IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 April 2023, accepted 16 May 2023, date of publication 1 June 2023, date of current version 7 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3281889

RESEARCH ARTICLE

Attention-Based Multi-Channel Gated Recurrent Neural Networks: A Novel Feature-Centric Approach for Aspect-Based Sentiment Classification

WAQAS AHMAD[1], HIKMAT ULLAH KHAN[1], TASSAWAR IQBAL[1], AND SAQIB IQBAL[2]
[1]Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah Cantt 47040, Pakistan
[2]Department of Software Engineering, Al Ain University, Al Ain, United Arab Emirates

Corresponding author: Hikmat Ullah Khan (Hikmat.ullah@ciitwah.edu.pk)

ABSTRACT Sentiment analysis is an active research domain of the current era, thanks to its vast applications. The main objective is to classify the polarity of the given text as negative, positive, or neutral. Thus, researchers' focus shifted towards the aspect or feature-based sentiment analysis because overall polarity does not determine the people's views towards certain features. Therefore, Aspect-Based Sentiment Analysis (ABSA) helps us to identify the sentiments about various aspects of different services and products. However, their accurate identification and extraction are still challenging for the research community due to the complex nature of natural languages. This paper presents a method named Attention-based Multi-Channel Gated Recurrent Neural Network (Att-MC-GRU), which extracts aspects and analyzes textual reviews to predict or classify their sentiments. It introduces the hybrid approach by combining word embedding, part of speech (POS) tags, and contextual position information. The main novelty lies in proposal of a Multi-Channel Gated Recurrent Neural Network (MC-GRU), in contrast to the existing studies that consider Recurrent Neural Networks (RNN) comprising only a single input channel. In addition, word embedding, POS tags, and contextual position information collectively improve the identification and prediction accuracy of aspects and their associated sentiments. Due to the application of the filtering by the attention mechanism that figured out first the significant words, which helps to determine entities' aspects related to the sentiment expressed. The empirical analysis proves the proposed approach's effectiveness compared to the existing techniques in the relevant literature using standard datasets. According to the empirical analysis, the proposed model performs better in the F1-measure, with an overall achievement of 94% in the task of aspect extraction and 93% in the classification of sentiment.

INDEX TERMS Attention mechanism, aspect extraction, contextual position information, deep learning multi-channel, gated recurrent unit, part of speech, sentiment analysis, word embedding.

I. INTRODUCTION

As the social web provides content generation facility to common people, its users share their views, experiences, and interact with other people around the globe using social media networks such as online forums, blogs, and social communities. In addition, the technological advancement of this modern era facilitates its users to share their feedback and reviews about various topics, products, events, political issues, etc. While articulating their opinions, the users are free to discuss different merits and demerits of topics or entities under discussion. Under these circumstances, social media platforms become a massive source of excessive increase in user-generated content. In the meantime, the most contributing facets which cause such content's gigantic production and sharing incorporate opinionated text and reviews regarding products and services. According to the satisfaction of consumers, these textual reviews, and comments exhibit users'

The associate editor coordinating the review of this manuscript and approving it for publication was Rute C. Sofia.

thoughts, behavior, and views polarity as positive, negative, and neutral. The significance of these opinionated texts analyzes by the fact that roughly 91% of people around the globe periodically or regularly read online reviews, of which approximately 84% prefer these reviews as compared the personal recommendation [1]. The exploration of sentiments from such opinionated data is advantageous for the individuals, social institutions, organizations, and business policy makers that revise their decisions from the users' perspective while assessing their reputation in the market. An unbiased automatic examination of these bulky textual contents that assist people in responding to people's reflections requires a system that collectively uses Natural Language Processing (NLP), linguistically computation, and text retrieval to classify feelings and their expressed sentiments. Thus, the analysis that utilizes NLP and its linguistic computations for subjectivity extraction from a textual context is known as Sentiment Analysis (SA) [2], [3]. It can also describe as a web-based unstructured textual content computational analysis, which can analyze the attitude, behavior, and thoughts criteria belonging to a person about a specific entity. In addition, it is a common practice to assign positive or negative opinions about various aspects of a topic during a single post, which makes traditional SA impractical because they express an opinion as a whole and cannot resolve such scenarios. However, a type of analysis that determines the relationship among opinionated targets and assigns it the corresponding polarity value is known as ABSA, which expresses opinions regarding each aspect identified during each post [4]. This task accomplishes in two parts. The first one identifies the potential aspects, whereas the other assigns the polarity of opinions corresponding to all extracted aspects [5]. These aspects are of two types. These aspects are either implicit, not prominently mentioned in textual reviews or explicit, clearly defined in the textual content. The relevant literature depicts various mechanisms and techniques of the past which perform the extraction and identification of aspects with their associating sentiment. This paper explains such contributions' opportunities and obstacles only belonging to explicit aspects because the implicit aspects are not in the scope of this paper.

In the literature, methodologies such as lexicon-based, topic modeling, and syntactic relation-based or rule-based approaches accomplish the task of explicit Aspect Term Extraction (ATE) and Sentiment Classification (SC). Meanwhile, the syntactic pattern techniques that extract aspects and their sentiment while utilizing language and grammatical dependent rules perform well in ABSA. However, most people do not care about following the syntactic pattern rules when producing their feedback. Furthermore, while sharing posts, feedback, and reviews by users, some websites ignore these rules, which influences the performance of these methodologies. In addition, their time complexity constraint and reliance on specialists for the creation of rules and lexicons become the cause of discouraging these approaches, which restricts them to a specific domain or language [6], [7]. On the other hand, machine learning-based supervised methodologies provide domain and language independence, which never demand any specialist to generate rules and lexicons during aspect identification and classification. Irrespective of these benefits, their reliance upon large volume labeled and annotated data expresses a bottleneck of supervised machine learning approaches [8], [9]. Therefore, the semi-supervised methodologies demand less labeled data during training while their complex mechanism for feature selection is time-consuming, which becomes the cause of declining such techniques. In the meantime, the manual feature engineering mechanism is a critical component of unsupervised methods. As a result, their extracted aspects and predicted sentiment quality are heavily reliant on manual feature engineering mechanisms, which affects their ability to adapt to a variety of real-life scenarios [10], [11]. Despite these approaches, there is another prominent branch of machine learning named Deep Learning (DL), which comprises an end-to-end method and performs an automatic feature selection without any manual intervention. These abilities improve their models' efficiency from the traditional preceding methods such as machine learning-based, linguistically rule/pattern based and lexicon-based procedures, whereas to some extent match the performance of human abilities.

Traditionally, extraction of aspects and prediction of sentiment from user reviews through manual annotation and rule creation strategies makes these tasks highly challenging and domain- and task-dependent. Therefore, PM-G&D [12] proposed a method to extract these targeted aspects automatically using a Convolutional Neural Network (CNN) along with two embedding layers based on GloVe. One of these embeddings is generally pre-trained, while the training of the second layer accomplishes according to a specific domain. Nowadays, DL algorithms that perform NLP tasks are massively dependent on either the word or character corresponding vector representations. In this regard, the MCNN [13] presented a Multi-Channel Convolutional Neural Network (MC-CNN) that merges three vector representations to perform the task of ABSA. Among them, the first two channels possess word embeddings based on GloVe and word2vec, while the last one includes the representation in the form of a one-hot character vector. In another contribution, MCNN+WV2+POS [14] proposed a methodology comprised of MC-CNN, which utilizes POS tags and word embedding as textual features. This approach comprises two channels for input. The first one takes word2vec embedding based upon Google-news pre-trained model as input. Meanwhile, the other channel receives POS tags as a one-hot-vector representation. The convolutional layer utilizes the knowledge of word embedding and sequences labeling information for predicting the potential terms as aspects.

On the other side, a deficiency detected in CNN while accomplishing the task of ABSA is the non-consideration of semantically localized features. As a result, Li et al. (III-A2) [15] proposed a novel approach for ATE, which utilizes a summary of opinion and aspect history for their detection. Their procedure has based on two Long-Short Term Memory Networks (LSTMs) for taking semantically localized and sequential information regarding aspect and opinion representations. HEA-LSTM [16] is another attention-based LSTM approach that extracts all highly potential features from the textual content based on their corresponding hints. In addition, existing methodologies lack insights into syntactic relations among aspects and opinion terms, which degrades their prediction ability of aspect-related opinions. Therefore, GMTCMLA [17] proposed a Gated Recurrent Unit (GRU) based multi-task learning method. It determines the implicit syntactic rules between the aspect and opinion terms then explicit syntactic rules have modelled with the help of global inference for predicting aspect and opinion terms. In the meantime, Wu et al. [18] presented a unsupervised hybrid methodology that combines a rules-based approach to train the GRU model for extracting targeted aspects from textual reviews.

ABSA is a fine-grain and unique genre of SA. Therefore, their algorithms should be capable enough that understands the complex and in-depth relations between aspect and opinion words. Among the DL algorithms, Recurrent Neural Networks (RNN) and CNN are the two most well-reputed algorithms. CNNs are famous for requiring fewer resources and parameters for extracting localized and position-invariant features from textual data with remarkable performance. In addition, long-term dependencies and sequence information handling raised difficulties within these algorithms. On the other side, the RNN variants such as LSTMs and GRUs are famous for handling long-term and semantic dependencies of sequential context. GRU's internal gating architecture is less complicated as compared to LSTM. Therefore, they are computationally inexpensive and faster to tune compared to LSTMs. In the meantime, the attention mechanism uses widely with GRUs for inferring high-level representations and capturing syntactic and semantic facts of sequential contexts. In the meanwhile, the attention mechanism enhances the algorithm's ability to determine the contribution of each word or phrase toward the associated sentiment. Therefore, the proposed methodology considers an attention mechanism with GRU to enhance the accuracy of ATE and SC.

In the relevant literature and best to our knowledge, RNN's family, such as LSTMs and GRUs, are observed with single-channel for input. Conversely, CNN discovers mostly along with multi-channels in literature, including our baseline techniques, which comprises multiple channels for input parameters. According to our knowledge, the proposed approach is the first to use the Multi-Channel Gated Recurrent Neural Network (MC-GRU) and a combination of novel features for identifying aspects and classifying their predicted sentiments. In addition, different positions of words possess variant influences in the context of the sentence. These influences describe through the contextual position information vectors, which ignores massively in the literature. The proposed model considers the attention mechanism and contextual position information to highlight the potential features along their sentiments. These fundamental facts become the cause of the proposed approach distinction compared to existing methodologies.

The aim of the proposed approach is to extract aspects and classify their sentiments from the textual reviews using MC-GRU, which uses three channels for input. Among them, the first channel utilizes Google News dataset pre-trained word2vec embedding to encode the words' syntactic and semantic information. In the meantime, the second channel supports sequential tagging through POS tag embedding. Furthermore, the implementation of position information depicts its importance with performance enhancement while accomplishing various NLP tasks. These inspirations motivate the proposed method to use the contextual position information vectors. Therefore, the proposed approach introduces a third and last channel, which acquires contextual position information embedding as position-information vectors. These three embeddings are then provided to GRU layers separately, and their outcome is then combined and delivered towards attention mechanism. The attention layer highlights and emphasizes the potential terms in the light of syntactic, semantic, contextual position information, and sequential information, which highly achieves the focus of the entire contextual information and then passes them towards the activation function for identifying aspects and classifying their sentiments. The proposed method evaluates through the publicly available standard datasets for ABSA tasks comprising SemEval and Twitter because these are vibrantly used datasets in literature. Below is a summary of the main contribution of this paper:

1) The deep learning algorithms consider word embedding due to preserving the semantic, syntactic meaning, and contextual information at reduced size vectors. Meanwhile, POS tags utilize for preserving the sequential information and depicting the role of each word within context. Furthermore, contextual information illustrates the impact of each position in relation to its surroundings. Therefore, the proposed approach proposes and analyzes the word2vec embedding, contextual position information vector, and POS tags' collaboration effect on the deep learning-based algorithm such as MC-GRU performance while identifying aspects and their classifying their associated sentiment.
2) The proposed approach delivers these syntactic, semantic, sequential information, and contextual position information of contextual phrases to the attention mechanism, which inclusively enhances its ability to determine, filter, and highlight the most prominent and influenced terms of context with their accurate placement knowledge.
3) The proposed approach utilizes a Multi-Channel Gated Recurrent Neural Network (MC-GRU), which is the

novelty of the proposed technique, along with an attention mechanism, the proposed model, simultaneously uses various deep features for the accurate extraction of aspects and their predicted sentiments classification.

4) The evaluation metrics, such as precision, recall, and F1 measure, demonstrate significant performance of the proposed approach on standardized datasets containing Twitter and SemEval. Consequently, the F1 indicates 94% performance for the proposed approach in the task of ATE and 93% in the SC.

