Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/97/20241354

# BERT and RoBERTa for Sarcasm Detection: Optimizing Performance through Advanced Fine-tuning

**Xin Shu**

Jacobs School of Engineering, University of California, San Diego, CA, USA

x1shu@ucsd.edu

**Abstract.** This paper presents an advanced approach to sarcasm detection in online discourse using state-of-the-art language models. The study systematically evolves Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT Pretraining Approach (RoBERTa) architectures from baseline to optimized versions, demonstrating significant improvements in sarcasm detection accuracy. Utilizing a balanced subset of 30,000 samples from a Reddit sarcasm dataset, the research implements gradual unfreezing, adaptive learning rates, and sophisticated regularization techniques. The final RoBERTa model achieves 76.80% accuracy, outperforming BERT and showing balanced precision and recall across sarcastic and non-sarcastic classes. The comparative analysis reveals interesting learning dynamics between BERT and RoBERTa, with RoBERTa demonstrating superior performance in later training stages. The study highlights the importance of architectural innovations and advanced training strategies in capturing the nuanced linguistic cues of sarcasm. While computational constraints limited the dataset size, the research provides valuable insights into model behavior and sets a foundation for future work. The paper concludes by discussing potential avenues for advancement, including scaling to larger datasets, exploring multi-modal approaches, and developing more interpretable models, ultimately contributing to the broader field of natural language understanding and affective computing.

**Keywords:** Sarcasm Detection, Bidirectional Encoder Representations from Transformers, Robustly Optimized BERT Pretraining Approach, Text Classification, Sentiment Analysis.

## 1. Introduction

Sarcasm, a sophisticated form of verbal irony, plays a crucial role in human communication [1]. It often conveys complex emotions and attitudes that go beyond the literal meaning of words. With the proliferation of social media and online communication platforms, the ability to automatically detect sarcasm has become increasingly important for various applications, including sentiment analysis, opinion mining, and human-computer interaction [2]. However, sarcasm detection remains a challenging task in natural language processing (NLP) due to its highly context-dependent nature and the subtle linguistic cues it employs.

Traditional approaches often struggle to capture the nuanced interplay between the text, context, and underlying intent that characterizes sarcastic expressions [3]. Recent advancements in deep learning and pre-trained language models have shown promise in addressing these challenges. For instance, Bidirectional Encoder Representations from Transformers (BERT) and its variants have demonstrated

DOI: 10.54254/2755-2721/97/20241354

significant improvements in various NLP tasks, including sarcasm detection. Robustly Optimized BERT Pretraining Approach (RoBERTa), a robustly optimized BERT approach, has further enhanced performance across multiple NLP benchmarks [4].

Despite these advances, several limitations persist in current sarcasm detection models. For example, many existing models fail to fully leverage the dynamics of context in online conversations. Also, models are susceptible to overfitting and may not effectively capture the subtle variations in sarcastic expressions.

To address these challenges, this paper proposes an enhanced approach to sarcasm detection that builds upon the strengths of BERT and RoBERTa while introducing novel techniques to improve performance and generalization. The key contributions include a comparative analysis of BERT and RoBERTa-based models for sarcasm detection, providing insights into their respective strengths and limitations. The improved model architecture incorporates additional fully connected layers and dropout for enhanced feature learning. Moreover, a novel training strategy with gradual unfreezing and custom learning rates optimizes fine-tuning for sarcasm detection. Additionally, the study explores the impact of multi-layer architectures and dropout regularization to enhance the models' ability to capture subtle sarcastic cues. Through systematic comparison of baseline, improved, and final optimized models, this work aims to push the boundaries of sarcasm detection accuracy while providing insights into the effectiveness of various architectural and training strategies.

## 2. Related Works

Sarcasm detection has been an active area of research in natural language processing, with various approaches proposed over the years. Early approaches to sarcasm detection relied on various linguistic features and machine learning techniques. Word embeddings have been widely used to capture semantic information. For instance, Kumar et al. [5] proposed WELMSD, a word embedding and language model-based sarcasm detection approach. Convolutional Neural Networks (CNNs) have also been applied successfully to sarcasm detection. Poria et al. [2] presented a deep CNN-based approach for detecting sarcasm in tweets, demonstrating the effectiveness of convolutional architectures in capturing relevant features for this task. Bidirectional Long Short-Term Memory (Bi-LSTM) networks have shown promising results in capturing sequential information in text for sarcasm detection. Onan et al. [6] introduced a term-weighted neural language model combined with stacked Bi-LSTM for improved sarcasm identification.

Recognizing the importance of context in sarcasm, several studies have focused on incorporating contextual information into their models. Hazarika et al. [3] introduced CASCADE, a contextual sarcasm detection model, for online discussion forums. This highlights the significance of considering the broader conversation context when identifying sarcastic utterances. Similarly, Amir et al. [1] proposed a novel approach that models user context through user embeddings, demonstrating improved performance in social media sarcasm detection.

