A Transformer-Based Multi-Task Learning Framework for Sentiment, Emotion, and Sarcasm Analysis

Rohan Rajendra Chimbaikar
Department of Information Technology
D.G. Ruparel College of Arts, Science and Commerce
Mumbai, India

---***---

Abstract - A major challenge in natural language processing is determining the emotional intent of text because sentiment, emotion, and sarcasm all work together. Most contemporary methods deal with these tasks individually, which leads to incomplete representations and greater computational costs. This study provides a transformer-based multitask learning system that jointly performs sentiment classification, fine-grained emotion recognition, and sarcasm detection within a single architecture. The framework uses a shared RoBERTa-based encoder with task-specific classification heads to capture the contextual dependencies and subtle linguistic cues. Experimental results reveal that the proposed multitask technique delivers a robust and balanced performance across all tasks, highlighting its efficacy for comprehensive affective text analysis in real-world applications such as social media monitoring and customer feedback analysis.

Key Words: Sentiment Analysis; Emotion Recognition; Sarcasm Detection; Transformer Models; Multi-Task Learning

1. INTRODUCTION

The growing development of user-generated textual material across digital platforms has intensified the demand for automated systems that can interpret human affect in language. Online reviews, social media posts, discussion forums, and customer feedback channels provide significant insights into opinions, emotions, and behavioural intentions. Accurately interpreting such affective information is essential for applications such as opinion mining, content moderation, mental health monitoring, and intelligent human–computer interaction. However, natural language expressions of affect are often subtle, context-dependent, and ambiguous, making reliable analysis challenging.

Traditionally, sentiment analysis has focused on identifying the overall polarity of a text as positive, negative, or neutral. While this provides a coarse understanding of opinion, it fails to capture the rich emotional diversity expressed in the natural language. Emotion recognition extends sentiment analysis by identifying specific emotional states, such as joy, anger, fear, or sadness, enabling a deeper understanding of user intent and psychological context. Sarcasm detection further complicates affective analysis because sarcastic expressions frequently invert surface-level sentiment and rely heavily on contextual and pragmatic cues. Treating these tasks independently often results in fragmented interpretations and reduced robustness, particularly in informal or conversational texts.

Conventional machine learning approaches for affective text analysis rely on handcrafted features and shallow classifiers, which struggle to model long-range dependency and implicit linguistic patterns. Recent advances in deep learning, particularly transformer-based architectures, have significantly improved performance by leveraging self-attention mechanisms to capture the contextual relationships within the text. Pre-trained transformer models, such as RoBERTa, have demonstrated strong generalization capabilities across a wide range of natural language processing tasks. Despite these advancements, many existing transformer-based solutions are designed for single-task learning and do not fully exploit the interrelated nature of sentiments, emotions, and sarcasm.

Multi-task learning offers a promising paradigm for affective analysis by enabling models to learn shared representations across related tasks. Jointly learning sentiment, emotion, and sarcasm allows the model to capture complementary signals, such as emotional intensity influencing sentiment polarity or sarcastic cues that alter emotional interpretation. By sharing a common encoder while maintaining task-specific output layers, multitask frameworks can improve learning efficiency, reduce redundancy, and enhance generalization performance.

In this study, we present a transformer-based multitask deep learning framework that jointly performs sentiment classification, fine-grained emotion recognition, and sarcasm detection within a unified architecture. The proposed approach utilizes a shared RoBERTa-based encoder to extract contextualized text representations, followed by independent classification heads that are optimized for each task. The framework supports both real-time inference on individual text inputs and batch-level processing for large-scale datasets, enabling its practical applicability in real-world analytical systems.

The main contributions of this study are as follows.

- A unified multi-task learning framework for joint sentiment, emotion, and sarcasm analysis using transformer-based representations.
- An effective modeling strategy for fine-grained emotion recognition, alongside sentiment and sarcasm detection, within a single architecture.
- An end-to-end system design that supports scalable inference and analytical visualization for both single-text and batch-based inputs

2. RELATED WORK

Research on affective text analysis has evolved significantly over the past decade, driven by advances in natural language processing and the increasing availability of large-scale textual data. This section reviews prior work in sentiment analysis, emotion recognition, sarcasm detection, and multi-task learning approaches that integrate multiple affective dimensions

A. Sentiment Analysis

Early approaches to sentiment analysis primarily relied on lexicon-based methods and traditional machine learning classifiers such as support vector machines and Naïve Bayes models. These techniques utilize handcrafted features, including n-grams, part-of-speech tags, and sentiment lexicons, to infer polarity. While effective for well-structured text, such methods struggle with contextual ambiguity, informal language, and domain transferability.

