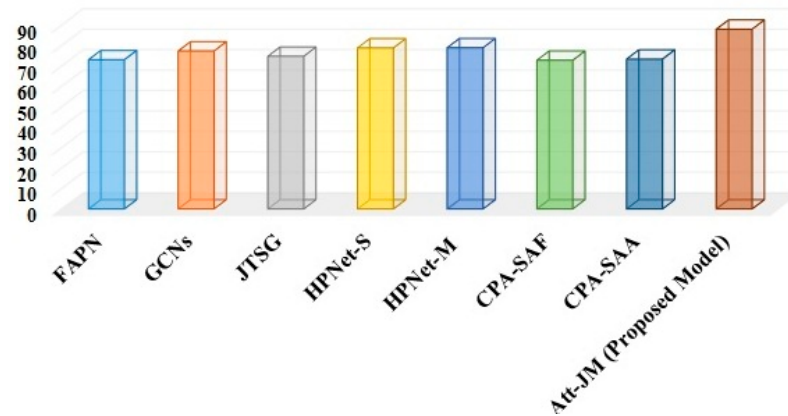


Figure 7. Model compression with baseline approaches on SemEval-14R's domain.

In addition, in the domain of the SemEval-15R dataset, the Att-JM obtains a performance improvement from the prescribed baselines, which is 90%. Consequently, the performance enhancement attains during the evaluation from baselines CPA-SAA, CPA-SAF, HPNet-M, HPNet-S, JTSG, GCNs, and FAPN is 29.85, 29.74, 11.06, 11.14, 15.34, 23.61, and 26.93, respectively. The proposed approach performs more reasonably than all the considered baselines, which is shown graphically in Figure 8.

Models' Sentiment classification performance comparison on SemEval-15R

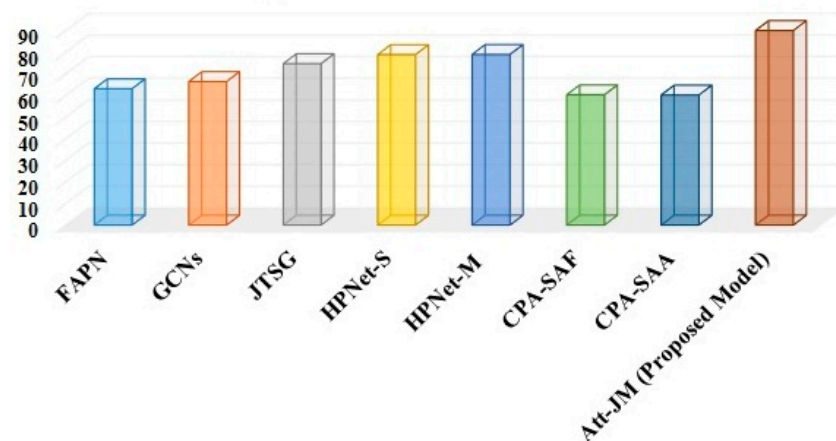


Figure 8. Model compression with baseline approaches on SemEval-15R's domain.

Furthermore, in the domain of the SemEval-16R dataset, the Att-JM achieved an enhancement, which is 92%. Thus it achieves improvements of 19.57%, 20.53%, 13.07%, 13.15%, 17.44, 16.57%, and 18.97% compared to CPA-SAA, CPA-SAF, HPNet-M, HPNet-S, JTSG, and GCNs, respectively. The collective consideration of contextual position information and syntactic and semantic information enhances the precision of the attention mechanism during the identification of potential terms. Moreover, the information sharing of hidden layers between the combined algorithms improves the prediction ability of the proposed approach. This performance enhancement depicts in Figure 9.

Models' Sentiment classification performance comparison on SemEval-16R

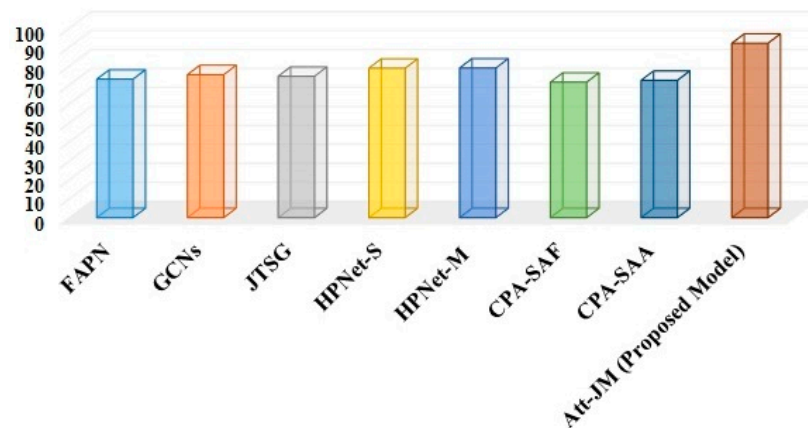


Figure 9. Model compression with baseline approaches on SemEval-16R's domain.

