Grenze International Journal of Engineering and Technology, June Issue

Sentiment Analysis on Social Media Data using Roberta-GRU Hybrid Model

Chandan Vishwakarma[1], Natthan Singh[2] and Sonam Srivastava[3]
[1-3]Department of Computer Science & Engineering, Institute of Engineering & Technology, Lucknow, India
Email: 2300521635003@ietlucknow.ac.in, natthan.singh@ietlucknow.ac.in, ssrivastava.csed.cf@ietlucknow.ac.in

***Abstract*— Sentiment analysis, a key technique natural language processing (NLP) determines the tone of emotion in text and express as positive or negative or any other personal feeling. The explosion of social media content has made sentiment analysis indispensable across various fields, including e-commerce, healthcare, politics and digital marketing. However, social media data provide unique analytical challenges due to its informal language, inherent noise, misspellings and the prevalence of code-mixed text. Traditional sequential models often fall short in addressing these complexities, demanding more robust and adaptable solutions. This paper introduces an enhanced RoBERTa-GRU hybrid model specifically designed for improved sentiment classification on noisy social media datasets. Our approach incorporates text normalization, tokenization and fine-tuning of RoBERTa embeddings to better capture the nuances of informal language. The incorporation of Gated Recurrent Units (GRU) enhances the model's capacity to capture sequential dependencies present in text, thereby improving the accuracy of sentiment classification. We assess the proposed model using real-world datasets characterized by noise, specifically consisting of comments from Twitter and reviews from YouTube. Preliminary findings indicate that our RoBERTa-GRU hybrid model significantly surpasses conventional approaches, achieving an accuracy rate of 90%. These outcomes highlight the model's proficiency in addressing the complexities associated with noisy, user-generated content. Our results emphasize the promise of this hybrid methodology in advancing sentiment analysis within social media contexts, facilitating more precise and contextually pertinent insights.**

***Keywords*— Sentiment Analysis, Deep Learning, Transformers, RoBERTa, GRU, NLP, Twitter Data.**

I. INTRODUCTION

The rise of internet technology plays a remarkable role in increasing the usage of social media and e-commerce platforms. Social media is become a useful tool for expression and communication in the current digital era. The exponential growth of social media platforms has resulted in an enormous amount of user-created content, providing valuable information about public opinion, consumer trends and emerging topics. However, the informal, unstructured and constantly changing nature of social media text makes analyzing this data difficult. Sentiment analysis (SA) seeks to extract opinions and emotions from text. Traditional machine learning and lexicon-based sentiment analysis methods often struggle to understand contextual nuances, particularly in the noisy social media environment with its abbreviations, slang and emojis.

Grenze ID: 01.GIJET.11.2.561_26
© Grenze Scientific Society, 2025

By employing contextual embeddings and sequential embeddings, recent developments in deep learning, particularly transformer models like Bert and RoBERTa, have greatly increased the accuracy of sentiment categorization [1]. However, these models frequently fail to capture long-range dependencies across sentences, which are essential for understanding sentiment changes in conversations.

The recent advancement in deep learning have significantly improve sentiment classification performance. The study in [2] demonstrated how Transformer-based models, such as Bidirectional Encoder Representations from Transformers (BERT) and its improved variant RoBERTa (Robustly Optimized BERT Pretraining Approach), have demonstrated remarkable accuracy in various NLP tasks. These models utilize self-attention mechanisms to learn contextual relationships between words. However, despite their effectiveness, transformers struggle to model sequential dependencies and sentiment transitions across sentences, which is crucial for understanding sentiment fluctuations in conversations. Machine learning broadly falls into two categories: traditional and deep learning-based approaches. Traditional machine learning techniques, encompassing both supervised and unsupervised learning [3], have been extensively utilized for text classification tasks. Supervised methods, which necessitate labeled datasets for model training, allow for rigorous performance evaluation against predefined sentiment categories on unseen data. Conversely, unsupervised methods operate on unlabeled data, enabling the discovery of inherent patterns and structures without explicit human guidance. However, these traditional methods often rely on feature extraction techniques like Bag-of-Words (BoW) and N-grams [4,5], which are inherently limited in their ability to capture contextual relationships and sequential dependencies within textual data. Consequently, they exhibit diminished effectiveness in analyzing the nuanced and noisy text characteristic of social media platforms.

To overcome these limitations, deep learning models have emerged, leveraging contextual embeddings and the explicit modeling of sequential dependencies to achieve enhanced sentiment classification accuracy. Specifically, our proposed RoBERTa-GRU hybrid model synergistically combines the robust contextual word embedding capabilities of RoBERTa with the long-range dependency modeling prowess of Gated Recurrent Units (GRU), thereby facilitating a more comprehensive analysis of sentiment trends. This integrated approach enables a deeper understanding of sentiment fluctuations and significantly outperforms traditional machine learning techniques in handling the complex and dynamic textual data prevalent in social media. Deep learning, as noted in sources [6,7], involves the application of artificial neural networks organized in various layers to tackle a range of tasks. This approach draws significant inspiration from the architecture of the human brain, which comprises a vast network of neurons that facilitate information processing. Deep learning techniques are primarily categorized into feedforward and recurrent neural networks. These neural networks are crucial at different levels for analyzing sentiment effectively.