The introduction of deep learning has significantly advanced sentiment analysis by enabling automatic feature extraction from raw text. Recurrent and convolutional neural networks improve performance by modelling sequential and local dependencies. Recently, transformer-based models have achieved state-of-the-art results by leveraging self-attention mechanisms to capture long-range contextual relationships. Pre-trained language models, particularly RoBERTa, have demonstrated strong generalization across diverse sentiment benchmarks owing to extensive pre-training on large corpora and improved optimization strategies.

B. Emotion Recognition

Emotion recognition in text extends beyond sentiment polarity to identify specific emotional states, often using multiclass classification frameworks. Early work in this domain relied on psychological emotion models and emotion lexicons, mapping words and phrases to predefined emotion categories. However, such approaches are limited by vocabulary coverage and the inability to capture contextual emotional shifts.

Deep learning-based emotion recognition models have addressed these limitations by learning the distributed representations of text. Transformer-based architectures are particularly effective in capturing nuanced emotional expressions, enabling fine-grained emotion classification. Despite these advancements, many studies have focused on a limited number of emotion categories and operated independently of sentiment and sarcasm, potentially overlooking shared affective cues.

C. Sarcasm Detection

Sarcasm detection remains one of the most challenging problems in affective computing because of its reliance on implicit meaning, contextual incongruity, and pragmatic understanding. Traditional approaches use surface-level features, such as punctuation patterns, sentiment contrast, and cue phrases. However, these methods often fail to generalize across domains and writing styles.

Neural approaches, including recurrent and transformer-based models, have improved sarcasm detection by modeling contextual dependencies and semantic incongruence in text. Pre-trained transformers have shown strong performance in capturing subtle contradictions between literal sentiment and intended meaning. However, sarcasm detection is frequently treated as an isolated task, limiting the ability to leverage emotional and sentiment-related signals that are inherently correlated with sarcastic expressions.

D. Multi-Task Learning for Affective Analysis

Multi-task learning has emerged as an effective strategy for jointly modeling related natural language processing tasks in recent years. By sharing representations across tasks, multitask frameworks can reduce overfitting, improve learning efficiency, and enhance generalization. Several studies have applied multitask learning to sentiment and emotion analyses, demonstrating improved performance compared with single-task baselines.

More recent studies have explored the joint learning of sentiment and sarcasm, or emotion and sarcasm, using shared neural architectures. Transformer-based multitask models have shown particular promise because of their ability to learn rich contextual representations. However, many existing approaches are limited to a subset of affective dimensions or employ shallow interactions. Comprehensive frameworks that jointly model sentiment, fine-grained emotion, and sarcasm within a unified transformer-based architecture remain relatively unexplored.

E. Research Gap

Although transformer-based models and multitask learning have individually advanced affective text analysis, there is a lack of unified frameworks that effectively integrate sentiment classification, fine-grained emotion recognition, and sarcasm detection within a single model. Existing approaches often focus on one or two affective dimensions, resulting in fragmented interpretations and redundant computations. This gap motivates the development of a transformer-based multitask framework capable of jointly capturing these interrelated affective signals.

3. PROPOSED METHODOLOGY

This section describes the proposed transformer-based multitask deep learning framework designed to jointly perform sentiment classification, emotion recognition, and sarcasm detection. The methodology focuses on leveraging shared contextual representations while preserving task-specific learning using independent output layers.

Problem Formulation

Let $x = \{w_1, w_2, \dots, w_n\}$ denote an input text sequence consisting of n tokens. The objective was to simultaneously predict the following:

- Sentiment label $y_s \in \{1, \dots, C_s\}$
- Emotion label $y_e \in \{1, \dots, C_e\}$
- Sarcasm label $y_c \in \{1, \dots, C_c$

where C_s, C_e, and $C_c$represent the number of sentiment, emotion, and sarcasm classes respectively. The model learns a joint mapping as follows:

$$f(x) \rightarrow (y_s, y_e, y_c) \qquad (1)$$

Shared Transformer Encoder

The framework employs a shared RoBERTa-based transformer encoder to generate contextual token representations. Given an input sequence, the encoder produces a set of hidden states as follows:

$$H = \{h_1, h_2, \dots, h_n\} \qquad (2)$$

where each hidden state $h_i \in \mathbb{R}^d$ represents the contextual embedding of the corresponding token. A pooled representation derived from the classification token was used as the global text representation.

