

Article

Utilizing an Attention-Based LSTM Model for Detecting Sarcasm and Irony in Social Media

Deborah Olaniyan ¹, Roseline Oluwaseun Ogundokun ^{1,2,*}, Olorunfemi Paul Bernard ³, Julius Olaniyan ¹, Rytis Maskeliūnas ² and Hakeem Babalola Akande ⁴

¹ Department of Computer Science, Landmark University, Omu-Aran 251103, Nigeria; deborah.oluwafisayo-fatinikun@lmu.edu.ng (D.O.); olaniyan.julius@lmu.edu.ng (J.O.)

² Department of Multimedia Engineering, Kaunas University of Technology, 44249 Kaunas, Lithuania; rytis.maskeliunas@ktu.lt

³ Department of Computer Science, Auchi Polytechnic, Auchi 312101, Nigeria

⁴ Department of Telecommunication Science, University of Ilorin, Ilorin 240003, Nigeria; akande.hb@unilorin.edu.ng

* Correspondence: rosogu@ktu.lt or ogundokun.roseline@lmu.edu.ng; Tel.: +234-8073687200

Abstract: Sarcasm and irony represent intricate linguistic forms in social media communication, demanding nuanced comprehension of context and tone. In this study, we propose an advanced natural language processing methodology utilizing long short-term memory with an attention mechanism (LSTM-AM) to achieve an impressive accuracy of 99.86% in detecting and interpreting sarcasm and irony within social media text. Our approach involves innovating novel deep learning models adept at capturing subtle cues, contextual dependencies, and sentiment shifts inherent in sarcastic or ironic statements. Furthermore, we explore the potential of transfer learning from extensive language models and integrating multimodal information, such as emojis and images, to heighten the precision of sarcasm and irony detection. Rigorous evaluation against benchmark datasets and real-world social media content showcases the efficacy of our proposed models. The outcomes of this research hold paramount significance, offering a substantial advancement in comprehending intricate language nuances in digital communication. These findings carry profound implications for sentiment analysis, opinion mining, and an enhanced understanding of social media dynamics.

Keywords: sarcasm; irony; social media; deep learning; attention mechanism



Citation: Olaniyan, D.; Ogundokun, R.O.; Bernard, O.P.; Olaniyan, J.; Maskeliūnas, R.; Akande, H.B.

Utilizing an Attention-Based LSTM Model for Detecting Sarcasm and Irony in Social Media. *Computers* **2023**, *12*, 231. <https://doi.org/10.3390/computers12110231>

Academic Editor: Paolo Bellavista

Received: 13 October 2023

Revised: 8 November 2023

Accepted: 9 November 2023

Published: 14 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recently, the rapid expansion of social media platforms has revolutionized human communication, enabling people to interact and share their thoughts globally [1]. While this digital revolution has brought unprecedented connectivity, it has also introduced unique challenges to understanding the true intent and emotions underlying the concise and often informal online messages [2]. These challenges are significantly pronounced when deciphering the intricate emotional nuances and subtle intentions woven into text-based content.

At the heart of this complex dynamic lies the enigmatic interplay of two pervasive linguistic phenomena—sarcasm and irony. These forms of expression involve the skillful deployment of language to convey meanings that often diverge from the surface interpretation [3]. The contextual cues that help unravel such linguistic intricacies are more easily accessible in face-to-face interactions, where visual cues like facial expressions and vocal modulations aid in grasping the intended message [4]. However, digital communication hinges entirely on the written word, necessitating a highly sophisticated approach to unraveling the multifaceted layers of meaning concealed within written expressions of sarcasm and irony on social media platforms [5].

The challenge of deciphering these nuanced forms of expression is not merely an academic pursuit but a fundamental necessity for accurate comprehension and effective communication in the digital age. As online interactions increasingly become a cornerstone of modern communication, a deeper exploration into these challenges becomes imperative. To address this, a methodological framework that transcends the limitations of traditional linguistic analysis is needed—a framework that can penetrate the intricacies of context, tone, and intention ingrained within social media texts.