The authors in reference [8] introduced a model that integrates recurrent and neural networks to assess the sentiment of brief texts sourced from social media platforms. In this framework, local features were derived using a convolutional neural network (CNN), while recurrent neural networks facilitated the learning of long-range dependencies, enhancing sentence-level feature representation. This synergistic approach ultimately achieved superior classification accuracy compared to existing models such as LSTM and gated recurrent units (GRUs), with accuracies of 82.28%, 51.50%, and 89.95% across three datasets. Likewise, references [9,10] presented an ensemble methodology that amalgamated 10 LSTMs and 10 CNNs through a soft voting mechanism to analyze sentiment from an English-language Twitter dataset. To address the imbalanced nature of the dataset, cross-entropy was employed as the loss function, implemented via TensorFlow. This experiment yielded promising and elevated results across five English subtasks, evaluated using accuracy and F-measure as performance metrics. Additionally, reference [11] utilized an ensemble strategy that incorporated both surface and deep features alongside various classifiers, analyzing movie reviews across six public Twitter datasets. Nonetheless, a gap in the literature exists concerning the application of ensemble learning for sentiment analysis within social media contexts. Consequently, the current research endeavors to fill this gap by employing an ensemble approach that combines various deep learning models to evaluate sentiment derived from social media data. The study also compares the performance of different classifiers utilized in the ensemble framework against specific performance metrics, aiming to identify the most effective deep learning strategy [12]. This methodology holds potential for accurately assessing sentiment across diverse social media platforms.

We leverage the advantages of Deep Learning, as numerous researchers have explored various types of Sentiment Analysis (SA) within artificial neural networks. Key benefits include Automated Feature Generation, which involves deriving new features from a limited set of training data features to enhance generalization, and scalability, whereby Deep Learning efficiently processes large datasets and performs extensive computations, yielding cost and time savings. In their work, [14] introduced a dynamic neural network, conducting experiments with various convolutional neural network (CNN) configurations and a dynamic max-pooling K operator. Socher

[15] employed this model in a different study, achieving results comparable to those of Deep Convolutional Neural Networks (DCNN). [16] offered a thorough review of prevalent approaches in deep learning, including CNNs and Long Short-Term Memory networks (LSTMs), specifically for aspect-level sentiment analysis. Liu and Shen [17] presented the Gated Alternate Neural Network (GANN), a novel neural network architecture designed to address several limitations in previously proposed models, particularly those related to noise in capturing meaningful sentiment expressions, resulting in enhanced outcomes. For research focused on financial sentiment, Akhtar [18] advocated for the integration of Multi-Layer Perceptrons (MLP) with deep learning and feature-based techniques, developing several models based on CNN, LSTM, and Gated Recurrent Units (GRU). In a separate investigation, Pan [19] adopted a hybrid methodology utilizing MLP for text localization, achieving favorable results. Karakuş [20] examined the performance of various deep learning models during both training and testing phases by constructing model variants with different layer sizes and word embedding strategies. They improved their results through the use of LSTM, CNN, Bidirectional LSTM (BILSTM), and CNN-LSTM architectures, and achieved a test accuracy of 78% on a movie review dataset using MLP. Additionally, another study applied MLP in the training process of a sentiment classification task using the IMDB review dataset, yielding promising results [21]. Most research in deep learning sentiment classification has focused on reviewed datasets to optimize outcomes. To achieve the best results during testing, a hybrid sentiment classification approach utilizing Twitter datasets is recommended. By harnessing the strengths of both machine learning (ML) and deep learning (DL), we can mitigate the limitations inherent in each method, leading to solutions that are both more accurate and computationally efficient [22]. Hybrid models demonstrate improved accuracy and enhance the efficiency and performance of systems.