Multi-Task Learning Objective

To enable joint optimization across tasks, the framework combines individual task losses into a single objective function. The cross-entropy loss is used for all tasks, and the total loss is defined as

$$\mathcal{L} = \lambda_s \mathcal{L}_s + \lambda_e \mathcal{L}_e + \lambda_c \mathcal{L}_c \qquad (3)$$

where $\mathcal{L}_s$, $\mathcal{L}_e$, and $\mathcal{L}_c$ denote sentiment, emotion, and sarcasm losses respectively, and λ terms are task-specific weighting coefficients.

Task – Specific Classification heads

To enable multitask learning, the shared encoder is followed by three independent classification heads, one for each task. Each head consists of a fully connected layer that maps the pooled encoder representation to the corresponding label space.

- Sentiment Head: Predicts sentiment polarity across three classes (positive, negative, and neutral) using a softmax activation function.
- Emotion Head: Performs fine-grained emotion classification across twenty-eight distinct emotion categories, enabling detailed affective interpretation.
- Sarcasm Head: Determines whether the input text is sarcastic or non-sarcastic, modelled as a binary classification task.

This design allows each task to learn specialized decision boundaries while benefiting from the shared linguistic features extracted by the encoder.

Inference and Deployment Workflow

During inference, the shared transformer encoder performs a single forward pass to simultaneously generate predictions for sentiment, emotion, and sarcasm. This unified inference mechanism significantly reduces the computational overhead compared with deploying separate models for each task. The framework supports both real-time analysis of individual text inputs and efficient batch-level processing of large datasets, making it well suited for scalable deployment in practical analytical systems.

4. SYSTEM ARCHITECTURE

This section presents the high-level architecture of the proposed transformer-based multitask affective analysis system, focusing on the inference pipeline and interaction between major functional components. The architecture was designed to support the efficient joint prediction of sentiment, emotion, and sarcasm while maintaining scalability and practical deploy ability.

Architectural Overview

The overall architecture of the proposed transformer-based multitask framework is illustrated in Fig. 1.

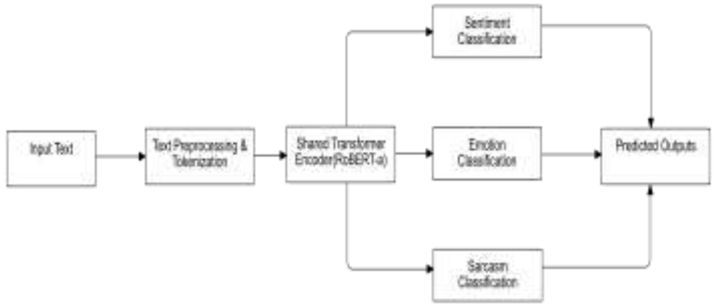

Fig. 1. *High-level architecture of the proposed transformer-based multi-task framework for joint sentiment, emotion, and sarcasm analysis*

The system follows a modular pipeline architecture in which user-provided text is processed through a sequence of well-defined stages: input handling, preprocessing, multitask model inference, and output generation. The architecture is intentionally lightweight and model-centric, ensuring efficient production of affective predictions in a single forward pass.

Input Handling and Preprocessing

Textual input may consist of individual sentences or batch datasets. All inputs were first passed through a preprocessing stage that performed text normalization and tokenization consistent with the transformer encoder requirements. The preprocessing component ensures a uniform input representation by applying sequence truncation and padding, thereby enabling stable and efficient batch processing.

Multi – Task Inference Engine

At the core of the system is the multi-task inference engine, which encapsulates the transformer-based model described in the proposed methodology. A shared RoBERTa-based encoder generates contextualized representations of the input text, which are simultaneously consumed by three task-specific classification heads for sentiment, emotion, and sarcasm prediction.

By performing joint inference within a single forward pass, the system minimizes computational redundancy and latency compared with deploying independent models for each task. This design is particularly advantageous for real-time applications and large-scale text analysis.