As a whole, the paper follows the following structure. The analysis of related literature regarding ATE and SC describes as ''Related Work'' in section II. This section describes the issues and challenges discovered within the existing contributions and becomes the cause of motivation while conducting this approach. While section III expresses the proposed methodology and its detail as the ''Proposed Methodology''. The experimental arrangement utilized to develop this model is described in Section IV as ''Experimental Arrangement''. In section V, ''Results and Discussions'', the performance comparisons of this model are discussed. The final section, known as section VI, sums up the whole approach with a discussion of its future directions as ''Conclusion''.

II. RELATED WORK

During the current era, the advancements in social media platforms (e.g., LinkedIn, Twitter, Instagram, and Facebook) and E-commerce websites (e.g., Amazon, Flipkart, eBay, and Alibaba) enable users to share their self-experience-based opinions and reviews. Thus, they elaborate their views regarding a tour trip, movie viewing experience in a Cineplex, product utilization, and hotel service experiences. Consequently, they became the cause of producing a large volume of unstructured data available on E-commerce websites and social media [19]. On the other side, various industries, firms and businesses' analysts broadly relies on customer-generated opinions while forecasting about future earnings of their corporations. Therefore, analyzing such unstructured data because of the disambiguation nature of natural languages becomes a crucial challenge for these organizations and industries. Hence, they demand an automated method that can summarize the users' and consumers' necessities, performs the purification of unessential data, and extracts appropriate data to make their organizational decision significant, practical, and according to the voice of people [20]. Therefore, the ability to automatically analyze or process online textual content available on social media platforms, such as microblogs and collaborative media, holds on to SA [21]. The mechanism of such analysis can classify the targeted textual contents as negative, positive, or neutral and assign the level of polarity as high, moderate, or low based upon the contextual analysis of feedback or reviews [22].

Two well-known types of sentiment analysis are document- and sentence-level sentiment analysis. These types focus only on an entire document or sentence-based opinion mining, which makes them inappropriate for daily life circumstances because the user is sometimes interested in a specific product or service's feature and wants to compare it with rivals effortlessly. Moreover, people want to enhance their purchase decisions based on the opinions of those users who have already experienced them earlier. In addition, daily life's textual review's single sentence may comprise the opinion about multiple entities or the multiple aspects of a single entity, which do not lie in the scope of the sentence and document-level sentiment analysis. Therefore, both before-mentioned SA types can never survive in those situations. They also never express liked and disliked aspects of an entity, whereas they emphasize the summarization of sentiments as positive, negative, and neutral. In this situation, ABSA fulfils this gap and handles these scenarios with enough ability which expresses the liked and disliked features of a targeted entity. It associates the sentiment polarity with specific aspects or features of any targeted entity [2], [21].

ABSA expresses user attitudes towards different aspects of an entity. Therefore, this task accomplishes in two phases. The first phase performs the identification and extraction task of potential features of the targeting entity. When all aspects are available, the second phase determines sentiment polarities. These sentiment polarities are then associated with the aspects corresponding to those sentiments. As the result of these two phases, ABSA conceives the overall sentiments corresponding to each aspect [5]. In the meantime, the accuracy of identified attributes of an entity and predicting their associated polarities is still an active area of research. It still demands the research community's attention. According to the literature, in the earlier days of ABSA, these aspects extract through handicraft features. These are complicated, laborious, and time-consuming techniques which demand a lot of effort from analysts [12], [14], [23], [24], [25]. Moreover, the feature extraction task accomplishes through machine learning approaches such as supervised [26], [27], semi-supervised [28], [29], and unsupervised [30], lexicons-based methods, e.g., SentiWordNet and domain-based lexicons [2], [31], [32], [33], [34], rule-based or pattern-based approaches [7], [35], [36], [37], [38], [39], [40], [41], topic modelling based techniques [42], [43], [44], and tree or graph-based approaches [45], [46]. In the present day, DL methods are known for ABSA tasks. However, future researchers will need to explore the possibility of incorporating reasoning, such as that of the human brain. In contrast to machine learning methods, DL algorithms never rely on manual procedures such as linguistically and experts' knowledge-based resources for identifying potential features. On the contrary, their models rely on a strategy named back-propagation. It utilizes the weights updating technique based on learning features and loss optimization mechanism, which enabled their automatic end-to-end training and enhanced their performance with bulky datasets.

Currently, the most of tasks of NLP accomplish using DL algorithms, which primarily rely on word or character-level embeddings. Therefore, MCNN [13] presented an approach comprising MC-CNN that merges three vector representations to perform the task of ABSA. These incorporate GloVe, word2vec, and a one-hot character-based embedding. During identifying aspects and their sentiments classification, the model performs remarkably. However, their approach relies only on syntactic, semantic and word co-occurrence information. In the meantime, the inclusion of prominent word position, contextual sequence, and inspirational terms knowledge can enhance the performance of their proposed methodology. In another contribution, A. Da'u and N. Salim [14] proposed a two input channel based MC-CNN model that utilizes POS tags and word embedding as textual features. Therefore, one of the channels takes word embedding, whereas the second one takes POS as input while identifying the potential features as aspects. It collectively utilizes word2vec embedding and POS representation, which enhances its performance while identifying targeted terms. In addition, their approach excluded the knowledge of long-term dependency and the significance of position information. The neglection of excluded features becomes the cause of effecting model prediction ability during aspect terms identification and extraction. In traditional approaches, contextual words and the relationship between aspect terms and contextual phrases are not considered. Therefore, Huang et al. [47] developed the CPA-SA approach, which performs ABSA while considering aspect-specific contextual location information. Their designed function adjusted context words' weights based on potential terms' positions to eliminate inference of potential terms on either side of conceivable terms and predict their polarity. Meanwhile, this technique involved no syntactic or semantic influence throughout the process of accomplishing this task. Consequence the above-discussed contributions, we analyze that CNN-based approaches work widely with multi-channels for input. As a result, these algorithms better analyze the under-discussed scenarios, which improves their ability to predict.

In the domain of NLP, RNNs' family LSTMs and GRUs are the most widely utilized algorithm because of their intrinsic ability to learn from sequential inputs. In addition, their ability to handle long-term and semantic dependencies of sequential data makes these algorithms more immaculate for ABSA task. Therefore, Li et al. [15] presented a framework for aspect terms extraction that comprises a summary of opinion and aspect history for their detection. This approach incorporates two LSTMs with a single channel for taking sequential information regarding aspect and opinion representations. However, they rely only on word2vec embedding while excluding the inclusion of contextual word position, contextual sequence, and inspirational terms knowledge can improve the performance of their approach. In another contribution, GMTCMLA [17] proposed a Bi-LSTM-based multi-task learning method with a single channel that determines the implicit syntactic rules among the opinion and aspect terms. Then explicit syntactic rules are derived with the help of global inference among opinion and aspect terms extraction. Their approach uses the rules for the extraction of possible terms. The creation of these rules makes this approach task-dependent, language-dependent and domain-specific. In addition, they rely only on word2vec embedding, whereas position invariance neglection reduces the performance of their proposed methodology. Therefore Chen et al. [48] proposed a novel technique named UniASTE that comprises BiLSTM with a single channel, which identifies aspects and their sentiments using the sentence's contextual information. It decomposes the entire task into three subtasks, which perform opinion, target, and sentiment tagging. The target and opinion tagging use BIO tags to detect the opinion expressions and targets. In addition, sentiment tagging determines potential opinion words and allocates them the sentimental tags. These factors improve the model's performance while predicting conceivable features. The proposed approach considers the context-based contextual information, whereas the consideration of contextual position influence is out of its extent. In addition, summarizing the above-discussed techniques, this study explores that RNN-based approaches widely contain a single channel for input. Meanwhile, they disregarded the multi-channel configuration, which enhances the algorithms' prediction capability within the relevant literature.

Additionally, in the relevant literature on SA, diverse hybrid approaches combine different algorithms of DL, such as CNN and RNN variants, to accomplish their respective analyses. They combine them either in a sequential or parallel manner. Furthermore, the relevant literature on SA also depicts contributions that utilize the combination of both methods. Thus, GU et al. [49] presented a hybrid methodology named MBGCV to enhance DL methods' performance while accomplishing the task of SA. MBGCV collectively utilizes both CNN and multi-channel BiGRU (that uses only word embedding) in a serialized manner, where the output of multi-channel BiGRU is provided to CNN to perform the SA task. On the other side, their approach is equipped with a variational information bottleneck and activation functions such as Maxout to reduce the vanishing gradient and over-fitting impact while training their model. They achieve good results with their model, but they lose beneficial information when they combine BiGRU and CNN in a serial manner. Regardless of the benefits of multi-channel configuration, it relies only on word embedding and has excluded other input features such as POS representation and the importance of contextual position information. In addition, online reviews and feedback, which express people's thoughts, opinions, and feelings about specific products and events, are sometimes accompanied by negative connotations. Therefore, Kanekar and Godbole [50] proposed a hybrid approach, which incorporates multi-channel CNN that integrates with BiGRU in a serialized manner. They utilize negation words and morphological GloVe-based word embedding as input for analyzing negated phrases. However, their method performs significantly, but the serialized integration becomes the cause of valuable contextual information loss and increases the model complexity. In addition, CNN is well known for capturing local feature information but faces difficulty handling long-term dependency information. In addition, the utilization of different features-based inputs can enhance further the accuracy of their method. In another parallel study, Cheng et al. [51] proposed a model consisting of attention-based Bi-GRU and attention-based MC-CNN that analyzes text sentiment. This framework utilizes only attention-based word2vec embeddings as input parameters for its algorithms (MCCNN and Bi-GRU). When accomplishing the classification, they ignore contextual positional information, dependent relations, and POS representation. The reference to these features can improve the prescribed framework's performance and accuracy. Moreover, CNN comprised multi-channels while Bi-GRU comprised only a single input channel which underutilizes the abilities of Bi-GRU.

According to the relevant literature, while extracting aspects and predicting sentiments, a fact about contextual words has analyzed that all their positions within textual context are not contributing equally. In addition, another fact observed in the literature is regarding the DL approaches, such as RNN-based methods while determining the association among variant aspects and sentiments, they always capture irrelevant information. That is due to the ability of DL algorithms to assign higher weight to upcoming terms of context, which causes them to extract inessential facts. Addressing this situation attention mechanism is provided as a solution. It can capture the effect of each word place of the sentence that influences the sentiments depicted within it. In addition, it also demonstrates them in the form of a dense vector, which represents the weight of each word within context. Besides, within this mechanism, instead of encoding information as a fixed-length vector, the algorithm revisits each part of the sentence to identify the most impactful contextual term [52], [53], [54], [55].

Summarizing the above-expressed discussions, we conclude that accurate extraction of aspects along their sentiments is the primary objective of the ABSA. Additionally, various techniques have been presented in the literature to accurately extract features through syntactic, semantic, POS, and attention-based mechanisms. In the meanwhile, fewer contributions address and recognizes the importance of utilizing contextual position information. Furthermore, a lack of those techniques observes in the relevant literature that uses all discussed parameters cumulatively within a single method in parallel. As a result, precise identification of aspects and their sentiment is still an issue for research contributions. DL methodologies proved their performance in NLP tasks. According to the literature, CNN is used in a multi-channel fashion to accomplish the task of ABSA. However, RNN's use in ABSA is lacking with multi-channels as a single-algorithm-based approach to perform these tasks. These inspirations motivate the study to propose an approach that comprises the GRU with three channels for input in parallel. These multi-channels for input comprise the word2vec embedding for semantic and syntactic information, POS tags embedding for sequential information, and contextual position information. However, the attention layer utilizes for the filtration of identified potential terms. The consideration of collective utilization of novel deep features and attention mechanism makes the proposed GRU-based model a comprehensive and novel approach which differentiates it from the previous contributions. The novelty of the proposed approach is that it presents a single-algorithm-centered procedure that addresses the field of ABSA. It utilizes GRU with multi-channels that collectively use parallel deep features such as word2vec, POS, and contextual position information provided on separate channels. As a result of these inputs, the proposed approach develops a comprehensive understanding of targeted contextual information to identify potential aspects and their expressed sentiments. The combined use of GRU with multiple channels equipped with various input deep features highlights the main novelty of the novel approach. In the meantime, integrating that algorithm with an attention mechanism to filter out unnecessary terms is a subpart of that novelty.

