ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2025.0518

AN IMPROVED GRU BASED ON RECURRENT ATTENTION UNIT AND SELF-ATTENTION TECHNIQUE FOR TEXT SENTIMENT ANALYSIS

Dhurgham Ali Mohammed[1] and Kalyani A. Patel[2]

[1]Department of Computer Science, Faculty of Education for Girl, University of Kufa, Iraq
[2]Department of Master of Science (CA&IT), K. S. School of Business Management and Information Technology, Gujarat University, India

Abstract

In text sentiment analysis, a crucial challenge is that conventional word vectors fail to capture lexical ambiguity. The Gated Recurrent Unit (GRU), an advanced variant of RNN, is extensively utilized in natural language processing tasks such as information filtering, sentiment analysis, machine translation, and speech recognition. GRU can retain sequential information, but it lacks the ability to focus on the most relevant features of a sequence. Therefore, this paper introduces a novel text sentiment analysis-based RNN approach, a Recurrent Attention Unit (RAU), which incorporates an attention gate directly within the traditional GRU cell. This addition enhances GRU's capacity to retain long-term information and selectively concentrates on critical elements in sequential data. Furthermore, this study integrates an improved Self-Attention technique (SA) with RA-GRU known as SA+RA-GRU. The improved self-attention technique is executed to reallocate the weights of deep text sequences. While attention techniques have recently become a significant innovation in deep learning, their precise impact on sentiment analysis has yet to be fully evaluated. The experimental findings show that the proposed approach SA+RA-GRU attains an accuracy of 92.17%, and 82.38% on the IMDB, and MR datasets, and outperformed traditional approaches. Moreover, the SA+RA-GRU model demonstrates excellent generalization and robust performance.

Keywords:

Sentiment Analysis, RNNs, GRU, Recurrent Attention Unit, Self-Attention Mechanism, Deep Learning

1. INTRODUCTION

Sentiment analysis, a subfield of NLP, emphasizes extracting and managing the expressive tone, opinion, or sentiment emotion in textual data. Its applications range from opinion mining from mainstream media and social media platforms to providing customer feedback analysis. The rapid growth of internet applications, including e-commerce and social media platforms, has significantly accelerated the generation and distribution of text data [1]. In this context, the issue of proficiently extracting emotional information from vast amounts of textual data and automating its analysis has emerged as a critical research area. Conventional approaches, which primarily depend on manually crafted features and statistical machine learning techniques, often fall short in feature representation and generalization capabilities [2].

Deep learning algorithms have revolutionized advanced sentiment analysis in recent years by leveraging complex procedures to extract complicated relations within textual information. These algorithms have demonstrated superior performance over typical machine learning methods, particularly in managing large-scale datasets and improving sentiment prediction accuracy [3].

Deep learning approaches, such as RNNs, LSTM, and GRU, have been extensively employed in sentiment analysis tasks to capture sequential dependencies in text. However, the LSTM approach has limitations in capturing contextual information within text sequences [4]. To address this issue, researchers introduced bidirectional LSTMs (BiLSTMs), simultaneously capturing information from both forward and backward directions. However, this improvement in LSTM also leads to increased computational complexity [5]. To mitigate this, Bidirectional GRUs (BI-GRUs) were developed, improving network efficacy by reducing model parameters and optimizing the gating mechanism, while still maintaining performance. Despite these advancements, current models continue to struggle with insufficient feature extraction parallelizability and capturing long-range dependencies effectively, particularly in the context of short texts [6].

Over time, the Transformer architecture and its variants such as GPT-3, GPT-3.5, GPT-4, and BERT have developed and efficiently performed sentiment analysis tasks. These large language models (LLMs) employ an attention mechanism and perform better context understanding by examining the contents in both directions. More recent advancements such as RoBERTa and ALBERT, further optimize BERT's performance by improving its training efficiency and reducing its complexity [7].

Attention mechanisms have been reported to address these drawbacks and enhance the model's ability to focus on crucial text parts while processing long sequences. Despite these improvements, attention mechanisms in sentiment analysis are still a hot area of research. Ongoing studies explore ways to further enhance the flexibility and effectiveness of these RNN models, chiefly by combining attention mechanisms with existing deep learning methods such as LSTMs and GRUs. Moreover, the drawback of understanding multi-lingual or code-mixed data remains a main concern for researchers in the field [8].

Although GRU is an efficient RNN variant, it has certain drawbacks: (1) It focuses solely on memorizing sequential information and does not assess the consequence of individual elements in the sequence; (2) GRU encounters difficulties when learning from long sequences, as illustrated in Section 3.1. Furthermore, the detailed association between various modalities is not effectively captured contents due to the minimal or complete lack of modules dedicated to modality interactions [9].

To address these issues, we proposed a novel RNN architecture, known as Recurrent Attention Unit (RAU) that merges the advantages of GRU and attention unit. In RAU, the recurrent attention unit is incorporated directly within the traditional GRU structure known as an attention gate. This gate increases GRU's capability to prioritize core information and discard less relevant content more efficiently.

The key contributions of the proposed approach are outlined as follows:

- We introduced a novel RNN technique, known as Recurrent Attention Unit (RAU), which incorporates an attention gate within the GRU cell called RA-GRU. This addition enables RA-GRU to automatically emphasize key information while filtering out less relevant details in sequential data.
- In this study, we first integrate the Self-Attention technique with RA-GRU, referred to as SA+RA-GRU permitting the model to capture hidden insights across different subspaces.
- By connecting two attention technques in the GRU network, the model effectively identifies critical information in the feature vector and improves the efficiency of learning modal features.
- The model proposed achieves strong sentiment analysis performance, demonstrating excellent scalability and practical applicability, making it highly suitable for real-world sentiment analysis tasks.

2. RELATED WORK

This section delivers related studies on sentiment analysis based on GRUs and attention techniques.

