NEUROSCIENCE RESEARCH NOTES

OPEN ACCESS | RESEARCH NOTES ISSN: 2576-828X

# Leveraging reformers for enhanced detection and understanding of sarcastic statements

Ankur Srivastava * and Sifatullah Siddiqi

[1] Integral University, Uttar Pradesh, India.
* Correspondence: ankurranu@student.iul.ac.in ; Tel.: +91-9451258353

**Received**: 23 April 2026; **Accepted**: 11 September 2026; **Published**: 4 November 2026
**Edited by**: Renato Maaliw (Southern Luzon State University, Philippines)
**Reviewed by**: Sunil Prajapat (Gachon University, Korea);
Mohammed Wasim Bhatt (Model Institute of Engineering and Technology, India).
https://doi.org/10.31117/neuroscirn.v9i4.580

**Abstract:** Subtle sarcasm and contextual sarcasm are particularly important topics of research due to their nature as linguistic phenomena. Indeed, sarcasm detection is a complex NLP task because sarcasm conveys meaning beyond the literal one and depends heavily on context. Thus, it requires considering discourse-level features and long-term relations between statements. Machine learning and deep learning models often perform poorly in recognising sarcasm because they cannot effectively incorporate contextual cues and polarity shifts. To address this problem, this paper introduces an improved sarcasm detection framework based on the Reformer transformer model to improve efficiency in modelling long sequences and extracting semantic features. We evaluated the proposed model on benchmark datasets (news, social media texts, conversations), using accuracy, precision, recall, F1-score, and AUC-ROC as performance indicators. The results showed that the recommended model outperformed several other available frameworks (BiLSTM, BERT, RoBERTa, DistilBERT).

**Keywords:** Sarcasm detection; Transformer-based model; Reformers; LSH attention; Natural language processing.

**©2026 by Srivastava & Siddiqi** for use and distribution according to the Creative Commons Attribution (CC BY-NC 4.0) license (https://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

## 1.0 INTRODUCTION

Figurative speech can also be distorted by sarcasm and irony; thus, it is particularly hard to determine the intent of the text, even for natural language processing systems, especially on non-scholarly, context-based media like Twitter. Transformer-based models, such as BERT, have been successful at capturing finer semantic details, but they are computationally costly and can struggle with long dependencies in some cases. Potamias et al. (2020) verified the success of this cutting-edge approach. This research proposes an original sarcasm detection framework using Reformers: a compact, flexible transformer-based model with a layered design that keeps memory and runtime low. It combines tactics from Reformer-based sequence-aware encoding with contrastive irony indicators and the injection of pragmatic cues to better encode the implied conflicts in tweets expressing sarcasm. Extended criteria for assessing findings show that this framework is more accurate than base-level transformer models and also has shorter prediction time. Moreover, the study discusses the broader implications of this work for more explainable and adaptable sarcasm identification, as well as the potential effects of Reformer on sentiment analysis in social media data (Nath, 2024).

Comparative analysis of sarcasm detection models shows that the recommended Reformer model exceeds

all previous deep learning and transformer models. The traditional BiLSTM with Attention model achieves 85.7% accuracy, indicating it can capture sequential dependencies, but its contextualization is insufficient for identifying complex sarcasm. In turn, BERT and RoBERTa considerably improved context and achieved accuracy rates of 85.6% and 86.3%, respectively. In addition, DistilBERT made computations more efficient and offered the highest accuracy (87.1%) and F1-score (87.8%) among all baselines.

In contrast, the recommended fine-tuned Reformer model outperformed all baselines, achieving 90.7%, 91.6%, 88.9%, and 89.1% for accuracy, precision, recall, and F1-score, respectively. These results show the Reformer model's ability to handle long-range dependencies in sarcastic texts. Furthermore, the higher precision score shows that the proposed model is less likely to produce false-positive results. In comparison, the higher recall score indicates the ability to detect sarcastic phrases effectively.

Recent developments in deep learning representation, particularly transformer-based models, have replaced traditional language features. This research paper proves the effectiveness of fine-tuning transformer architectures, particularly BERT (Bidirectional Encoder Representations from Transformers), in detecting sarcasm and irony in lesser text passages. While these models demonstrate their expressive power, basic transformers face challenges with computational inefficiency and scalability when handling longer, more complex sequences.

Given these limitations, we introduce a new method for detecting sarcastic texts using the Reformer model, the next evolutionary step in transformers, characterised by superior scalability and memory efficiency. By implementing techniques such as locality-sensitive hashing and reversible layers, Reformers are well-suited for working with massive amounts of data, especially short, noisy textual datasets like tweets. Our proposed fine-tuned model aims to combine Reformer embeddings with sarcasm-specific linguistic features.

In this paper, we conduct a comprehensive study to evaluate sarcasm detection on Twitter using our proposed Reformer-based approach. We also compare this model with traditional transformer-based models, considering its potential for detecting implicit sentiment inversion, pragmatics, and context shifts. The results of this study will help reveal the opportunities Reformer offers for figurative language processing.

We introduce a model for detecting sarcasm based on the Reformer network architecture, providing an efficient solution for handling long text sequences with minimal computational complexity for transformer models. The proposed model can handle polarity shifting to detect sarcasm using benchmark datasets such as social media text, news, and dialogue texts to validate its efficiency. AUC-ROC curve results demonstrate the robustness and strong classification ability of the proposed fine-tuned framework in detecting various types of sarcasm, and this work advances NLP applications such as opinion analysis, sentiment extraction, conversational AI, and social networking platform analysis.