III. PROPOSED METHODOLOGY

The proposed approach aims to extract aspects and their sentiments from the textual reviews using MC-GRU, which comprises three channels for input. According to relevant literature, the existing methods have widely utilized CNN with multiple input channels. On the contrary, GRU-based models contain a single channel for input. In addition, CNN-based models use variant input features in the association of word2vec embedding. These models mainly comprise two input channels, which often disregard attention mechanisms and contextual positional information. In the meantime, this becomes the cause of losing valuable information about aspect identification and SC. All these causes motivate this proposed approach to utilize the Attention-based Multi-Channel GRU (Att-MC-GRU) involving multiple deep-novel features comprised of contextual position information, word2vec and POS tags in parallel. According to the relevant literature, word2vec embedding alone cannot handle the rich contextual knowledge of the targeted context [13]. Therefore, to improve the effectiveness of word2vec embedding, additional features, such as POS representations and contextual position information, have been combined with these vectors. Each DL algorithm has distinct capabilities. Therefore, these input feature selections are based on the abilities of the targeted DL algorithms to perform a specific task. Therefore, the first channel uses word2vec embedding to preserve the semantic and syntactic information of the context, whereas the second channel uses POS tagging to encode the context-based sequential information. In addition, the words' contextual position along their associated positional weights performed significantly in the tasks of NLP that encourages the proposed approach to add position information vectors within the Att-MC-GRU. Hence, the Att-MC-GRU introduces the third and last channel that uses

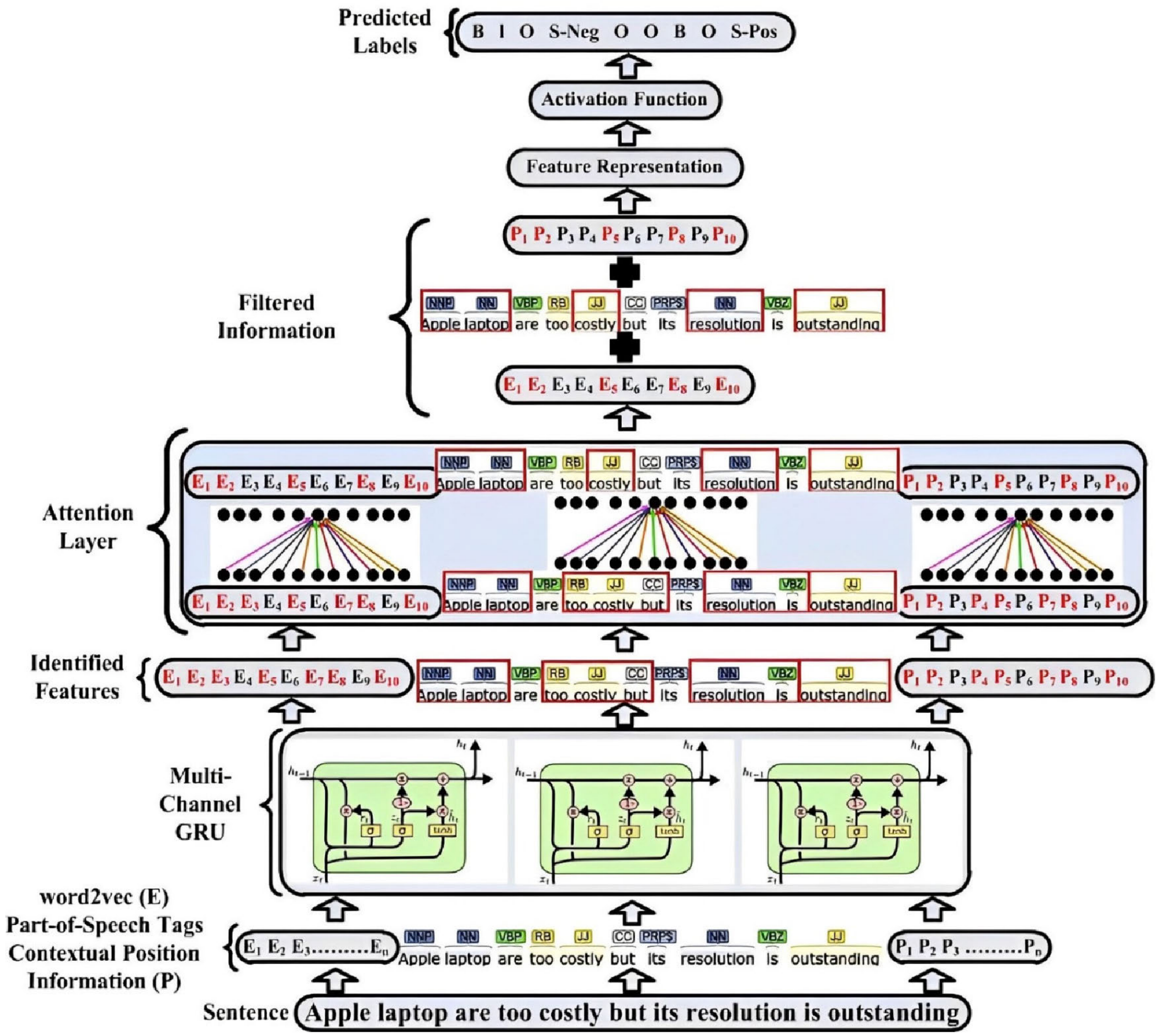


FIGURE 1. A framework of attention-based MC-GRU's for accomplishing the tasks of ABSA.

contextual position information. These three embeddings are provided towards the GRU layer, whereas their output is combined and delivered toward the attention layer. It highlights and emphasizes those terms that depict the focus of the whole context in the light of word2vec embedding, contextual position information, and POS tags and then assign them to the softmax activation function, which identifies aspects and classifies their sentiments.

Word embedding provides the semantic and syntactic information of the targeted sentence, whereas POS tags provide the sequential information for identifying the potential terms as aspects. Different positions of words possess variant influences in the context of the sentence. The contextual position information vectors depict these influences. Collectively such significant information has been provided to MC-GRU. According to the relevant literature, previous techniques are insufficient to utilize these novel features while identifying potential terms with GRU. In the meanwhile, RNNs' family is famous for handling long-term dependencies. In the meantime, the provision of multidimensional inputs provides sufficient facts to GRU that assists the model while predicting the possible terms based on these long-term dependencies. These distinctions make the approach novel and significant compared to those which provide a single feature as input to GRU, which is the main contribution of the proposed approach. In addition, these learned features reinforce the attention mechanism while identifying the highly effective terms of the targeted sentence. Then has classified the targeted features and their associated sentiments as positive, negative, and neutral. Thus, the classification process categorizes the attributes' sentiment into their associated polarity. The detailed description of the proposed methodology framework has given in Fig.1.

A. DETAILED DESCRIPTION OF PROPOSED MODEL

Initially, GRU came to the fore in the year 2014. Their main objective is to apprehend the time-based variations and overcome the vanishing and exploding gradient issues. Generally,

TABLE 1. Utilized symbols in the procedure.

Symbol	Interpretation	Symbol	Interpretation
W	Weight matrix	f	Activation function
x	Input term	l	Sequence length
b	Bias	j	Targeted term
p	Word position	u	Update vector
H	GRUs combined output	POS	Part-of-Speech vector
θ	Ratio of influence	c	Length of sentence
i	Current sentence	$\odot$	Element-wise multiplication
α	Score of Attention	S	Concatenation
CPI	Vector of position information	WE	Word embedding

the GRUs and LSTMs are identical in functionality, whereas they differ only in the number of utilized gates. Thus, the GRU comprises only two types of gates. Those named as update gate Z and reset gate R. The update gate Z updates the hidden state h, which assists the decision-making process that indicates the previous state's influence ratio on the current state while the updates occur. In the meantime, the reset gate R helps the model to decide the forgotten percentage of the past state's information within the current state. These gated units, such as the update and reset gates, are merged into a single gate unit named forget gate in a minimal gated version. Consequently, GRU utilizes a fewer number of parameters compared to LSTM, which reduces the computation time and computational complexity. The following equation set (01-04) [56], [57] demonstrates the functionality of the GRU layer with parameters. The symbols we used to elaborate our procedure are shown in Table 1.

$$Z_t = f(W_z h_t + U_z h_{t-1} + b_z) \tag{01}$$
$$R_t = f(W_r h_t + U_r h_{t-1} + b_r) \tag{02}$$
$$\hat{h}_t = f(W_t h_t + U_t (R_t \odot h_{t-1}) + b_t) \tag{03}$$
$$h_t = (1 - Z_t) \odot h_{t-1} + Z_t \odot \hat{h}_t \tag{04}$$

In the above-expressed equations, both W and U represent the weight matrix. While f and $\odot$ represent the activation function (such as Sigmoid, ReLU, Leaky ReLU, and Tanh) and element-wise multiplication, respectively. The term Z_t expresses the update vector, which defines the degree of influence of the previous state during the calculation of the current state. These are inherited gating mechanisms from their predecessors, such as LSTM. However, GRU modifies them and merges both the input and forget gates into a single unit named the update gate. Therefore, it becomes the cause of reducing the time complexity and improves the convergence during training and testing. These modifications lead to the described algorithm being the most well-known recurrent model among its siblings. Meanwhile, R_t expresses the reset gate and is parametrically similar to the Z_t gate. However, the $\hat{h}_t$ represents the previously hidden state, whereas the h_t describes the current state of the GRU model, and b is a bise term.

1) PRE-PROCESSING FOR DATA

Pre-processing steps are performed in the initial phases of Att-MC-GRU to filter data and transform it into a processable form. In subsequent sections, we will discuss specifics about these pre-processing steps and data acquisitions.

2) MULTI-CHANNEL GATED RECURRENT UNIT

In multi-dimensional vector space, word embeddings, such as GloVe and word2vec, have been utilized, which express the vocabulary words as real-valued vectors [13]. Word embedding has proven its paranormal abilities while extracting word regularities through semantic and syntactic constraints in machine learning and its subsidiaries. This inspiration has forced the proposed method strategy to consider such a procedure while extracting targeted aspects and their associated sentiments. As a result, the first channel utilizes word2vec embedding to preserve the semantic and syntactic information of the words and phrases. Hence, pre-trained word2vec embedding on Google News dataset, which contains three hundred dimensional vectors, is used. In addition, padding applies to acquire a constant length of sentences. Therefore, zero-padding is utilized and mapped to each word's corresponding n-dimensional embedding. Meanwhile, equation (05) expresses the semantic feature s_l of length l related to each sentence S.

$$\lim_{1\to l} S = (s_1, s_2, \ldots\ldots\ldots\ldots s_l) \tag{05}$$

Here S is the targeted sentence, s_l is the semantic feature, and l is the length of each semantic feature s_l. A feature-based weight matrix W_{WE} utilizes for word embedding. Afterwards, the following equations (06-09) use to identify and extract potential features.

$$Z_{WE} = f(W_{WE} x_t + U_{z-1} h_{WE-1} + b_z) \tag{06}$$
$$R_{WE} = f(W_{WE} x_t + U_{r-1} h_{WE-1} + b_r) \tag{07}$$
$$\hat{h}_{WE} = f(W_{WE} x_t + U_{WE-1} (R_{WE} \odot h_{WE-1}) + b_h) \tag{08}$$
$$h_{WE} = (1 - Z_{WE}) \odot h_{WE-1} + Z_{WE} \odot \hat{h}_{WE} \tag{09}$$

Here, the symbol f depicts the activation function. Thus, as an activation function Leaky ReLU uses in equations

(06-07). In the meantime, equation (08) utilizes Tanh. The rest of the parameters in equation (06-09) are identical as described in equation (01-04). In addition, multiple existing contributions from NLP highly rely on word2vec embedding, which becomes unsuccessful during the seizure of sequential relations among the words. On the other side, POS tags demonstrate their abilities while producing parse trees to map targeted terms to corresponding source words and capturing sequential relations among the combination of words. Thus this tagging method assigns tags to each word based on their contextual details. In the meantime, based on words' utilization and placement in context, a single word may possess multiple tags, which makes the task more challenging and critical. According to the relevant literature on ABSA, POS tags such as nouns and noun phrases are those parts of context primarily anticipated as potential aspect terms. In addition, verbs, adverbs, and adjectives have expected as conceivable sentiment terms [22], [58]. Therefore, the proposed approach utilizes these tags as authentic evidence for identifying and extracting potential aspect terms and their sentiments.