Most recent developments in sentiment analysis have reported that the GRU network and attention mechanisms or techniques captured nuanced patterns in text data. GRU, is a well-suited and simplified structure for NLP tasks due to its efficient gating strategies, while attention mechanisms help to emphasize the most valuable information in the sequence, making this combination extremely successful in sentiment analysis. Deep learning has been widely utilized in NLP fields, such as text classification, speech recognition, keyword extraction, spam filtering, machine translation, and information retrieval [10]. These advancements have significantly impacted sentence classification, one of the most frequent tasks in NLP. In 2020, Kumar et al. [11] established a GRU-based design for sentiment analysis on product reviews, improving with pre-trained word embeddings. Their approaches leveraged GRU's competence to sustain contextual information over longer sequences, obtaining notable improvements in accuracy on the Amazon reviews dataset. Fan et al. [12] developed a new text classification framework by integrating BERT and CNN, in the methodology authors employed the BERT model as an embedded layer within CNN. This combination outperformed the standalone BERT and CNN models.

In 2022, Li and Choi [13] referred to a GRU approach combined with CNN layers for multi-method sentence analysis, that incorporated textual and visual information. This methodology was primarily effective for social media datasets, as the CNN layers extracted visual features, while the GRU layer managed text sequence data. Their approach meaningfully improved sentiment classification on image-text paired datasets, like those from Twitter and Instagram. Moreover, Zeng *et al*. [14] introduced a classification technique that employed a hierarchical mechanism with the CNN layers. Furthermore, they utilized an enhanced Word2vec method and TF-IDF method to extract text features for effective classification.

Moreover, Zhang *et al.* [15] introduced a GRU network with attention mechanisms to concentrate on sentiment-rich words within sentences, rather than treating each word equally. They reported attention mechanism-enhanced GRU performance than the conventional GRU model, particularly on datasets with varying sentence lengths and structures, such as movie reviews and online forum posts. This study highlighted the effectiveness of integrating GRU with attention technique to context-sensitive sentiment prediction. Lu [16] employed LSTM and BERT approaches to create a keyword classification procedure for research articles, attaining better results than existing approaches. Furthermore, BERT, a deep and narrow neural network model introduced by Google [17], enhances text sentence prediction by incorporating context across all neural network layers during training.

A more recent work [18] established a hierarchical attention technique in combination with the GRU network, prioritizing substantial phrases and sentences at diverse hierarchical levels within documents. This hierarchical GRU-attention approach was predominantly effective for long-form text classification, as it first applied attention within individual sentences, then across sentence sequences. Ahmed *et al*. [19] developed a GRU model with an attention layer to design a robust network for sentence classification on code-mixed data, a common feature in multilingual texts. The model retained superior accuracy on multilingual sentiment benchmarks, making it highly applicable for social media sentiment analysis in diverse linguistic environments.

Lastly, Gupta and Kumar [20] proposed a stacked GRU architecture with cross-attention mechanism. This dual approach enabled the model to focus on intra-sequence and inter-sequence relationships, capturing deeper contextual meanings in complex datasets. The stacked GRU-attention model demonstrated high accuracy in sentiment prediction on large-scale datasets, underscoring its scalability and effectiveness for large-scale sentiment tasks.

2.1 GATED RECURRENT UNIT (GRU)

GRU is a progressive and simple variant of LSTM introduced by Chung et al. [21]. GRU is an effective approach for learning from sequential information and has been extensively used in numerous fields of NLP such as text summarization, keyword extraction [22], speech recognition [23], sentiment analysis [24], spam detection, and neural machine translation [25]. Figure 1 demonstrates the architecture of a GRU network, which comprises two main gates: the update gate and the reset gate, these gates are responsible for updating the hidden state at each time step. The input at each time step t is denoted as x_t, and the hidden state is h_t. GRU processes these as follows:

- **Update gate (z_t):** this gate controls how much information from the previous hidden state h_{t-1} is engaged.

$$z_t = \sigma(W_z \cdot x_t + U_z \cdot h_{t-1} + B_z) \quad (1)$$

where W_z, U_z, and B_z are the weight parameters and bias for the update gate, and σ is the sigmoid activation function.

- **Reset gate (r_t):** This gate controls how much past information to forget or rely on previous hidden state h_{t-1}:

$$r_t = \sigma(W_r \cdot x_t + U_r \cdot h_{t-1} + B_r) \quad (2)$$

where W_r, U_r, and B_r are the weight matrices and bias for the reset gate.

- **Candidate state** $\left(\tilde{h}_t\right)$**:** this state computes the new content influenced by the reset gate.

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot x_t + U_{\tilde{h}} \cdot (r_t \cdot h_{t-1}) + B_{\tilde{h}}) \quad (3)$$

- **Final hidden state (h_t):** The final hidden state is derived from the combination of the preceding hidden state h_{t-1} and the candidate's activation $\tilde{h}_t$ which is modulated by the update gate's regulatory influence.

$$h_t = z_t \cdot h_{t-1} + (1 - z_t) \cdot \tilde{h}_t \quad (4)$$

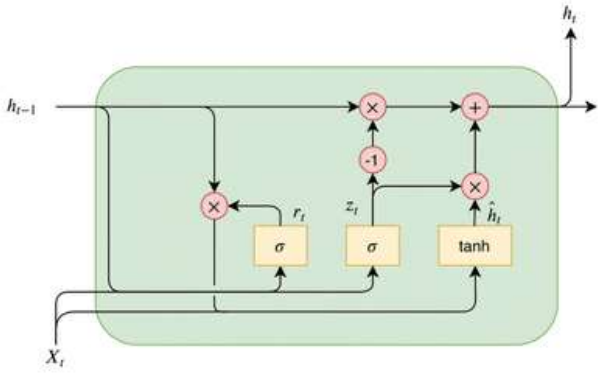

Fig.1. Traditional architecture of the GRU cell

In this context, when r_t equals 1, $\tilde{h}_t$ incorporates both the current input r_t and the previous hidden state h_{t-1}. On the other hand, if $r_t = 0$, $\tilde{h}_t$ depends entirely on the current input x_t, without influence from h_{t-1}, allowing units that extract short-term dependencies to activate reset gates. The update gate z_t manages how much of the previous hidden state h_{t-1} Influences the current state h_t. If z_t=1, The GRU propagates the similar contents at each step, independent of x_t. This results in units focusing on long-term dependencies and maintaining an active update gate. Together, the reset and update gates help GRUs manage short and long-term information effectively, reducing issues like vanishing or exploding gradients that commonly affect standard RNNs.

