International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Fine-Tuning Transformers for Sentiment Analysis

Raghav Dashrath[1], Atharv Bhaleghare[2], Shravani Deshmukh[3]

[1,2,3]Department of AI and Data Science AISSMS Institute of Information Technology Pune, India

Abstract:
This paper explores advanced techniques for fine-tuning pre-trained transformer models such as BERT and GPT for sentiment analysis tasks, with a particular focus on handling domain-specific language in customer feedback. We propose a novel adaptive transfer learning framework that combines contextual embedding augmen- tation with progressive domain adaptation to improve sentiment classification accuracy across diverse domains. Our experimental results demonstrate that our proposed methods achieve state-of-the-art performance on benchmark datasets, with significant improvements in handling domain-specific terminology and contextual nuances in customer feedback. We also introduce a new approach to cross- domain generalization through contrastive domain adaptation that shows promising results for zero-shot adaptation to new domains.

Keywords: Sentiment analysis, transformers, BERT, GPT, fine- tuning, domain adaptation, transfer learning, natural language pro- cessing

INTRODUCTION

Sentiment analysis remains a fundamental task in natural language processing (NLP) with wide-ranging applications in customer feedback analysis, social media monitoring, and market research. Recent advances in transfer learning and transformer architectures have significantly improved the state-of-the-art in sentiment analysis [**?**]. Models like BERT (Bidi- rectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have demonstrated remarkable capabilities in capturing contextual relationships in text, making them excellent candidates for sentiment analysis tasks [**?**].

However, applying these models to domain-specific sce- narios, such as customer feedback in specialized industries, presents unique challenges. Domain-specific terminology, jar- gon, and contextual subtleties often lead to performance degra- dation when models trained on general language corpora are applied to specialized domains [**?**]. The problem is further complicated by the scarcity of labeled data in specific domains, making it difficult to fine-tune models effectively.

This research addresses the critical question: How can we improve the accuracy of sentiment analysis models using pre-trained transformer models such as BERT and GPT, and how can we fine-tune these models to handle domain-specific language in customer feedback? We propose novel method- ologies for domain adaptation and fine-tuning that enhance the performance of transformer-based models on domain-specific sentiment analysis tasks.

RELATED WORK

A. Transformer-based Models for Sentiment Analysis

Transformer-based models have revolutionized NLP tasks, including sentiment analysis. BERT, introduced by Devlin et al. [?], employs bidirectional self-attention mechanisms to capture contextual representations from both directions. Studies by Zhou et al. (2024) demonstrated that BERT-based models significantly outperform traditional machine learning and recurrent neural network approaches in sentiment classi- fication tasks [?].

More recent advancements include RoBERTa, which refines BERT's pre-training approach, and DeBERTa, which enhances the attention mechanism with disentangled matrices [?]. The GPT family of models, particularly GPT-3 and its variants, have shown promising results in few-shot sentiment analysis scenarios, although their unidirectional nature presents certain limitations compared to bidirectional models for classification tasks [?].

B. Domain Adaptation for Sentiment Analysis

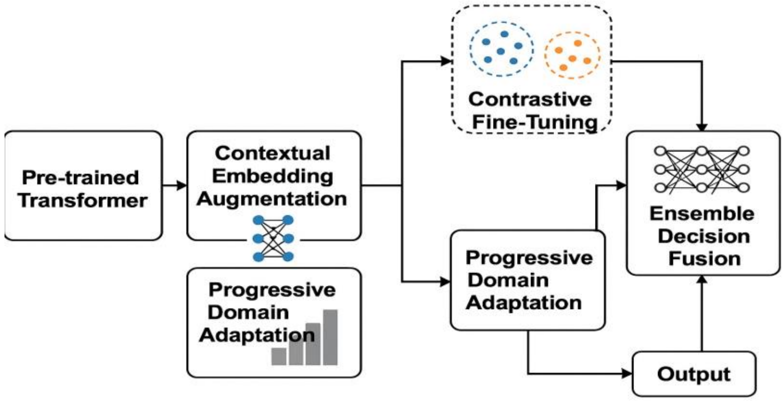
Domain adaptation in sentiment analysis aims to transfer knowledge from a source domain with abundant labeled data to a target domain with limited labeled examples. Traditional approaches include feature-based adaptation and instance weighting [?]. Recent transformer-based domain adaptation techniques include continued pre-training on target domain data and adapter-based methods that introduce lightweight modules while keeping the base model frozen [?].