In the state of these motivations, the proposed methodology considers POS tag embeddings, as input, upon the second input channel to handle the context sequential information. Hence, the POS tagging process accomplishes through the Stanford tagger and a weight matrix W_{POS} utilizes for related terms. Furthermore, a one-hot-vector mechanism uses to transform these tagged sentences into n-dimensional vectors. Subsequently, the proposed approach uses the following equations (10-13) that identify and extract potential features and their sentiments. These equations (10-13) are equivalent to equations (01-04) in relations of their comprised functionality and elements, which has provided as under:

$$Z_{POS} = f(W_{POS}x_t + U_{z-1}h_{POS-1} + b_z) \tag{10}$$

$$R_{POS} = f(W_{POS}x_t + U_{r-1}h_{POS-1} + b_r) \tag{11}$$

$$\hat{h}_{POS} = f\left(W_{POS}x_t + U_{POS-1}\left(R_{POS} \odot h_{POS-1}\right) + b_h\right) \tag{12}$$

$$h_{POS} = (1 - Z_{POS}) \odot h_{POS-1} + Z_{POS} \odot \hat{h}_{POS} \tag{13}$$

Here, the symbol f depicts the activation function. Therefore, as activation functions, Leaky ReLU and Tanh use within equations (10-11) and equation (12), respectively. Leaky ReLU is a significant determinant of the inclusion and exclusion of information within the reset and update gates that reduces the impact of vanishing gradients and dying ReLU during the decision process. As a result, when calculating the final memory contents, Tanh significantly regulates the selected information of current gates and the previous state information. Hence, the proposed approach utilizes these activation functions to gain a deeper understanding of contextual information and identify potential terms. The performance enhancement of NLP's tasks illustrates the significance of utilizing contextual position information. The contextual position information represents the influence of a targeted term within a sentence through vector representation. The following equation (14) [59] illustrates these representations with the help of mathematical notation:

$$CPI_i = \lim_{j \to n} \left(\theta_j + (1 - \theta_j)\left(\frac{p_j}{n}\right)\right), 0 \leq \theta \leq 1, \tag{14}$$

CPI_i represents the vector representation of contextual position information for the dataset's i^{th} sentence in equation (14). Furthermore, the term's location-based presence ratio, θ_j, is described in the [0, 1] range. It illustrates how the place of the jth word's occurrence affects its context. Additionally, p_j indicates where a word is placed in the targeted sentence, whereas n indicates sentence length. Equation (01) determines their appropriate position-oriented influence if the j^{th} word is present in the vocabulary. If the term is not available, then zero is added. As a result, the Att-MC-GRU introduces the third and last channel comprises contextual position information embedding as position-information vectors. As a result, contextual positional information vectors are delivered onto the third channel of the MC-GRU with their respective weight matrix W_{CPI}.The equations (14-17) identify and extracts targeted features with the help of position information

$$Z_{CPI} = f(W_{CPI}x_t + U_{z-1}h_{CPI-1} + b_z) \tag{15}$$

$$R_{CPI} = f(W_{CPI}x_t + U_{r-1}h_{CPI-1} + b_r) \tag{16}$$

$$\hat{h}_{CPI} = f\left(W_{CPI}x_t + U_{CPI-1}\left(R_{CPI} \odot h_{CPI-1}\right) + b_h\right) \tag{17}$$

$$h_{CPI} = (1 - Z_{CPI}) \odot h_{CPI-1} + Z_{POS} \odot \hat{h}_{CPI} \tag{18}$$

Hence, in the above equations, the symbol f indicates the activation function. In this way, in equations (15-16), the Leaky ReLU is used as an activation function. In the meantime, equation (17) uses Tanh as an activation function. Additionally, equations (15-18) are composed of the same elements listed in equations (01-04). The attention mechanism assigns the influential score regarding each targeted word based on the interaction relations between targeted and context words [60]. Another advantage of the attention layer is the filtration of unrequired terms for improving the accuracy of identified targeted terms. The methodology of attention mechanism obtained these inspirations from the human brain, which focuses only on the words of interest while ignoring other terms while reading any text of an article. As a result, the attention mechanism transforms the whole targeted sentence or phrase into an equivalent real-valued vector. These values depict the influence score of each sentence's term based on its context. The elementwise multiplication of these scores with relevant terms indicates the model's most prominent influenced terms of the targeted sentence [61].

3) ATTENTION LAYER

Additionally, based on the context of the sentence, different words can have different influence scores. Recently, NLP methods widely used attention mechanisms to highlight prominent words and identify possible features. Based on thorough critical analysis from the existing studies in the relevant literature, attention mechanisms improve the model's performance while filtering out irrelevant terms. Moreover, we have compared attention with other machine learning based approaches in our preliminary experimentation, and attention provided better results, therefore, we consider that mechanism.

Thus, to determine the highly effective terms in the targeted sentence, the Att-MC-GRU has considered an attention mechanism. The Att-MC-GRU enhances the attention mechanism while providing syntactic, sequential, semantic, and contextual position information. These factors together enhance its ability to determine the relationship between aspects, sentiment, and contextual words. There are typically two modules in the attention mechanism named Encoders and Decoders. The encoder's function is to convert the target sentence into a real-valued, equal-length vector that contains semantic details for each term. The decoder module's purpose is to produce output following an encoded vector transformation. Equations for the attention layer are presented in (19-21) [62] as given below:

$$u_p = \tanh\left(W_p H_p + b_p\right) \tag{19}$$

$$\alpha_p = \frac{exp\left(u_p^T u_w\right)}{\sum_p exp\left(u_p^T u_w\right)} \tag{20}$$

$$H_p = \sum\nolimits_p \alpha_p H_p \tag{21}$$

The updated context vector is expressed as u_w in the equations above. Meanwhile, u_p depicts the resultant vector that comprises the hidden layer's vectors H_p, b_p, and W_p. Whereas b_p and W_p are the bias, and weight terms of the attention layer, respectively. The final term, α_p, denotes the sentence's every word attention score. Additionally, the terms h_{WE}, h_{POS}, and h_{CPI} represent the output of the input channels obtained from the equation (09),(13), and (18). After that, the Att-MC-GRU applies attention-mechanism on these outputs to identify those terms that are highly noticeable in each sentence. The attention-mechanism highlights them based on word2vec embedding, POS tags, and contextual position information. All sentences' most noteworthy features are eventually filtered out by the attention mechanism.

$$h_{Att-WE} = Attention(h_{WE}) \tag{22}$$

$$h_{Att-POS} = Attention(h_{POS}) \tag{23}$$

$$h_{Att-CPI} = Attention(h_{CPI}) \tag{24}$$

Here h_{WE}, h_{POS}, and h_{CPI} are, respectively, the outcomes of word2vec-based, POS, and contextual position information-based channels.

4) FEATURES CONCATENATION AND OUTPUT LAYER

As a final step, the output layers are concatenated into a final layer H containing semantic, contextual position, syntactic, and sequential information about the potential words. In addition, h_{Att-WE}, $h_{Att-POS}$, and $h_{Att-CPI}$ are the attention-based filtered features of h_{WE}, h_{POS}, and h_{CPI}, respectively. These three outputs are combined using equation (25).

$$H = f(h_{Att-WE}, h_{Att-POS}, h_{Att-CPI}) \tag{25}$$

In the above equation, H is the combined finalized output of preceding layers, whereas f is the function that combines the outcome of equations (22-24). Then these highlighted terms are given to the softmax function to identify and extract the targeted aspects from textual reviews.

$$Predicted\ aspects = softmax\ (W_H.H + b_H) \tag{26}$$

Here H, W_H, and b denote combined features, the term's weight, and the bias term's representation, respectively. For prediction and classification, these outcomes have been submitted to the softmax activation function. In this way, all targeted terms are classified as either aspects or non-aspects. As a consequence of manipulating equation (26) the predicted aspects are categorized as aspects or non-aspects. Accordingly, the aspect terms are admissible, but the non-aspect terms are not. In the next step, AttMC-GRU determines the polarity of the remaining terms as either positive, negative, or neutral. Thus, their corresponding sentiments are determined by combining these aspects with their contexts. The process is illustrated in equation (27).

Henceforth, the softmax obtains the filtered predicted terms, in addition to the equation (25) resulting and categorizes them as negative, positive, or neutral based on their sentiment. The mathematical notation of this procedure is given in equation (28). Fig. 2 illustrates the proposed method detailed architecture diagram.

$$S = (Predicted\ aspects, H) \tag{27}$$

$$Predicted\ sentiments = softmax(S) \tag{28}$$

IV. EXPERIMENTAL SETUP

A description of the experimental environment is provided in this section of the study. After that, the datasets used in these experiments are described. Thereafter, pre-processing steps are explained for filtering those datasets, and finally, baselines are expressed for comparison to the proposed approach. Since precision, recall, and F1 measure metrics are most extensively utilized metrics in the ABSA relevant literature, these metrics are used in this study to evaluate that approach.

The Windows 07 operating system has been utilized as the structured environment for conducting evaluations concerning the Att-MC-GRU and assessing their performance. The hardware consists of a GeForce GTX 1060 GPU, an Intel Xeon W3530 2.8 GHz CPU, and 16GB of DDR3 RAM. Additionally, the Keras 2.1.0, Python-based GPU Tensoflow 2.0, and the PyCharm IDE for Python 3.7 tool are used in the implementation software.

A. DATASET

The seven standard datasets have been used to assess the proposed model's performance. SemEval organizers make these test and training datasets publicly available. In the first dataset, we have reviews from the restaurant domain and the

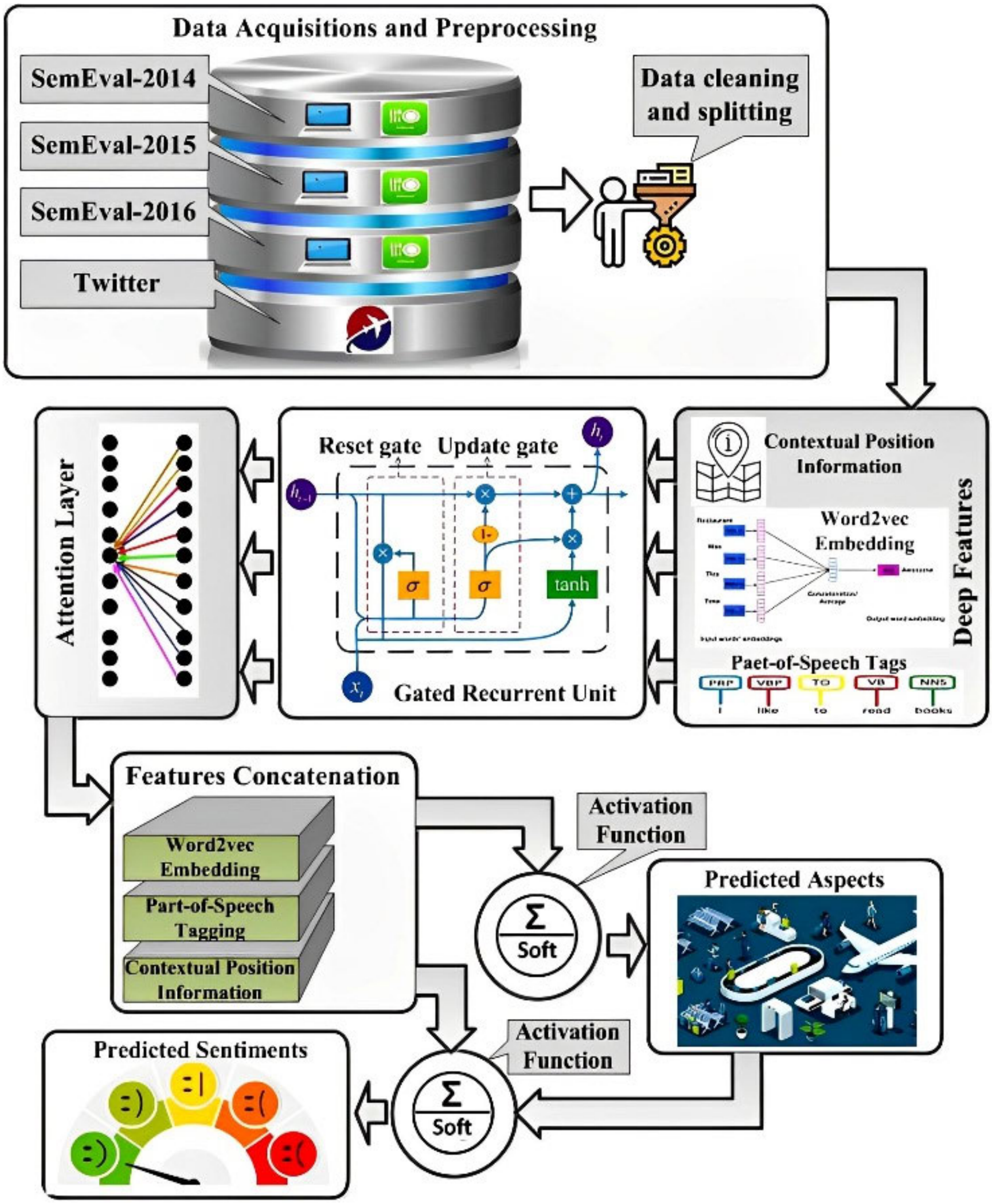


FIGURE 2. Proposed model of attention-oriented multi-channel gated recurrent neural network.

laptop domain of SemEval-14 [63] dataset. The third and fourth datasets are also related to laptops and restaurants of SemEval-15 [64] dataset.

SemEval-16 [65] has the restaurant and the laptop domains as its fifth and sixth datasets. In the last dataset, we have the Airline Twitter Sentiment dataset, which is based on a survey of leading U.S airlines. There are positive, negative, and neutral tweets among this Twitter data [66], [67]. Since 2020, Crowd Flower has been making this dataset available to the public. It was updated in the years of 2021 and 2022. Twitter is the primary source of public opinion; thus, the experimental study uses its dataset to analyze the effectiveness of the proposed method. These underlying datasets are summarized in Table 2.

B. PRE-PROCESSING

The proposed approach produces a clean and structured textual dataset by using various pre-processing steps on the textual reviews. These preprocessing steps are as follows:

- All words in English sentences are converted to lower case.
- The paragraphs of the textual reviews are broken down into sentences by the full-stop symbol.
- White spaces are used to break all sentences into tokens.
- Removed all the punctuation words from textual sentences.
- All alphanumeric words that are impure alphabets have been removed from sentences.
- All stop words within phrases have been removed.