Several GRU enhancements have been developed to improve its learning capabilities. Li et al. [26] presented the independent RNN (IndRNN), which replaces the matrix product with the Hadamard product in the recurrent input and employs an activation function to control long sequences. Furthermore, Józefowecz et al. [27] introduced three GRU variations: incorporating tanh within the gate for added nonlinearity, eliminating dependence on the hidden state within the gate, and combining these adjustments in the hidden state computations. More recent, Zhang and Cheen [28] addressed the issue of quick memory decline in GRU from a theoretical angle, proposing the selective GRU (SGRU), which employs a tensor discriminator to dynamically decide if the hidden state should be updated at each time step, thus enabling learning with rapidly changing data. However, like the standard GRU, these variations lack an attention mechanism for sequence learning. To overcome this limitation, we introduce a new attention-based RNN network that incorporates an attention gate, allowing the RNN to disregard irrelevant information and emphasize essential data during sequential processing.

3. PROPOSED METHODOLOGY

The proposed methodology for sentiment analysis is based on the Recurrent Attention Unit (RAU) and an Improved Self-Attention mechanism to enhance feature extraction and improve accuracy in classifying sentiment in text data. The Fig.3 illustrates the overall architecture of the proposed SA+RA-GRU approach for sentiment analysis introduced in this study. The entire framework is composed of two components: (1) an RAU known as an Attention Gate, and (2) an Improved Self-Attention mechanism. To ensure clarity, first, we explain the attention gate, that is added in the existing GRU cell in Section 3.1, followed by a description of the self-attention mechanism in Section 3.2.

3.1 ATTENTION GATE

The proposed attention gate, integrated into the standard GRU cell, enhances the model's ability to concentrate on the most relevant parts of a sequence during processing, which is mainly valuable in tasks such as sentence classification where specific words or phrases heavily influence the outcome. Each cell processes sequential data in the standard GRU by updating its hidden state based on past information. However, it lacks a mechanism to selectively prioritize certain features within the sequence, which can lead to limitations in controlling long dependencies and identifying substantial contextual information.

To incorporate the attention gate, we implemented a new set of weights and attention scores that compute the relevance of each element in the input sequence relative to the current hidden state. These attention scores are then utilized to modulate the influence of each time step's input on the cell's hidden state. The key computational equations of the traditional GRU cell are modified to incorporate the attention gate. The following equations detail each step involved:

- **Attention Scores Calculation:** Based on Eq.(1) and Eq.(2) the attention mechanism computes a score, α_t representing the importance of the current input x_t. The attention score is derived from the hidden state and the current input, calculated as follows:

$$e_t = \tanh(W_\alpha \cdot x_t + U_\alpha \cdot h_{t-1} + B_\alpha) \quad (5)$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{i=1}^{T} \exp(e_i)} \quad (6)$$

where W_α and U_α are weight matrices, B_α is the bias term, and α_t signifies the regularized attention score, providing a probability distribution over time steps.

- **Applying the Attention Gate:** The candidate's hidden state, $\tilde{h}_t$ is modified by the attention score α_t to focus on the most relevant information, while ReLU is the rectified linear unit:

$$\tilde{h}_t = \text{ReLU}(W_h \cdot (r_t \cdot h_{t-1}) + U_h \cdot x_t + B_h) \quad (7)$$

$$\hat{h}_t = \alpha_t \cdot \tilde{h}_t \quad (8)$$

where $\hat{h}_t$ is the attention-weighted candidate hidden state, and * refers to Hadamard product. By merging the current time step's input with the previous time step, the attention gate enables the network to automatically identify which feature dimensions are essential for the current and future hidden cells.

- **Final Hidden State Update:** The final hidden state h_t is then updated by combining the attention-weighted candidate state $\hat{h}_t$ and the previous hidden state h_{t-1} controlled by the update gate z_t:

$$h_t = z_t \cdot h_{t-1} + (1 - z_t) \cdot \tilde{h}_t + \hat{h}_t \tag{9}$$

It is important to note that the attention gate strategy in the traditional GRU differs from prior approaches such as hierarchical-attention network (HAN), feature attention mechanism, self-attention, and soft attention [29] combined with RNNs, as those approaches are employed outside the memory cell of RNN network. In this way, the attention gate allows the GRU cell to automatically concentrate on important information within the sequence, reducing the influence of less relevant information. This modified GRU cell structure leverages both the memory capabilities of the GRU and the selective focus enabled by the attention mechanism, leading to more efficient and contextually aware sequence processing.

The Fig.2 refers to the modified architecture of a traditional GRU cell, known as an RA-GRU cell, where the red-linked section represents the attention gate. This gate produces α_t weight vector based on the prior hidden cell h_{t-1} and the current input x_t. Consequently, the attention-modified hidden state, $\hat{h}_t$, Computed by in Eq.(9). Moreover, similar to GRU, the modified GRU cell known as RA-GRU is a differentiable structure that simplifies optimization.