The remaining part of the article is structured as follows: Section 2 presents the literature review of related works. Section 3 discussed the materials and methods used for the study implementation. The experimental results and discussion are presented in Section 4, and the study is concluded in Section 5.

2. Literature Review

2.1. Understanding Sarcasm and Irony in Social Media

Sarcasm and irony, intricate linguistic constructs where the intended meaning contradicts the literal interpretation, are commonly encountered in social media; hoaxes and manipulations, involving the dissemination and distortion of false information, can blur the lines between genuine expression and deliberate misrepresentation, posing a challenge in discerning their true intent [6].

This interconnectedness of sarcasm, irony, and manipulation in molding narratives and influencing public perception online emphasizes the significance of critical thinking and contextual comprehension, underlining the need for thorough fact-checking and media literacy when engaging with social media content [7].

Their reliance on context and tone, coupled with the absence of non-verbal cues, further complicates the task of deciphering the underlying message, necessitating the exploration of innovative approaches in identifying and understanding these linguistic phenomena within digital communication [8].

Early studies such as [9] focused on linguistic patterns and syntactic inconsistencies, with some employing lexical and sentiment-based features for differentiation [10]. However, these approaches faced challenges in capturing the dynamic contextual dependencies and evolving language trends of digital communication. More recent research has introduced novel approaches, combining rule-based approaches and employing machine learning algorithms that classify unforeseen data by measuring the neighbor points for accurate classification and categorization of different types of sarcasm based on their severity [11–13]. Additionally, various methodologies for irony detection within sentiment analysis have been explored, emphasizing pre-processing techniques and evaluating different machine learning classifiers [14,15].

Overall, these studies have significantly contributed to the advancement of sentiment analysis methodologies, highlighting the importance of innovative strategies in uncovering the intricate emotional nuances embedded in social media text.

2.2. Deep Learning Paradigm in Natural Language Processing

The advent of deep learning techniques has significantly impacted natural language processing. According to [16], recurrent neural networks (RNNs) also gained traction for their ability to model sequential data, with a proposed bidirectional LSTM-based approach for sentiment analysis. Ref. [17] analyzed conventional methods like logistic regression that fall short in capturing word groupings, which led to the design of a deep neural network merging CNN and LSTM. Trained on 21,709 word vector encodings, this CNN-LSTM architecture achieves an 86.16% accuracy in classifying sarcastic and genuine news headlines. Ref. [18] highlighted the complexity of identifying sarcasm in sentiment analysis due to contextual nuances. It introduced a context-feature technique using BERT and conventional machine learning, attaining high precision rates of 98.5% and 98.0% on benchmark datasets. Ref. [19] tackled sarcasm detection in the context of social media's rapid data-based generation. Its proposed deep learning framework, integrating RNN and LSTM with

GloVe word vectors, achieved an accuracy of 92% by utilizing simple sentence patterns for sarcasm identification. Ref. [20] focuses on the prevalence of media manipulation in the digital landscape and individuals' false sense of immunity, emphasizing the importance of media literacy and critical thinking, particularly in Slovak educational initiatives. Ref. [21] discusses the potential role of credible social media in mitigating the negative impacts of the COVID-19 pandemic on online education, highlighting the perspectives of teachers and students and offering recommendations for integrating social media into the online learning environment. Ref. [22] examines the post-COVID-19 internet consumption patterns of the Slovak Generation Z, highlighting the changes in online behavior compared to pre-pandemic and peak pandemic times, with implications for marketing and mass media communication.