5.3. Proposed Approach Performance and Other Metrics

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn. The estimation of the performance enhancement regarding Att-JM comprises variant metrics and different datasets' domains. Therefore, the Att-JM's performance assesses through precision metrics using the previously discussed various datasets that express its significant performance improvement while conducting experiments. Moreover, the maximum precision improvement of the Att-JM ATE task observes in two datasets, i.e., SemEval-14R's domain and SemEval-15L's domain, which is 97% in both fields. On the other hand, the Att-JM's minimum precision observes in SemEval-15R's domain dataset is 89%. Furthermore, the SC's task obtained the maximum precision value on the Twitter Data-US Airline domain, which is 91%. Additionally, its minimum precision value observes that the SemEval-15L and SemEval-16L domain dataset is 82%. During experiments, the Att-JM's precision-based performance on all other datasets represents in Figure 10.

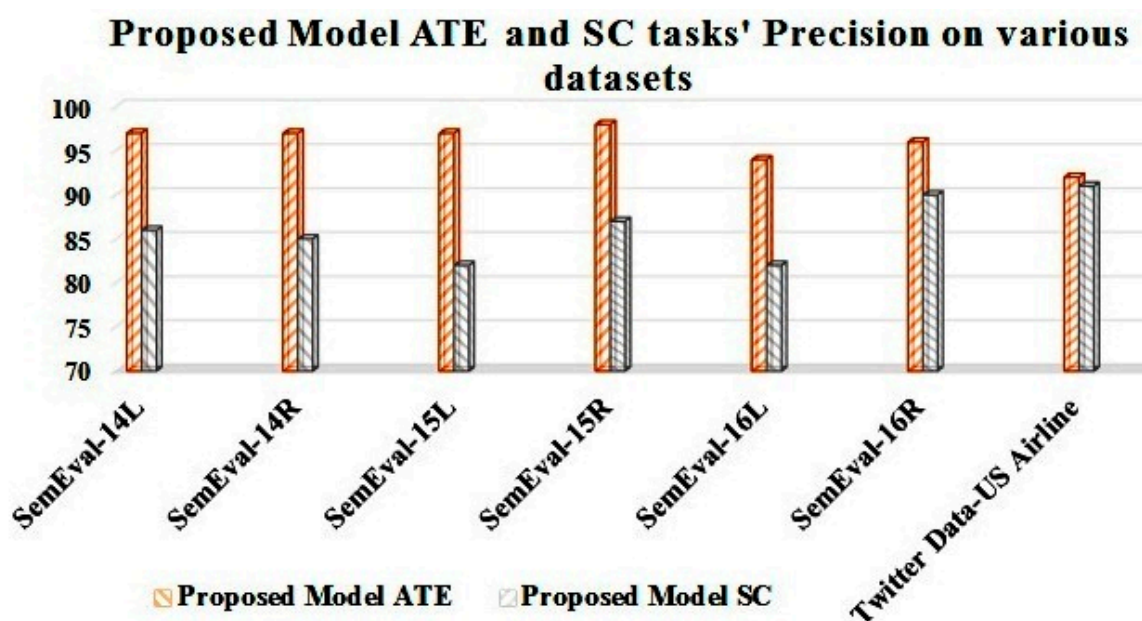


Figure 10. Model precision-based performance on variant datasets.

In addition, Att-JM's performance analysis through the recall metrics upon previously described datasets also depicts its significant performance improvement. Therefore, the ATE task's maximum recall value from SemEval-15R's domain dataset is 98%. Furthermore, the minimum improvement related to recall metrics found from the SemEval-15L's domain dataset is 82%. On the other hand, SC's task's maximum recall is 96% on SemEval-16R's domain dataset. Moreover, the minimum recall value observes on SemEval-15L and SemEval-16L's domain dataset is 90%. All other datasets' recall-based observations regarding the Att-JM experiments can analyze in Figure 11.