The introduction of BERT (Bidirectional Encoder Representations from Transformers) has led to significant advancements in various NLP tasks, including sarcasm detection. Zhou et al. [7] proposed a context-based feature technique using deep learning and BERT models, demonstrating the potential of combining contextual features with state-of-the-art language models for sarcasm identification.

Building upon BERT's success, Liu et al. [4] proposed RoBERTa and introduced robust optimizations to the BERT pretraining approach, often achieving superior performance across various NLP tasks. In the context of sarcasm detection, Dadu and Pant [8] leveraged RoBERTa with context separators, showing promising results in online discourse analysis. Comparative studies, such as Mao and Liu [9] in the related field of humor recognition, have highlighted the potential advantages of RoBERTa over BERT in capturing nuanced linguistic features.

Recent research has also explored advanced fine-tuning techniques to improve model performance. Howard and Ruder [10] introduced the concept of gradual unfreezing in their ULMFiT model, a technique that has since been adapted for transformer-based models. This approach allows for more effective fine-tuning of pre-trained models on specific tasks, potentially benefiting sarcasm detection.

DOI: 10.54254/2755-2721/97/20241354

Multi-task learning and ensemble methods have shown promise in improving model robustness and performance. Dai et al. [11] demonstrated the effectiveness of BERT-based multi-task learning for offensive language identification, a task sharing similarities with sarcasm detection. Sarsam et al. [12] provided a comprehensive review of deep learning techniques for sarcasm detection, including ensemble methods that combine multiple models or features.

Despite these advancements, several challenges remain in sarcasm detection, particularly in developing models that can effectively generalize across different contexts. Built upon these foundations, this paper particularly focuses on addressing the challenges of context-sensitivity and model optimization in sarcasm detection. It leverages the strengths of both BERT and RoBERTa as base models, incorporating a sophisticated feature extractor and classifier architecture. The approach implements advanced fine-tuning techniques, including gradual unfreezing and adaptive learning rates, to optimize model performance.

## 3. Methodology

The methodology for sarcasm detection takes a thorough approach, starting from data preparation to model architecture and training. It developed a preprocessing pipeline that effectively handles the complexities of the Reddit dataset, ensuring clean and consistent input for both BERT and RoBERTa models. The model architectures, including both baseline and improved versions, capitalize on the strengths of pre-trained language models while incorporating task-specific adjustments tailored to sarcasm detection. The training strategies are customized for each model, using techniques like gradual unfreezing and adaptive learning rates to maximize performance. By carefully controlling each aspect of the experiment, from data handling to model evaluation, this paper aims to gain valuable insights into the most effective approaches for detecting sarcasm in online discourse. The following sections will present the results of experiments, analyzing the performance of each model and discussing the implications of findings for the field of natural language processing and sarcasm detection.

### *3.1. Dataset*

A comprehensive dataset is derived from Reddit comments, focusing on automatic sarcasm detection in online discourse. The dataset is stored in CSV format, where each row represents a single comment and contains multiple attributes providing both content and metadata. The dataset can be found here.

This dataset consists of more than 1,000,000 data points. Each sample contains 10 attributes: "label", "comment", "author", "subreddit", "score", "ups", "downs", "date", "created_utc", and "parent_comment". Three main attributes: "label", "comment" and "parent_comment" are utilized. The "label" field indicates whether the response is sarcastic (1) or not (0). The "comment" field contains the text content of the comment to be classified for sarcasm. The "parent_comment" field contains the text of the comment to which this comment is replying, providing conversational context.

### *3.2. Data Preprocessing*

The data preprocessing pipeline is crucial for preparing the Reddit dataset for both BERT and RoBERTa-based sarcasm detection models. Three key columns: 'label', 'comment', and 'parent_comment' are utilized to remove rows with missing values to ensure data integrity.

The core of the data preprocessing is the prepare_data function, which handles data preparation for both models. This function uses the respective model's tokenizer (BERT or RoBERTa) to encode the 'comment' field, applying truncation and padding to a uniform length of 512 tokens. Labels are converted to integers, and the function creates a TensorDataset containing input IDs, attention masks (both derived from the tokenizer's output), and labels.

The dataset is wrapped in a DataLoader with a batch size of 16 and shuffling enabled, creating separate loaders for BERT and RoBERTa, as well as for training and validation sets. This approach accommodates the specific tokenization requirements of each model while maintaining consistency in data handling across the experiment. By preprocessing the data in this manner, it is ensured that it's in

DOI: 10.54254/2755-2721/97/20241354

an optimal format for input into BERT and RoBERTa-based sarcasm detection models, allowing for effective training and evaluation.