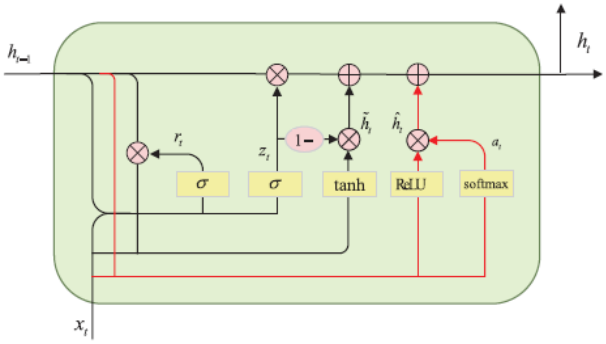

Fig.2. The modified architecture of the GRU cell includes an attention gate, highlighted by red links.

The attention mechanism, with minimal extra parameters, enhances RA-GRU's ability to balance past information with present input and to prioritize crucial elements in the sequences.

3.2 SELF-ATTENTION TECHNIQUE

In sentiment analysis tasks, each word contributes to the classification outcome to varying extents. To differentiate the significance of each word, a self-attention layer is added to assign weights to the output vector produced by RA-GRU. This addition allows the RA-GRU cell to influence the strengths of attention by adaptively weighting elements in the sequence, ensuring that the most relevant features are prioritized in each timestep. The self-attention technique, a distinct variant of the attention mechanism, provides this functionality. To clarify how the self-attention technique operates, we first examine the computational steps of the general attention technique. Essentially, the attention technique is comprised of multiple *Query* and *Key-Value* pairs. To apply attention, each input x_t is transformed into three vectors: a query Q, a key K, and a value V, with calculations are outlined as follows:

$$Q = W_q \cdot X, \quad K = W_k \cdot X, \quad V = W_v \cdot X \tag{10}$$

where W_q, W_k, and W_v are learnable weight matrices for queries, keys, and values.

The attention score $A_{i,j}$ between the query of x_t and key of x_j is computed using the scaled dot-product approach, which measures the relevance between each pair of inputs:

$$A_{i,j} = \frac{Q_i \cdot K_j^T}{\sqrt{d_k}} \tag{11}$$

where, d_k is the dimension vectors, and softmax is applied over the scores to normalize:

$$\text{Self-Attention} = \alpha_{i,j} = \text{softmax}(A_{i,j}) \tag{12}$$

The output of the attention layer for each input x_t is a weighted sum of the values V, where the weights are given by the normalized scores:

$$\text{Self-Attention}_i = \sum_{j=1}^{n} \alpha_{i,j} \cdot V_j \tag{13}$$

- **Integration with the modified RA-GRU Cell:** In the modified RA-GRU cell, this attention output is then integrated into the updated hidden state. At each timestep *t*, the attention-enhanced hidden state $\hat{h}_t$ is computed by combining the traditional GRU hidden state h_t with the contextual attention vector:

$$\hat{h}_t = f(h_t, \text{Self-Attention}_t) \tag{14}$$

where, f denotes a non-linear function that adjusts the hidden state based on the weighted significance of the sequence features. The updated hidden state $\hat{h}_t$ passes through the modified RA-GRU cell's recurrent update and reset gates, allowing the model to extract both the temporal dependencies and the valuable information highlighted by the self-attention mechanism.

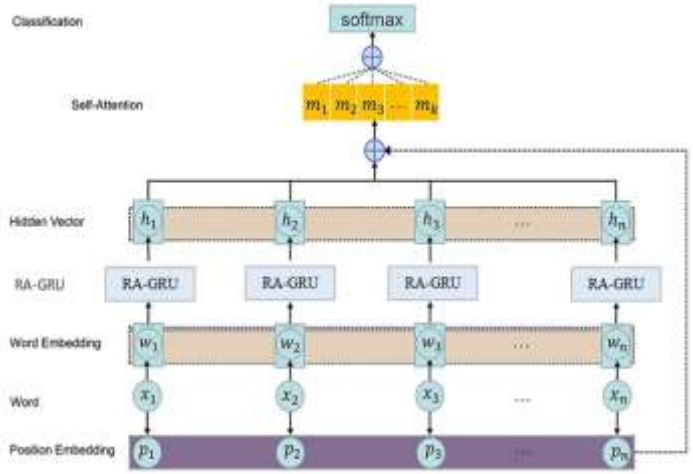

Fig.3. Overall architecture of the proposed methodology

This combination enables the modified RA-GRU cell with the Improved Self-Attention mechanism to concentrate important sequence information, manage complex dependencies, and

deliver accurate representations of sequential data for sentiment analysis.

The Fig.3 illustrates the complete framework of the proposed SA+RA-GRU, which integrates an attention gate in the traditional GRU cell, and a combined Self-Attention (SA) mechanism. The proposed framework comprises six main components: position embedding, a word embedding layer, a modified RA-GRU cell, a hidden vector, a self-attention layer, and finally, utilized softmax as a final classification layer.

4. EXPERIMENTAL DESIGN AND RESULTS ANALYSIS

4.1 DATASET DESCRIPTION

For experimental analysis, the IMDB dataset, initially provided by Andrew Maas [30], consists of 50,000 binary-labeled review samples from the Internet movie database, specifically tailored for sentiment analysis, and an additional 50,000 unlabeled samples useful for unsupervised learning. On the other hand, The Movie Review (MR) dataset is an extensively utilized benchmark dataset in NLP for sentiment analysis. It contains 10,662 short movie reviews collected from Rotten Tomatoes, divided into 5,295 positive reviews while 5,295 portray negative reviews. Its simplicity and balance (approximately equal distribution of positive and negative samples) have made it a standard dataset for evaluating machine learning models' performance in binary sentiment analysis.