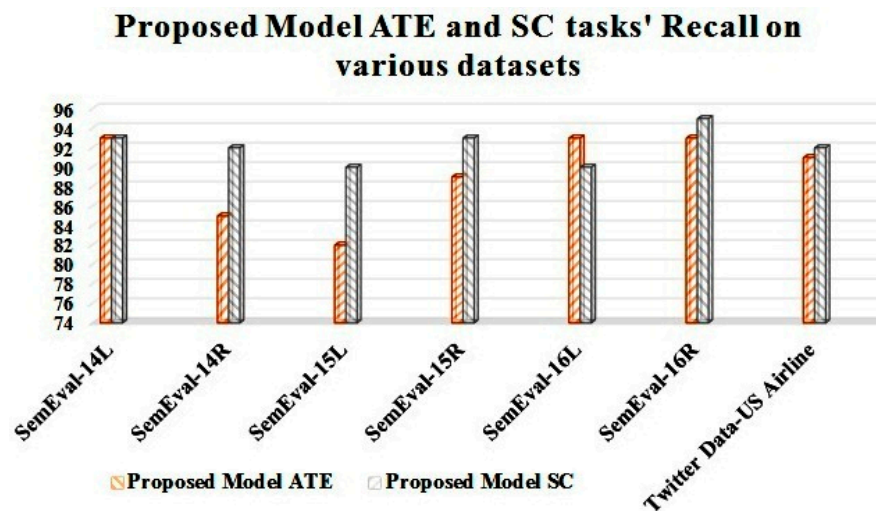


Figure 11. Model recall-based performance on variant datasets.

In this way, the Att-JM's performance improvement analyzes variant datasets using the F1 measure metric. Thus, the ATE task's maximum F1 measure score observed on the SemEval-14L's domain dataset is 95%. On the other hand, the minimum F1 measure score analysis on SemEval-15L's domain dataset is 89%. In addition, SC's task's maximum F1 score is 92% on SemEval-16R's domain dataset, whereas its minimum F1 score found on the SemEval-15L and SemEval-16L domain datasets is 86%. Therefore, during experiments, all other datasets' F1 measure scores observed regarding Att-JM have depicted in Figure 12.

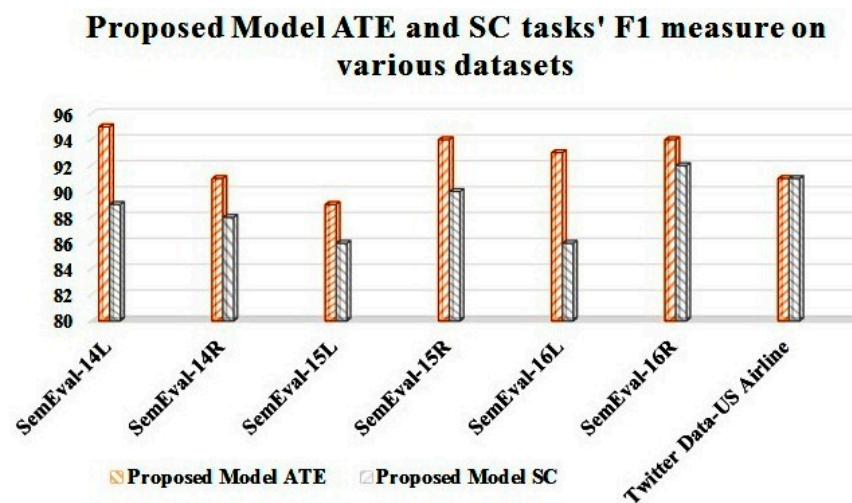


Figure 12. Model F1-measure score-based performance on variant datasets.

A frequently used visual tool known as the precision–recall (PR) curve evaluates the performance of proposed approaches regarding their discrimination capabilities between two classes. Therefore, the Att-JM uses the PR curve to assess its ability to discriminate among different categories. The area under the curve demonstrates the significant performance of the Att-JM while classifying multi-class classification, which depicts in Figure 13.