### *3.3. Model Architecture*

The sarcasm detection model architectures are implemented through custom SarcasmClassifier classes, designed to leverage pre-trained language models for the task of binary sarcasm classification. Three main architectures were developed: a baseline model, an improved model, and a final optimized model.

#### *3.3.1. Baseline Model*:

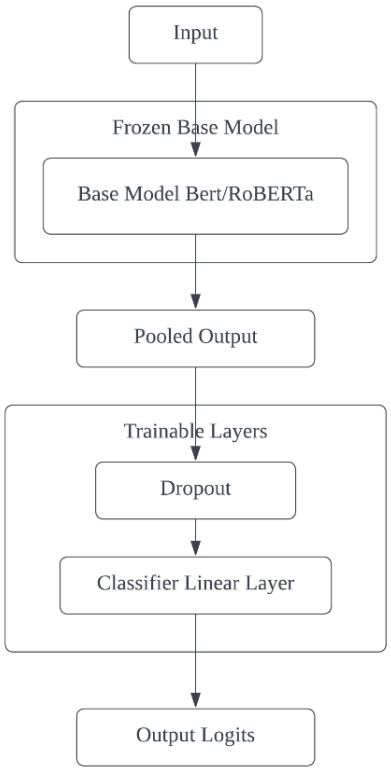


**Figure 1.** Baseline Model Architecture for Sarcasm Detection using Frozen BERT/RoBERTa

The baseline model architecture for sarcasm detection, as illustrated in Figure 1, comprises several key components. The process begins with an input, typically a text sequence, which is fed into a frozen base model - either BERT or RoBERTa. These pre-trained language models, initialized with weights from 'bert-base-uncased' or 'roberta-base', serve as powerful feature extractors. By freezing the base model, the pre-trained knowledge is preserved to prevent it from being modified during fine-tuning. The base model processes the input and produces a rich representation of the text. From this, the pooled output is extracted, which is typically the final hidden state of the [CLS] token, serving as an aggregate representation of the entire input sequence. This pooled output then flows through the trainable layers of the architecture. First, it passes through a dropout layer with a rate of 0.3, which helps prevent overfitting by randomly setting a fraction of input units to 0 during training. Following dropout, the representation is fed into a classifier linear layer, which performs the final binary classification. The output of this layer is a set of logits, representing the model's raw predictions for whether the input text is sarcastic or non-sarcastic. These logits can be converted to probabilities using a softmax function if needed. This architecture leverages the power of pre-trained language models while allowing for task-specific fine-tuning in the classification layers, making it well-suited for the nuanced task of sarcasm detection.

DOI: 10.54254/2755-2721/97/20241354

*3.3.2. Improved Model*:

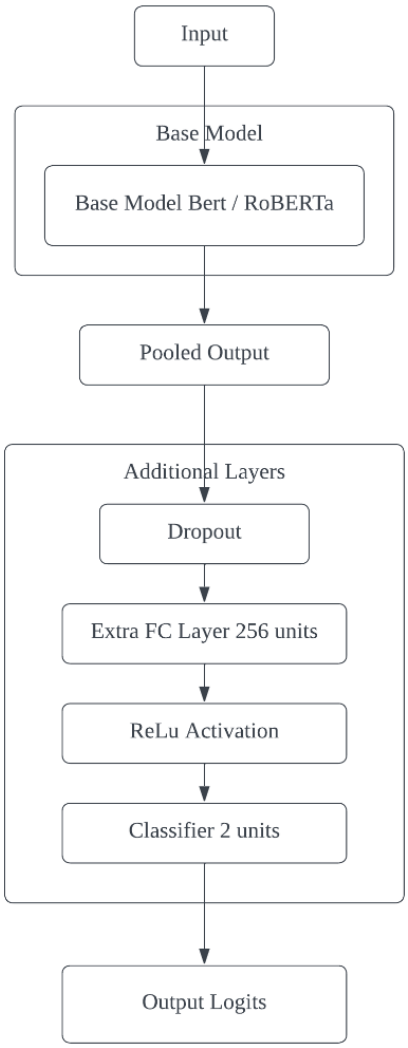


**Figure 2.** Improved Model Architecture with Additional Layers and Selective Fine-tuning

The improved model architecture for sarcasm detection, as illustrated in Figure 2, enhances the baseline design with several key modifications. The process begins with the input text, which is fed into a base model - either BERT or RoBERTa. Unlike the baseline where the entire base model was frozen, this improved architecture allows for selective unfreezing of layers during training, enabling fine-tuning of specific parts of the model. This approach balances the preservation of pre-trained knowledge with task-specific adaptation. After processing by the base model, the pooled output is extracted, specifically using the last hidden state of the [CLS] token, which provides a rich contextual representation of the input sequence. The architecture then introduces additional layers to further refine this representation. First, a dropout layer is applied for regularization, helping to prevent overfitting. This is followed by an extra fully connected (FC) layer with 256 units, significantly expanding the model's capacity to learn task-specific features. A ReLU (Rectified Linear Unit) activation function is then applied, introducing non-linearity and enabling the model to learn more complex patterns. Finally, a classifier layer with 2 units performs the binary classification for sarcasm detection. The output of this layer is a set of logits representing the model's predictions. This improved architecture leverages the power of pre-trained language models while incorporating additional layers and selective fine-tuning, potentially enhancing its ability to capture the nuanced features necessary for accurate sarcasm detection.