This paper provides a detailed analysis of an emoji-based dataset, showing that our hybrid model outperforms standard transformer-based models in classifying noisy sentiment data. The results demonstrate that the RoBERTa-GRU hybrid is a powerful tool for practical sentiment analysis applications. Social media platforms produce a diverse array of content, including alphabetic characters, punctuation marks, stop words, usernames, icons, and web links (URLs). However, these elements do not play a significant role in the sentiment analysis process. For instance, usernames do not assist algorithms in effectively classifying tweets as positive or negative. such that we implement Robust text normalization techniques are implemented, including noise removal, tokenization and handling of emojis and abbreviations, to improve data quality for sentiment analysis. RoBERTa is utilized to generate high-quality contextual embeddings, capturing semantic relationships between words and enhancing sentiment classification performance. A GRU network is employed to effectively capture long-range dependencies across sentences, improving sentiment trend analysis in conversational data [1]. To address the challenge of traditional models performing poorly with unstructured datasets containing alphabetic characters like punctuation, numbers, stop words, usernames, icons, web links (URLs), abbreviations, and emojis, we propose a Sentiment Analysis deep learning hybrid model: RoBERTa-GRU. This model is designed to leverage the strengths of both RoBERTa and GRU architectures. RoBERTa extracts rich contextual representations, while GRU effectively captures sequential dependencies, leading to a more robust understanding of sentiment fluctuations. Our approach improves both accuracy and the interpretability of sentiment trends in large social media datasets. By combining RoBERTa linguistic comprehension with GRU's sequential modeling abilities, our model achieves better performance compare to other traditional machine learning and deep learning models. We also use the different datasets that are collected form Twitters and used by the previous Authors in their research study to check the model performance of proposed framework.

This paper is organized into seven sections, which are laid out in the following order, In Section II describes related work on sentiment analysis in social media data, different methods used and their result. Section III discusses details proposed methodology for sentiment analysis. Section IV describes the experimental setup and implementation in details. Section V present the results Comparison between others approaches and our findings. Finally, Section VI contains conclusion of the paper and Section VII outlines the future work and research directions.

II. LITERATURE REVIEW

Sentiment analysis has been extensively researched across various domains, leading to the development of hybrid deep learning models, feature engineering techniques and transformer-based architectures to improve classification performance. This section reviews recent research work focused on hybrid models, contextual embeddings and deep learning techniques that have contributed to advancements in sentiment analysis. A study by Wongkar and Angdresey compared the performance of Naive Bayes, Support Vector Machine (SVM), and K-Nearest Neighbour (KNN) machine learning algorithms. They analyzed tweets related to the 2019 Indonesian presidential candidates, which were labeled as positive or negative. Prior to training, the tweet data was preprocessed through text parsing, tokenization, and text mining. The dataset was split into 80% for training and 20% for testing. The findings indicated that the Naïve Bayes classifier achieved the highest accuracy at 75.58%, followed by KNN at 73.34%, and SVM at 63.99%. In a separate study, the authors also compared Support Vector Machine, Multinomial Naïve Bayes, Long Short-Term Memory (LSTM), and BERT for sentiment classification using a larger dataset of 1.6 million positive and negative tweets. This dataset underwent similar preprocessing steps, including tokenization, stemming, lemmatization, and the removal of stop words and punctuation. Again, an 80% training and 20% testing split was used. The results of this study showed that BERT was the most effective model, reaching an accuracy of 85.4%.

Similarly, in the work by Salur and Aydin [25] conducted a study on Turkish sentiment analysis, examining the performance of hybrid models combining CNN with LSTM, BiLSTM, or GRU. A Turkish tweets dataset focused on GSM operators was utilized. Character-level embeddings were used as input for the CNN, while preprocessed text (case folding, URL/unnecessary word removal) represented as FastText/word2vec was used for LSTM, BiLSTM, and GRU. The features extracted from both CNN and LSTM/BiLSTM/GRU were concatenated and classified with a SoftMax layer. The CNN-BiLSTM-FastText model achieved the best accuracy, reaching 82.14%. In a comparative study, the RoBERTa model was found to achieve an accuracy of 90.67%, which, while respectable, was lower than the proposed RoBERTa-GRU model that reached an accuracy of 91.28%. This indicates that while RoBERTa performs well, there is room for improvement when combined with other architectures like GRU for enhanced performance in sentiment analysis tasks [26].

Md. Sagar Hossen et al. [27] utilized a Gated Recurrent Unit (GRU) model for sentiment analysis on hotel reviews, achieving an accuracy of 84%. This model was part of a broader study that compared various algorithms, highlighting the effectiveness of GRU in capturing customer sentiments from textual data examines restaurant reviews, emphasizing the utility of machine learning algorithms in sentiment analysis. Additionally, authors explore sentiment analysis in Hindi film reviews. The review underscores the growing reliance on online reviews for decision-making, indicating a trend towards using artificial intelligence for future business predictions. Tan et al. (2023, 2022) [1] have demonstrated the efficacy of hybrid deep learning models, combining RoBERTa with GRU and LSTM, for sentiment analysis. Their models, evaluated on three benchmark datasets, achieved superior performance compared to existing methods. The integration of RoBERTa, a powerful transformer model applicable to various NLP tasks, highlights a significant improvement in sentiment analysis [28]. These related works highlight the evolution of sentiment analysis models, demonstrating the effectiveness of hybrid architectures, contextual embeddings, and specialized preprocessing techniques. The RoBERTa-GRU model proposed in this research builds on these advancements, optimizing both contextual representation and sequential modeling for robust sentiment classification in social media data.