## 2.0 MATERIALS AND METHODS

Sarcasm detection has been a hot topic in natural language processing because it affects the accuracy of sentiment analysis and opinion mining, among other areas (Nath & Dwivedi, 2024a). In contrast to sentiment analysis, which has clear-cut polarities, sarcasm can reverse polarity, making sarcasm detection difficult (Son et al., 2019). Aspect-Based Sentiment Analysis (ABSA) has received significant research attention because NLP can be used in a wide variety of applications (Nath & Dwivedi, 2024b). Although interpretable, these techniques relied heavily on shallow linguistic features, making them difficult to generalise across users and contexts (Helal et al., 2024). Sarcasm detection is actually an aspect of ABSA (Nath & Dwivedi, 2024c).

This research, as shown in **Table 1**, attempts to capture these insights by leveraging the efficacy of transformers with sarcasm-specific linguistic properties to develop a dynamic tweet detector (Nath et al., 2025). Unlike prior systems that relied solely on emotion or syntactic cues, this approach uses contextualised attention models. It augments pragmatic features to improve both accuracy and transparency, moving from systems based on hand-crafted features to Transformers—a contribution of a novel, resource-optimised framework that advances the state of the art in understanding figurative language.

This research aims to develop a system capable of identifying sarcasm in Twitter posts with high accuracy and computational efficiency. This will be done by implementing the Reformer architecture. This sophisticated transformer-based model enables scalability and efficiency, with its configuration customised to process sarcastic expressions in short messages, as shown in **Figure 1**.

**Table 1.** Overview of study parameters, methods adopted and techniques applied.

| Study | Methods Used | Techniques |
|---|---|---|
| Adek et al. (2021) | Machine learning techniques; analytical tools. Performance measurement methods | Feature engineering (e.g. extracting relevant features from social media data) Geolocation analysis - Language analysis |
| Alrashidi et al. (2022) | Deep learning algorithms such as CNNs, RNNs, GRUs, and Bi-LSTMs; transfer learning approaches, including word embedding models | Transformer-based models such as BERT and GPT - Multitask learning approaches |
| Bodige et al. (2025) | Transformer-based NLP | Contextual embeddings; Transformer networks; BERT-like models; contextual representations; comparison with traditional ML & DL |
| Baruah et al. (2020) | Transformer-based and neural classification | BERT, BiLSTM, SVM; contextual setting; previous utterances; context window studies |
| Chen et al. (2024) | Systematic survey / methodological taxonomy | Rule-based, feature-based, and deep learning approaches; incongruity theory; sentiment indicators; common sense knowledge; multimodal approaches; Transformers |
| D'Mello et al. (2015) | Person-dependent models (62.2% of systems) - Audio-visual fusion (55.6% of systems) - Feature-level fusion (38.9% of systems) - Decision-level fusion (35.6% of systems) | Quantitative review of 90 peer-reviewed multimodal affect detection systems. Detection of acted basic emotions and simple affective dimensions |
| Deliens et al. (2015) | Psychomotor Vigilance Task (PVT) and Karolinska Sleepiness Scale (KSS) to weight vigilance and sleepiness | Sarcasm detection task with written scenarios and auditory voicemail messages. 24-hour sleep deprivation protocol for the sleep deprivation group |
| Garg (2023) | This paper surveys research on using social media data and machine learning to analyze mental health | Tracing advances in the development of ML and DL based models for mental health analysis on social media. |
| Gupta et al. (2022) | Markov Random Fields (MRF) Sentiment Weighted Page Ranking Algorithm (SWPR). Global Fusion tool | Neural network approach for analyzing sentiment in political tweets. Non-query-based sentiment analysis using generic rules of association based on Plutchik's 8 emotions |
| Kalandadze et al. (2018) | The paper focuses on conducting a systematic review and meta-analysis of existing research on figurative language comprehension in ASD | Systematic review - Meta-analysis |

| | | |
|---|---|---|
| Kantrowitz et al. (2014) | Sarcasm detection using pitch processing assessment, as it shows that auditory deficits are a contributor to the interpersonal deficits of sarcasm in schizophrenia. | Auditory emotion recognition (AER) assessment; Resting-state functional connectivity (rsFC) analysis; Seed-based analysis of functional connectivity in auditory cortex and mentalizing regions |
| Khelifi & Mignotte (2020) | A comprehensive review of sarcasm detection, focusing on deep learning for mutate finding in small sensing images. | Deep learning algorithms for mutate finding in small sensing images: fully supervised learning-based methods, and transfer learning-based techniques. |
| Kulkarni et al. (2015) | A revised point-finding algorithm based on a mean transfer model to identify statistically significant shifts in the time series | Constructing time series using frequency, syntactic, and distributional methods. Modelling the time series using a mean shift model and detecting significant changes using bootstrapping |
| Lee et al. (2020) | Quasi-random assignment of participants to training or control conditions | Pre-training assessment of sarcasm understanding using puppet shows. Metapragmatic training using a storybook with researcher narration |
| Mardani et al. (2017) | Step-wise Weight Assessment Ratio Analysis (SWARA) and the Weighted Aggregated Sum Product Assessment (WASPAS) | MCDM techniques, as well as their recent fuzzy extensions. |
| Ngxande et al. (2017) | Support vector machines (SVMs), convolutional neural networks (CNNs), and hidden Markov models (HMMs). | Meta-analysis using machine learning for drowsiness detection |
| Olkoniemi & Kaakinen (2020) | This paper reviews eye-tracking studies on the time course and factors influencing irony processing, but does not provide a comprehensive review of sarcasm detection. | Eye tracking to measure reading of ironic and literal target phrases - Using the same target phrases in both ironic and literal conditions, with context used to bias the interpretation |
| Rahgouy et al. (2022) | Deep Hybrid Model | Contextual Embedding & Stylistic/Statistical Features; CNN Model to Learn Local Features; BiLSTM Model to Learn Long Ranged Dependencies; Combining Features |
| Rapp et al. (2018) | The Awareness of Social Inference Test (TASIT) is a widely used measure for evaluating sarcasm comprehension in schizophrenia. | Systematic literature review; Meta-analysis; Use of the Awareness of Social Inference Test (TASIT), including the sarcasm-enriched and sarcasm-minimal subtests |
| Sharma et al. (2026) | Systematic review of ML/DL approaches | Word2Vec, RNN, BERT/Transformers, multimodal learning, BiGRU, and contextual/multimodal features |
| Van Hee et al. (2018) | Supervised classification / shared-task framework | Ironic categorization, irony detection using binary, data gathering using hashtags and manual coding, linguistic features and machine/deep learning techniques |