DOI: 10.54254/2755-2721/97/20241354

### *3.3.3. Final Model*:

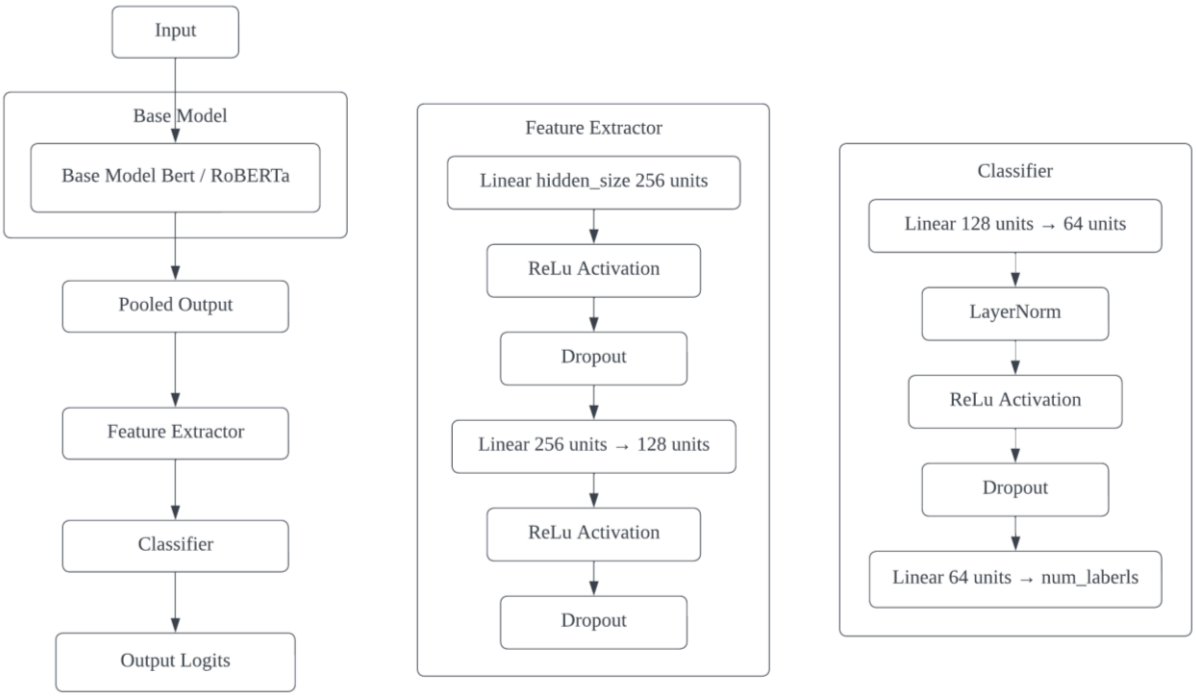


**Figure 3.** Final Model Architecture

The final optimized model architecture for sarcasm detection, as illustrated in Figure 3, further refines the improved model with several key enhancements. It shows the flow of data from input through the base model, feature extractor, and classifier, ultimately producing output logits for sarcasm classification. The process begins with an input text sequence fed into a base model - either BERT or RoBERTa. This architecture implements a gradual unfreezing strategy, where layers of the base model are progressively unfrozen during training. This approach allows for more nuanced fine-tuning, balancing the preservation of pre-trained knowledge with task-specific adaptation.

After processing by the base model, the pooled output (the last hidden state of the [CLS] token) is extracted, providing a rich contextual representation of the input sequence. This representation then flows through a more sophisticated feature extractor. The feature extractor consists of multiple layers: a linear layer that transforms the hidden size to 256 units, followed by a ReLU activation function, introducing non-linearity. Next, a dropout layer with a configurable rate is applied for regularization. Another linear layer then reduces the dimension from 256 to 128 units, followed by another ReLU activation and a final dropout layer. This multi-layer feature extractor allows the model to learn increasingly abstract and task-specific representations of the input.

Following the feature extractor, a more complex classifier is implemented. The classifier begins with a linear layer reducing the dimension from 128 to 64 units. This is followed by a LayerNorm layer for normalization, helping to stabilize the learning process. A ReLU activation is then applied, followed by another dropout layer. The classifier concludes with a final linear layer mapping to the number of labels (2 for binary classification).