III. METHODOLOGY

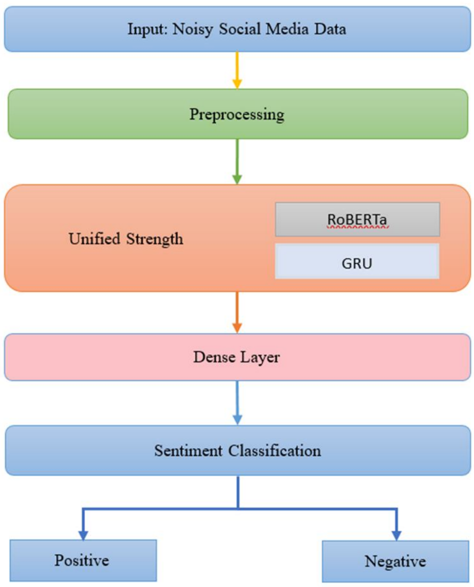

Figure 1. Architecture of the proposed RoBERTa-GRU model

In this section, we discuss the methodology of the proposed RoBERTa-GRU hybrid model designed to address the challenges of sentiment analysis on noisy social media data. This methodology consists of six key phases Firstly, Data collection, advanced preprocessing techniques, model construction, Dense Layer, Sentiment Classification as positive and negative, Subsequently The proposed model is carefully developed to address the complex nature of social media content, including emojis, misspellings, slang, informal language and abbreviations. our proposed methodology is illustrated in Figure. 1.

A. Data Collection

The evaluation of the proposed RoBERTa-GRU model was conducted using two publicly accessible sentiment analysis datasets Sentiment140 and ONEkemojitweets_senti_posneg. The Sentiment140 dataset developed by Stanford University [1], consists of approximately 1.6 million tweets categorized as positive and negative. Furthermore, the ONEkemojitweets_senti_posneg dataset, comprising 1,000 tweets also labeled with positive and negative class, offers a valuable resource due to its inclusion of emojis and emotive expressions. These datasets are crucial for rigorously evaluating sentiment analysis models, as they encapsulate the inherent complexities of real-world social media language, such as informal syntax, slang, misspellings, abbreviations, and the prevalent use of emojis and emoticons [9]. These linguistic features contribute to the challenging yet realistic nature of the sentiment classification task. The distribution of sentiment classes within both datasets is visually depicted in Figure 2, illustrating their diversity and balance.