4.2 PARAMETER SETTINGS

The model study was conducted on a computer with high-performance specifications, including an Intel Core i7-3770K @ 3.40GHz CPU, 32GB of RAM machine with DDR4 and 500GB SSD. The framework is implemented in popular deep learning libraries such as TensorFlow, Numpy Sklearn, Scipy, Pandas, and Keras packages that are carried out in Python 3.12, which facilitate the data preprocessing and manipulation. Compared to TensorFlow alone, Keras offers a clear and intuitive API, simplifying model construction. This framework supports commonly used CNN and RNN architectures, as well as their combinations, making it easy to build more complex models. Additionally, Keras can run seamlessly on both CPUs, allowing full utilization of hardware capabilities to accelerate model training and inference. The Table.1 presents the optimal hyperparameter settings that were employed to train the proposed model.

4.3 EVALUATION METRICS

The evaluation metrics for the developed approach are designed to evaluate its performance on sentiment analysis tasks. Usually, several evaluation metrics such as accuracy, precision, recall, and F1-score are utilized, each of which measures specific aspects of model effectiveness.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{16}$$

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN} \tag{17}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{18}$$

where *TP*, *TN*, *FP*, and *FN* refer to the True Positive, True Negative, False Positive, and False Negative, respectively.

Table.1. Optimal setting of the proposed model

Parameter	Value
Word vector dimension	764
Activation function	ReLU/tanh
Loss Function	Cross-entropy
Optimizer	Adam
Num-self-attention	12
Learning Rate	0.001
Batch size	128
Dropout	0.3
Learning rate decay	0.01
Network epochs	30

5. EXPERIMENTAL RESULTS

This section presents an exhaustive experimental evaluation of the proposed mechanism on two sentiment analysis datasets. The primary objective is to assess the robustness and efficacy of our developed approach compared to existing deep learning approaches. To achieve this, we conducted a comprehensive analysis using various evaluation metrics to validate the performance of our proposed model. The Table.4 presents all the evaluation metrics for the proposed framework.

5.1 FIRST DATASET (IMDB) RESULTS

The proposed approach shows significant improvements in sentiment analysis performance on the IMDB dataset compared to several baseline and comparative methods, results presented in Table.2. Our approach attained impressive precision scores of 92.68% and 92.58% for positive and negative reviews, respectively. Similarly, the recall scores for negative and positive reviews were 92.52% and 92.84%, respectively. These results culminated in an F1 score of 92.042% and an accuracy of 92.17%. Notably, our proposed approach outperformed traditional approaches, including the cognition-based attention (CBA) method proposed by Long et al. [31], which achieved an accuracy of 90.10% on the IMDB dataset. Although their approach combined two various attention frameworks with an LSTM network, it failed to yield satisfactory results. In contrast, our proposed mechanism demonstrated superior accuracy without relying on complex architectures. Furthermore, our approach outperformed the hybrid CNN-LSTM model proposed by Collados and Pilehvar [32], which accomplished an accuracy of 88.9% on the IMDB dataset. Despite their model's application of several cleaning techniques and the combination of CNN and LSTM approaches, it failed to attain better results. Moreover, Peng et al. [33] investigated three deep learning models, evaluated

the performance using the IMDB dataset, and reported that CNN achieved a superior accuracy of 88.22%.

Meanwhile, Fu et al. [34] explored ALE-LSTM and WALE-LSTM methods, achieving an accuracy of 89.30%, and 89.50%, respectively. Despite employing several attention techniques, their results were only marginally satisfactory. In comparison, our proposed framework demonstrated significant improvement in accuracy over the ALE-LSTM and WALE-LSTM methods. Kardakis et al. [35] investigated various attention-based methods combined with LSTM and GRU for sentence classification and highlighted the accuracy of 89.71%, and 87.92% respectively. Furthermore, Lio Bing [36] combined a "CBOW language mechanism with a deep CNN network for sentiment analysis on the IMDB dataset, reporting an accuracy of 87.20%. In contrast, Zulqarnain et al. [37] introduced a normalized auto-encoder GRU framework for sentiment analysis and reported an accuracy of 91.32% on the IMDB dataset. Yohong et al. [38] introduced a feature-based method with an attention mechanism (FARNN-Att), achieving an accuracy of 89.22%. In contrast, based on the self-attention layers and recurrent attention gate, our proposed approach yielded exceptional results compared to baseline approaches. In conclusion, our proposed approach, based on a self-attention mechanism and recurrent attention gate, demonstrated outstanding performance on the IMDB dataset, outperforming existing studies [31-38] characterized by more intricate architectures. Evaluation outcomes are illustrated in Table.3.

Table.2. Performance of the proposed approach on both datasets

Datasets	Normalized Confusion Matrices			Evaluation Parameters			Average Values		
	Predict class ← Actual class ↓	0	1	Precision (%)	Recall (%)	F1 Score (%)	Test size	Accuracy (%)	F1 score (%)
IMDB	0	0.9024	0.0976	92.68	92.52	92.59	CV	92.17	92.042
	1	0.9053	0.0947	92.58	92.84	92.71	CV		
MR	0	0.8202	0.1798	81.87	82.93	81.52	CV	82.38	82.451
	1	0.2187	0.7813	82.90	83.71	83.30	CV		

Table.3. Comparison with traditional studies on the IMDB dataset

Methods	Model Complexity	Accuracy (%)
CNN [33]	number of layers, filter sizes, strides, and input dimensions, leading to a computational cost-proportional	88.22
FARNN-Att [38]	Attention mechanism and adversarial training with BiLSTM	89.22
WALE-LSTM [34]	The combination of lexicon and attention layers with LSTM	89.50
CNN+LSTM [32]	Combination of CNN and LSTM with different cleaning processes	88.90
Self-Att-LSTM+word2vec [35]	Implemented Self-attention mechanism with LSTM based on word2vector	89.71
Hierarchical-Att-GRU+Dropout [35]	Investigated Hierarchical-attention mechanism with GRU based on word2vector	87.92
CBOW+D-CNN [36]	Combined the CBOW method with the CNN algorithm	87.20
LSTM+CBA+LA [31]	Integrated of two different feature attention frameworks with LSTM	90.11
BERT [44]	Bidirectional Encoder Representations from Transformers	85.83
XLNeT [44]	Transformer-XL with a two-stream attention mechanism	91.10
NAE-GRU [37]	Combined Auto-encoder with GRU through batch normalization	91.32
SA-RA+GRU (Proposed)	**Self-Attention with Recurrent Attention-GRU**	**92.17**