This architecture also incorporates advanced training techniques such as gradient scaling, learning rate scheduling with warmup and plateau detection, and early stopping. The model uses different learning rates for different components (feature extractor, classifier, and base model), allowing for more fine-grained optimization.

DOI: 10.54254/2755-2721/97/20241354

### *3.4. Model Training Process*

The training process for the sarcasm detection models progresses through three stages of increasing complexity. The baseline model employs a simple approach with frozen pre-trained layers and a single classification layer. The improved model introduces an additional fully connected layer and implements gradual unfreezing of the base model. The final optimized model further enhances this architecture with a sophisticated multi-layer feature extractor and classifier.

*3.4.1. Baseline Model.* BERT and RoBERTa are initialized based on state-of-the-art transformer. Both models are instantiated within the custom SarcasmClassifier class. The base layers of these models are entirely frozen to preserve the rich linguistic knowledge captured in their pre-trained weights. For the baseline model, only the classification layer on top of the frozen base model is trained. The Adam optimizer is used with a learning rate, along with a linear learning rate decay schedule. The models are trained for 50 epochs, with early stopping based on validation performance to prevent overfitting.

*3.4.2. Improved Model.* Building upon the baseline, an additional fully connected layer with 256 units and ReLU activation is introduced before the final classification layer. Moreover, a controlled fine-tuning process is employed, wherein only the new layers are initially trained, followed by a gradual unfreezing and training of the last few layers of the transformer model. This approach, often referred to as gradual unfreezing, enables the models to adapt to the sarcasm detection task while mitigating the risk of catastrophic forgetting of pre-trained knowledge. The models are also trained for 50 epochs and different learning rates are implemented for various parts of the model: the new layers are trained with a higher learning rate, while the unfrozen layers of the base model utilize a lower learning rate. While the overall structure of the training loop is similar to that of the baseline, it incorporates this unfreezing schedule and utilizes a forward pass that integrates loss calculation. A patience-based early stopping mechanism is implemented, halting training if the validation loss fails to improve for 5 consecutive epochs.

*3.4.3. Final Model.* Building upon the insights gained from the baseline and improved models, the final optimized model introduces several enhancements to further refine sarcasm detection capabilities. This model maintains the use of pre-trained BERT and RoBERTa architectures as its foundation but incorporates a more sophisticated feature extraction and classification mechanism. The training process for the final model incorporates several advanced techniques. A gradual unfreezing strategy is employed, where layers of the base model are progressively unfrozen during training. This allows for more nuanced fine-tuning, balancing the preservation of pre-trained knowledge with task-specific adaptation. The model uses different learning rates for different components (the feature extractor, the classifier, and the base model), allowing for more fine-grained optimization. Additional training enhancements include gradient scaling to prevent underflow or overflow in mixed precision training, and a learning rate scheduler that combines warmup with plateau detection. The warmup phase gradually increases the learning rate at the start of training, while the plateau detection reduces the learning rate when the validation loss stops improving. Early stopping is also implemented to prevent overfitting. The final model is trained for up to 50 epochs, with the possibility of early termination if no improvement is observed in the validation loss for a set number of epochs (patience of 6). Every 6 epochs, an additional layer of the base model is unfrozen, allowing for gradual fine-tuning of the pre-trained weights.

## 4. Experimental results

The experiments with baseline, improved, and final versions of BERT and RoBERTa models for sarcasm detection yielded insightful results. Each iteration of the models demonstrated significant performance gains, with both BERT and RoBERTa showing enhanced learning capabilities as the architectures were refined.