TABLE 2. Datasets' statistics.

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

- Special characters have been removed from sentences.
- All words with a length of one character or less have been terminated.

These pre-processing techniques provide the proposed model with an easy-to-understand and usable form of data while filtering the unnecessary parts of data for determining potential terms and analyzing their associated sentiments. Meanwhile, incorporating these techniques in the initial phases of the model pipeline yields accurate results.

C. BASELINES

The effectiveness of the proposed approach assesses with the comparison to the following baseline models. We select the baseline methods that addresses ABSA's aspect extraction and sentiment classification task. While their evaluation procedure uses same dataset and metrics as we use in our methodology. Table 3 contains information regarding these baseline models in tabular form.

V. RESULTS AND DISCUSSIONS

Based on the proposed methodology, we explore how parallel word2vec embedding, sequential information of context, and position information collectively enhance the precision of identifying aspects and their SCs in experiments. The reasoning behind this is that the POS tagging process preserves the sequence order information, the relationship between words, and the role of each word within the context of targeted sentences. These facts improve the model's interpretation of contextual acquaintance. In addition, word2vec embedding conserves the targeted sentence semantic and syntactic information. Additionally, position information in context illustrates the impact of each term within its surroundings. Therefore, these factors together improve the performance of the proposed approach during experiments. Furthermore, the filtration of prominent terms of context through the attention mechanism becomes concise due to the combined effect of these novel deep features. The empirical evaluation results of F1-measure, precision, and recall metrics prove the collective impact of novel features on models' performance in the subsequent sections.

The best performance and accuracy of NLP tasks highly depend on word2vec embedding. In addition, pre-training is considered a substantial performance factor of word2vec embedding. Meanwhile, word2vec embedding achieves the best pre-training through training on massive volume datasets. Therefore, the proposed approach attains this purpose while utilizing pre-trained word2vec embedding based on the Google News dataset. In addition, POS tags provide sequential detail, whereas contextual position information indicates the impactful terms within context. Thus the combined effect of these input parameters makes this approach most robust and comprehensive. Meanwhile, incorporating Twitter and SemEval datasets from the airline, restaurant, and laptop domains, this study evaluates the significance of the Att-MC-GRU. According to our knowledge, this methodology is among the first to use MC-GRU for extracting aspects and classifying sentiments. A maximum sentence length is specified during the training phase of the model for the textual review, which assumes the review's single sentence to have the maximum possible length. There have been zero paddings added to all sentences whose conceived size is under the maximum length. In Table 4, the parameters of the proposed model are described in detail.

The Att-MC-GRU thoroughly analyzed the baseline studies' experimental arrangements. Accordingly, the Att-MC-GRU determines which parameter configuration increases or decreases performance. Therefore, the Att-MC-GRU deduces the appropriate parameter settings from a thorough analysis of the baseline approaches, as depicted in Table 3, and the parametric optimization technique. These parameter settings have illustrated in Table 4. Henceforth, a F1 measure-based comparison of the baseline approaches with the proposed method while conducting the ATE task is shown in Table 5, whereas Table 6 demonstrates their classification-based performance. As evidenced by these tabular notations, the Att-MC-GRU has indicated significant improvement in all areas of interest when evaluated through standardized

TABLE 3. The baseline methods' summary.

Baseline Approaches/ Reference	Baseline Models' Novelty	ATE / SC	Year of Publication	Dataset
WDEmb [68]	They uses a dependency-based method and the CRF to fulfil the task.	ATE	2016	SemEval-14R, 15R and 14L
MTMN [69]	They utilizes attention-based model comprised of multilayers	ATE	2017	SemEval-14R, 15R and 14L
MIN [70]	They uses dependency rules and lexicon within a model that comprised of two LSTMs.	ATE	2018	SemEval-14R, 15R and 14L
BiDTreeCRF [71]	Consider Bi-LSTM and CRF along a tree-based method.	ATE	2019	SemEval-14R, 15R and 14L
MCNN+WV+POS [14]	Uses general word2vec embedding and to encode the rich contextual Information they considers POS tagging.	ATE	2019	SemEval-14R, 15R and 14L
ESGCN [72]	Aspect and opinion pair extraction is improved by encoding the graph edges relating to syntactic dependencies.	ATE	2021	SemEval-14R, 15R and 14L
FAPN [11]	CNN-based FAPN captures the word-level relationship between a phrase and its context using phrase-aware CNNs.	SC	2021	SemEval-14L, 14R, 15R, and 16R
HPNet-S [9]	An inference framework is proposed that makes use of pre-trained languages hierarchy models.	SC	2021	SemEval-14L, 14R, 15R, and 16R
UniASTE [48]	The proposed novel approach named UniASTE comprises BiLSTM, which identifies aspects and their sentiments using the sentence contextual information.	SC	2022	SemEval-14L, 14R, 15R, and 16R
SSJE [3]	The proposed framework comprises Gated CNN, which extracts aspect and sentiment from sentences dealing with complex relationships between aspect and opinion terms in an end-to-end manner.	SC	2022	SemEval-14L, 14R, 15R, and 16R
CPA-SAA [47]	Determines the potential terms relative weights of according to their position and relies on inferences among both sides of conceivable terms.	SC	2022	SemEval-14L, 14R, 15R, and 16R

TABLE 4. Parametric setting for proposed method.

Proposed Model parameters	Value
Word Embedding Vector's Dimensions	300
POS Embedding Vector's Dimensions	45
Size of GRU Layer	64,128,256
Recurrent Activation Function in GRU Layer	Leaky ReLU
Size of Batch	50
Epochs	50
Optimizer	Adam
Optimizer's Learning Rate	0.001
Function for Loss	Categorical Cross entropy
Function regarding Prediction	Softmax
Dropout	0.5

datasets compared with the acquisition of baseline approaches.

A. ASPECT TERM EXTRACTION PERFORMANCE COMPARISON

The Att-MC-GRU expressed a significant performance increase in the F1 measure on the SemEval-14L domain dataset. Therefore, it achieves a reasonable margin of success from all baseline approaches in F1 score performance. Such as 25.31% from ESGCN, 13.37% from MCNN+WV2+POS, 14.16% from MCNN+WV+POS, 13.43% from BiDTreeCRF, 16.42% from MIN, 24.86% from MTMN, and 18.85% from WDEmb. This performance improvement depicts graphically in Fig. 3(a).

The evaluation process of Att-MC-GRU, instead of restricting to a single domain, comprises different domain datasets. Thus the evaluation process estimates the accurate performance of the proposed model in various domains and scenarios. In these circumstances, the Att-MC-GRU evaluates on SemEval-14R domain dataset against the mentioned baselines. As a result, the evaluation mechanism observes remarkable performance in the F1 measure score. Such as it achieves 14.78% from ESGCN, 6.31% from MCNN+WV+POS, 4.11% from MCNN+WV2+POS, 6.17% from BiDTreeCRF, 5.71% from MIN, 17.74% from MTMN, and 6.03% from WDEmb. The evaluation of the Att-MC-GRU demonstrates its better performance on the SemEval-14R domain dataset compared to the baseline approaches. Its performance enhancement can observe in Fig. 3(b).

The evaluation process of Att-MC-GRU does not depend on a single dataset. Therefore, SemEval-15R's domain dataset utilizes to evaluate that approach. During the evaluation, the Att-MC-GRU expresses a significant performance

TABLE 5. The proposed approach performance comparison using F1 measure to the baseline approaches while extracting aspect terms.

Model	SemEval-14(Laptop)	SemEval-14(Restaurant)	SemEval-15(Restaurant)
WDEmb	75.15	84.97	69.73
MTMN	69.14	73.26	71.31
MIN	77.58	85.29	70.73
BiDTreeCRF	80.57	84.83	70.83
MCNN+WV+POS	79.84	84.69	72.84
MCNN+WV2+POS	80.63	86.89	72.65
ESGCN	68.69	76.22	68.34
Att-MC-GRU (Proposed Model)	94.00	91.00	92.00

TABLE 6. The proposed approach performance comparison using f1 measure to the baseline approaches while classifying sentiment polarities.

Model	SemEval-14(Laptop)	SemEval-14(Restaurant)	SemEval-15(Restaurant)	SemEval-16(Restaurant)
UniASTE	47.11	62.73	54.73	62.58
SSJE	60.41	72.26	65.05	71.38
FAPN	72.91	73.07	63.07	73.03
HPNet-S	68.94	78.87	78.85	78.84
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-MC-GRU (Proposed Model)	86.00	87.00	93.00	90.00

improvement against the baseline approaches in the F1 measure score. Hence the Att-MC-GRU attains a maximum F1 measure score of 92%, compared to baselines. In this way, the Att-MC-GRU obtains a notable edge on baselines. Such as 23.66% from ESGCN, 19.16% from MCNN+WV+POS, 19.35% from MCNN+WV2+POS, 21.17% from BiDTreeCRF, 21.27% from MIN, 20.69% from MTMN, and 22.27% from WDEmb. This performance improvement depicts in Fig. 3(c).

B. PERFORMANCE COMPARISON WHILE CLASSIFYING SENTIMENTS

The Att-MC-GRU achieves a notable performance of 86% in the F1 measure score while classifying the identified aspects of the SemEval-14L dataset. It achieves a significant margin of success in F1 compared to all benchmarked approaches. Accordingly, its performance gain is 15.13%, 14.5%, 17.06%, 16.89%, 13.09%, 25.59%, and 38.89% compared to CPA-SAF, CPA-SAA, HPNet-S, HPNet-M, FAPN, SSJE, and UniASTE, respectively. This strong performance is due to the Att-MC-GRU's ability to model the long-term dependencies between words in a sentence and capture the global semantic representation of the sentence. This helps the model to identify the key aspects of a sentence and accurately classify them. Fig. 4(a) shows these performance improvements.

Meanwhile, the Att-MC-GRU gained a significant F1 measure score of 87% in categorizing the aspects identified in the SemEval-14R dataset. Thus, it achieved 14.19%, 13.62%, 8.05%, 13.93%, 14.74%, and 24.27% compared to CPA-SAF, CPA-SAA, HPNet-S, HPNet-M, FAPN, SSJE, and UniASTE, respectively.

Fig. 4(b) provides a pictorial representation of these performance improvements over the baseline approaches.

In addition, when assessing the Att-MC-GRU performance over the prescribed baselines based on the SemEval-15R dataset, a 93% value of the F1 measure score has been obtained. The Att-MC-GRU improves its performance over CPA-SAF, CPA-SAA, HPNet-S, HPNet-M, FAPN, SSJE, and UniASTE by 32.74, 32.85, 14.15, 14.06, 29.03, 27.95, and 38.27, respectively, when evaluated against the baselines. The Att-MC-GRU performs better than all the baselines considered, which can be analyzed in Figure 4(c). Moreover, the Att-MC-GRU obtained a 90% enhancement in the SemEval-16R dataset in the F1 measure score. Hence, it attains improvements of 18.53%, 17.57%, 11.07%, 16.97%, 11.16%, 18.62%, and 27.42% compared to CPA-SAF, CPA-SAA, HPNet-S, HPNet-M, FAPN, SSJE, and UniASTE, respectively. The inclusion of sequential facts, contextual position information, and semantic and syntactic information collectively for the identification of potential terms and classifying their sentiments enhances the precision of the