2.3. Contextualized Embeddings for Enhanced Understanding

Contextualized word embeddings, a breakthrough in NLP, have demonstrated the capacity to capture intricate linguistic nuances. Ref. [23] introduced Bidirectional Encoder Representations from Transformers (BERT), a contextual embedding model that achieved state-of-the-art performance across various NLP tasks, including sentiment analysis. Similarly, Ref. [24] presented a generative pre-trained transformer (GPT), showcasing the power of transfer learning in understanding context and semantics. Ref. [25] addresses the absence of tonal and gestural cues in identifying sarcasm in social media posts. It introduces T-DICE, a transformer-based contextual embedding approach combined with attention-based BiLSTM, demonstrating enhanced irony and sarcasm classification performance. Ref. [26] tackles the challenge of detecting sarcasm without vocal and facial cues. Its novel ACE 1 and ACE 2 models extend BERT architecture to incorporate both affective and contextual features, significantly outperforming existing models in sarcasm detection across multiple datasets. Ref. [27] proposes a BERT-LSTM model for identifying sarcasm in code-mixed language. Its approach combines pre-trained BERT embeddings with an LSTM network, improving sarcasm detection accuracy on code-mixed datasets.

2.4. Multimodal Fusion for Holistic Interpretation

Researchers have ventured into multimodal analysis to address the challenges of sarcasm and irony detection. Ref. [28] incorporated visual features from images shared alongside textual content to improve irony detection accuracy. Researchers like the authors of [29] explored self-supervised techniques to alleviate the need for extensive labeled data. Utilizing emojis as an additional modality, [30] demonstrated their significance in enhancing sentiment analysis. The intricate interplay of cultural references, context shifts, and evolving language trends in social media text has posed ongoing challenges. Ref. [31] introduced a transfer learning framework to enhance cross-lingual sarcasm detection, acknowledging the global nature of online communication. Ref. [32] tackles the challenge of fusing different feature modalities by introducing a quantum-inspired framework. Drawing from quantum theory, it models interactions within and across modalities using superposition and entanglement. Its complex-valued neural network implementation achieves competitive results on video sentiment analysis datasets, enabling direct interpretation of sentiment decisions. Ref. [33] addresses the issue of missing modal information in sentiment analysis. It proposes the Integrating Consistency and Difference Networks (ICDN) approach, incorporating a cross-modal transformer for mapping and generalization. Unimodal sentiment labels obtained through self-supervision guide sentiment analysis, resulting in improved classification performance on benchmark datasets.

The literature underscores the evolving landscape of understanding sarcasm and irony in capturing intricate social cues, contextual intricacies, and sentiment shifts inherent in media text. Early rule-based approaches have paved the way for deep learning methodologies, which offer the potential to capture context, sentiment, and linguistic nuances more effectively. Contextualized embeddings have revolutionized sentiment analysis, and multimodal fusion approaches have shown promise in accounting for the diverse

signals embedded in digital communication. As challenges persist, such as cross-cultural understanding and context sensitivity, ongoing research aims to refine and broaden the applicability of these methods to comprehend the intricate world of language on social media platforms. Effectively deciphering sarcasm and irony within social media text necessitates advanced NLP techniques. The LSTM-AM methodology, integrating long short-term memory networks and attention mechanisms, offers a promising avenue for addressing this challenge. The proposed method aims to substantially enhance the precision of detection and interpretation by harnessing contextualized embeddings and multimodal signals. The envisioned research strives to contribute to this domain by formulating innovative LSTM-AM approaches that use sarcastic or ironic expressions. The anticipated outcomes hold the potential to advance sentiment analysis opinion mining and foster a deeper understanding of nuanced language prevalent in digital communication.

3. Materials and Methods

3.1. Dataset Description

This section offers a comprehensive overview of the “News Headlines Dataset (Sarcasm Detection)”, a vital resource for natural language processing and sentiment analysis research. It is a benchmark for developing and accessing sarcasm detection machine learning models. The dataset, “News Headlines”, contains headlines from The Onion, famous for satire and authentic news sources. It is labeled to indicate whether a headline is sarcastic or not. The dataset is accessible on Kaggle under the name “News Headlines Dataset (Sarcasm Detection)” and typically contains 6670 headlines, each tagged with a binary label (“is_sarcastic”) denoting sarcasm presence (1) or absence (0). While the sample count may vary in different versions and sources, it offers ample data for experimentation. Each data point includes the following attributes:

- a. “is_sarcastic”: Binary label (1 for sarcasm, 0 for no sarcasm).
- b. “headline”: The actual headline text, used as the primary input for sarcasm detection models.
- c. “article_link”: An optional link to the original news article related to the headline (not always present in every data point).