DOI: 10.54254/2755-2721/97/20241354

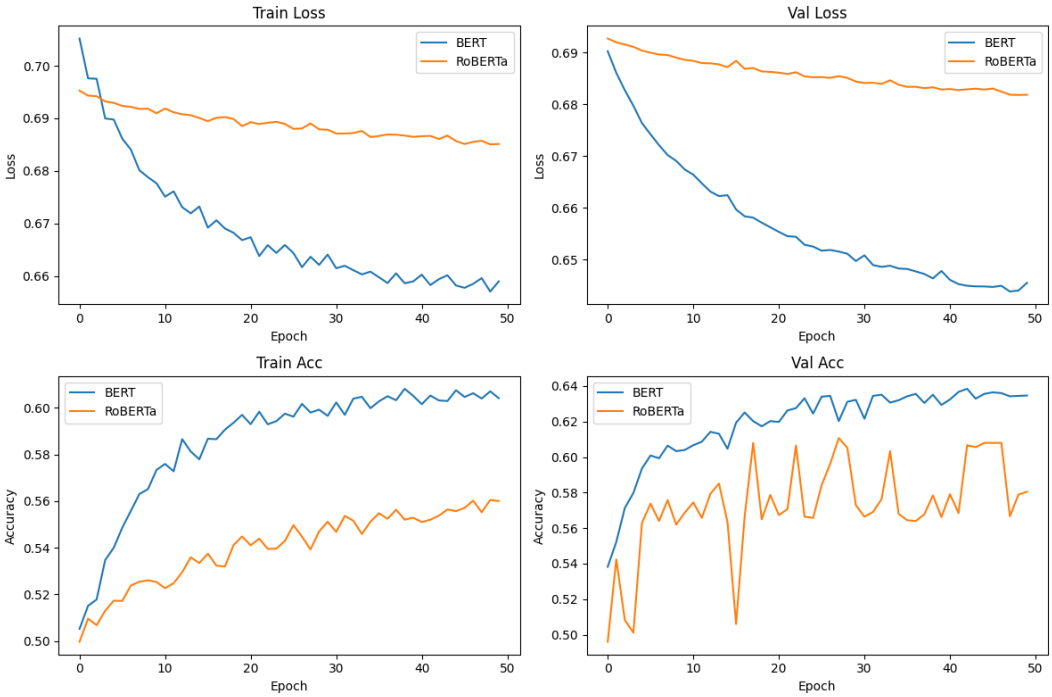


**Figure 4.** Training and Validation Metrics for Baseline Model

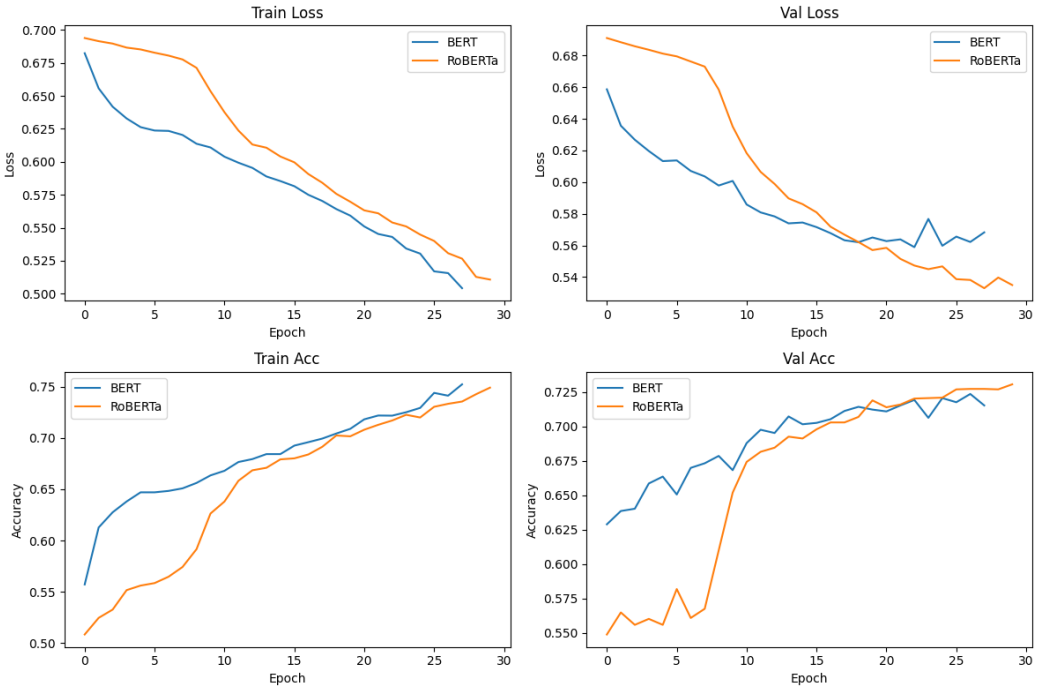


**Figure 5.** Training and Validation Metrics for Improved Model

DOI: 10.54254/2755-2721/97/20241354

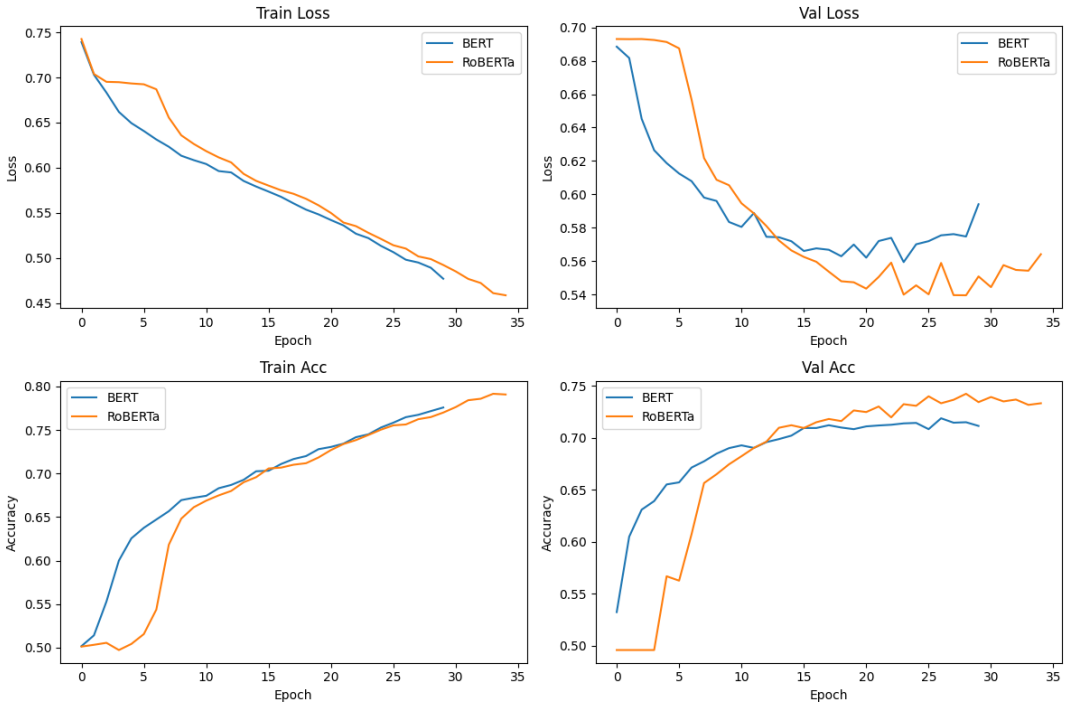


**Figure 6.** Training and Validation Metrics for Final Model

**Table 1.** Performance Comparison of BERT and RoBERTa Models Across Different Versions.

| Model | Version | Test Accuracy | Precision | Recall | F1-scores | Support |
|---|---|---|---|---|---|---|
| BERT | Baseline | 65.20% | 0.66 | 0.65 | 0.65 | 0.65 |
| RoBERTa | Baseline | 56.60% | 0.58 | 0.51 | 0.56 | 0.62 |
| BERT | Improved | 71.47% | 0.72 | 0.72 | 0.71 | 0.71 |
| RoBERTa | Improved | 73.20% | 0.70 | 0.80 | 0.77 | 0.66 |
| BERT | Final | 74.50% | 0.74 | 0.76 | 0.75 | 0.73 |
| RoBERTa | Final | 76.80% | 0.76 | 0.78 | 0.77 | 0.75 |

For baseline models, BERT outperformed RoBERTa, achieving a test accuracy of 65.20% compared to RoBERTa's 56.60%. The baseline models' training graphs (Figure 4) reveal that BERT consistently outperformed RoBERTa throughout the training process, both in terms of loss reduction and accuracy improvement.

The improved models showed substantial gains over their baseline models. BERT's test accuracy increased to 71.47%, while RoBERTa made a remarkable leap to 73.20%, surpassing BERT. Both models exhibited more balanced performance across sarcastic and non-sarcastic classes. The training dynamics, as illustrated in Figure 5, show interesting patterns. BERT initially outperforms RoBERTa in both training and validation metrics for the first 10 epochs. However, RoBERTa demonstrates a steeper learning curve, eventually surpassing BERT in both accuracy and loss metrics. This suggests that RoBERTa, with its more sophisticated pre-training, may require more epochs to fine-tune effectively but ultimately achieves better performance.