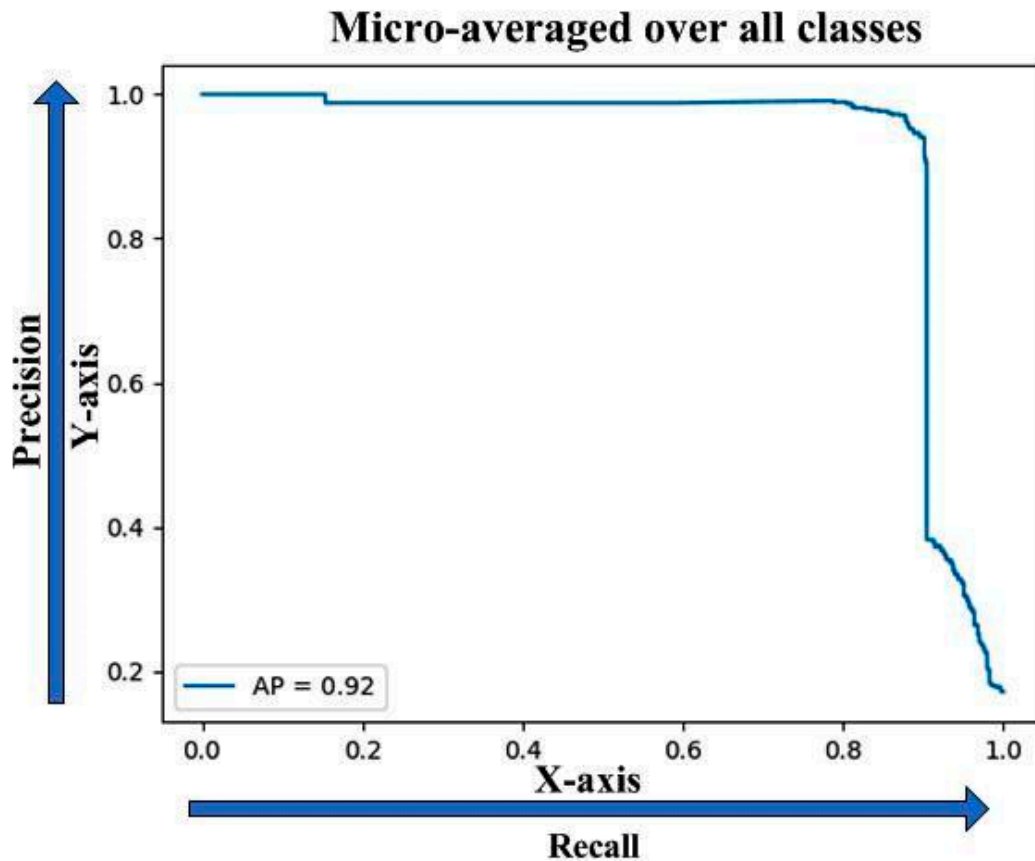


Figure 13. Model's precision–recall curve.

5.4. Discussions

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn. The Att-JM depicts a unique methodology that identifies aspects instead of acquiring them explicitly as input. Based on these identified aspects proposed approach predicts their sentiment from the corresponding context. Consequently, Figure 14 summarizes the overall performance of the Att-JM concerning different domains of standardized datasets during aspect extraction and SC with variant metrics perspectives such as precision, recall, and F1 measures. The framework significantly outperforms all datasets' fields compared to baseline approaches. This pictorial representation expresses that the proposed approach obtains a better precision value while accomplishing the task of ATE, whereas the SC task performs better in the recall. The F1 measure depicts a mixed trend of achievement. The main reason behind better performance regarding this approach is the consideration of two different algorithms and sharing of their hidden layers between them. In addition, the utilization of contextual positional information vectors and word2vec embedding assists the attention mechanisms during the filtration of targeted aspects along with their sentiments with precise accuracy. These quantitative achievements on various domains' datasets express the significance of deep features and algorithms' combined utilization.

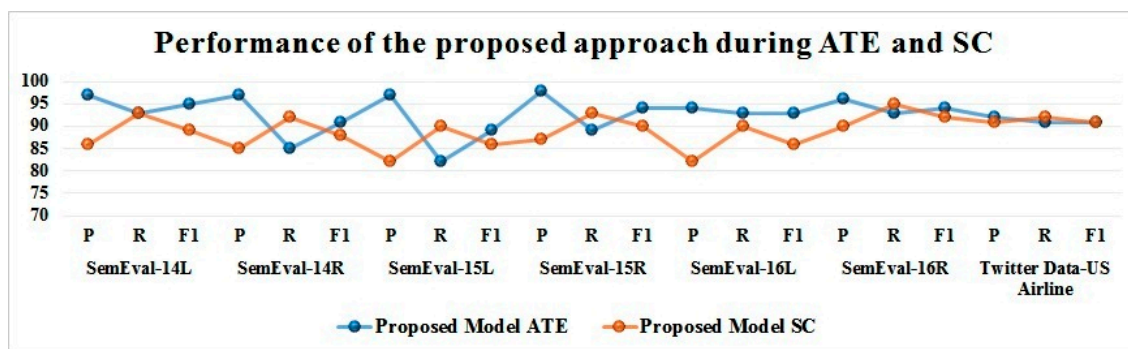


Figure 14. Model performance on variant datasets.