5.2 SECOND DATASET (MR) RESULTS

The performance of our developed approach on the MR dataset is illustrated in Table.4. Our mechanism achieved precision scores of 81.87% and 82.90% for positive and negative observations, respectively. The recall score was 82.93% and 83.71%, respectively. Based on these metrics, our developed model obtained an F1 score of 82.451% and an accuracy of 82.38%. We compared the effectiveness of our developed framework with existing studies and highlighted their complexities. Chen et al. [39] reported BiLSTM with provisional random fields to improve sentiment analysis and attained an accuracy of 82.30%. Furthermore, Fu et al. [34] introduced an attention mechanism (ALE-LSTM) built upon a lexicon-enhanced LSTM method, yielding 79.90% accuracy on the MR dataset. Notably, our developed model performed better than both ALE-LSTM and WALE-LSTM approaches on the MR dataset. Zhang et al. [40] referred to a novel architecture combining BiGRU with CNN (BiGRU+CNN), achieving an accuracy of 78.30%". Our model demonstrated a 4.08% improvement over their approach. Usama et al. [41] presented a multi-level and multi-type feature extraction approach, integrating GRU, LSTM, and CNN, with reported accuracies of 79.80% and 80.20%, respectively. In conclusion, our proposed approach based on a self-attention mechanism and recurrent attention gate demonstrated exceptional performance on the IMDB dataset, outperforming existing studies [35], [39]-[43] characterized by more intricate architectures, evaluation results illustrated in Table.4. Furthermore, this section discussed an experiment of the proposed model and traditional approaches based on the loss rate and epochs. To provide a more intuitive understanding of the training process for each network, we plotted the accuracy and loss rate variation curves of the proposed approach with five

traditional methods on the Shopping validation set, as shown in Fig.4. An examination of the loss value curves over 25 epochs reveals that the self-attention integrated approach exhibits accelerated convergence, enhanced training efficacy, and improved stability during the training process, surpassing standard methods including LSTM, Bi-GRU, and BERT. This superior performance can be attributed to the Self-Attention layers and Recurrent Attention Gate in the GRU, which enables more effective optimization of internal structural information representation in text data.

Table.4. Comparison with traditional studies on MR dataset.

Methods	Model Complexity	Accuracy (%)
BiGRU+CNN [40]	Sequential combination of BiGRU and CNN framework	78.30
MV+RNN [42]	The joint strategy of analyzing trees and the RNN model	79.00
WALE+LSTM [34]	Implementation of WALE and attention mechanism jointly with LSTM	79.90
Self-Att-LSTM+word2vec [35]	Implemented Self-attention mechanism with LSTM based on word2vector	79.93
Hierarchical-Att-GRU+Dropout [35]	Investigated Hierarchical-attention mechanism with GRU based on word2vector	79.97
CNN-GRU-multilevel and multitype fusion [41]	Proposed combined methodology in terms of multilevel and multi-type features based on CNN and GRU models	80.20
BiLSTM-CRF [39]	Fusion of bidirectional LSTM with CRT combined with CNN	82.30
LR+BiLSTM [43]	Illustrated sentence-level representation based on LR, and combined with BiLSTM	82.10
SA-RA+GRU (Proposed)	**Self-Attention with Recurrent Attention-GRU**	**82.38**

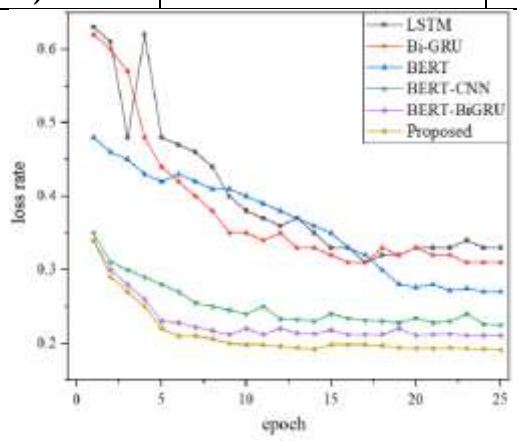

Fig.4. Loss curve proposed and comparative approaches

6. CONCLUSION AND FUTURE DIRECTION

This paper introduces a novel recurrent neural network "(RNN) architecture, dubbed the Recurrent Attention Unit (RAU), known as RA. By embedding the recurrent attention technique within the core of the Gated Recurrent Unit (GRU) cell, RAU enables dynamic focus on pertinent sequential information, fostering adaptive processing and enhanced performance. Moreover, this study combined an improved Self-Attention technique (SA) with RA-GRU known as SA+RA-GRU. The improved self-attention technique is implemented to reallocate the weights of long text sequences. While attention mechanisms have recently become a significant innovation in deep learning, their precise impact on sentiment analysis has yet to be fully evaluated. We evaluated our developed method on two benchmark datasets IMDB, and MR, and efficiently predicted sentiment polarity using an efficient attention mechanism. The experimental outcomes demonstrate that the developed method, incorporating attention mechanisms, achieves outstanding performance than traditional approaches, in terms of accuracy, particularly on the IMDB and MR datasets making them a powerful tool for textual sentiment analysis and classification tasks. Furthermore, the proposed model can perform superior on other tasks of NLP such as machine translation, text summarization, and speech recognition. Future research directions could include exploring the application of attention mechanisms to extremely long sequences and investigating the performance of local attention and other attention variants on sequential data. Additionally, the research could explore the effectiveness of self-attention, global-attention, and hierarchical-attention mechanisms within various deep learning algorithms and training configurations, including the impact of varying dropout rates," and could provide further insights.