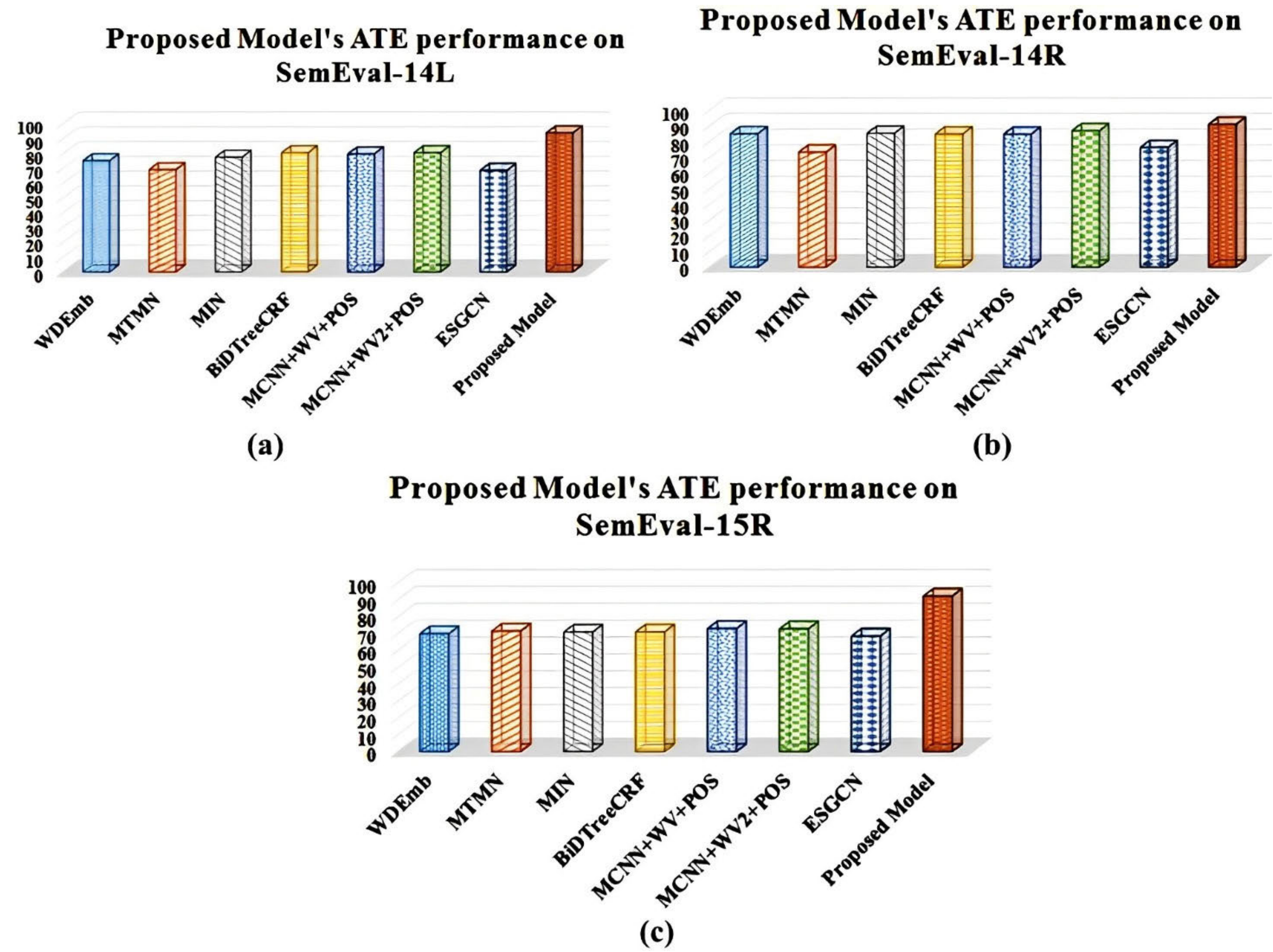


FIGURE 3. Model's ATE performance compression with baseline approaches (a) Using SemEval-14L (b) Using SemEval-14R (c) Using SemEval-15R.

attention mechanism. This means that by combining all of the different types of information, such as the context and meaning of the text, the attention mechanism is able to more accurately identify the terms and their sentiment, which leads to a better overall understanding of the text. As shown in Fig. 4(d), this performance gain is visible graphically.

C. PERFORMANCE COMPARISON OF PROPOSED APPROACH BASED ON OTHER METRICS

In the meantime, the evaluation process of Att-MC-GRU is not limited to a single metric. It includes various metrics along different domains of standard datasets. Therefore, the evaluation process analyzes the training performance of the Att-MC-GRU using precision, recall, and F1 measures. Thus, Att-MC-GRU's ATE task obtains a maximum precision of 95% on the SemEval-14L dataset, and the minimum precision on the SemEval-16L dataset is 92%. On the other side, the task of SC obtains maximum precision on the SemEval-15R domain, which is 90%, whereas the minimum precision is 80% on the SemEval-16L domain dataset. Fig. 5(a) represents precision-based observations of Att-MC-GRU experiments on different domain datasets. However, while analyzing the Att-MC-GRU training performance through the recall metric, the ATE task gains the highest recall value of 91% on the SemEval-14L dataset, and the minimal recall on the Twitter Data-US Airline domain dataset is 85%. In the meantime, SC's maximum recall value is 94% on the dataset SemEval-15R, whereas the recall minimum value is 85% on the dataset SemEval-15L. Recall-based experimental observations of Att-MC-GRU on datasets from diverse domains have shown in Fig. 5(b). During the evaluation, the performance of Att-MC-GRU is also observed through the F1 measure while performing ATE and SC tasks. Thus, the task of ATE obtains the F1 measure's maximum value of 93% on the SemEval-14L dataset and its minimum value of 88% on the Twitter Airline dataset. On the other side, the SC task gains a maximum F1 measure score on SemEval-15R's domain dataset, which is 92%. In the meantime, its minimum F1 value on the datasets of SemEval-16L and SemEval-15L is 84%. Fig. 5(c) depicts the performance of Att-MC-GRU's F1 measure score analyzed through experiments on different domains of datasets.

Consequently, the evaluation process analyzes the Att-MC-GRU's testing performance in terms of precision on

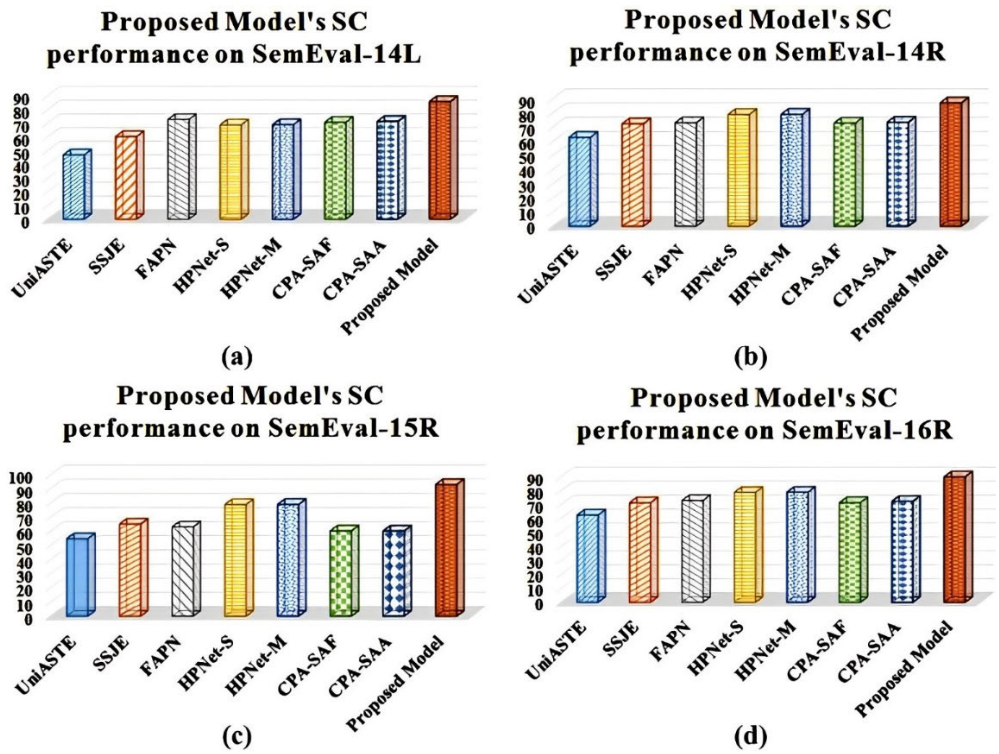


FIGURE 4. Compression of Att-MC-GRU's SC performance and baseline approaches (a) Using SemEval-14L (b) Using SemEval-14R (c) Using SemEval-15R (d) Using SemEval-16R.

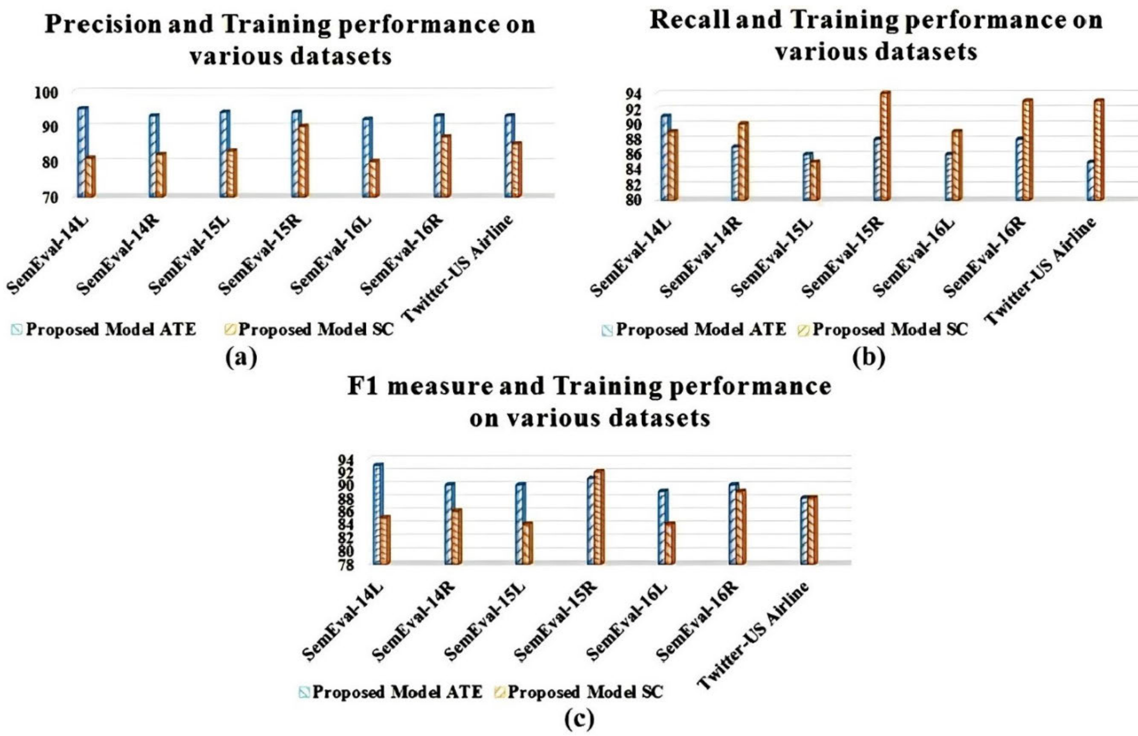


FIGURE 5. Model's training performance on variant datasets(a) Precision-based experimental observations (b) Recall-based experimental observations (c) F1 measure-based experimental observations.

standard datasets. In this manner, the maximum precision gain of ATE's task observed in the dataset of SemEval-14L is 96%. Meanwhile, the minimum discovered precision on the dataset of SemEval-16L is 93%. Subsequently, the task of SC obtained a maximum precision value of 91% on the dataset of SemEval-15R. On the dataset of SemEval-16L's,

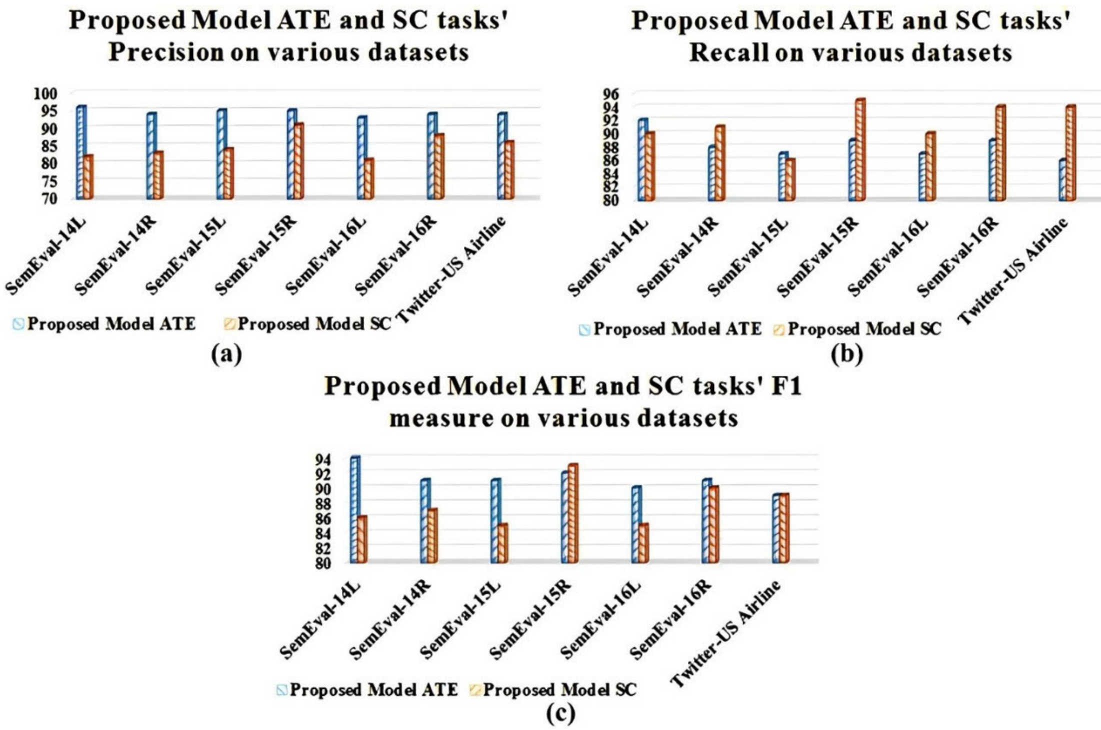


FIGURE 6. Model's performance on variant datasets (a) Precision-based experimental observations (b) Recall-based experimental observations (c) F1 measure-based experimental observations.

its observed minimum precision value is 81%. On different datasets, Fig. 6(a) shows precision-based observations of Att-MC-GRU experiments.