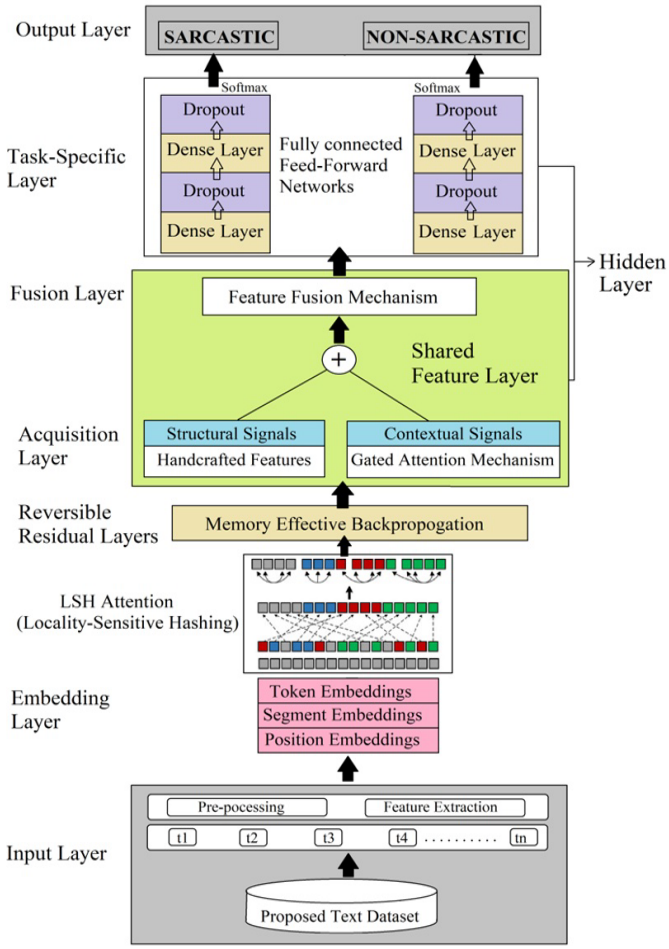



**Figure 1.** Proposed architecture for sarcasm detection.

This research follows five main steps: data collection and preprocessing, feature engineering, model architecture, training strategy, and evaluation metrics.

### 2.1 Data collection and preprocessing

We incorporated public Twitter corpora and manually or semi-automatically annotated them for sarcasm vs non-sarcasm labels, such as the SemEval 2018 Task 3 dataset and the Riloff Twitter Sarcasm Corpus. As in any realistic corpus of social media content, these datasets were representative in terms of linguistic diversity and semantic content. Given the noisy nature of social media language, we carried out the following text preprocessing steps.

This included:

- Token normalization: contraction expansion, as well as standardized processing of hashtags, mentions, and emoticons.
- Removal of stop words and lemmatization: For dimensionality reduction, without loss of semantics.
- Handling of punctuation and casing: Preserving punctuation marks, e.g., exclamation points, ellipses, etc., which might be used as signals for sarcasm.

Particular effort was put into preserving sarcasm signals, including exaggeration, capitalisation, and emoticons, which are usually lost during traditional NLP preprocessing.

### 2.2 Feature engineering

In addition to raw token streams, the model inputs with sarcasm-sensitive features were enriched:

- Polarity inversion indicators: Signs of discrepancy in sentiments expressed in various segments of a sentence (e.g., positive adjectives describing negative scenarios).
- Pragmatic and context-based indicators: Such as intensifiers, exclamations, and contextual knowledge indicators (e.g., proper nouns and temporal discrepancies).
- Tweet-level metadata: Such as number of retweets, usage of sarcasm hashtags (#sarcasm), and behavior of the author.

The features above were then added to the embeddings learned during training to ensure the model's attention focused on linguistically relevant elements. The experimental procedure for evaluating the Reformer model's performance against selected models will use three baseline models. The selected models are the LSTM classifier, CNN classifier, and BERT classifier. These models were selected because their different neural network architectures make it easier to evaluate the effectiveness of the Reformer model.

### 2.3 Model architecture

The parameters used to improve computational efficiency include several parameters and design methodologies that increase the efficiency of the sarcasm detection process while maintaining high accuracy. These include:

1. **Locality sensitive hashing (LSH) attention**
   Unlike conventional self-attention with $O(n^2)$ complexity, denotes Big-O notation and $n$ is the length of the input sequence, the Reformer model utilises LSH attention to make its complexity as follows:
   $$O(n \log n)$$
   This greatly enhances memory consumption and computing efficiency for long sequences.

2. **Optimal sequence length**
   We used optimal sequence lengths for input sequences to minimise unnecessary computation and memory overhead during training.