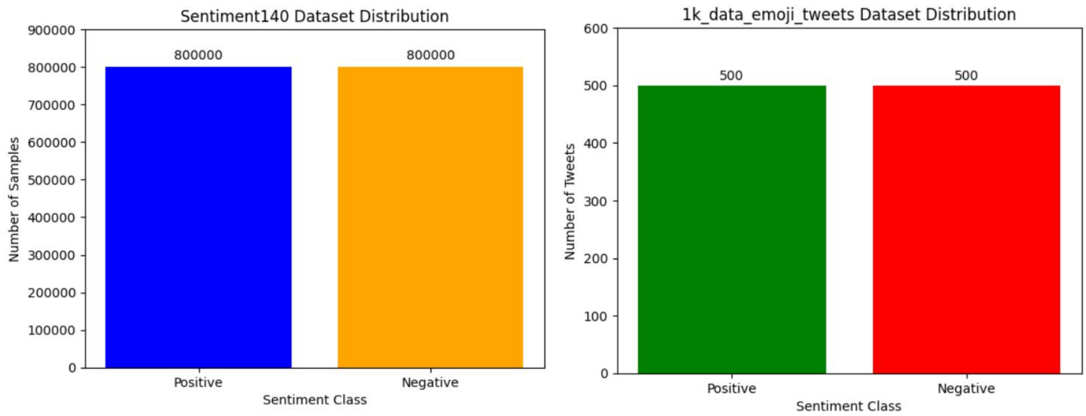

Figure 2. Sample of the dataset's distribution labelled

B. Text Preprocessing

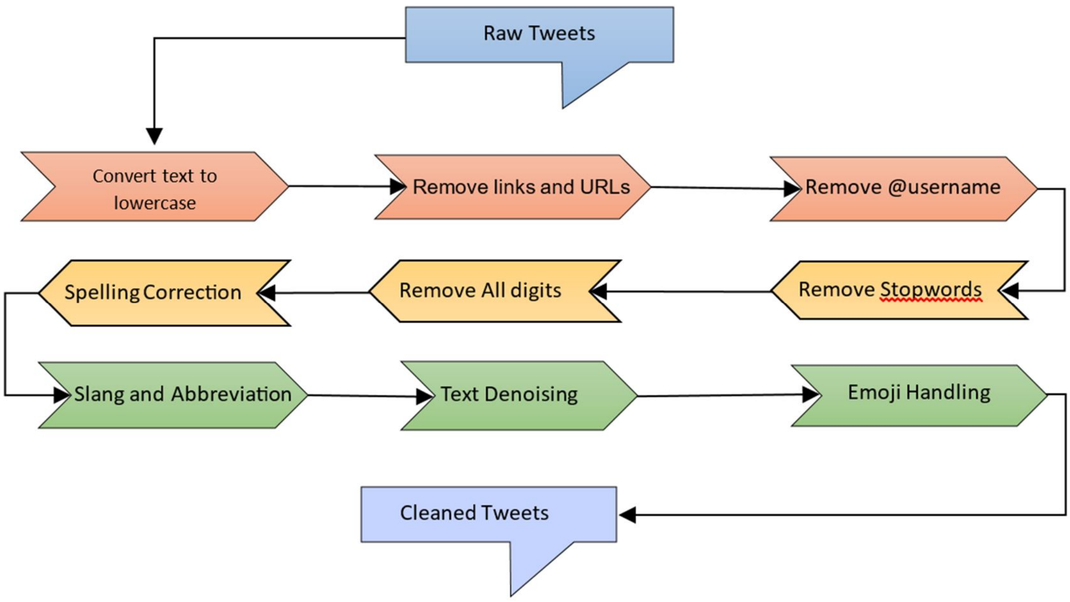

Figure 3. The Data preprocessing pipeline for datasets

Text preprocessing involves using various techniques to clean, format and prepare raw textual data for optimal utilization in natural language processing (NLP) or machine learning (ML) tasks. The primary objective of text preprocessing is to enhance the quality and usefulness of the text data, making it ready for subsequent modelling and analysis [10]. Throughout the preprocessing steps, we performed various operations that refined the dataset by addressing specific challenges associated with raw text. To prepare text for sentiment analysis and ensure reliable model performance, a comprehensive preprocessing pipeline is implemented [11]. In this pipeline, text lowercased, irrelevant elements such as URLs, usernames, and numbers are removed, stop words are removed, spelling correction is applied, slang and abbreviations are expanded, unwanted characters and symbols are removed and emojis are handled carefully by removing unnecessary ones and converting sentiment-relevant emojis into keywords. By ensuring clean, consistent and optimized input text, this painstaking preprocessing step reduces noise and improves the model's capacity to precisely interpret and categorize sentiment. In Figure 3 shows the Text Preprocessing pipeline used during experiments.

After RoBERTa generates contextual embeddings, a Gated Recurrent Unit (GRU) is used to capture temporal dependencies in sequential text. GRUs excel at processing noisy social media data by leveraging their gating mechanisms to retain relevant information and discard noise [1][2]. This allows GRUs to handle fragmented and diverse sentence structures, capturing sentiment shifts within posts or threads. The integration of RoBERTa and GRU provides a powerful hybrid approach for sentiment analysis, particularly in handling the challenges of noisy social media data. Roberta robust contextual embeddings, fine-tuned on domain-specific datasets, effectively capture the semantic nuances and informal language features prevalent in social media posts. GRU complements this by modeling sequential dependencies and understanding sentiment evolution across text, even in fragmented or noisy sequences. Together, this hybrid model bridges the gap between contextual understanding and temporal sentiment flow, offering a scalable and robust solution for real-world sentiment analysis. While the RoBERTa-GRU hybrid model demonstrates high accuracy and resilience to noisy data, challenges such as computational complexity and further exploration of multilingual adaptability remain areas for future research [12]. By addressing these limitations and extending evaluations to more diverse datasets, the proposed approach can set a new benchmark for sentiment analysis in dynamic and informal data environments, making it a valuable tool for real-world applications [13].

A powerful transformer-based model for sentiment analysis and natural language interpretation is called the Robustly Optimized BERT Pretraining Approach (RoBERTa). By using methods like dynamic masking, bigger batch sizes, and in-depth training on a variety of datasets, RoBERTa outperforms BERT. According to Tan et al. (2023) [1], these developments make it a strong option for managing noisy and unstructured social media data [14]. Fine-Tuning RoBERTa is further refined on domain-specific noisy datasets such as Sentiment140, Twitter Airline Sentiment, and YouTube comments after being pre-trained on extensive general-purpose datasets. RoBERTa's settings are modified during this adaptation process in order to better capture the colloquial language, emoticons, misspellings, and abbreviations found in social media content [2][15]. For example, RoBERTa learns vocabulary and hashtag usage patterns from Twitter data, and it adjusts to conversational tones and sentiment-related emojis from YouTube reviews. One of RoBERTa's main advantages is its capacity to produce dynamic word representations depending on context. For example, the term "lit" might be neutral in one context ("lit the candle") but positive in another ("this party is lit 🔥"). Since feelings are frequently inferred from minor linguistic clues in noisy social media data, this contextual understanding made possible by RoBERTa's transformer architecture is essential [16]. By leveraging RoBERTa’s ability to generate rich contextual embeddings tailored to noisy social media data, the model provides a robust foundation for sequential analysis, which is further enhanced by the GRU layer to capture temporal dependencies and sentiment evolution. Figure 4 illustrates the main components of the GRU unit.