6. Conclusions

The Att-JM that combines MC-Bi-GRU and MC-CNN extracts the targeted aspect and classifies their sentiments from the textual reviews as positive, negative, and neutral. In addition, both algorithms have acquired the same number of input parameters, such as word2vec-based embedding and contextual position information, which becomes the prominent cause of the proposed model's distinctions and incredible performance. The performance results of the Att-JM endorsed the envisioned hypothesis regarding parallel fused algorithms that equally and simultaneously input parameters provision can enhance the accuracy of targeted aspects and their predicted sentiments. The information sharing between the merged algorithms enables the model to obtain accurate information regarding the context and the prominent terms that influence each phrase within the context. In the future, we aim to extend this work to determine the implicit polarities relating to targeted aspects.

Author Contributions: Conceptualization, T.I.; Methodology, H.U.K., U.T. and J.-h.C.; Software, U.T.; Validation, M.A.K.; Formal analysis, W.A. and J.-h.C.; Investigation, H.U.K.; Resources, T.I. and J.-h.C.; Data curation, H.U.K.; Writing—original draft, W.A.; Writing—review & editing, W.A. and T.I.; Supervision, H.U.K.; Project administration, T.I. and M.A.K.; Funding acquisition, M.A.K., U.T. and J.-h.C. All authors contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the “Human Resources Program in Energy Technology” of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resources from the Ministry of Trade, Industry & Energy, Republic of Korea. (No. 20204010600090).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this work are publicly available.

Acknowledgments: The authors extend their appreciation to the “Human Resources Program in Energy Technology” of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resources from the Ministry of Trade, Industry & Energy, Republic of Korea. (No. 20204010600090).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Peng, Y.; Xiao, T.; Yuan, H. Cooperative gating network based on a single BERT encoder for aspect term sentiment analysis. *Appl. Intell.* **2021**, *52*, 5867–5879. [CrossRef]
2. Yang, J.; Yang, R.; Lu, H.; Wang, C.; Xie, J. Multi-Entity Aspect-Based Sentiment Analysis with Context, Entity, Aspect Memory and Dependency Information. *ACM Trans. Asian Low-Resour. Lang. Inf. Process. (TALLIP)* **2019**, *18*, 1–22. [CrossRef]
3. Yu, S.; Liu, D.; Zhu, W.; Zhang, Y.; Zhao, S. Attention-based LSTM, GRU and CNN for short text classification. *J. Intell. Fuzzy Syst.* **2020**, *39*, 1–8. [CrossRef]