Output Generation and Aggregation

The outputs generated by the task-specific classification heads consisted of predicted class labels along with their corresponding confidence scores for sentiment, emotion, and sarcasm analyses. For batch-level inputs, individual predictions were aggregated to compute summary statistics, such as class-wise distributions, frequency counts, and overall prediction trends. These aggregated results enable large-scale analytical insights and facilitate comparative analyses across datasets. Structured outputs can be seamlessly integrated with downstream analytical pipelines or visualization modules, supporting applications such as reporting dashboards, trend analysis, and decision-support systems.

Architectural Characteristics

The proposed architecture demonstrates several key characteristics that enhance its effectiveness and practical applicability.

- **Efficiency:** A single-pass multi-task inference mechanism significantly reduces the computational overhead by eliminating the need for separate models for each task.
- **Scalability:** The framework supports real-time analysis of individual text inputs and batch-level processing of large datasets.
- **Modularity:** A clear separation between preprocessing, model inference, and output generation enables easier maintenance and extensibility of the system.
- **Deploy ability:** The architecture is designed for seamless integration into real-world analytical pipelines using service-based or API-driven interfaces.

5. EXPERIMENTAL SETUP

This section describes the datasets, training configuration, evaluation metrics, and implementation details used to assess the performance of the proposed transformer-based multitask framework for sentiment, emotion, and sarcasm analyses.

Datasets

The model was trained and evaluated on annotated textual datasets containing labels for sentiment polarity, emotional categories, and sarcasm. Each input sample was associated with three labels corresponding to the respective affective tasks. The sentiment classification task consists of three classes (positive, negative, and neutral), whereas the emotion recognition task involves twenty-eight fine-grained emotion categories. Sarcasm detection is formulated as a binary classification problem with sarcastic and non-sarcastic labels.

The set of fine-grained emotion categories used for emotion recognition is summarized in Table I.

TABLE I. EMOTION CATEGORIES USED FOR FINE –GRAINED EMOTION RECOGNITION

Emotion Classes
Admiration, amusement, anger, annoyance
Approval, caring, confusion, curiosity
Desire, disappointment, disapproval, disgust
Embarrassment, excitement, fear, gratitude
Grief, joy, love, nervousness
Optimism, pride, realization, relief
Remorse, sadness, surprise, neutral

The datasets comprise user-generated textual content collected from diverse sources, including conversational text and informal language, enabling the model to learn robust representations across different linguistic styles. Standard preprocessing steps were applied to ensure consistency across samples.

The datasets employed in this study were selected to comprehensively support the three affective analysis tasks addressed by the proposed multitask framework. Benchmark sentiment datasets were used to capture sentiment polarity in both formal reviews and informal social media texts. For emotion recognition, large-scale emotion-annotated corpora were incorporated to enable fine-grained classification across diverse emotional categories. Sarcasm detection was supported using multiple sarcasm- and irony-labeled datasets derived from social media and online discussion platforms, including benchmark tweet-based datasets and a Reddit-based sarcasm corpus. This combination enables the model to learn sarcastic expressions across different discourse styles and conversational contexts.

The datasets used for sentiment analysis, emotion recognition, and sarcasm detection are listed in Table II.

TABLE II. DATASETS USED FOR SENTIMENT, EMOTION AND SARCASM ANALYSIS

Task	Dataset Name(s)	Domain	Classes
Sentiment Analysis	Amazon Reviews Polarity, Sentiment140, SST-5	Reviews, Tweets	3
Emotion Recognition	Go Emotions	Social media	28
Sarcasm Detection	TweetEval (Irony), Reddit Sarcasm Dataset	Tweets	2

Training Configuration

The proposed framework utilizes a RoBERTa-based encoder initialized with pretrained weights. The shared encoder is fine-tuned jointly with three task-specific classification heads using a multitask learning objective. The input sequences were tokenized using a subword tokenizer consistent with the encoder vocabulary and truncated or padded to a fixed maximum length.

The model was trained using the AdamW optimizer with a linear learning rate schedule. Minibatch training was employed to ensure stable convergence. Task-specific losses were combined using weighted summation to balance learning across sentiment, emotion, and sarcasm tasks.

To prevent overfitting, regularization techniques, such as dropout and early stopping, were applied during training.