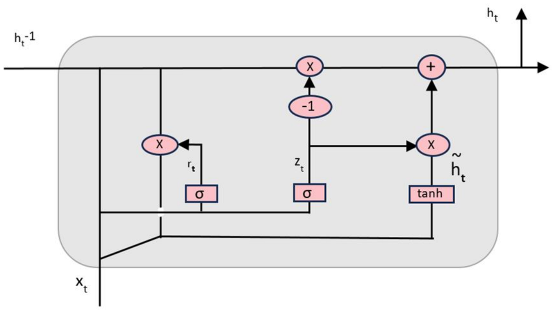

Figure 4. The architecture of Gated Recurrent Unit diagram

The calculations within the GRU unit at each time step t are as follows:

$$zt = \sigma(Wzxt + Uzht-1 + bz)$$

$$rt = \sigma(Wr\, xt + Urht-1 + br)$$

$$ct = tanh(Wcxt + Uc(rt * ht-1) + bc)$$

$$ht = zt * ht-1 + (1 - zt) * ct$$

The reset gate (rt) in a GRU determines the extent to which past information is disregarded when computing the candidate hidden state (ct), while the update gate (zt) regulates the balance between incorporating new inputs and retaining historical data [1]. This mechanism enables the GRU to effectively model long-term dependencies and adapt to noisy social media input through its final hidden state (ht), which integrates both the candidate and previous hidden states. Subsequently, this output is propagated to subsequent layers for further processing. GRUs process the embeddings generated by RoBERTa to capture temporal dependencies and sentiment evolution.

The gating mechanisms employed in this model are highly effective at managing noisy and unstructured data, which makes them particularly well-suited for analyzing social media text. This integration leverages the benefits of both transformer and recurrent neural network (RNN) architectures. In the final step, a dense layer with a softmax activation function classifies the processed sequential data into different sentiment categories. This layer combines the features extracted from RoBERTa and Gated Recurrent Units (GRUs), transforming them into probability distributions for each sentiment class.

To assess the model's performance on noisy benchmark datasets, we employ metrics such as accuracy, precision, recall, and F1-score. Following the GRU's processing of the sequential embeddings from RoBERTa, the dense layer integrates these features within a fully connected network. The softmax activation function then converts the resulting scores into probability distributions across the sentiment categories. By fine-tuning its weights, the dense layer can adapt to the complexities of noisy social media text, enhancing the model's ability to accurately classify sentiment, even in informal and fragmented data.

C. Model Evaluation

In the evaluation of transformer-based models for sentiment analysis a range of assessment criteria are commonly utilized to determine model efficacy and effectiveness. These metrics derived from the provided data, enable the quantification of the model's predictive performance with respect to software quality attributes. The assessment typically encompasses the following measures:

Accuracy: It gauges how accurate the model's predictions are overall.

$$Accuracy\ (A) = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Sensitivity (Recall): It measures the percentage of accurate positive predictions among all the model's positive predictions.

$$Recall\ (R) = \frac{TP}{(TP + FN)}$$

Precision: It calculates the proportion of true positive predictions in relation to all actual positive cases in the dataset.

$$Precision\ (P) = \frac{TP}{(TP + FP)}$$

F1-Score: It offers an objective evaluation of a model's effectiveness by showcasing the harmonic mean of precision and recall.

$$F1 = \frac{(2 \times P \times R)}{(P + R)}$$

Where TP, TN, FP, and FN refer to True Positive, False Positive, True Negative, and True Negative, respectively.

IV. EXPERIMENT SETUPS AND PARAMETERS

This Experiments were conducted in a Jupyter Notebook environment and utilizing a high-performance computing environment with a 96-core Intel Xeon Platinum 8168 CPU and NVIDIA GPU acceleration Model development implemented in Python the model was trained using the Adam optimizer, 256 embedding dimensions, and a batch size of 64 for 20 epochs (maximum 30). To avoid overfitting, we implemented a dropout rate of 0.25 along with early stopping. The model was trained using a learning rate of 1e-5, applying ReLU activation in the fully connected layer and SoftMax for sentiment classification. You can find the detailed settings of the model parameters in Table 3.

TABLE III. MODEL HYPERPARAMETERS

Parameter	Values
Optimizer	Adam
Dimension	256
Batch size	64
Dropout	0.25
Max epochs	20
Learning rate	1e-5
Activation function (fully connected layer)	ReLu
Activation function (classification layer)	SoftMax

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we present the outcome of the experiment results of the BERT, RoBERTa, BERT-GRU and RoBERTa-GRU on the Sentiment140, and ONEkemojitweets_senti_posneg Sentiment datasets that demonstrate the proposed model achieves high accuracy in sentiment classification. Specifically, the model attains an accuracy of 90% and Precision 90.61%, Sensitivity 90.34% also we get the f1-score is 66.0% on the ONEkemojitweets_senti_posneg dataset, highlighting its effectiveness in detecting sentiment from social media text. This performance exceeds that of conventional machine learning methods and closely aligns with the latest advancements in deep learning models. For a detailed comparison of the experimental results, please refer to Tables 4 and 5.