4. Mahmood, A.; Khan, H.U.; Ramzan, M. On Modelling for Bias-Aware Sentiment Analysis and Its Impact in Twitter. *J. Web Eng.* **2020**, *19*, 1–28.
5. Jihan, N.; Senarath, Y.; Tennekoon, D.; Wickramarathne, M.; Ranathunga, S. Multi-Domain Aspect Extraction using Support Vector Machines. In Proceedings of Proceedings of the 29th Conference on Computational Linguistics and Speech Processing (ROCLING 2017), Taipei, Taiwan, 27–28 November 2017; pp. 308–322.
6. Hegde, R.; Seema, S. Aspect based feature extraction and sentiment classification of review data sets using Incremental machine learning algorithm. In Proceedings of the 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Chennai, India, 27–28 February 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 122–125.
7. Xiang, Y.; He, H.; Zheng, J. Aspect term extraction based on MFE-CRF. *Information* **2018**, *9*, 198. [[CrossRef](#)]
8. Shams, M.; Baraani-Dastjerdi, A. Enriched LDA (ELDA): Combination of latent Dirichlet allocation with word co-occurrence analysis for aspect extraction. *Expert Syst. Appl.* **2017**, *80*, 136–146. [[CrossRef](#)]
9. Das, S.J.; Chakraborty, B. An Approach for Automatic Aspect Extraction by Latent Dirichlet Allocation. In Proceedings of the 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), Morioka, Japan, 23–25 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
10. Wan, C.; Peng, Y.; Xiao, K.; Liu, X.; Jiang, T.; Liu, D. An association-constrained LDA model for joint extraction of product aspects and opinions. *Inf. Sci.* **2020**, *519*, 243–259. [[CrossRef](#)]
11. Liao, C.; Feng, C.; Yang, S.; Huang, H.-Y. A hybrid method of domain lexicon construction for opinion targets extraction using syntax and semantics. *J. Comput. Sci. Technol.* **2016**, *31*, 595–603. [[CrossRef](#)]
12. Mowlaei, M.E.; Abadeh, M.S.; Keshavarz, H. Aspect-based sentiment analysis using adaptive aspect-based lexicons. *Expert Syst. Appl.* **2020**, *148*, 113234. [[CrossRef](#)]
13. Wai, M.S.; Aung, S.S. Simultaneous opinion lexicon expansion and product feature extraction. In Proceedings of the 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 24–26 May 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 107–112.
14. Liu, Q.; Gao, Z.; Liu, B.; Zhang, Y. Automated rule selection for opinion target extraction. *Knowl.-Based Syst.* **2016**, *104*, 74–88. [[CrossRef](#)]
15. Kang, Y.; Zhou, L. Rule-based methods for extracting product features from online consumer reviews. *Inf. Manag.* **2017**, *54*, 166–176. [[CrossRef](#)]
16. Asghar, M.Z.; Khan, A.; Zahra, S.R.; Ahmad, S.; Kundi, F.M. Aspect-based opinion mining framework using heuristic patterns. *Clust. Comput.* **2019**, *22*, 7181–7199. [[CrossRef](#)]
17. Kushwaha, A.; Chaudhary, S. Review highlights: Opinion mining on reviews: A hybrid model for rule selection in aspect extraction. In Proceedings of the 1st International Conference on Internet of Things and Machine Learning, Liverpool, UK, 17–18 October 2017; pp. 1–6.
18. Ruskanda, F.Z.; Widiantoro, D.H.; Purwarianti, A. Comparative Study on Language Rule Based Methods for Aspect Extraction in Sentiment Analysis. In Proceedings of the 2018 International Conference on Asian Language Processing (IALP), Bandung, Indonesia, 15–17 November 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 56–61.
19. Matsuno, I.P.; Rossi, R.G.; Marcacini, R.M.; Rezende, S.O. Aspect-based Sentiment Analysis using Semi-supervised Learning in Bipartite Heterogeneous Networks. *JIDM* **2016**, *7*, 141–154.
20. Rana, T.A.; Cheah, Y.-N. A two-fold rule-based model for aspect extraction. *Expert Syst. Appl.* **2017**, *89*, 273–285. [[CrossRef](#)]
21. Marcacini, R.M.; Rossi, R.G.; Matsuno, I.P.; Rezende, S.O. Cross-domain aspect extraction for sentiment analysis: A transductive learning approach. *Decis. Support Syst.* **2018**, *114*, 70–80. [[CrossRef](#)]
22. Shu, L.; Xu, H.; Liu, B. Lifelong learning crf for supervised aspect extraction. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics 2017, Vancouver, BC, Canada, 30 July–4 August 2017; pp. 148–154.
23. Khan, H.U.; Nasir, S.; Nasim, K.; Shabbir, D.; Mahmood, A. Twitter trends: A ranking algorithm analysis on real time data. *Expert Syst. Appl.* **2021**, *164*, 113990. [[CrossRef](#)]
24. Al-Smadi, M.; Talafha, B.; Al-Ayyoub, M.; Jararweh, Y. Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 2163–2175. [[CrossRef](#)]
25. Guo, L.; Zhang, D.; Wang, L.; Wang, H.; Cui, B. CRAN: A hybrid CNN-RNN attention-based model for text classification. In Proceedings of the International Conference on Conceptual Modeling, Xi'an, China, 22–25 October 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 571–585.
26. Lai, S.; Xu, L.; Liu, K.; Zhao, J. Recurrent convolutional neural networks for text classification. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.
27. Maitama, J.Z.; Idris, N.; Abdi, A.; Shuib, L.; Fauzi, R. A Systematic Review on Implicit and Explicit Aspect Extraction in Sentiment Analysis. *IEEE Access* **2020**, *8*, 194166–194191. [[CrossRef](#)]
28. Quan, W.; Chen, Z.; Gao, J.; Hu, X.T. Comparative Study of CNN and LSTM based Attention Neural Networks for Aspect-Level Opinion Mining. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 2141–2150.
29. Yang, C.; Zhang, H.; Jiang, B.; Li, K. Aspect-based sentiment analysis with alternating coattention networks. *Inf. Process. Manag.* **2019**, *56*, 463–478. [[CrossRef](#)]