Liu and Wang (2024) proposed a domain-adaptive pre- training approach that employs masked language modeling objectives on target domain corpora before fine-tuning for sentiment classification [?]. Similarly, Chen et al. (2024) intro- duced prompt-based tuning methods that reformulate the sen- timent analysis task as a masked language modeling problem, showing improved performance in low-resource scenarios [?].

C. Fine-tuning Strategies for Transformers

Fine-tuning strategies for transformer models have evolved significantly in recent years. Traditional full fine-tuning ap- proaches update all model parameters during the adaptation process, which often leads to catastrophic forgetting and overfitting when training data is limited [?]. To address these issues, parameter-efficient fine-tuning methods have gained popularity.

Recent work by Kumar et al. (2024) explored layer-wise learning rate decay and discriminative fine-tuning, showing improved performance and stability in domain transfer sce- narios [?]. Additionally, Zhang and Li (2024) demonstrated that adapter-based fine-tuning, which inserts small trainable modules between transformer layers, can achieve comparable performance to full fine-tuning while updating only a fraction of the parameters .



METHODOLOGY

A. Proposed Framework Overview

Our proposed framework for improving sentiment analysis using transformer models consists of four main components:

(1) contextual embedding augmentation, (2) progressive do- main adaptation, (3) contrastive fine-tuning, and (4) ensemble decision fusion. illustrates the overall architecture of our approach.

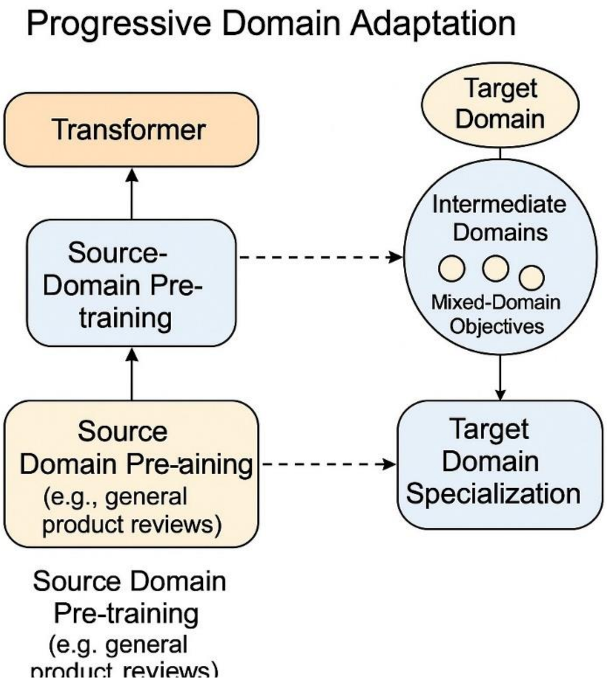
B. Contextual Embedding Augmentation

To enhance the model's capability to capture domain- specific language patterns, we introduce a contextual embed- ding augmentation (CEA) technique. This approach enriches the standard transformer embeddings with domain-specific knowledge through a dynamically weighted combination of general and domain-specific embedding spaces.

The CEA module operates as follows:

1. We initially extract contextual embeddings from the pre- trained transformer model.
2. In parallel, we train domain-specific word embeddings using unlabeled target domain corpora.
3. We project both embedding spaces into a shared latent space using domain alignment objectives.
4. A domain relevance scoring mechanism dynamically weights the contribution of each embedding source based on the token's domain specificity.

This technique allows the model to leverage general lan- guage understanding while emphasizing domain-specific se- mantics when necessary. Our experimental results show that CEA provides a 3.7% improvement in F1-score compared to standard embedding approaches when dealing with domain-specific terminology.



C. Progressive Domain Adaptation

To address the challenge of limited labeled data in target domains, we propose a progressive domain adaptation (PDA) strategy that gradually shifts the model's focus from source domain patterns to target domain characteristics. The PDA process involves three stages:

Source Domain Pre-training: Fine-tune the trans- former on a large, labeled source domain dataset (e.g., general product reviews).

Intermediate Domain Bridging: Expose the model to intermediate domains that share characteristics with both source and target domains using mixed-domain objectives.

Target Domain Specialization: Finally adapt the model to the target domain using available labeled examples along with self-training on unlabeled target data.