TABLE IV. THE COMPARISON RESULTS ON THE ONEKEMOJITWEETS_SENTI_POSNEG.

Methods	Accuracy	Precision	Recall	F1-Score
KNN	60.39	66	60	57
BERT	79.82	69	61	61
GRU	87.88	88	88	88
RoBERTa-LSTM	85.69	86	86	86
RoBERTa-GRU	90.00	90	90	90

TABLE V. THE COMPARISON RESULTS ON THE SENTIMENT140 DATASET

Methods	Accuracy	Precision	Recall	F1-Score
Naive Bayes	76.57	77	77	77
KNN	60.39	66	60	57
GRU	78.96	78	78	78
LSTM	79.10	79	79	79
BiLSTM	78.53	78	78	78
RoBERTa-GRU	91.59	91	91	91

Table 5 provides a comparison of various machine learning and deep learning techniques applied to the Sentiment140 dataset. Notably, the RoBERTa-GRU model achieves the highest accuracy at 91%, outperforming all other methods. This indicates that the RoBERTa-GRU model is particularly adept at predicting sentiment in text compared to its counterparts. One key factor contributing to the RoBERTa-GRU's exceptional performance is that RoBERTa has been pretrained on a significantly larger corpus and exposed to a wider variety of pretraining tasks than models like LSTM and GRU. This extensive training enhances its ability to generate better sentence embeddings and understand context, allowing the RoBERTa-GRU to identify more subtle patterns in the data and deliver more precise predictions. Additionally, the GRU's gating mechanism is designed to selectively update its hidden state based on both the current input and the previous hidden state. Research has shown that GRUs often outperform LSTMs when dealing with data characterized by short-term dependencies.

Given that the Sentiment140 dataset consists primarily of brief tweets, GRUs are particularly well-suited for this task.

VI. CONCLUSIONS

This research study presents the RoBERTa-GRU model, an innovative approach that effectively combines two leading-edge deep learning techniques for natural language processing: RoBERTa and GRU. RoBERTa, part of the Transformer family, uses dynamic attention masking to create meaningful text embeddings, which significantly boosts the model's capacity to generalize. On the other hand, the GRU model excels at capturing long-term dependencies in text sequences, with its gating mechanism successfully tackling the vanishing gradient problem often encountered by traditional Recurrent Neural Networks (RNNs). Furthermore, the RoBERTa-GRU model utilizes RoBERTa's contextual embedding strategy across various datasets, resulting in a more nuanced understanding of word meanings that change with context. Experimental outcomes from the Sentiment140 and ONEkemojitweets_senti_posneg datasets reveal that the RoBERTa-GRU model outperforms all other comparison methods, achieving remarkable accuracy scores of 91% and 90%. This powerful integration of RoBERTa and GRU leads to a highly effective model for sentiment analysis, making it a promising option for numerous NLP tasks.