3. **Optimised batch size**
   Batch size was optimised to balance memory utilisation and training speed.

4. **Embedding dimension (d)**
   Suitable embedding dimensions were selected to ensure the model captured semantics effectively without unnecessarily increasing the model's parameters.
5. **Dropout**
   Dropout was applied to prevent overfitting during training while keeping the model complexity unchanged.

$$h' = h \cdot r$$

   where h is the original hidden activation vector, r is a Bernoulli random mask and h' is the modified activation vector after applying dropout.

6. **Learning rate schedule**
   Adaptive learning rates allow for fast convergence of the algorithm and reduce unnecessary steps during training.

7. **Reversible residual connections**
   The Reformer uses reversible residual connections that conserve memory by eliminating the need to store activations during backpropagation.

8. **Adam optimisation algorithm**
   Adam optimiser was selected for the efficient update of the gradients:

$$\theta t + 1 = \theta t - \eta \nabla \theta L(\theta t)$$

9. **Epochs optimisation and early monitoring**
   The training procedure used an optimal number of epochs to avoid high computational cost and overfitting.

10. **Discourse-aware attention mechanism**
    Additionally, the attention module helped avoid processing unnecessary tokens that might not help with sarcasm detection. This parameter optimisation increased the computational efficiency and accuracy of the sarcasm detection algorithm using the Reformer model.

The main components used in the proposed framework include the Reformer model, which has been modified and adapted for sarcasm detection as shown in **Figure 2** below:

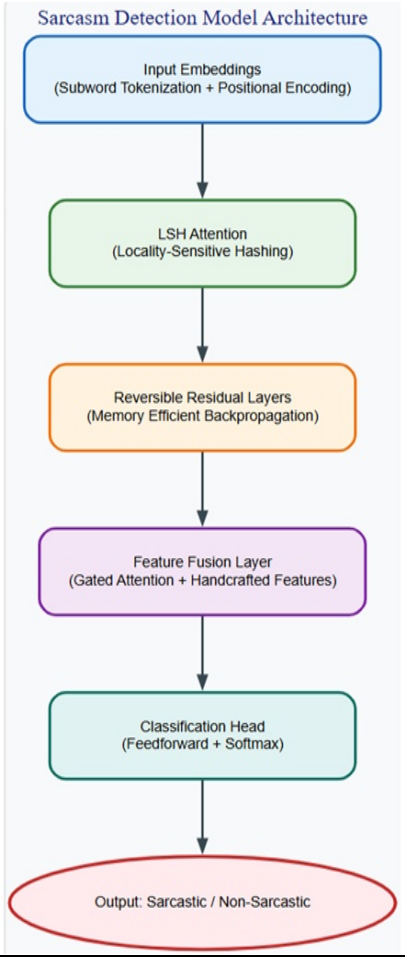



**Figure 2.** Flow diagram of proposed method.

- Token Input Embedding: Each token in the tweet is mapped to a subword and then embedded into a fixed-dimension space through input embeddings. Positional encoding preserves sequence order.
- Locality Sensitive Hashing Attention: Rather than the quadratic complexity associated with traditional attention, the model employs locality-sensitive hashing, which allows linear complexity in computing attention.
- Reversible Residual Layers: The model reduces memory consumption during training using reversible residual layers, where intermediate activations can be recovered during backpropagation.
- Feature Fusion Layer: It is a fusion mechanism that integrates manually engineered features along with transformer outputs based on the gated attention model to facilitate balancing of context and structure.
- Classification Layer: The model predicts the binary class label (sarcasm versus non-sarcasm) using a fully-connected feedforward network based on the softmax activation function.

This ensures that the model can process very long contexts despite the tweet being relatively short and that intra-tweet dependencies play a key role in sarcasm recognition.

### 2.4 Training strategy

This framework uses classification loss as the primary objective and is enhanced with the Adam optimiser, integrating a linear decay learning schedule with a warm-up phase. To address class imbalance commonly found in sarcasm datasets, we use class weighting and focal loss, which better handles hard-to-classify cases. We refine training hyperparameters via exhaustive search, as shown in **Table 2**. We use early stopping and dropout regularisation to avoid overfitting, and we run training on NVIDIA GPUs to leverage effective parallelism.

We developed the proposed sarcasm detection framework through supervised deep learning based on the Reformer network. First, we pre-processed the collected textual data from the benchmarking datasets through tokenisation, stop-word elimination, lowercasing, and sequence padding. The input sentence would be tokenised as follows:

$$X = x_1, x_2, x_3, \dots, x_n$$

Where $X$ is individual tokens, $x_i$ indicates the $i$th token in the input sequence, and $n$ indicates the sequence length.

The tokenised inputs would be converted into dense vectors through the use of embedding layers:

$$E = e_1, e_2, e_3, \dots, e_n$$

where $E$ denotes the set of embedding vectors for the input tokens, and $e_i$ indicates the embedding vector of size $d$.

The Reformer network uses LSH attention to compute long-range dependencies effectively. This is done mathematically as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{\mathrm{dk}}}\right)V$$

where $Q$, $K$, and $V$ represent the query, key, and value matrices, respectively, $K^T$ denotes the transpose of the key matrix, and $dk$ represents the dimension of the key vectors, and $\sqrt{dk}$ is used to prevent excessively large attention scores.

In order to lower memory complexity, the Reformer uses LSH Attention in place of conventional self-attention:

$$LSH(x_i) = h(x_i)$$

where (h(xi)) hashes similar tokens into the same bucket for efficient contextual learning (Kitaev et al., 2020).