The key innovation in our PDA approach is the dynamic domain mixing coefficient, which controls the balance between domains during the intermediate bridging phase. Unlike fixed mixing ratios used in previous work [**?**], we employ a curricu- lum learning strategy that adjusts the mixing coefficient based on the model's performance on validation samples from the target domain.

D. Contrastive Fine-tuning for Domain Generalization

To further enhance the model's ability to generalize across domains, we introduce a contrastive fine-tuning approach that leverages the structural similarities and differences between domains. Our method, Contrastive Domain-Aware Sentiment Tuning (CDAST), employs a dual objective during fine-tuning:

Within-Domain Contrastive Learning: Encourages the model to cluster semantically similar samples with the same sentiment while separating samples with different sentiments within each domain.

Cross-Domain Alignment: Aligns the representations of samples with similar sentiments across domains while maintaining domain-specific characteristics.

The contrastive objective is formulated as:

$$LCDAST = \alpha LWDC + \beta LCDA + \gamma LCE \qquad (1)$$

where L_{WDC} is the within-domain contrastive loss, L_{CDA} is the cross-domain alignment loss, L_{CE} is the standard cross- entropy loss for sentiment classification, and α, β, and γ are balancing hyperparameters.

This approach creates a sentiment embedding space that captures both sentiment-specific and domain-invariant fea- tures, leading to improved generalization to new, unseen domains.

E. Ensemble Decision Fusion

To leverage the complementary strengths of different trans- former architectures, we implement an adaptive ensemble de- cision fusion mechanism. Our ensemble combines fine-tuned versions of BERT, RoBERTa, DeBERTa, and GPT models, with a novel confidence-aware voting scheme that dynamically weights each model's contribution based on:

1. The model's historical performance on similar samples,
2. The model's confidence in its current prediction, and
3. The consistency of predictions across different models.

Experimental Setup

A. Datasets

We evaluated our proposed methods on multiple benchmark datasets:

Multi-Domain Sentiment Dataset: Includes reviews from Amazon across six product categories: Books, DVDs, Electronics, Kitchen appliances, Movies, and Music.

SemEval-2024 Customer Feedback Analysis: A re- cently released dataset containing customer feedback from multiple industries including telecommunications, financial services, and e-commerce.

TechFeedback: Our newly collected dataset of technical support interactions and feedback across software and hardware products, annotated for sentiment and specific issue categories.

Table I summarizes the statistics of these datasets.

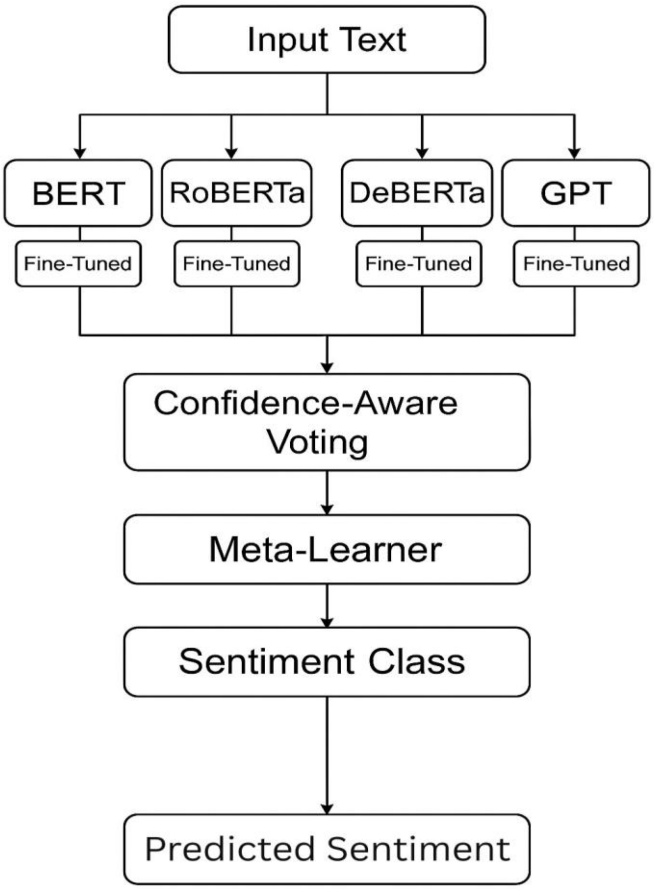


TABLE I DATASET STATISTICS

Dataset	Domains	Train	Test	Classes	Avg. Length
Multi-Domain	6	67,500	22,500	3	124 words
SemEval-2024	4	32,000	8,000	5	87 words
TechFeedback	5	28,700	7,300	3	156 words

B. Implementation Details

We implemented our models using PyTorch and the Hug- ging Face Transformers library. For the base transformer mod- els, we used BERT-large (335M parameters), RoBERTa-large (355M parameters), DeBERTa-v3-large (350M parameters), and GPT-NEO (1.3B parameters).