REFERENCES

[1] K. L. Tan, C. P. Lee, and K. M. Lim.: RoBERTa-GRU Hybrid Model Performance on Sentiment Datasets. Applied Sciences, vol. 13, no. 6, p. 3915, Jan. 2023.
[2] M. M. Rahman, S. A. Islam, Y. Watanobe, and M. A. Alam.: RoBERTa-BiLSTM: A Context-Aware Hybrid Model for Sentiment Analysis. arXiv (Cornell University), Jun.2024.
[3] Hernández-Rubio M, Cantador I, Bellogín A. A comparative analysis of recommender systems based on item aspect opinions extracted from user reviews. User Model User-Adapt Interact 2019;29(2):381–441.
[4] Jianqiang Z, Xiaolin G, Xuejun Z. Deep convolution neural networks for twitter sentiment analysis. IEEE Access 2018;6:23253–60.
[5] Chauhan UA, Afzal MT, Shahid A, Moloud A, Basiri ME, Xujuan Z. A comprehensive analysis of adverb types for mining user sentiments on amazon product reviews. World Wide Web 2020;23(3):1811–29.
[6] Zhang, L.; Wang, S.; Liu, B.: Deep learning for sentiment analysis: a survey. WIREs Data Min. Knowl. Discov. 8(4), e1253 (2018). https://doi.org/10.1002/widm.1253.
[7] Kamruzzaman, M.M.: Arabic sign language recognition and generating Arabic speech using convolutional neural network. Wireless Commun. Mob. Comput. 2020, 3685614 (2020). https://doi.org/10.1155/2020/3685614.
[8] Heikal, M.; Torki, M.; El-Makky, N.: Sentiment analysis of Arabic tweets using deep learning. Proc. Comput. Sci. 142, 114–122 (2018). https://doi.org/10.1016/j.procs.2018.10.466.
[9] Cliche, M.: BB_twtr at SemEval-2017 task 4: twitter sentiment analysis with CNNs and LSTMs. In: The Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), Association for Computational Linguistics, Vancouver, Canada, pp. 573–580 (2017) https://doi.org/10.18653/v1/S17-2094
[10] Wang, X.; Jiang, W.; Luo, Z.: Combination of convolutional and recurrent neural network for sentiment analysis of short texts. In: The Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, The COLING 2016 Organizing Committee, Osaka, Japan, pp. 2428–2437 (2016)
[11] Araque, O.; Corcuera-Platas, I.; Sánchez-Rada, J.F.; Iglesias, C.A.: Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Syst. Appl. 77, 236–246 (2017). https://doi.org/10.1016/j.eswa.2017.02.002.
[12] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, ''RoBERTa-LSTM: A hybrid model for sentiment analysis with transformer and recurrent neural network,'' IEEE Access, vol. 10, pp. 21517–21525, 2022.
[13] Biswas S. Advantages of deep learning, plus use cases and examples.https://www.width.ai/post/advantages-of-deep learning. Accessed 10 Nov 2021.
[14] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188, 2014.
[15] Socher R, et al. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing; 2013.
[16] Do HH, et al. Deep learning for aspect-based sentiment analysis: a comparative review. Expert Syst Appl. 2019;118:272–99.
[17] Liu N, Shen B. Aspect-based sentiment analysis with gated alternate neural network. Knowl-Based Syst. 2020;188: 105010.
[18] Akhtar MS, et al. A multilayer perceptron based ensemble technique for fine-grained financial sentiment analysis. In: Proceedings of the 2017 conference on empirical methods in natural language processing. 2017.
[19] Pan Y-F, Hou X, Liu C-L. Text localization in natural scene images based on conditional random field. In: 2009 10th international conference on document analysis and recognition. IEEE; 2009.

[20] Ay Karakuş B, et al. Evaluating deep learning models for sentiment classification. Concurrency Computat Pract Exper. 2018;30(21): e4783.
[21] Hong J, Fang M. Sentiment analysis with deeply learned distributed representations of variable length texts. Stanford: Stanford University Report; 2015. p. 1–9.
[22] Bhattacharya A. Deep hybrid learning—a fusion of conventional ML with state of the art DL. https://towardsdatascience.com/deep-hybrid-learning-a-fusion-of-conventional-ml-with-state-of-the-art-dl-cb43887fe14. Accessed 26 Jul 2020.
[23] M. Wongkar and A. Angdresey, ''Sentiment analysis using naive Bayes algorithm of the data crawler: Twitter,'' in Proc. 4th Int. Conf. Informat. Comput. (ICIC), Oct. 2019, pp. 1–5.
[24] K. Dhola and M. Saradva, ''A comparative evaluation of traditional machine learning and deep learning classification techniques for sentiment analysis,'' in Proc. 11th Int. Conf. Cloud Comput., Data Sci. Eng.,Jan. 2021, pp. 932–936.
[25] M. U. Salur and I. Aydin, ''A novel hybrid deep learning model for sentiment classification,'' IEEE Access, vol. 8, pp. 58080–58093, 2020.
[26] Trang, Nguyen Thi Mai. "A Hybrid Model of RoBERTa and Bidirectional GRU for Enhanced Sentiment Analysis." International Journal of Open Information Technologies 11.11 (2023): 19-24.
[27] Hossen, M.S.; Jony, A.H.; Tabassum, T.; Islam, M.T.; Rahman, M.M.; Khatun, T. Hotel review analysis for the prediction of business using deep learning approach. In Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021; pp. 1489–1494.
[28] Li, X., Wang, H., Xiu, P., Zhou, X., & Meng, F. (2022, August). Resource usage prediction based on BiLSTM-GRU combination model. In 2022 IEEE international conference on joint cloud computing (JCC) (pp. 9-16). IEEE.
[29] K. L. Tan, C. P. Lee, and K. M. Lim, "RoBERTa-GRU: A Hybrid Deep Learning Model for Enhanced Sentiment Analysis," Applied Sciences 2023, Vol. 13, Page 3915, vol. 13, no. 6, p. 3915, Mar. 2023, doi: 10.3390/APP13063915.
[30] Saad Ahmed Sazan, M. Ahmed, T. B. Saad, and M. Roy.: Advanced Natural Language Processing Techniques for Efficient Sentiment Analysis of US Airline Twitter Data: A High-Performance Framework for Extracting Insights from Tweets. pp. 01–06, May 2024. https://doi.org/10.1109/iceeict62016.2024.10534511
[31] Wang, Yicai. "Chinese Text Classification Based on A hybrid Model of CNN and BiGRU." Proceedings of the 5th International Conference on Computer Information and Big Data Applications. 2024.