Additionally, the performance of Att-MC-GRU is analyzed through another metric named recall on different datasets comprising various domains. Thus, the maximum recall score of ATE's task discovered in the dataset of SemEval-14L is 92%. Meanwhile, the minimum recall observed on the dataset of Twitter Data-US Airlines is 86%. On the flip side, the task of SC achieved a maximum recall value of 95% on the dataset of SemEval-15R. Meanwhile, a minimum recall value of 86% is observed on the dataset of SemEval-15L. The recall-based experimental observations on diverse domains' datasets regarding the Att-MC-GRU are depicted in Fig. 6(b). The performance of Att-MC-GRU is also explored through the F1 measure score on different datasets containing different domains. Thus, the maximum F1 measure score of the ATE task discovered in the dataset of SemEval-14L is 94%. Meanwhile, the minimum F1 measure value observed in the Twitter dataset of US Airlines is 89%. On the domain dataset of SemEval-15R, SC's task achieved a maximum F1 score of 93%. Meanwhile, the Att-MC-GRU's minimum F1 measure value on the datasets of SemEval-16L and SemEval-15L is 85%. A summary of the performance of Att-MC-GRU's F1 measure score in all other datasets has presented in Fig. 6(c).

In addition, Att-MC-GRU uses a parallel combination of word embedding, POS tags, and contextual position information. Thus, this study also analyzes the performance of each input channel separately and compares their effectiveness with multi-channel GRU performance to discriminate the advantages of the parallel fusion of various input novel deep features. Furthermore, the experiments demonstrate that multi-channel GRU outperforms single-feature utilization. On the other hand, the hypothesis that contextual position information improves word embedding and POS tag performance also validates itself through experiments. Furthermore, the parallel use of novel deep features with GRU shows the superiority of parallel-combined inputs when dealing with multimodal data. Experiments show that the parallel combination of contextual position information, word embedding, and POS tags strengthens multi-channel GRU. In addition, it examines that using multiple deep features produces better results than single-feature utilization when dealing with multimodal data. Based on Fig. 7, we observed that the parallel fusion of novel deep features with GRU performs better than using a single channel or relying on a single input feature. The reason is that the parallel use of input parameters combines the advantages of each input and uses their strengths to capture multimodal data more accurately than either one alone.

The Precision-Recall (PR) curve, a commonly used visual tool, is used to assess how well proposed approaches perform in terms of their ability to distinguish between two classes. In order to measure how well the attention-based multi-channel gated recurrent neural network distinguishes between several categories, it utilizes the PR curve. The area under the curve shows the attention-based multi-channel gated recurrent neural network's substantial performance when categorizing multi-class classification, as shown in Fig. 8.

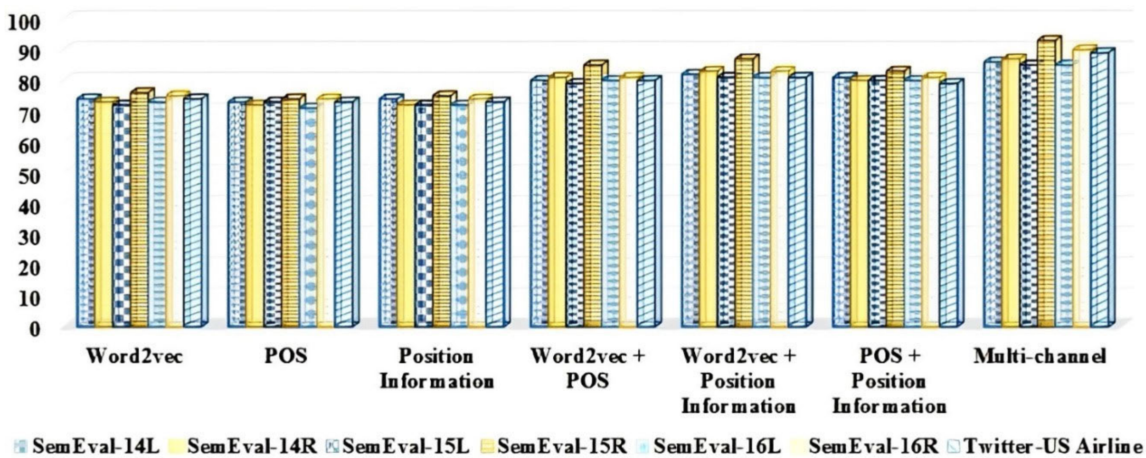


FIGURE 7. Impact of single and multi-channels on proposed model.

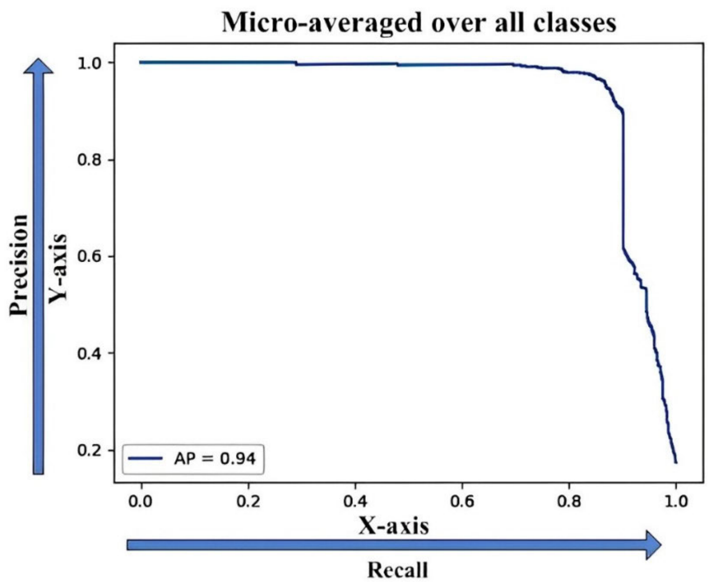


FIGURE 8. Model's precision-recall curve.

VI. CONCLUSION

The proposed method considers an attention-based multi-channel gated recurrent neural network while extracting targeted aspects from textual reviews and classifying their associated sentiments. The utilization of contextual position information and consideration of word embedding and POS tags contribute to the remarkable capabilities of the proposed methodology. A distinctive feature of the Att-MC-GRU is that it identifies aspects rather than obtaining them explicitly. Using these identified aspects, the proposed method predicts their sentiment based on the context in which they are encountered. Hence, we summarize the overall performance of Att-MC-GRU on different domains of standardized datasets during ATE and SC tasks from a variety of metrics perspectives, including F1, precision, and recall. Based on the results from all datasets, the framework performs significantly better than baseline approaches. According to this illustration, the Att-MC-GRU achieves a better precision value when it performs the task of ATE, while the SC task gains a higher recall value. According to the F1 measure, their performance is mixed. Multiple input channels for the GRU algorithm have been the reason for the approach's promising performance. Furthermore, word embedding, POS tags, and positional information maximize accuracy in filtering targeted aspects and their sentiments by collaborating with attention mechanisms. This illustrates the efficacy of deep feature collaboration and the significance of multi-channel inputs in GRUs in empirical quantitative studies. Further work on implicit polarity regarding targeted aspects will be undertaken in the future.

REFERENCES

[1] T. Chen, P. Samaranayake, X. Cen, M. Qi, and Y.-C. Lan, "The impact of online reviews on consumers' purchasing decisions: Evidence from an eye-tracking study," *Frontiers Psychol.*, vol. 13, p. 2723, Jun. 2022.

[2] M. L. Yadav and B. Roychoudhury, "Effectiveness of domain-based lexicons vis-à-vis general lexicon for aspect-level sentiment analysis: A comparative analysis," *J. Inf. Knowl. Manage.*, vol. 18, no. 3, Sep. 2019, Art. no. 1950033.