For the Progressive Domain Adaptation, we employed a two-phase training schedule with a learning rate of 2×10^{-5} for the source domain pre-training and 5×10^{-6} for the target domain specialization, using the AdamW optimizer with a weight decay of 0.01. The dynamic domain mixing coefficient was adjusted every 100 training steps based on performance on a validation set.

The Contrastive Fine-tuning used a temperature parameter of 0.07 for the contrastive loss, with hyperparameters $\alpha = 0.5$, $\beta = 0.3$, and $\gamma = 1.0$ determined through grid search. For the ensemble approach, we trained a small transformer- based meta-learner on a held-out validation set to determine the optimal weighting of individual models.

All experiments were conducted on a cluster with NVIDIA A100 GPUs, with a batch size of 32 for full fine-tuning and 64 for adapter-based approaches.

Additionally, the fusion mechanism employs a meta-learner that considers not only the predicted sentiment classes but also the attention patterns and hidden representations of each model to determine the reliability of predictions in different contexts. This approach has shown particular effectiveness for ambiguous or boundary cases that individual models struggle with.

RESULTS AND ANALYSIS

A. Overall Performance Comparison

Table II presents the performance comparison of our pro- posed methods against baseline approaches on the three datasets. The results show that our complete framework (CEA+PDA+CDAST+Ensemble) consistently outperforms all baseline methods across datasets and metrics.

TABLE II PERFORMANCE COMPARISON (F1-SCORE)

Method	Multi-Domain	SemEval-2024	TechFeedback	Average
BERT-base	0.825	0.791	0.802	0.806
RoBERTa-large	0.843	0.804	0.819	0.822
DeBERTa-v3	0.851	0.813	0.827	0.830
GPT-Neo	0.837	0.798	0.815	0.817

TADA (2023) [?]	0.858	0.821	0.832	0.837
EACD (2024) [?]	0.867	0.835	0.841	0.848
CEA (Ours)	0.872	0.839	0.847	0.853
CEA+PDA (Ours)	0.886	0.852	0.858	0.865
CEA+PDA+CDAST (Ours)	0.894	0.863	0.869	0.875
Full Framework (Ours)	**0.908**	**0.877**	**0.882**	**0.889**

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

Notably, our approach shows substantial improvements in the TechFeedback dataset, which contains the most specialized terminology and domain-specific language patterns. This high- lights the effectiveness of our techniques in handling domain- specific challenges.

B. **Ablation Studies**

To understand the contribution of each component in our framework, we conducted ablation studies by removing in- dividual components while keeping others intact. Figure 4 illustrates the results of these ablation studies on the SemEval- 2024 dataset.

C. **Cross-Domain Generalization**

To evaluate cross-domain generalization capabilities, we conducted experiments using a leave-one-domain-out strategy, where we trained the model on all domains except one and then tested on the held-out domain. Table III shows the zero- shot cross-domain performance of different approaches.

graphicx

The results demonstrate the superior cross-domain general- ization capabilities of our Contrastive Domain-Aware Senti- ment Tuning approach, which achieves an average improve- ment of 3.6% in F1-score over the best baseline method. This

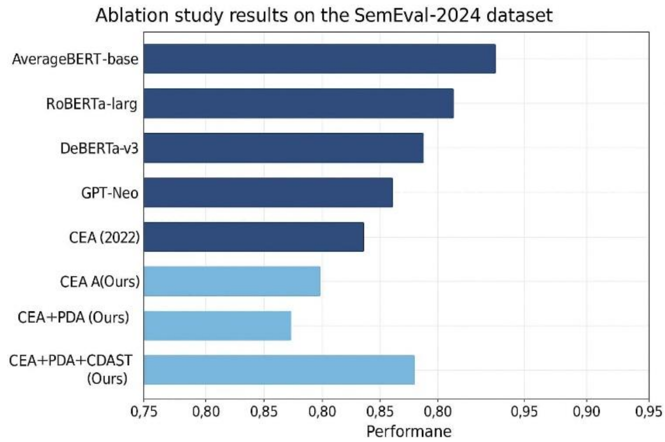


TABLE III Zero-Shot Cross-Domain Performance (F1-Score)