REFERENCES

[1] J. Khan, N. Ahmad, S. Khalid, F. Ali and Y. Lee, "Sentiment and Context-Aware Hybrid DNN with Attention for Text Sentiment Classification", *IEEE Access*, Vol. 11, pp. 28162-28179, 2023.

[2] K.L. Tan, C.P. Lee and K.M. Lim, "A Survey of Sentiment Analysis: Approaches, Datasets and Future Research", *Applied Sciences*, Vol. 13, No. 7, pp. 1-6, 2023.

[3] M. Zulqarnain, R. Ghazali, S.H. Khaleefah and A. Rehan, "An Improved Performance of the GRU Model based on Batch Normalization for Sentence Classification", *International Journal of Computer Science and Network Security*, Vol. 19, No. 9, pp. 176-186, 2019.

[4] N.L. Rane, S.K. Mallick, O. Kaya and J. Rane, "Machine Learning and Deep Learning Architectures and Trends: A Review", *Applied Machine Learning and Deep Learning: Architectures and Techniques*, pp. 1-38, 2024.

[5] S. Rani and A. Jain, "Aspect-based Sentiment Analysis of Drug Reviews using Multi-Task Learning based Dual BiLSTM Model", *Multimedia Tools and Applications*, Vol. 83, No. 8, pp. 22473-22501, 2024.

[6] M.M. Rahman, A.I. Shiplu, Y. Watanobe and M.A. Alam, "RoBERTa-BiLSTM: A Context-Aware Hybrid Model for Sentiment Analysis", *Artificial Intelleigence*, pp. 1-6, 2024.

[7] J. Lak, R. Boostani, F.A. Alenizi, A.S. Mohammed and S.M. Fakhrahmad, "RoBERTa, ResNeXt and BiLSTM with Self-Attention: The Ultimate Trio for Customer Sentiment Analysis", *Applied Soft Computing*, Vol. 164, pp. 1-6, 2024.

[8] M. Zulqarnain, R. Sheikh, S. Hussain, M. Sajid, S.N. Abbas, M. Majid and U. Ullah, "Text Classification using Deep Learning Models: A Comparative Review", *Cloud Computing and Data Science*, pp. 80-96, 2024.

[9] A. Alslaity and R. Orji, "Machine Learning Techniques for Emotion Detection and Sentiment Analysis: Current State, Challenges, and Future Directions", *Behaviour and Information Technology*, Vol. 43, No. 1, pp. 139-164, 2024.

[10] M. Wankhade, A.C.S. Rao and C. Kulkarni, "A Survey on Sentiment Analysis Methods, Applications and Challenges", *Artificial Intelligence Review*, Vol. 55, No. 7, pp. 5731-5780, 2022.

[11] A. Kumar, K. Srinivasan, W.H. Cheng and A.Y. Zomaya, "Hybrid Context Enriched Deep Learning Model for Fine-Grained Sentiment Analysis in Textual and Visual Semiotic Modality Social Data", *Information Processing and Management*, Vol. 57, No. 1, pp. 1-7, 2020.

[12] F. Wei and F. Li, "News Text Classification based on Hybrid Model of Bidirectional Encoder Representation from Transformers and Convolutional Neural Network", *Journal of Physics: Conference Series*, Vol. 2005, No. 1, pp. 1-6, 2021.

[13] Z. Li, B. Xu, C. Zhu and T. Zhao, "CLMLF: A Contrastive Learning and Multi-Layer Fusion Method for Multimodal Sentiment Detection", *Computation and Language*, pp. 1-6, 2022.

[14] Q. Zeng, "Design of Intelligent Sentiment Classification Model based on Deep Neural Network Algorithm in Social Media", *IEEE Access*, pp. 1-6, 2024.

[15] K. Zhang, Y. Geng, J. Zhao, J. Liu and W. Li, "Sentiment Analysis of Social Media Via Multimodal Feature Fusion", *Symmetry*, Vol. 12, No. 12, pp. 1-6, 2022.

[16] W. Lu, P. Li, G. Zhang and Q. Cheng, "Recognition of Lexical Functions in Academic Texts: Automatic Classification of Keywords based on BERT Vectorization", *Journal of the China Society for Scientific and Technical Information*, Vol. 39, No. 12, pp. 1320-1329, 2020.

[17] A. Bello, S.C. Ng and M.F. Leung, "A BERT Framework to Sentiment Analysis of Tweets", *Sensors*, Vol. 23, No. 1, pp. 1-6, 2023.

[18] D. Chen, W. Su, P. Wu and B. Hua, "Joint Multimodal Sentiment Analysis based on Information Relevance", *Information Processing and Management*, Vol. 60, No. 2, pp. 1-7, 2023.

[19] W. Ahmad, H.U. Khan, T. Iqbal and S. Iqbal, "Attention-based Multi-Channel Gated Recurrent Neural Networks: A Novel Feature-Centric Approach for Aspect-based Sentiment Classification", *IEEE Access*, Vol. 11, pp. 54408-54427, 2023.

[20] T. Gupta and E. Kumar, "Fusion of Bi-GRU and Temporal CNN for Biomedical Question Classification", *International Journal of Computers and Applications*, Vol. 45, No. 6, pp. 460-470, 2023.

[21] J. Chung, C. Gulcehre, K. Cho and Y. Bengio, "Gated Feedback Recurrent Neural Networks", *Proceedings of International Conference on Machine Learning*, pp. 2067-2075, 2015.