Evaluation Metrics

The performance of the proposed multitask framework was evaluated separately for each affective analysis task using standard classification metrics. For sentiment analysis and sarcasm detection, which are formulated as multi-class and binary classification problems, respectively, accuracy, precision, recall, and F1-score were used to assess the overall predictive performance and class-wise discrimination capability.

For emotion recognition, which involves twenty-eight fine-grained emotion categories, evaluation is performed using macro-averaged precision, recall, and F1-score. Macro-averaging assigns equal importance to each emotion class, making it particularly suitable for imbalanced label distributions and ensuring that the performance on less frequent emotion categories is not dominated by majority classes.

Precision measures the proportion of correctly predicted instances among all predictions for a given class, whereas recall quantifies the model's ability to identify all relevant instances. The F1-score represents the harmonic mean of precision and recall, providing a balanced measure of classification effectiveness. Accuracy is reported, where applicable, to summarize the overall correctness.

Together, these evaluation metrics provide a comprehensive assessment of the model's capability to capture sentiment polarity, recognize fine-grained emotional states, and detect sarcastic expressions across diverse textual inputs.

Implementation Details

The proposed multitask learning framework was implemented using a modern deep learning library that supports transformer-based architectures and efficient GPU acceleration. The shared transformer encoder was initialized with pretrained RoBERTa weights and fine-tuned jointly with task-specific classification heads.

All experiments were performed using fixed random seeds to ensure reproducibility across multiple runs. Input sequences are processed in mini-batches, and batch-level inference is employed during the evaluation to efficiently handle large-scale datasets while maintaining stable performance across varying input lengths. The implementation supports both single-instance inference and batch processing, enabling flexible deployment in real-world analytical environments.

Experimental Protocol

The datasets were partitioned into **training**, **validation**, and **test** sets according to standard evaluation practices. The training set was used to optimize the model parameters, whereas the validation set was employed to monitor the performance during training and guide the hyperparameter selection, including the learning rate and training duration. The final performance metrics were reported exclusively on the held-out test set to ensure an unbiased evaluation.

The same experimental protocol was consistently applied across all affective analysis tasks to maintain fairness and comparability. The original partitions were preserved for datasets that provided predefined splits. In cases where predefined splits were not available, the data were randomly divided following a fixed ratio, with random seeds held constant to ensure reproducibility. This uniform evaluation strategy ensures that performance comparisons across sentiment, emotion, and sarcasm tasks are both reliable and methodologically sound.

6. RESULTS AND DISCUSSION

This section presents the experimental results obtained using the proposed transformer-based multitask framework and discusses its performance in sentiment analysis, emotion recognition, and sarcasm detection tasks.

Overall Performance

The proposed multitask model demonstrated consistent performance across all three affective analysis tasks, validating the effectiveness of the shared contextual representations learned through joint training. Despite the differing complexity levels of sentiment, sarcasm, and fine-grained emotion recognition, the framework maintained a balanced performance without significant degradation in any individual task.

Sentiment Analysis Results

The performance of the model on the three-class sentiment classification task is presented in Table III. The model achieved an **overall accuracy of 77.69%,** indicating reliable sentiment polarity detection across diverse textual inputs.

TABLE III. SENTIMENT CLASSIFICATION PERFORMANCE

Metric	Value
Accuracy	0.7769
Precision (Macro)	0.7703
Recall (Macro)	0.7876
F1-score (Macro)	0.7764

Class-wise analysis showed balanced precision and recall across sentiment categories, with particularly strong recall for the neutral and negative sentiment classes. These results suggest that the shared encoder effectively captures sentiment cues, even when emotional or sarcastic signals are present. The integration of emotion and sarcasm supervision during training contributes to improved contextual robustness compared to that of isolated sentiment models.

Sarcasm Detection Results
Table IV presents the evaluation results for sarcasm detection. The model achieves an **accuracy of 75.85%,** demonstrating its capability to identify sarcastic expressions that often rely on implicit meaning and contextual contrast.

TABLE IV. SARCASM DETECTION PERFORMANCE

Metric	Value
Accuracy	0.7585
Precision (Macro)	0.7776
Recall (Macro)	0.7585
F1-score (Macro)	0.7543

The results indicate strong recall for non-sarcastic instances and competitive precision for sarcastic cases. The inclusion of both tweet-based and Reddit-based sarcasm datasets improves generalization across conversational styles, highlighting the advantage of multi-source training. Joint learning with sentiment and emotion tasks further aids sarcasm detection by enabling the model to capture polarity reversals and emotional inconsistencies.