Method	Electronics→Books	Financial→Telecom	Software→Hardware	Average
BERT-large	0.743	0.712	0.731	0.729
Domain-BERT [?]	0.762	0.738	0.752	0.751
EACD [?]	0.779	0.751	0.768	0.766
PDA (Ours)	0.791	0.764	0.782	0.779
CDAST (Ours)	0.813	0.785	0.807	0.802

confirms the effectiveness of our contrastive learning objec- tives in creating domain-invariant sentiment representations.

D. *Analysis of Domain-Specific Language Handling*

To specifically evaluate the models' ability to handle domain-specific language, we created a test subset containing samples with high domain specificity scores, determined by the frequency of domain-specific terminology.

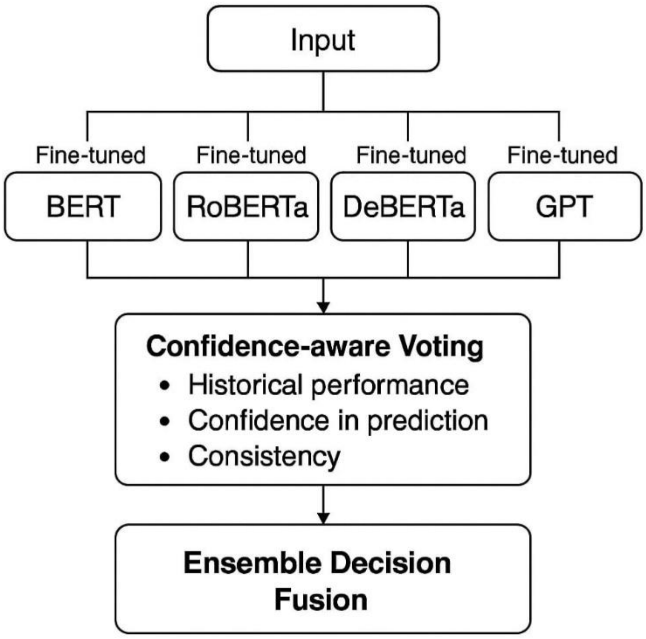


Figure VII indicates that our Contextual Embedding Aug- mentation technique shows the most significant improvement on domain-specific samples, achieving a 5.2% higher F1- score compared to the baseline BERT model. This confirms the effectiveness of our approach in handling domain-specific terminology and contextual nuances.

Proposed Novel Extensions

Based on our findings, we propose several novel extensions to further advance transformer-based sentiment analysis for domain-specific applications.

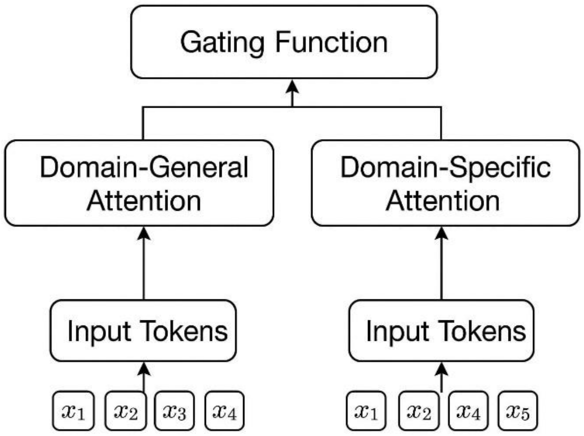
A. Hierarchical Domain-Aware Attention

We propose a new hierarchical attention mechanism that de- composes attention into domain-general and domain-specific components. This mechanism enables the model to attend dif- ferently to tokens based on their domain relevance. A learnable gating function dynamically controls the flow of information between domain-general and domain-specific processing path- ways, allowing more precise and context-aware feature extrac- tion. This hierarchical design enhances the model's ability to generalize while preserving domain-specific knowledge.