[3] Y. Li, Y. Lin, Y. Lin, L. Chang, and H. Zhang, "A span-sharing joint extraction framework for harvesting aspect sentiment triplets," *Knowl.-Based Syst.*, vol. 242, Apr. 2022, Art. no. 108366.
[4] A. Ishaq, S. Asghar, and S. A. Gillani, "Aspect-based sentiment analysis using a hybridized approach based on CNN and GA," *IEEE Access*, vol. 8, pp. 135499–135512, 2020.
[5] P. Karagoz, B. Kama, M. Ozturk, I. H. Toroslu, and D. Canturk, "A framework for aspect based sentiment analysis on Turkish informal texts," *J. Intell. Inf. Syst.*, vol. 53, no. 3, pp. 431–451, Dec. 2019.
[6] I. P. Matsuno, R. G. Rossi, R. M. Marcacini, and S. O. Rezende, "Aspect-based sentiment analysis using semi-supervised learning in bipartite heterogeneous networks," *J. Inf. Data Manag.*, vol. 7, no. 2, pp. 141–154, 2016.
[7] T. A. Rana and Y.-N. Cheah, "A two-fold rule-based model for aspect extraction," *Expert Syst. Appl.*, vol. 89, pp. 273–285, Dec. 2017.
[8] R. M. Marcacini, R. G. Rossi, I. P. Matsuno, and S. O. Rezende, "Cross-domain aspect extraction for sentiment analysis: A transductive learning approach," *Decis. Support Syst.*, vol. 114, pp. 70–80, Oct. 2018.
[9] D. Xiao, F. Ren, X. Pang, M. Cai, Q. Wang, M. He, J. Peng, and H. Fu, "A hierarchical and parallel framework for end-to-end aspect-based sentiment analysis," *Neurocomputing*, vol. 465, pp. 549–560, Nov. 2021.
[10] L. Shu, H. Xu, and B. Liu, "Lifelong learning CRF for supervised aspect extraction," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, vol. 2, 2017, pp. 148–154.
[11] W. Liao, J. Zhou, Y. Wang, Y. Yin, and X. Zhang, "Fine-grained attention-based phrase-aware network for aspect-level sentiment analysis," *Artif. Intell. Rev.*, vol. 55, no. 5, pp. 1–20, 2021.
[12] P. Barnaghi, G. Kontonatsios, N. Bessis, and Y. Korkontzelos, "Aspect extraction from reviews using convolutional neural networks and embeddings," in *Proc. Int. Conf. Appl. Natural Lang. Inf. Syst.*, 2019, pp. 409–415.
[13] D.-H. Pham and A.-C. Le, "Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis," *Int. J. Approx. Reasoning*, vol. 103, pp. 1–10, Dec. 2018.
[14] A. Da'u and N. Salim, "Aspect extraction on user textual reviews using multi-channel convolutional neural network," *PeerJ Comput. Sci.*, vol. 5, p. e191, May 2019.
[15] X. Li, L. Bing, P. Li, W. Lam, and Z. Yang, "Aspect term extraction with history attention and selective transformation," in *Proc. 27th Int. Joint Conf. Artif. Intell. (IJCAI)*, Jul. 2018, pp. 4194–4200.
[16] M. Ahmed, Q. Chen, Y. Wang, and Z. Li, "Hint-embedding attention-based LSTM for aspect identification sentiment analysis," in *Proc. Pacific Rim Int. Conf. Artif. Intell.*, 2019, pp. 569–581.
[17] J. Yu, J. Jiang, and R. Xia, "Global inference for aspect and opinion terms co-extraction based on multi-task neural networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 1, pp. 168–177, Jan. 2019.
[18] C. Wu, F. Wu, S. Wu, Z. Yuan, and Y. Huang, "A hybrid unsupervised method for aspect term and opinion target extraction," *Knowl.-Based Syst.*, vol. 148, pp. 66–73, May 2018.
[19] V. Gupta, V. K. Singh, P. Mukhija, and U. Ghose, "Aspect-based sentiment analysis of mobile reviews," *J. Intell. Fuzzy Syst.*, vol. 36, no. 5, pp. 4721–4730, May 2019.
[20] S. Y. Ho, K. W. S. Choi, and F. F. Yang, "Harnessing aspect-based sentiment analysis: How are tweets associated with forecast accuracy?" *J. Assoc. Inf. Syst.*, vol. 20, no. 8, p. 2, 2019.
[21] M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh, and O. Qawasmeh, "Enhancing aspect-based sentiment analysis of Arabic Hotels' reviews using morphological, syntactic and semantic features," *Inf. Process. Manage.*, vol. 56, no. 2, pp. 308–319, Mar. 2019.
[22] S. Rida-E-Fatima, A. Javed, A. Banjar, A. Irtaza, H. Dawood, H. Dawood, and A. Alamri, "A multi-layer dual attention deep learning model with refined word embeddings for aspect-based sentiment analysis," *IEEE Access*, vol. 7, pp. 114795–114807, 2019.
[23] N. Liu, B. Shen, Z. Zhang, Z. Zhang, and K. Mi, "Attention-based sentiment reasoner for aspect-based sentiment analysis," *Hum.-Centric Comput. Inf. Sci.*, vol. 9, no. 1, p. 35, Dec. 2019.
[24] D. Zeng, Y. Dai, F. Li, J. Wang, and A. K. Sangaiah, "Aspect based sentiment analysis by a linguistically regularized CNN with gated mechanism," *J. Intell. Fuzzy Syst.*, vol. 36, no. 5, pp. 3971–3980, May 2019.
[25] M. Al-Smadi, B. Talafha, M. Al-Ayyoub, and Y. Jararweh, "Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 8, pp. 2163–2175, Aug. 2019.
[26] N. Jihan, Y. Senarath, D. Tennekoon, M. Wickramarathne, and S. Ranathunga, "Multi-domain aspect extraction using support vector machines," in *Proc. 29th Conf. Comput. Linguistics Speech Process. (ROCLING)*, 2017, pp. 308–322.
[27] R. Hegde and S. Seema, "Aspect based feature extraction and sentiment classification of review data sets using incremental machine learning algorithm," in *Proc. 3rd Int. Conf. Adv. Electr., Electron., Inf., Commun. Bio-Inform. (AEEICB)*, Feb. 2017, pp. 122–125.
[28] Z. Yan, M. Xing, D. Zhang, and B. Ma, "EXPRS: An extended pagerank method for product feature extraction from online consumer reviews," *Inf. Manage.*, vol. 52, no. 7, pp. 850–858, Nov. 2015.
[29] A. K. Samha, Y. Li, and J. Zhang, "Aspect-based opinion extraction from customer reviews," *Comput. Sci. Inf. Technol.*, vol. 4, no. 4, pp. 149–160, 2014.
[30] Y. Xiang, H. He, and J. Zheng, "Aspect term extraction based on MFE-CRF," *Information*, vol. 9, no. 8, p. 198, Aug. 2018.
[31] C. Liao, C. Feng, S. Yang, and H.-Y. Huang, "A hybrid method of domain lexicon construction for opinion targets extraction using syntax and semantics," *J. Comput. Sci. Technol.*, vol. 31, no. 3, pp. 595–603, May 2016.
[32] M. E. Mowlaei, M. S. Abadeh, and H. Keshavarz, "Aspect-based sentiment analysis using adaptive aspect-based lexicons," *Expert Syst. Appl.*, vol. 148, Jun. 2020, Art. no. 113234.
[33] M. S. Wai and S. S. Aung, "Simultaneous opinion lexicon expansion and product feature extraction," in *Proc. IEEE/ACIS 16th Int. Conf. Comput. Inf. Sci. (ICIS)*, May 2017, pp. 107–112.
[34] F. Xianghua, L. Guo, G. Yanyan, and W. Zhiqiang, "Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon," *Knowl.-Based Syst.*, vol. 37, pp. 186–195, Jan. 2013.
[35] Q. Liu, Z. Gao, B. Liu, and Y. Zhang, "Automated rule selection for opinion target extraction," *Knowl.-Based Syst.*, vol. 104, pp. 74–88, Jul. 2016.
[36] Y. Kang and L. Zhou, "RubE: Rule-based methods for extracting product features from online consumer reviews," *Inf. Manage.*, vol. 54, no. 2, pp. 166–176, Mar. 2017.
[37] M. Z. Asghar, A. Khan, S. R. Zahra, S. Ahmad, and F. M. Kundi, "Aspect-based opinion mining framework using heuristic patterns," *Cluster Comput.*, vol. 22, no. S3, pp. 7181–7199, May 2019.
[38] A. Kushwaha and S. Chaudhary, "Review highlights: Opinion mining on reviews: A hybrid model for rule selection in aspect extraction," in *Proc. 1st Int. Conf. Internet Things Mach. Learn.*, Oct. 2017, pp. 1–6.
[39] Q. Liu, Z. Gao, B. Liu, and Y. Zhang, "Automated rule selection for aspect extraction in opinion mining," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 1291–1297.
[40] T. A. Rana and Y.-N. Cheah, "Hybrid rule-based approach for aspect extraction and categorization from customer reviews," in *Proc. 9th Int. Conf. IT Asia (CITA)*, Aug. 2015, pp. 1–5.
[41] F. Z. Ruskanda, D. H. Widyantoro, and A. Purwarianti, "Comparative study on language rule based methods for aspect extraction in sentiment analysis," in *Proc. Int. Conf. Asian Lang. Process. (IALP)*, Nov. 2018, pp. 56–61.
[42] M. Shams and A. Baraani-Dastjerdi, "Enriched LDA (ELDA): Combination of latent Dirichlet allocation with word co-occurrence analysis for aspect extraction," *Expert Syst. Appl.*, vol. 80, pp. 136–146, Sep. 2017.
[43] S. J. Das and B. Chakraborty, "An approach for automatic aspect extraction by latent Dirichlet allocation," in *Proc. IEEE 10th Int. Conf. Awareness Sci. Technol. (iCAST)*, Oct. 2019, pp. 1–6.
[44] C. Wan, Y. Peng, K. Xiao, X. Liu, T. Jiang, and D. Liu, "An association-constrained LDA model for joint extraction of product aspects and opinions," *Inf. Sci.*, vol. 519, pp. 243–259, May 2020.
[45] G. Ansari, C. Saxena, T. Ahmad, and M. N. Doja, "Aspect term extraction using graph-based semi-supervised learning," *Proc. Comput. Sci.*, vol. 167, pp. 2080–2090, Jan. 2020.
[46] S. W. Jeon, H. J. Lee, H. Lee, and S. Cho, "Graph based aspect extraction and rating classification of customer review data," in *Proc. Int. Conf. Database Syst. Adv. Appl.*, 2019, pp. 186–199.
[47] B. Huang, R. Guo, Y. Zhu, Z. Fang, G. Zeng, J. Liu, Y. Wang, H. Fujita, and Z. Shi, "Aspect-level sentiment analysis with aspect-specific context position information," *Knowl.-Based Syst.*, vol. 243, May 2022, Art. no. 108473.
[48] F. Chen, Z. Yang, and Y. Huang, "A multi-task learning framework for end-to-end aspect sentiment triplet extraction," *Neurocomputing*, vol. 479, pp. 12–21, Mar. 2022.

[49] T. Gu, G. Xu, and J. Luo, "Sentiment analysis via deep multichannel neural networks with variational information bottleneck," *IEEE Access*, vol. 8, pp. 121014–121021, 2020.

[50] B. Kanekar and A. Godbole, "Multi-channel CNN-BiGRU neural network for sentiment analysis," in *Machine Intelligence and Smart Systems*. Singapore: Springer, 2021, pp. 527–536.

[51] Y. Cheng, L. Yao, G. Xiang, G. Zhang, T. Tang, and L. Zhong, "Text sentiment orientation analysis based on multi-channel CNN and bidirectional GRU with attention mechanism," *IEEE Access*, vol. 8, pp. 134964–134975, 2020.

[52] A. Kumar, A. S. Veerubhotla, V. T. Narapareddy, V. Aruru, L. B. M. Neti, and A. Malapati, "Aspect term extraction for opinion mining using a hierarchical self-attention network," *Neurocomputing*, vol. 465, pp. 195–204, Nov. 2021.

[53] G. S. Chauhan, Y. K. Meena, D. Gopalani, and R. Nahta, "A two-step hybrid unsupervised model with attention mechanism for aspect extraction," *Expert Syst. Appl.*, vol. 161, Dec. 2020, Art. no. 113673.

[54] D.-G. Shao, M.-F. Zhang, Y. Xiang, R. Hu, and T. Lu, "Non-local attention based CNN model for aspect extraction," *J. Inf. Sci. Eng.*, vol. 37, no. 6, pp. 1419–1433, 2021.

[55] D. Gu, J. Wang, S. Cai, C. Yang, Z. Song, H. Zhao, L. Xiao, and H. Wang, "Targeted aspect-based multimodal sentiment analysis: An attention capsule extraction and multi-head fusion network," *IEEE Access*, vol. 9, pp. 157329–157336, 2021.

[56] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. NIPS Workshop Deep Learn.*, 2014, pp. 1–9.

[57] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," in *Proc. 8th Workshop Syntax, Semantics Struct. Stat. Transl. (SSST)*, 2014, pp. 103–111.

[58] M. S. Akhtar, D. Gupta, A. Ekbal, and P. Bhattacharyya, "Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis," *Knowl.-Based Syst.*, vol. 125, pp. 116–135, Jun. 2017.

[59] Y. Jin, H. Zhang, and D. Du, "Incorporating positional information into deep belief networks for sentiment classification," in *Proc. Ind. Conf. Data Mining*, 2017, pp. 1–15.

[60] X. Wang, G. Xu, J. Zhang, X. Sun, L. Wang, and T. Huang, "Syntax-directed hybrid attention network for aspect-level sentiment analysis," *IEEE Access*, vol. 7, pp. 5014–5025, 2019.

[61] Y. Zhu, X. Gao, W. Zhang, S. Liu, and Y. Zhang, "A bi-directional LSTM-CNN model with attention for aspect-level text classification," *Future Internet*, vol. 10, no. 12, p. 116, Nov. 2018.

[62] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2016, pp. 1480–1489.

[63] D. Kirange, R. R. Deshmukh, and M. Kirange, "Aspect based sentiment analysis SemEval-2014 task 4," *Asian J. Comput. Sci. Inf. Technol*, vol. 4, no. 8, pp. 72–75, 2014.

[64] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "SemEval-2015 task 12: Aspect based sentiment analysis," in *Proc. 9th Int. Workshop Semantic Eval. (SemEval)*, 2015, pp. 486–495.

[65] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, and G. Eryigit, "SemEval-2016 task 5: Aspect based sentiment analysis," in *Proc. Int. Workshop Semantic Eval.*, 2016, pp. 19–30.

[66] R. Xiang, E. Chersoni, Q. Lu, C. Huang, W. Li, and Y. Long, "Lexical data augmentation for sentiment analysis," *J. Assoc. Inf. Sci. Technol.*, vol. 72, no. 11, pp. 1432–1447, Nov. 2021.

[67] P. V. Tijare and J. R. Prathuri, "Correlation between K-means clustering and topic modeling methods on Twitter datasets," in *Cyber Security and Digital Forensics*. Singapore: Springer, 2022, pp. 459–477.

[68] Y. Yin, F. Wei, L. Dong, K. Xu, M. Zhang, and M. Zhou, "Unsupervised word and dependency path embeddings for aspect term extraction," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2016, pp. 2979–2985, 2016.

[69] W. Wang, S. J. Pan, and D. Dahlmeier, "Multi-task memory networks for category-specific aspect and opinion terms co-extraction," in *Proc. Assoc. Adv. Artif. Intell.*, 2017, pp. 3316–3322.

[70] X. Li and W. Lam, "Deep multi-task learning for aspect term extraction with memory interaction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 2886–2892.

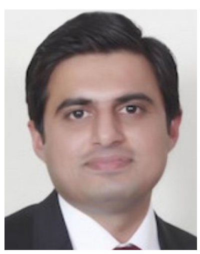
[71] H. Luo, T. Li, B. Liu, B. Wang, and H. Unger, "Improving aspect term extraction with bidirectional dependency tree representation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 7, pp. 1201–1212, Jul. 2019.

[72] S. Wu, H. Fei, Y. Ren, B. Li, F. Li, and D. Ji, "High-order pair-wise aspect and opinion terms extraction with edge-enhanced syntactic graph convolution," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 2396–2406, 2021.

WAQAS AHMAD received the master's degree in computer science from the Department of Computer Science, COMSATS University Islamabad, Attock Campus, Pakistan. He is currently a Ph.D. Scholar with the Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah Cantt, Pakistan. He has published several research works in top journals and conferences. His research interests include data mining, machine learning, and sentiment analysis.

HIKMAT ULLAH KHAN received the master's and Ph.D. degrees in computer science from International Islamic University, Islamabad. He has been an active researcher for the last ten years. He was an Associate Professor and the Head of the Data Science Research Group, Department of Computer Science, COMSATS University Islamabad, Wah Cantt, Pakistan, for about 15 years. He is currently a Professor with the Department of Computer Science, Namal University Mianwali, Pakistan. He has authored several research articles in top peer-reviewed journals and international conferences. His research interests include social web mining, semantic web, data science, information retrieval, and scientometrics.

TASSAWAR IQBAL received the master' (Hons.) and M.S. (CS) degrees from COMSATS University Abbottabad Campus, in 2007, and the Ph.D. degree from the Vienna University of Technology, in 2012. He is currently an Associate Professor with the Department of Computer Science, COMSATS University Islamabad, Wah Campus, Pakistan. He has several articles published in reputable journals and conferences. His current research interests include the studying impact of technologies in society, human–computer interaction, social network analysis, and scholarly network analysis.



SAQIB IQBAL received the M.Sc. degree in software engineering from the Queen Mary University of London, U.K., in 2007, and the Ph.D. degree in software engineering from the University of Huddersfield, in 2013. He has taught in various postgraduate universities for over ten years and has worked in the industry as a software engineer for more than three years. He is actively involved in research in areas particularly related to requirements engineering, software design, model-transformation, and software testing.