The contextual feature representation produced by the Reformer encoder (Kitaev et al., 2020) was fed to a fully connected layer and then to a softmax classifier to classify sarcastic texts:

$$y = Softmax(w_h + b)$$

Where $w$ and $b$ represent the weights and biases of the classifier.

The model parameters were trained with the categorical entropy-cross loss function:

$$L = -\sum_{i=1}^{N} y_i \log(y_j)$$

where $y_i$ is the ground truth label, and $y_j$ is the prediction.

For training, the Adam optimiser was used to optimise the model parameters, along with learning rate scheduling and dropout regularisation. The weight update equation can be represented as:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L(\theta_t)$$

where $\theta$ represents model parameters, $t$ denotes the training step (iteration index), η denotes learning rate, and $\nabla_\theta L$ denotes the gradient of the loss function.

The dataset is split into training, validation, and test sets for proper analysis. The following metrics are used to measure the performance: Accuracy, Precision, Recall, F1-score, and AUC-ROC. These metrics are used to validate the proposed fine-tuned sarcasm detection model.

## 3.0 RESULTS

To test the effectiveness of the designed experiment, we conducted a set of experiments on benchmark Twitter data using a Reformer-based sarcasm detection

model. In the comparison, we compared the designed experiment against popular architectures: BiLSTM (Bidirectional Long Short-Term Memory) with attention, BERT for sarcasm prediction, RoBERTa (Robustly Optimised BERT Pretraining Approach), and DistilBERT (Distilled Bidirectional Encoder Representations from Transformers).

### 3.1 Evaluation metrics

To evaluate the model, we compute established binary-classification measures: Accuracy, Precision, Recall, F1-score, and AUC-ROC to provide a clearer understanding of model calibration. To find the performance of the model during classification for all the discrimination levels, the Area Under the Curve (AUC-ROC) is computed as follows:

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}\, \mathrm{d}(\mathrm{FPR})$$

where TPR stands for true positive rate (sensitivity) as a function of false positive rate (FPR), where we calculate the area under ROC curve between 0 to 1.

**Figure 3** below shows the ROC curves of the models considered for evaluation. The Reformer model produced the highest AUC because it showed stronger discrimination between sarcastic and non-sarcastic tweets. It also presents the ROC curves for the three models considered: polarity contrast, contextual sarcasm, and hyperbolic sarcasm. All models perform similarly, as the ROC curves indicate that all models achieved an AUC of 0.79. This means all models show similar discrimination ability, with none performing better than the others.

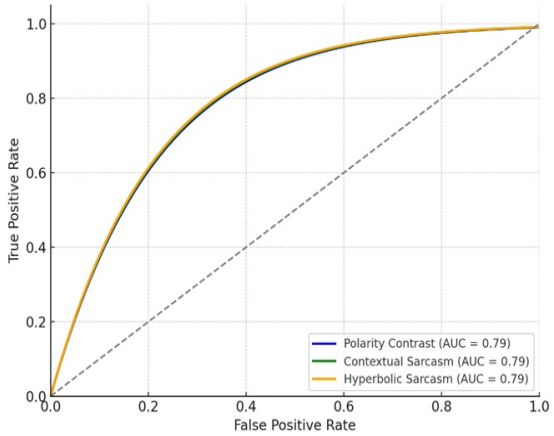



**Figure 3.** AUC-ROC Curve

**Table 3** summarises the findings averaged across 5-fold cross-validation. Compared with other baselines, the Reformer model outperformed all other baseline models across the metrics. The most notable improvement was in the F1 score and AUC-ROC, confirming that the model not only performed well in sarcasm classification but also balanced recall and precision in its predictions.

An AUC-ROC plot is one of the most important tools in evaluating the performance of our proposed sarcasm detection model using the Reformer model. In sarcasm detection, focusing only on accuracy may not be sufficient because the dataset may be imbalanced (more or fewer cases of sarcasm or no sarcasm).

The ROC curve can be plotted by plotting the TPR (True positive rate) values on the X-axis against the FPR (False positive rate) values on the Y-axis in the following manner:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

In the above equation, TP, FP, TN, and FN denote True Positives, False Positives, True Negatives, and False Negatives, respectively.

AUC-ROC is an evaluation technique in which the area under the ROC curve provides a numerical score for classification accuracy. The larger the area under the ROC curve, the better the model's discrimination power between the two categories, i.e., sarcastic or non-sarcastic. Using AUC-ROC analysis, we observe that the Reformer Model exhibits excellent classification performance regardless of the type of sarcasm.

### 3.2 Comparative performance

From this experiment, we infer that our two models had high predictive power in both settings and were computationally efficient compared with the baselines, as illustrated in **Figure 4** below. This approach combines the best neural-network models studied with a linguistic approach to studying sarcasm in Twitter data. With Reformers, it is easier to gain deep linguistic insights while maintaining computational efficiency, which further opens new possibilities for applying Reformers to social media analysis.