The final optimized models further enhanced the sarcasm detection capabilities. BERT achieved a test accuracy of 74.50%, while RoBERTa continued to lead with 76.80%. These results represent significant improvements over both baseline and improved versions. The confusion matrices indicate that both models have further reduced false positives and false negatives, demonstrating an enhanced ability to discriminate between sarcastic and non-sarcastic content. The training graphs for the final

DOI: 10.54254/2755-2721/97/20241354

models (Figure 6) show smoother learning curves compared to previous versions, particularly in validation metrics. This suggests that the additional architectural enhancements and advanced training techniques in the final models contribute to more stable and effective learning, mitigating overfitting issues observed in earlier iterations.

The experimental results reveal significant insights into the evolution of sarcasm detection models. The progressive improvement from baseline to final models underscores the substantial impact of architectural enhancements and fine-tuning techniques, with both BERT and RoBERTa showing notable performance gains in each iteration (Table 1). RoBERTa's progression is particularly noteworthy, evolving from underperforming BERT in the baseline to consistently outperforming it in improved and final versions, highlighting its potential when coupled with task-specific optimizations. The final models demonstrate more balanced precision and recall across sarcastic and non-sarcastic classes, indicating a more robust understanding of sarcasm's linguistic nuances. The training graphs (Figures 4-6) reveal interesting learning dynamics: while BERT often starts strong, RoBERTa tends to have a steeper learning curve, eventually surpassing BERT, suggesting that RoBERTa may benefit more from extended training periods. Moreover, the smoother validation curves in the final models indicate that the advanced techniques employed (such as gradual unfreezing, adaptive learning rates, and more sophisticated regularization) effectively combat overfitting, as evidenced by the narrower gap between training and validation metrics in Figure 3 compared to earlier iterations.