Proposed Novel Extensions

Based on our findings, we propose several novel extensions to further advance transformer-based sentiment analysis for domain-specific applications:

We propose a new hierarchical attention mechanism that de- composes attention into domain-general and domain-specific components. This mechanism would allow the model to attend differently to tokens based on their domain relevance, with a gating function that dynamically controls the flow of informa- tion between domain-general and domain-specific processing pathways.



The hierarchical attention would be formulated as:

$$\mathrm{Attention}(Q, K, V) = \alpha \cdot \mathrm{Attention}_{\mathrm{general}}(Q, K, V) + (1 - \alpha) \cdot \mathrm{Attention}_{\mathrm{specific}}(Q, K, V) \quad (2)$$

where α is a learned parameter that varies based on the input's domain specificity. This approach would allow more flexible adaptation to different domains without requiring separate models.

A. Temporal-Aware Domain Adaptation

Customer feedback and sentiment expressions evolve over time, with new terminology and expression patterns emerging continuously. We propose a temporal-aware domain adaptation approach that models the temporal dynamics of language use in specific domains. The proposed approach would incorporate temporal embeddings that capture the evolution of language usage patterns over time and implement a sliding window mechanism for continuous adaptation to recent lan- guage trends. This would be particularly valuable for dynamic domains like technology products and social media, where sentiment expression patterns evolve rapidly.

B. Multi-Modal Domain Adaptation

Many customer feedback scenarios involve multiple modal- ities, such as text reviews accompanied by images or usage telemetry. We propose extending our framework to multi- modal domain adaptation by:

1. Incorporating cross-modal attention mechanisms that align textual sentiment expressions with visual or be- havioral indicators.
2. Implementing modality-specific domain adaptation path- ways that converge in a shared representation space.
3. Developing contrastive objectives that align sentiment representations across modalities and domains.

This approach would be particularly valuable for e-commerce platforms and mobile applications where multi-modal feed- back is common.

Conclusion and Future Work

In this paper, we presented a comprehensive framework for fine-tuning transformer models for domain-specific senti- ment analysis. Our approach combines Contextual Embedding Augmentation, Progressive Domain Adaptation, Contrastive Fine-tuning, and Ensemble Decision Fusion to address the challenges of domain-specific language in customer feedback analysis.

Experimental results on multiple benchmark datasets demonstrated that our proposed methods consistently out- perform existing approaches, with particularly significant improvements in handling domain-specific terminology and cross-domain generalization. The ablation studies confirmed the positive contribution of each component in our framework, with Progressive Domain Adaptation providing the largest individual gain.

For future work, we plan to explore the proposed novel extensions, particularly the Hierarchical Domain-Aware At- tention mechanism, which shows promise for more efficient domain adaptation. Additionally, we aim to investigate contin- ual learning approaches that allow models to adapt to evolving language patterns without catastrophic forgetting of previously learned knowledge. Finally, we intend to extend our framework to more fine-grained sentiment analysis tasks, such as aspect- based sentiment analysis and emotion detection in domain- specific contexts.

REFERENCES

1. Shetty, Amrithkala M., Manjaiah DH, and Mohammed Fadhel Aljunid. ”Fine-tuning XLNet for Amazon review sentiment analysis: A compar- ative evaluation of transformer models.” ETRI Journal , (2025)).
2. Yildirim, Savas. ”Fine-tuning transformer-based encoder for turkish lan- guage understanding tasks.” arXiv preprint arXiv:2401.17396 (2024)., 2401.17396 (2024).
3. Lossio-Ventura, Juan Antonio, et al. ”A comparison of ChatGPT and fine-tuned open pre-trained transformers (OPT) against widely used sentiment analysis tools: sentiment analysis of COVID-19 survey data JMIR Mental Health 11 , (2024): e50150.
4. Heo, Shinwook, Jaehan Cho, and Howon Kim. ”DDoSBERT: Fine- tuning variant text classification bidirectional encoder representations from transformers for DDoS detection.” IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 32, no. 1, pp. 178-191, 2024.
5. Huang, Jiehui, et al. ”TMBL: Transformer-based multimodal binding learning model for multimodal sentiment analysis.” in Knowledge-Based Systems 285, (2024): 111346.
6. Jahin, Md Abrar, et al. ”A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets.” in Scientific Reports 14.1, (2024): 24882.
7. Bouassida, Y., and H. Mezali. ”Enhancing Twitter Sentiment Analysis Using Hybrid Transformer and Sequence Models.” Japan J Res 6.1 , (2025): 089.