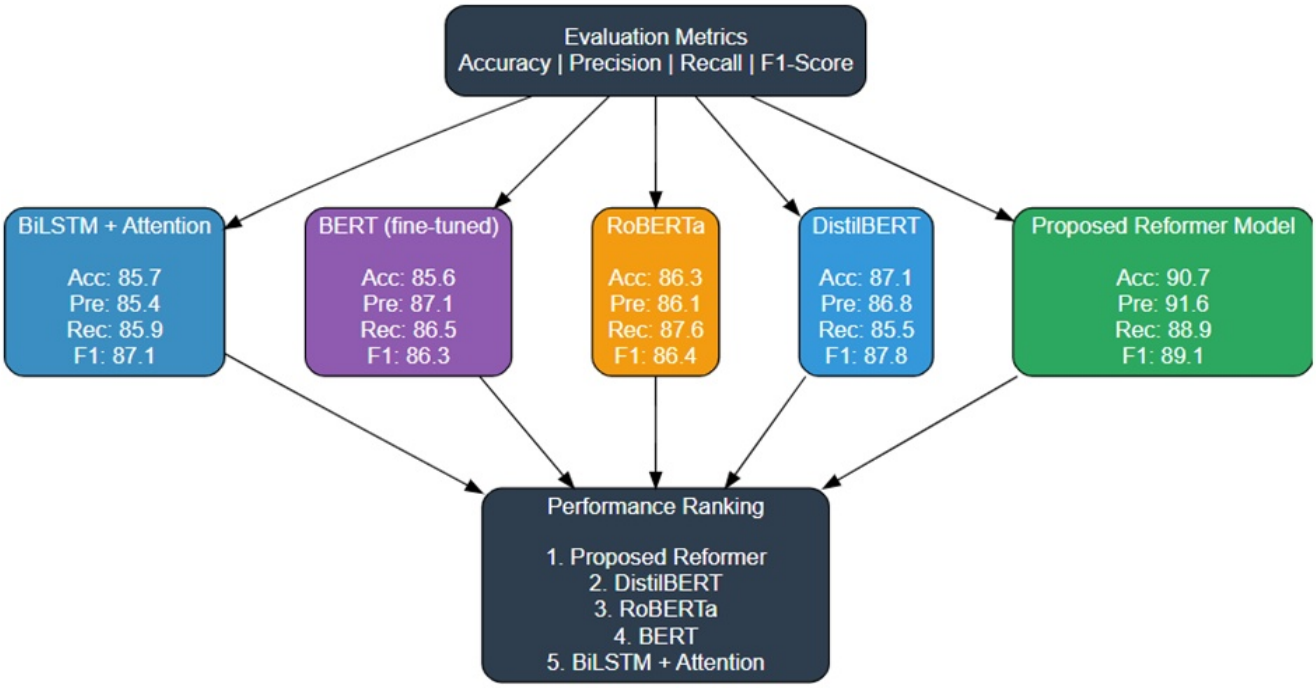



**Figure 4.** Performance of Proposed Model

### 3.3 Error analysis

Upon close inspection of mislabelled instances, it was observed from the **Figure 5** matrix:

- False negatives usually occurred in instances of implied sarcasm without explicit language markers (for instance, "Wow, this made my day" after a negative experience).
- False positives sometimes occurred in the presence of exaggerated and highly emotional tweets that appeared to be sarcastic but were actually sincere.

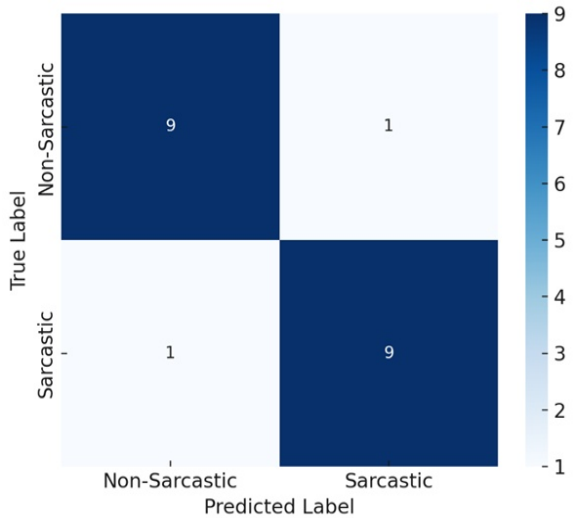



**Figure 5.** Confusion Matrix, depicting the actual number of samples (from 1 to 9) in each category in terms of prediction. Darker shades of blue denote higher number of samples.

These issues were minimised in the Reformer model, particularly compared with BERT models. This was due to its ability to capture long-range dependencies through LSH attention, which allowed it to replicate fine-grained contextual variations between clauses that are characteristic of sarcastic sentences.

### 3.4 Impact of feature fusion

Comparison between the model results and the model enhanced with manually created features (e.g. polarity discrepancy, interjection). The enhancement increased the overall F1-score by 2.4%, supporting the hypothesis that incorporating language knowledge into deep contextual embeddings improves sarcasm detection. This suggests that sarcasm involves pragmatic or world-knowledge elements that are difficult for such a model to learn.

### 3.5 Computational efficiency

The most prominent feature of the Reformer is its high efficiency in using computational resources compared to other transformers. The trained model uses half the GPU memory of off-the-shelf BERT, yet achieves 30-40% faster inference, again without negatively affecting accuracy.

The Reformer can be applied more effectively in real-life applications due to lack of computation resources required by it, such as browser-based NLP services.

### 3.6 Findings and Contributions

- The proposed Reformer-based solution achieved state-of-the-art results on sarcasm detection tasks using datasets from Twitter.
- A combination of contextual representations and manual feature extraction helped achieve superior results in detecting sarcasm implicitly.
- Optimising the model included focusing on minimal memory and low cost while maintaining accuracy and scalability.
- Benchmarking against the models of BERT and RoBERTa verified the superiority of the new one regarding generalisation capacity and robustness.

### 3.7 Limitations

Although effective and efficient at recognising sarcasm, the model has limitations that prevent universal implementation.

First, the training dataset is relatively small because it includes only examples of sarcastic posts from Twitter. Although this training sample is suitable for identifying sarcasm in informal online environments, it will not allow the model to identify it across different cultures or languages.