30. Tang, F.; Fu, L.; Yao, B.; Xu, W. Aspect based fine-grained sentiment analysis for online reviews. *Inf. Sci.* **2019**, *488*, 190–204. [[CrossRef](#)]
31. Gallego, F.O.; Corchuelo, R. Torii: An aspect-based sentiment analysis system that can mine conditions. *Softw. Pract. Exp.* **2020**, *50*, 47–64. [[CrossRef](#)]
32. Han, Y.; Liu, M.; Jing, W. Aspect-level Drug Reviews Sentiment Analysis based on Double BiGRU and Knowledge Transfer. *IEEE Access* **2020**, *8*, 21314–21325. [[CrossRef](#)]
33. Al-Smadi, M.; Al-Ayyoub, M.; Jararweh, Y.; Qawasmeh, O. Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features. *Inf. Process. Manag.* **2019**, *56*, 308–319. [[CrossRef](#)]
34. Karagoz, P.; Kama, B.; Ozturk, M.; Toroslu, I.H.; Canturk, D. A framework for aspect based sentiment analysis on turkish informal texts. *J. Intell. Inf. Syst.* **2019**, *53*, 431–451. [[CrossRef](#)]
35. Zeng, D.; Dai, Y.; Li, F.; Wang, J.; Sangaiah, A.K. Aspect based sentiment analysis by a linguistically regularized CNN with gated mechanism. *J. Intell. Fuzzy Syst.* **2019**, *36*, 3971–3980. [[CrossRef](#)]
36. Liu, N.; Shen, B.; Zhang, Z.; Zhang, Z.; Mi, K. Attention-based Sentiment Reasoner for aspect-based sentiment analysis. *Hum.-Cent. Comput. Inf. Sci.* **2019**, *9*, 35. [[CrossRef](#)]
37. Yu, J.; Jiang, J.; Xia, R. Global inference for aspect and opinion terms co-extraction based on multi-task neural networks. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2018**, *27*, 168–177. [[CrossRef](#)]
38. Ye, H.; Yan, Z.; Luo, Z.; Chao, W. Dependency-tree based convolutional neural networks for aspect term extraction. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Jeju, Republic of Korea, 23–26 May 2017; Springer: Berlin/Heidelberg, Germany, 2017; pp. 350–362.
39. Xu, H.; Liu, B.; Shu, L.; Yu, P.S. Double embeddings and cnn-based sequence labeling for aspect extraction. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics 2018, Volume 2: Short Papers, Melbourne, Australia, 15–20 July 2018; pp. 592–598.
40. Shu, L.; Xu, H.; Liu, B. Controlled CNN-based Sequence Labeling for Aspect Extraction. *arXiv* **2019**, arXiv:1905.06407.
41. Da'u, A.; Salim, N. Aspect extraction on user textual reviews using multi-channel convolutional neural network. *PeerJ Comput. Sci.* **2019**, *5*, e191. [[CrossRef](#)]
42. Li, X.; Bing, L.; Li, P.; Lam, W.; Yang, Z. Aspect term extraction with history attention and selective transformation. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18) 2018, Stockholm, Sweden, 13–19 July 2018.
43. Saraiva, F.Z.R.; da Silva, T.L.C.; de Macêdo, J.A.F. Aspect Term Extraction Using Deep Learning Model with Minimal Feature Engineering. In Proceedings of the International Conference on Advanced Information Systems Engineering, Grenoble, France, 8–12 June 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 185–198.
44. Tran, T.U.; Hoang, H.T.T.; Huynh, H.X. Aspect Extraction with Bidirectional GRU and CRF. In Proceedings of the 2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF), Danang, Vietnam, 20–22 March 2019; pp. 1–5.
45. Wu, C.; Wu, F.; Wu, S.; Yuan, Z.; Huang, Y. A hybrid unsupervised method for aspect term and opinion target extraction. *Knowl.-Based Syst.* **2018**, *148*, 66–73. [[CrossRef](#)]
46. Chauhan, G.S.; Meena, Y.K.; Gopalani, D.; Nahta, R. A two-step hybrid unsupervised model with attention mechanism for aspect extraction. *Expert Syst. Appl.* **2020**, *161*, 113673. [[CrossRef](#)]
47. Meng, W.; Wei, Y.; Liu, P.; Zhu, Z.; Yin, H. Aspect Based Sentiment Analysis With Feature Enhanced Attention CNN-BiLSTM. *IEEE Access* **2019**, *7*, 167240–167249. [[CrossRef](#)]
48. Liu, N.; Shen, B. Aspect-based sentiment analysis with gated alternate neural network. *Knowl.-Based Syst.* **2020**, *188*, 105010. [[CrossRef](#)]
49. Zhu, Y.; Gao, X.; Zhang, W.; Liu, S.; Zhang, Y. A bi-directional LSTM-CNN model with attention for aspect-level text classification. *Future Internet* **2018**, *10*, 116. [[CrossRef](#)]
50. Akhtar, M.S.; Garg, T.; Ekbal, A. Multi-task Learning for Aspect Term Extraction and Aspect Sentiment Classification. *Neurocomputing* **2020**, *398*, 247–256. [[CrossRef](#)]
51. Zhang, C.; Li, Q.; Cheng, X. Text Sentiment Classification Based on Feature Fusion Text Sentiment Classification Based on Feature Fusion. *Rev. Intell. Artif.* **2020**, *34*, 515–520.
52. Zhang, J.; Liu, F.a.; Xu, W.; Yu, H. Feature Fusion Text Classification Model Combining CNN and BiGRU with Multi-Attention Mechanism. *Future Internet* **2019**, *11*, 237. [[CrossRef](#)]
53. Cheng, Y.; Yao, L.; Xiang, G.; Zhang, G.; Tang, T.; Zhong, L. Text sentiment orientation analysis based on multi-channel CNN and bidirectional GRU with attention mechanism. *IEEE Access* **2020**, *8*, 134964–134975. [[CrossRef](#)]
54. Huang, B.; Guo, R.; Zhu, Y.; Fang, Z.; Zeng, G.; Liu, J.; Wang, Y.; Fujita, H.; Shi, Z. Aspect-level sentiment analysis with aspect-specific context position information. *Knowl.-Based Syst.* **2022**, *243*, 108473. [[CrossRef](#)]
55. Feng, A.; Zhang, X.; Song, X. Unrestricted Attention May Not Be All You Need—Masked Attention Mechanism Focuses Better on Relevant Parts in Aspect-Based Sentiment Analysis. *IEEE Access* **2022**, *10*, 8518–8528. [[CrossRef](#)]
56. Liao, W.; Zhou, J.; Wang, Y.; Yin, Y.; Zhang, X. Fine-grained attention-based phrase-aware network for aspect-level sentiment analysis. *Artif. Intell. Rev.* **2021**, *55*, 3727–3746. [[CrossRef](#)]
57. Jin, Y.; Zhang, H.; Du, D. Incorporating positional information into deep belief networks for sentiment classification. In Proceedings of the Industrial Conference on Data Mining, New York, NY, USA, 12–13 July 2017; pp. 1–15.