[22] R. RoselinKiruba, S. Sowmyayani, S. Anitha, J. Kavitha, R. Preethi and C.S. Jothi, "Text Summarization based on Feature Extraction using GloVe and B-GRU", *Proceedings of International Conference on Sustainable Computing and Smart Systems*, pp. 517-522, 2024.

[23] D. Kumar and S. Aziz, "Performance Evaluation of Recurrent Neural Networks-LSTM and GRU for Automatic Speech Recognition", *Proceedings of International Conference on Computer, Electronics and Electrical Engineering and their Applications*, pp. 1-6, 2023.

[24] M. Zulqarnain, S.A. Ishak, R. Ghazali, N.M. Nawi, M. Aamir and Y.M.M. Hassim, "An Improved Deep Learning Approach based on Variant Two-State Gated Recurrent Unit and Word Embeddings for Sentiment Classification", *International Journal of Advanced Computer Science and Applications*, Vol. 11, No. 1, pp. 594-603, 2020.

[25] P. Eswaraiah and H. Syed, "A Hybrid Deep Learning GRU based Approach for Text Classification using Word Embedding", *EAI Endorsed Transactions on Internet of Things*, Vol. 10, pp. 1-7, 2024.

[26] S. Li, W. Li, C. Cook, C. Zhu and Y. Gao, "Independently Recurrent Neural Network (IndRNN): Building A Longer and Deeper RNN", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 5457-5466, 2018.

[27] R. Jozefowicz, W. Zaremba and I. Sutskever, "An Empirical Exploration of Recurrent Network Architectures", *Proceedings of International Conference on Machine Learning*, pp. 2342-2350, 2015.

[28] W. Zheng and G. Chen, "An Accurate GRU-based Power Time-Series Prediction Approach with Selective State Updating and Stochastic Optimization", *IEEE Transactions on Cybernetics*, Vol. 52, No. 12, pp. 13902-13914, 2021.

[29] S. Kardakis, I. Perikos, F. Grivokostopoulou and I. Hatzilygeroudis, "Examining Attention Mechanisms in Deep Learning Models for Sentiment Analysis", *Applied Sciences*, Vol. 11, No. 9, pp. 1-7, 2021.

[30] A. Maas, R.E. Daly, P.T. Pham, D. Huang, A.Y. Ng and C. Potts, "Learning Word Vectors for Sentiment Analysis", *Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 142-150, 2011.

[31] Y. Long, Q. Lu, R. Xiang, M. Li and C. R. Huang, "A Cognition based Attention Model for Sentiment Analysis", *Association for Computational Linguistics*, pp. 1-6, 2017.

[32] J. Camacho-Collados and M.T. Pilehvar, "On the Role of Text Preprocessing in Neural Network Architectures: An Evaluation Study on Text Categorization and Sentiment Analysis", *Proceedings of the EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pp. 1-54, 2017.

[33] X. Ouyang, P. Zhou, C.H. Li and L. Liu, "Sentiment Analysis using Convolutional Neural Network", *Proceedings of International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing;Pervasive Intelligence and Computing*, pp. 2359-2364, 2015.

[34] X. Fu, J. Yang, J. Li, M. Fang and H. Wang, "Lexicon-Enhanced LSTM with Attention for General Sentiment Analysis", *IEEE Access*, Vol. 6, pp. 71884-71891, 2018.

[35] S. Kardakis, I. Perikos, F. Grivokostopoulou and I. Hatzilygeroudis, "Examining Attention Mechanisms in Deep Learning Models for Sentiment Analysis", *Applied Sciences*, Vol. 11, No. 9, pp. 1-6, 2021.

[36] B. Liu, "Text Sentiment Analysis based on CBOW Model and Deep Learning in Big Data Environment", *Journal of Ambient Intelligence and Humanized Computing*, Vol. 11, No. 2, pp. 451-458, 2020.

[37] M. Zulqarnain, A.K.Z. Alsaedi, R. Sheikh, I. Javid, M. Ahmad and U. Ullah, "An Improved Gated Recurrent Unit based on Auto Encoder for Sentiment Analysis", *International Journal of Information Technology*, Vol. 16, No. 1, pp. 587-599, 2020.

[38] Y. Ma., H, Fan and C. Zhao, "Feature-based Fusion Adversarial Recurrent Neural Networks for Text Sentiment Classification", *IEEE Access*, Vol. 7, pp. 132542-132551, 2019.

[39] T. Chen, R. Xu, Y. He and X. Wang, "Improving Sentiment Analysis Via Sentence Type Classification using BiLSTM-CRF and CNN", *Expert Systems with Applications*, Vol. 72, pp. 221-230, 2017.

[40] D. Zhang, L. Tian., M. Hong, F. Han, Y. Ren and Y. Chen, "Combining Convolution Neural Network and Bidirectional Gated Recurrent Unit for Sentence Semantic Classification", *IEEE Access*, Vol. 6, pp. 73750-73759, 2018.

[41] M. Usama, W. Xiao, B. Ahmad, J. Wan, M.M. Hassan and A. Alelaiwi, "Deep Learning based Weighted Feature Fusion Approach for Sentiment Analysis", *IEEE Access*, Vol. 7, pp. 140252-140260, 2019.

[42] R. Socger, B. Huval, C.D. Manning and A.Y. Ng, "Semantic Compositionality through Recursive Matrix-Vector Spaces", *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 201-1211, 2012.

[43] Q. Qian, M. Huang, J. Lei and X. Zhu, "Linguistically Regularized LSTMs for Sentiment Classification", *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 1-7, 2016.

[44] M.M. Danyal, S.S. Khan, M. Khan, S. Ullah, F. Mehmood and I. Ali, "Proposing Sentiment Analysis Model based on BERT and XLNet for Movie Reviews", *Multimedia Tools and Applications*, pp. 1-25, 2024.