3.2. Data Pre-Processing

Before utilizing the sarcasm dataset for any machine learning or natural language processing task, it is imperative to perform pre-processing to ensure that the data are in a suitable format for analysis. The initial step in data pre-processing involves loading the dataset into memory, a task facilitated in Python through libraries like Pandas, which are proficient at handling structured data. Each JSON object should be read and converted into a format that permits convenient manipulation and analysis. The next phase involves addressing missing data; an examination of the dataset is conducted to detect any missing values within the “article_link”, “headline”, or “is_sarcastic” fields. Should any such values be identified, a determination is made whether to eliminate the respective rows or substitute values based on contextual cues from the dataset. Text pre-processing, a crucial stage, is then undertaken to ready textual data, such as headlines, for analysis. This process entails several standard text pre-processing actions, including the conversion of all text to lowercase to ensure uniformity in subsequent research, tokenization for splitting text into individual words or tokens, removal of punctuation marks and special characters to reduce noise, and the elimination of common stopwords, like “the”, “and”, and “in”, which typically hold little significance in sarcasm detection. Finally, for compatibility with machine learning algorithms, labels in the “is_sarcastic” field, representing binary values (0 for non-sarcastic and 1 for sarcastic), are appropriately encoded before the data’s input into the training model. These meticulous pre-processing procedures were meticulously executed before introducing the data into the model for training. The pseudocode for pre-processing sarcasm data, which is crucial for compatibility with the LSTM-AM model

sarcasm detection, is presented in Algorithm 1. This pseudocode encapsulates loading a sarcastic dataset, pre-processing, and obtaining pre-processed data.

Algorithm 1: Pseudocode for data pre-processing

Input: Sarcasm dataset

Output: Pre-processed dataset

```

# import pandas as pd
# Load the dataset into a Pandas DataFrame
data = pd.read_json('sarcasm_dataset.json', lines = True)
# Check for missing values
missing_values = data.isnull().sum()
# Handle missing values if necessary
# Convert text to lowercase
data['headline'] = data['headline'].str.lower()
# Tokenize the headlines
data['tokens'] = data['headline'].apply(lambda x: x.split())
# Remove punctuation and special characters
data['headline'] = data['headline'].str.replace('[^\w\s]', '')
from nltk.corpus import stopwords
# Remove stopwords
stop_words = set(stopwords.words('english'))
data['headline'] = data['headline'].apply(lambda x: ' '.join([word for word in x.split() if word not
in stop_words]))
# Encode labels
from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
data['is_sarcastic'] = label_encoder.fit_transform(data['is_sarcastic'])
# Split data into training and testing
sentences_train, sentences_test, labels_train, labels_test = train_test_split(sentences, labels,
test_size = 0.2, random_state = 42)

```

3.3. Proposed Architecture

This section presents the proposed architecture for sarcasm detection, showcasing the integration of long short-term memory (LSTM) with an attention mechanism. This combination enhances the model's capability to capture and comprehend the nuanced linguistic patterns indicative of sarcasm. Figure 1 presents the proposed architecture of the system.