Emotion Recognition Results
Emotion recognition performance is evaluated using micro- and macro-averaged metrics due to the **28-label fine-grained emotion space**. As shown in Table V, the model achieves a **Micro-F1 score of 0.3860** and a **Macro-F1 score of 0.2554,** reflecting the increased difficulty of fine-grained emotion classification.

TABLE V. EMOTION RECOGNITION PERFORMANCE

Metric	Value
Micro-F1	0.3860
Macro-F1	0.2554
Micro-AP	0.3746
Macro-AP	0.2647

Per-class average precision analysis reveals that emotions such as *gratitude*, *admiration*, and *amusement* are detected with relatively high confidence, while subtle or infrequent emotions such as *grief*, *pride*, and *relief* remain challenging. This performance trend is consistent with prior findings in fine-grained emotion recognition and highlights the impact of class imbalance and semantic overlap among emotion categories

Discussion
The experimental results demonstrate that the proposed transformer-based multi-task framework effectively captures interrelated affective signals within a unified architecture. The consistent performance observed across sentiment analysis, emotion recognition, and sarcasm detection highlights the benefits of shared representation learning when modeling complementary affective tasks. By jointly learning these tasks, the framework is able to exploit shared linguistic and contextual cues that would otherwise be underutilized in single-task models.

Sentiment analysis and sarcasm detection, in particular, benefit significantly from the multi-task learning strategy. Sarcasm often involves polarity reversal or implicit sentiment, and the integration of sentiment supervision during training helps the model better resolve such inconsistencies. Similarly, emotion-aware representations provide additional contextual signals that enhance sentiment classification robustness, especially in emotionally rich or informal text. These interactions explain the stable and competitive performance achieved in both tasks despite their differing classification complexities.

Emotion recognition remains the most challenging task due to the fine-grained nature of the label space and the presence of class imbalance across emotion categories. While the overall Macro-F1 score reflects this difficulty, the model demonstrates strong performance on frequently expressed emotions such as gratitude, admiration, and amusement. Lower performance on subtle or infrequent emotions, such as grief or relief, is consistent with prior findings in fine-grained emotion recognition and underscores the inherent challenges associated with sparse annotations and semantic overlap among emotion classes. Nevertheless, the observed performance indicates that shared contextual representations learned through multi-task training contribute positively to emotion recognition despite these challenges.

The results further validate the effectiveness of using a shared transformer encoder with task-specific classification heads. This design allows the framework to maintain task specialization while benefiting from common linguistic representations, resulting in improved learning efficiency and reduced computational redundancy. Compared to deploying separate models for each task, the proposed architecture achieves a favorable balance between performance and efficiency, making it suitable for real-world analytical scenarios.

Overall, the findings confirm that jointly modeling sentiment, emotion, and sarcasm leads to richer contextual understanding and improved generalization across affective analysis tasks. The observed performance trends support the suitability of multi-task learning for comprehensive affective text analysis, particularly in applications involving informal, user-generated content where affective signals are closely intertwined.

7. CONCLUSION

This paper presented a transformer-based multi-task deep learning framework for the joint modeling of sentiment analysis, fine-grained emotion recognition, and sarcasm detection. By leveraging a shared contextual encoder combined with task-specific classification heads, the proposed framework effectively captures multiple interrelated affective dimensions within a unified architecture. This design enables the model to learn rich contextual representations and implicit linguistic cues that are essential for accurately interpreting affective intent in natural language.

Experimental results demonstrate that the proposed approach achieves strong and reliable performance in sentiment classification and sarcasm detection, while also providing meaningful predictive capability for fine-grained emotion recognition across twenty-eight emotion categories. The findings highlight the effectiveness of shared representation learning in exploiting complementary affective signals, such as polarity shifts and emotional intensity, which are particularly important in informal and expressive text. Although emotion recognition remains challenging due to label granularity and class imbalance, the framework exhibits stable and consistent behavior across all tasks, indicating robust generalization.

Overall, the results confirm that transformer-based multi-task learning offers an efficient and scalable solution for comprehensive affective text analysis. Compared to single-task approaches, the proposed framework reduces computational redundancy while maintaining strong task-specific performance. These characteristics make it well suited for real-world applications, including social media analytics, customer feedback interpretation, and intelligent human–computer interaction systems, where multiple affective cues must be analyzed simultaneously.