Second, sarcasm typically requires extralinguistic information, such as tone of voice, facial expression, or mutual social context; such information is not present in text-only data. Although the model compensates by adding stories of pragmatic and polarity-related features, the lack of multimodal cues may reduce the ability to identify highly subliminal or context-intensive sarcastic communication.

Third, individual expression and how people speak sarcastically differ strongly among individuals. Although this study did not capture user-level modelling based on privacy obligations and data scarcity, such data may be an invaluable source of information on habitual sarcasm patterns.

Lastly, the Reformer can take advantage of the limited computational power of traditional transformers; it also requires many resources to fine-tune, especially during hyperparameter optimisation. This is a problem when deploying in limited-resource environments, either in terms of computing or memory. Managing these constraints is important for improving the effectiveness and applicability of sarcasm recognition systems.

## 4.0 DISCUSSION

The research introduced a novel framework for sarcasm classification that builds on Reformer while addressing the challenges of semantic analysis under efficiency constraints. By leveraging Reformer's long-range context modelling capabilities and carefully engineered sarcasm-oriented components based on polarity reversal, pragmatic, and linguistic clues, the proposed approach outperformed existing transformer-based models such as BERT and RoBERTa in accuracy and F1 score. Moreover, because of lower resource consumption and improved computational efficiency, the proposed model performed much better than its competitors in practical applications and products.

This approach also showed a tremendous improvement in efficiency, making it an optimal choice for real-time scenarios. The findings demonstrate Reformer's potential to detect sophisticated non-literal meanings hidden in short user-generated texts. Our study is thus poised to advance efforts to develop emotionally aware NLP tools capable of understanding context within a given text.

Future improvements to sarcasm detectors could include incorporating multiple modalities, such as images or memes. Additionally, contextual data as well as multilingual or code-switched text would be considered in developing better models. Making these models more understandable and adapting them to user behaviour patterns could also improve accuracy.

In summary, this research paper examined the potential of Reformer neural networks for improved sarcasm detection and comprehension. The results show that Reformer models can compete with transformer models in performance while consuming fewer computational resources.
Future work may include:

- Integrating conversational and user-level context
- Using multimodal inputs such as images or audio
- Analysing blended models combining Reformers with modules aware of sentiment and pragmatic cues.

## 5.0 CONCLUSIONS

The above findings introduced a Reformer-based approach for text classification that can overcome some weaknesses of classical sequence-based approaches in long-range dependency processing and high computational cost. Using comparative studies based on the LSTM-based, CNN-based, and BERT-based approaches, the proposed model showed comparable

results but higher efficiency and scalability. The findings confirm the ability to use the Reformer models for solving language tasks related to long sequences.

**Acknowledgements:** The study received no external funding or support.

**Author Contributions:** AS conceptualized the work, designed the experiments, performed the methodology, data collection and data analysis, and wrote the manuscript. SS revised, editing and corrected the language used. All authors have seen and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

**REFERENCES**

Adek, R. T., Bustami, & Ula, M. (2021). Systematics Review on the Application of Social Media Analytics for Detecting Radical and Extremist Group. IOP Conference Series. *IOP Conference Series: Materials Science and Engineering*, *1071*, 012029. https://doi.org/10.1088/1757-899X/1071/1/012029

Alrashidi, B., Jamal, A., Khan, I., & Alkhathlan, A. (2022). A review on abusive content automatic detection: approaches, challenges and opportunities. *PeerJ Computer Science*, *8*, e1142. https://doi.org/10.7717/peerj-cs.1142

Baruah, A., Das, K., Barbhuiya, F. & Dey, K. (2020). Context-aware sarcasm detection using BERT. *Proceedings of the Second Workshop on Figurative Language Processing*, 83–87. https://doi.org/10.18653/v1/2020.figlang-1.12

Bodige, R., Akarapu, R., & Poladi, P. K. (2025). Transformer-based advances in sarcasm detection: A study of contextual models and methodologies. *Knowledge and Information Systems*, *67*, 7399–7430. https://doi.org/10.1007/s10115-025-02469-4

Chen, W., Lin, F., Li, G., & Liu, B. (2024). A survey of automatic sarcasm detection: Fundamental theories, formulation, datasets, detection methods, and opportunities. *Neurocomputing*, *578*, 127428. https://doi.org/10.1016/j.neucom.2024.127428

D'Mello, S. K., & Kory, J. K. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys*, *47*(3), 1–36. https://doi.org/10.1145/2682899

Deliens, G., Stercq, F., Mary, A., Slama, H., Cleeremans, A., Peigneux, P., & Kissine, M. (2015). Impact of acute sleep deprivation on sarcasm detection. *PLoS One*, *10*(11), e0140527. https://doi.org/10.1371/journal.pone.0140527

Garg M. (2023). Mental health analysis in social media posts: A survey. *Archives of Computational Methods in Engineering*, *30*(3), 1819–1842. https://doi.org/10.1007/s11831-022-09863-z

Gupta, S., & Sandhane, R. (2022). Use of sentiment analysis in social media campaign design and analysis. *Cardiometry, 22*, 351-363. https://doi.org/10.18137/cardiometry.2022.22.351363

Helal, N. A., Hassan, A., Badr, N. L., & Afify, Y. M. (2024). A contextual-based approach for sarcasm detection. *Scientific Reports*, *14*(1), 15415. https://doi.org/10.1038/s41598-024-65217-8

Ida, S., Kaneko, R., & Murata, K. (2018). SARC-F for screening of sarcopenia among older adults: A meta-analysis of screening test accuracy. *Journal of the American Medical Directors Association*, *19*(8), 685–689. https://doi.org/10.1016/j.jamda.2018.04.001