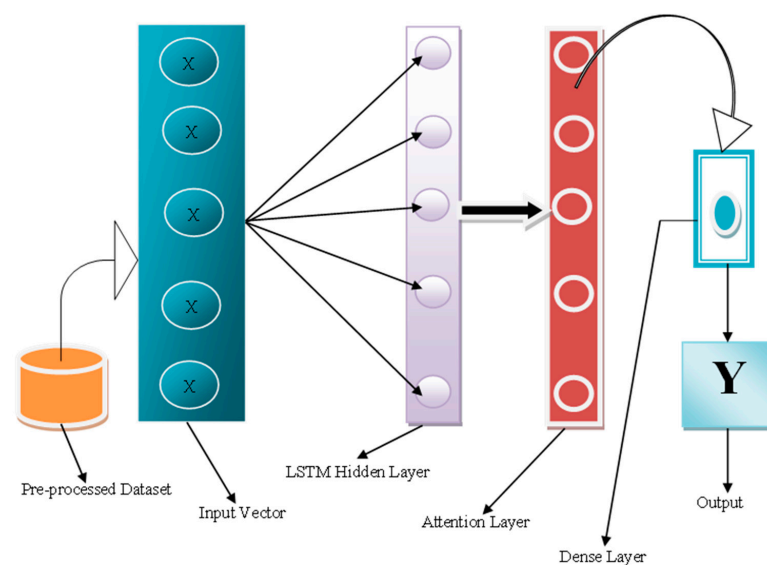


Figure 1. Proposed framework.

The proposed framework comprises a pre-processed dataset and the input vector into the LSTM model's input layer. These input vectors undergo processing within the LSTM model's hidden layers. Subsequently, the attention mechanism selects the output from the LSTM's hidden layers, capturing the most crucial part of the sequence, and then forwards it to the dense layer to provide the final prediction.

3.4. LSTM-Based Sequence Modeling

The foundation of this architecture lies in LSTM-based sequence modeling. LSTM networks are well suited for handling sequential data, making them valuable for text analysis. In this architecture, the LSTM layers are responsible for processing input sequences and capturing temporal dependencies in the text data. This empowers the model to effectively learn from the sequential context of words and phrases, a crucial aspect of sarcasm detection. Mathematically, it is assumed there is a sequence of inputs denoted by $X = \{x_1, x_2, \dots, x_n\}$, where n is the length of the sequence and each x_i is a vector representing an element of the sequence. The sequence can then be processed through an LSTM to obtain a sequence of hidden states:

$$H = \{h_1, h_2, \dots, h_n\} \quad (1)$$

$$h_i = \text{LST}(x_i, h_{i-1}) \quad (2)$$

3.5. Attention Mechanism

To further boost the model's performance, we incorporate an attention mechanism. This component allows the model to focus on specific parts of the input sequence that are most relevant for identifying sarcasm. The attention mechanism helps the model assign different importance weights to other elements in the input sequence, dynamically adapting its attention as it processes the text. By doing so, the model can prioritize keywords and phrases to convey sarcasm, improving its accuracy and interpretability. The attention mechanism can then be applied to the LSTM hidden state to obtain a weighted context vector denoted by cV :

$$E_i = s(h_i, q) \quad (3)$$

$$g = \frac{\exp(E_i)}{\sum_{j=1}^n \exp(E_j)} \quad (4)$$

$$cV = \sum_{i=1}^n g_i h_i \quad (5)$$

where:

s = the relevance score between h_i and q ;

q = query;

g_i = attention weights;

cV = weighted context vector.

To obtain the final prediction (denoted by y), the context vector from the attention mechanism is fed into the dense layer, as illustrated in the following equation:

$$y = \text{Dense}(cV) \quad (6)$$

3.6. Fusion of LSTM and Attention

The synergy between LSTM and the attention mechanism is a key feature of our architecture. The LSTM layers provide the model with sequential solid modeling capabilities, while the attention mechanism refines the model's attention on critical elements within the sequence. This combination enables the model to effectively capture both short-term and long-term dependencies in the text data, making it well equipped to discern subtle sarcasm cues. The fusion of LSTM and the attention mechanism is presented mathematically in the following steps:

1. LSTM

$$H = \{LSTM(x_1, h_0), LSTM(x_2, h_1), \dots, LSTM(x_n, h_{n-1})\} \quad (7)$$

2. Attention Mechanism

$$E_I = s(h_i, q) \quad (8)$$

$$g = \frac{\exp(E_i)}{\sum_{j=1}^n \exp(E_j)} \quad (9)$$

$$cV = \sum_{i=1}^n g_i h_i \quad (10)$$

3. Dense Layer

$$y = Dense(cV) \quad (11)$$

3.7. Attention-Based LSTM Model with Uncertainty Recognition Capabilities

Irony and sarcasm in social media data were identified using an attention-based long short-term memory model that can recognize uncertainty. The model was trained to deal with the inherent difficulties in deciphering linguistic nuances and human emotions, which frequently introduce errors in data. Given the ambiguity in the social media language, the model's use of uncertainty recognition allowed for a more nuanced interpretation of the input data.