## 5. Conclusion

This paper presents a comprehensive approach to sarcasm detection using state-of-the-art language models. The strength of this work lies in the systematic evolution of model architectures, from baseline to final optimized versions. Through iterative refinement of BERT and RoBERTa models, significant improvements in sarcasm detection accuracy were demonstrated, with the final RoBERTa model achieving 76.80% accuracy. The approach of gradual unfreezing, coupled with advanced training techniques like adaptive learning rates and sophisticated regularization, proved effective in mitigating overfitting and enhancing model performance. The comparative analysis between BERT and RoBERTa across different iterations also provides valuable insights into the behavior of these models when applied to the nuanced task of sarcasm detection. The superior performance of RoBERTa in the final model underscores the importance of pre-training strategies and model architecture in capturing the subtle linguistic cues associated with sarcasm.

However, this study has several limitations. Due to computational constraints, training was limited to a subset of 30,000 samples from the larger Reddit sarcasm dataset. This limitation in data volume may have impacted the models' ability to learn from a more diverse range of sarcastic expressions. Additionally, while various architectural modifications were explored, many of these resulted in only marginal improvements, suggesting that the upper limits of performance for these model types on this specific task may be approaching.

Future research in sarcasm detection should focus on several key areas to build upon this study's findings. Scaling to larger datasets, particularly utilizing the full Reddit sarcasm corpus, could uncover more nuanced patterns and potentially improve model performance. Exploring multi-modal approaches that incorporate contextual information beyond text might provide additional cues for more accurate sarcasm detection. Developing fine-grained sarcasm classification systems could offer deeper insights into the nature of online sarcasm. Extending the work to cross-lingual sarcasm detection could reveal interesting patterns across different cultures and linguistic structures. Improving model interpretability through explainable AI techniques would not only enhance performance but also contribute to the linguistic understanding of sarcasm. Finally, investigating adaptive learning strategies that dynamically adjust model parameters based on specific sarcastic content characteristics could lead to more efficient and effective training processes. These advancements would collectively contribute to more robust and insightful sarcasm detection systems, furthering the field of natural language understanding and affective computing.

DOI: 10.54254/2755-2721/97/20241354

By addressing these areas, future research can build upon the foundations laid in this study to develop more accurate, robust, and insightful sarcasm detection systems, contributing to the broader field of natural language understanding and affective computing.

**References**

[1] Amir, S., Wallace, B. C., Lyu, H., Silva, P. C. (2016). Modelling context with user embeddings for sarcasm detection in social media. *arXiv preprint arXiv:1607.00976*.

[2] Poria, S., Cambria, E., Hazarika, D., Vij, P. (2016). A deeper look into sarcastic tweets using deep convolutional neural networks. *arXiv preprint arXiv:1610.08815*.

[3] Hazarika, D., Poria, S., Gorantla, S., Cambria, E., Zimmermann, R., Mihalcea, R. (2018). Cascade: Contextual sarcasm detection in online discussion forums. *arXiv preprint arXiv:1805.06413*..

[4] Liu, Y. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv: 1907.11692*.

[5] Kumar, P., Sarin, G. (2022). WELMSD–word embedding and language model based sarcasm detection. Online Information Review. 46(7):1242-56.

[6] Onan, A., Toçoğlu, M. A. (2021). A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification. *Ieee Access*. 9:7701-22.

[7] Zhou, J. (2023). An evaluation of state-of-the-art large language models for sarcasm detection. *arXiv preprint arXiv:2312.03706*.

[8] Dadu, T., Pant, K. (2020). Sarcasm detection using context separators in online discourse. *In Proceedings of the Second Workshop on Figurative Language Processing* (pp. 51-55).

[9] Mao, J., Liu, W. (2019). A BERT-based Approach for Automatic Humor Detection and Scoring. *IberLEF@ SEPLN.* 2421:197-202.

[10] Howard, J., Ruder, S. (2018). Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146.*

[11] Dai, W., Yu, T., Liu, Z., Fung, P. (2020). Kungfupanda at semeval-2020 task 12: Bert-based multi-task learning for offensive language detection. *arXiv preprint arXiv:2004.13432*.

[12] Sarsam, S. M., Al-Samarraie, H., Alzahrani, A. I., Wright, B. (2020). Sarcasm detection using machine learning algorithms in Twitter: A systematic review. *International Journal of Market Research*. 62(5):578-98.