58. LeCun, Y.; Bengio, Y. Convolutional networks for images, speech, and time series. *Handb. Brain Theory Neural Netw.* **1995**, *3361*, 1995.
59. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. In Proceedings of the NIPS 2014 Workshop on Deep Learning 2014, Montreal, QC, Canada, 11 December 2014.
60. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. In Proceedings of the SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation 2014, Doha, Qatar, 25 October 2014; pp. 103–111.
61. Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; Hovy, E. Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 1480–1489.
62. Wang, W.; Pan, S.J. Syntactically Meaningful and Transferable Recursive Neural Networks for Aspect and Opinion Extraction. *Comput. Linguist.* **2020**, *45*, 705–736. [[CrossRef](#)]
63. Xiang, R.; Chersoni, E.; Lu, Q.; Huang, C.R.; Li, W.; Long, Y. Lexical data augmentation for sentiment analysis. *J. Assoc. Inf. Sci. Technol.* **2021**, *72*, 1432–1447. [[CrossRef](#)]
64. Tijare, P.V.; Prathuri, J.R. Correlation Between K-means Clustering and Topic Modeling Methods on Twitter Datasets. In *Cyber Security and Digital Forensics*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 459–477.
65. Xiao, D.; Ren, F.; Pang, X.; Cai, M.; Wang, Q.; He, M.; Peng, J.; Fu, H. A hierarchical and parallel framework for End-to-End Aspect-based Sentiment Analysis. *Neurocomputing* **2021**, *465*, 549–560. [[CrossRef](#)]
66. Li, Z.; Li, L.; Zhou, A.; Lu, H. JTSG: A joint term-sentiment generator for aspect-based sentiment analysis. *Neurocomputing* **2021**, *459*, 1–9. [[CrossRef](#)]
67. Dai, A.; Hu, X.; Nie, J.; Chen, J. Learning from word semantics to sentence syntax by graph convolutional networks for aspect-based sentiment analysis. *Int. J. Data Sci. Anal.* **2022**, *14*, 17–26. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.