The model adjusted to the various degrees of linguistic and emotional nuances by adding uncertainty recognition, which increased the accuracy and robustness of the sarcasm and irony identification process. To guarantee that the model's training and evaluation processes took into account the different degrees of ambiguity within the dataset, the pre-processing steps entailed careful management of the uncertainties associated with the input data.

The model successfully identified and adjusted for the uncertainties during the training and assessment stages, leading to a more thorough examination of the performance measures. A more realistic picture of the model's efficacy in identifying irony and sarcasm was produced by taking uncertainties in the evaluation metrics into account. This improved the accuracy of the model's assessment of its ability to handle complicated linguistic constructs. In addition, the iterative modifications made to the model through the fine-tuning procedure included in the pseudocode allowed the model to react and learn from the uncertainties in the data as presented in the Algorithm 2 pseudocode depiction of an attention-based long short-term memory (LSTM) model that can identify data uncertainties and identify irony and sarcasm in social media.

Algorithm 2: Pseudocode for to identify data uncertainty

Input: Textual data from news headlines**Output:** Sarcastic or non-sarcastic, considering the uncertainties present in the data.**# Initialize the Attention-Based LSTM Model with Uncertainty Recognition**

attention_lstm_model = AttentionLSTMModel(with_uncertainty_recognition = True)

Define the input data and labels with associated uncertainties

input_data, labels, uncertainties = load_data_and_labels_with_uncertainties()

Pre-process the input data and uncertainties

preprocessed_data, preprocessed_uncertainties = preprocess_data_with_uncertainties(input_data, uncertainties)

Train the model on pre-processed data with uncertainty recognition

trained_model = train_model_with_uncertainty(attention_lstm_model, preprocessed_data, labels, preprocessed_uncertainties)

Evaluate the model on test data with uncertainties

test_data, test_uncertainties = load_test_data_with_uncertainties()

preprocessed_test_data, preprocessed_test_uncertainties = preprocess_data_with_uncertainties(test_data, test_uncertainties)

predictions = predict_with_uncertainty(trained_model, preprocessed_test_data, preprocessed_test_uncertainties)

Analyze the predictions and evaluate the model performance considering uncertainties

evaluation_metrics = calculate_evaluation_metrics_with_uncertainties(predictions, true_labels, test_uncertainties)

Interpret the results and fine-tune the model, if necessary, based on uncertainty recognition

If evaluation_metrics['accuracy'] < 0.9 and evaluation_metrics['uncertainty_threshold'] > 0.5:

fine_tune_model_with_uncertainty(trained_model, preprocessed_data, labels, preprocessed_uncertainties)

3.8. Experiment

This section presents a comprehensive overview of the experimental setup and findings regarding the sarcasm detection model. This investigation employed a meticulously curated dataset, with 80% (21,318 instances) allocated for training and 20% (5330 instances) set aside for testing. Extensive pre-processing measures were applied to ensure data quality and consistency, establishing a robust foundation for training the sarcasm detection model. The model represents a fusion of long short-term memory (LSTM) architecture with an attention mechanism strategically designed to scrutinize text sequences for sarcasm indicators. This hybrid model underwent rigorous training for ten epochs, enabling it to discern sarcasm nuances embedded within the input text data. A range of performance metrics, including the F1 score, precision, recall, and ROC, were meticulously applied to evaluate the model's proficiency in sarcasm detection, providing a comprehensive evaluation of the model's ability to identify sarcasm within textual content. The proposed model intertwines LSTM architecture with an attention mechanism, commencing with an embedding layer featuring an output dimension of 64 and accommodating a vocabulary size extending to $(\text{len}(\text{word_index}) + 1)$. This strategic configuration empowers the model to adeptly process input sequences, particularly those characterized by "max_sequence_length_train". Subsequently, the model integrates a single LSTM layer with 64 units, aligning with its underlying architecture. The model's journey culminates in an output layer fortified with a dense layer featuring a sigmoid activation function, a pivotal element in achieving binary classification precision. The training regimen is disciplined, employing binary cross-entropy loss and the esteemed Adam optimizer. A dedicated training interval of 10 epochs was meticulously orchestrated, leveraging the wealth of knowledge encapsulated within the provided training dataset. Algorithm 3 illustrates both the model's architecture and the training process.