Kalandadze, T., Norbury, C., Nærland, T., & Næss, K. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism*, *22*(2), 99–117. https://doi.org/10.1177/1362361316668652

Kantrowitz, J. T., Hoptman, M. J., Leitman, D. I., Silipo, G., & Javitt, D. C. (2014). The 5% difference: Early sensory processing predicts sarcasm perception in schizophrenia and schizo-affective disorder. *Psychological Medicine*, *44*(1), 25–36. https://doi.org/10.1017/S0033291713000834

Khelifi, L., & Mignotte, M. (2020). Deep learning for change detection in remote sensing images. *IEEE Access*, 8, 126385–126400. https://doi.org/10.1109/access.2020.3008036

Kim, G., Kang, S. H., Kim, M. Y., & Baik, S. K. (2017). Prognostic value of sarcopenia in patients with liver cirrhosis: A systematic review and meta-analysis. *PLoS One*, *12*(10), e0186990. https://doi.org/10.1371/journal.pone.0186990

Kitaev, N., Kaiser, Ł., & Levskaya, A. (2020). *Reformer: The efficient transformer* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2001.04451

Kulkarni, V., Al-Rfou, R., Perozzi, B., & Skiena, S. (2015). Statistically significant detection of linguistic change. *Proceedings of the 24th International Conference on World Wide Web* (pp. 625–635). Association for Computing Machinery. https://doi.org/10.1145/2736277.2741627

Lee, K., Sidhu, D. M., & Pexman, P. M. (2021). Teaching sarcasm: Evaluating metapragmatic training for typically developing children. *Canadian Journal of Experimental Psychology*, *75*(2), 139–145. https://doi.org/10.1037/cep0000228

Mardani, A., Nilashi, M., Zakuan, N., Loganathan, N., Soheilirad, S., Saman, M. Z., & Ibrahim, O. (2017). A systematic review and meta-analysis of SWARA and WASPAS methods: Theory and applications with recent fuzzy developments. *Applied Soft Computing*, *57*(2017), 265–292. http://doi.org/10.1016/j.asoc.2017.03.045

Nath, D. (2024). Unveiling the impact of Indian government policies using aspect-based sentiment analysis with multi-criteria decision-making and hybrid deep learning. *International Journal of Intelligent Systems and Applications in Engineering*, *12*(22s), 72–82.

Nath, D., & Dwivedi, S. K. (2024a). Aspect category detection with attention mechanism. *Proceedings of the 2024 IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies (INSPECT)* (pp. 1–6). IEEE. https://doi.org/10.1109/INSPECT63485.2024.10896198

Nath, D., & Dwivedi, S. K. (2024b). Aspect-based sentiment analysis: Approaches, applications, challenges and trends. *Knowledge and Information Systems*, *66*, 7261–7303. https://doi.org/10.1007/s10115-024-02200-9

Nath, D., & Dwivedi, S. K. (2024c). Transforming aspect-based sentiment analysis to domain-specific mining by integrating a machine learning ensemble (MLE) model. *Proceedings of the 2024 International Conference on Signal Processing and Advance Research in Computing (SPARC)* (pp. 1–6). IEEE. https://doi.org/10.1109/SPARC61891.2024.10828959

Nath, D., Srivastava, A., Pandey, R. K., Chauhan, M., & Satyarthi, A. (2025). A multi-aspect model for identifying people's opinion on Indian government policies. *AIP Conference Proceedings*, *3261*, 050011. https://doi.org/10.1063/5.0258608

Ngxande, M., Tapamo, J., & Burke, M. (2017). Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques. *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics Conference* (pp. 156–161). IEEE. https://doi.org/10.1109/RoboMech.2017.8261140

Olkoniemi, H., & Kaakinen, J. K. (2021). Processing of irony in text: A systematic review of eye-tracking studies. *Canadian Journal of Experimental Psychology*, *75*(2), 99–106. https://doi.org/10.1037/cep0000216

Potamias, R. A., Siolas, G., & Stafylopatis, A. (2020). A transformer-based approach to irony and sarcasm detection. *Neural Computing and Applications*, *32*, 17309–17320. https://doi.org/10.1007/s00521-020-05102-3

Rahgouy, M., Giglou, H. B., Rahgooy, T., & Seals, C. (2022). NULL at SemEval-2022 Task 6: Intended sarcasm detection using stylistically fused contextualized representation and deep learning. *Proceedings of the 16th International Workshop on Semantic Evaluation*, 862–870. https://doi.org/10.18653/v1/2022.semeval-1.120

Rapp, A. M., Purr, F., & Felsenheimer, A. K. (2018). Sarcasm comprehension as a social cognition measure in schizophrenia - – A systematic literature search and meta-analysis on the use of the TASIT. *Schizophrenia Bulletin*, *44*(Suppl 1), S136. https://doi.org/10.1093/schbul/sby016.334

Sharma, E., Gondhi, N. K., Chaahat, Murugappan, M., Naik, G., & Murugan, L. (2026). Toward effective sarcasm detection in social media: A review of artificial intelligence-based automated approaches. *Expert Systems*, *43(3)*, e70203. https://doi.org/10.1111/exsy.70203

Son, L. H., Kumar, A., Sangwan, S. R., Arora, A., Nayyar, A., & Abdel-Basset, M. (2019). Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network. *IEEE Access*, *7*, 23319–23328. https://doi.org/10.1109/access.2019.2899260

Van Hee, C., Lefever, E., & Hoste, V. (2018). SemEval-2018 Task 3: Irony detection in English tweets. *Proceedings of the 12th International Workshop on Semantic Evaluation* (pp. 39–50). Association for Computational Linguistics. https://doi.org/10.18653/v1/S18-1005