8. FUTURE WORK

While the proposed transformer-based multi-task framework demonstrates effective performance across sentiment analysis, emotion recognition, and sarcasm detection, several directions remain for further improvement and extension. A key area for future research involves addressing class imbalance and label sparsity in fine-grained emotion recognition. Advanced optimization strategies, such as focal loss, class-balanced loss, or targeted data augmentation techniques, may help improve predictive performance for low-frequency emotion categories.

Another important direction is the extension of the framework to multilingual and code-mixed text, which is increasingly common in real-world communication. Integrating multilingual transformer models and cross-lingual learning strategies could substantially enhance the framework's applicability across diverse languages and cultural contexts.

Future studies may also explore parameter-efficient fine-tuning approaches, including adapter-based methods and low-rank adaptation techniques, to reduce computational overhead while preserving model performance. Such approaches would support deployment in resource-constrained environments and enable more efficient real-time inference. In addition, investigating lightweight transformer architectures may further improve inference efficiency without significant loss in accuracy.

From a modeling perspective, deeper task interactions could be achieved by incorporating cross-task attention mechanisms, allowing predictions for sentiment, emotion, and sarcasm to explicitly inform one another. Such task-aware modeling strategies may improve the framework's ability to resolve complex affective phenomena, particularly in emotionally ambiguous or sarcastic text.

Finally, future work may include comprehensive benchmark comparisons with strong single-task and multi-task baselines, as well as user-centric evaluations to assess practical effectiveness. Extending the framework to downstream applications such as mental health monitoring, content moderation, and conversational agents represents a promising direction for applied affective computing research.

REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, Minneapolis, MN, USA, 2019, pp. 4171–4186. Available: https://doi.org/10.18653/v1/N19-1423

[2] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," arXiv:1907.11692, 2019. Available: https://arxiv.org/abs/1907.11692

[3] R. Caruana, "Multitask learning," *Machine Learning*, vol. 28, no. 1, pp. 41–75, 1997, Available: https://doi.org/10.1023/A:1007379606734

[4] E. Cambria, S. Poria, A. Gelbukh, and M. Thelwall, "Sentiment analysis is a big suitcase," *IEEE Intelligent Systems*, vol. 32, no. 6, pp. 74–80, Nov./Dec. 2017. Available: https://doi.org/10.1109/MIS.2017.4531228

[5] S. M. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, "SemEval-2018 task 1: Affect in tweets," in *Proc. Int. Workshop on Semantic Evaluation (SemEval)*, New Orleans, LA, USA, 2018, pp. 1–17. Available: https://aclanthology.org/S18-1001/

[6] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A dataset of fine-grained emotions," in *Proc. ACL*, Online, 2020, pp. 4040–4054.
Available: https://aclanthology.org/2020.acl-main.372/

[7] F. Barbieri, J. Camacho-Collados, L. Neves, and L. Espinosa-Anke, “TweetEval: Unified benchmark for tweet classification,” in *Proc. EMNLP*, Online, 2020, pp. 1644–1650.
Available:https://aclanthology.org/2020.findingsemnlp.148.pdf

[8] A. Ghosh and T. Veale, “Fracking sarcasm using neural networks,” in *Proc. Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA)*, San Diego, CA, USA, 2016, pp. 161–169.
Available: https://aclanthology.org/W16-0425.pdf

[9] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in *Proc. Int. Conf. on Learning Representations (ICLR)*, 2019.
Available: https://arxiv.org/abs/1711.05101

[10] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, “Attention is all you need,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 5998–6008.
Available: https://doi.org/10.48550/arXiv.1706.03762

[11] T. Wolf, L. Debut, V. Sanh, *et al.*, “Transformers: State-of-the-art natural language processing,” in *Proc. EMNLP: System Demonstrations*, 2020, pp. 38–45. Available: https://aclanthology.org/2020.emnlp-demos.6/

[12] D. I. Hernández Farías, V. Patti, and P. Rosso, “Irony detection in Twitter: The role of affective content,” *ACM Trans. Internet Technol.*, vol. 16, no. 3, Article 19, Aug. 2016.
Available: https://doi.org/10.1145/2930663