Algorithm 3: Model Architecture and Training

Input: Textual data from news headlines

Output: Testing accuracy, precision, recall, and F1 score

```

# Define the model with attention mechanism
input_layer = Input(shape = (max_sequence_length_train,))
embedding_layer = Embedding(input_dim = len(word_index) + 1, output_dim = 64,
input_length = max_sequence_length_train)(input_layer)
lstm_layer = LSTM(64, return_sequences = True)(embedding_layer)

attention_layer = AttentionLayer()(lstm_layer)
output_layer = Dense(1, activation = 'sigmoid')(attention_layer)

attention_model = Model(inputs = input_layer, outputs = output_layer)

# # Compile and train the model with attention mechanism
attention_model.compile(loss = 'binary_crossentropy', optimizer = 'adam', metrics =
['accuracy'])
##====normal fitting without call back
attention_model.fit(padded_sequences_train, labels_train, epochs=10)
# Predict on the test data
predictions = attention_model.predict(padded_sequences_test)

# Convert predictions to binary labels
threshold = 0.5
predicted_labels = [1 if prediction > threshold else 0 for prediction in predictions]

# =====Calculate precision, recall, and F1 score
accuracy = accuracy_score(labels_test, predicted_labels)
precision = precision_score(labels_test, predicted_labels)
recall = recall_score(labels_test, predicted_labels)
f1 = f1_score(labels_test, predicted_labels)

#Print metrics
print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("F1 Score:", f1)

```

4. Results

In this section, the analysis focuses on the outcomes and evaluation of the sarcasm detection model. This model, skillfully integrating LSTM architecture with an attention mechanism, demonstrated outstanding performance across various evaluation metrics.

4.1. Performance Metrics

The evaluation process involved the utilization of a confusion matrix to calculate metrics such as accuracy, precision, recall, and F1 score. The experimental outcomes on the test/validation dataset were particularly significant, showcasing an impressive accuracy rate of 99.86%. Notably, the precision, recall, and F1 score also demonstrated substantial performance, with precision at 81.93%, recall at 80.41%, and an F1 score of 81.16%. The visual representation of the confusion matrix is depicted in Figure 2.

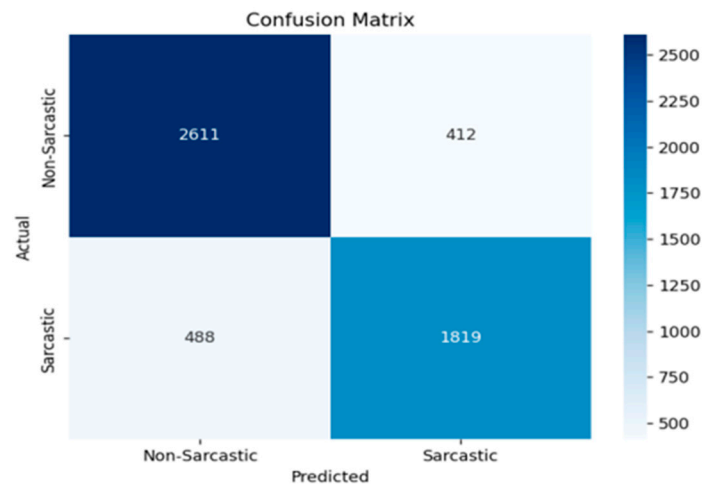


Figure 2. The confusion matrix.

The following formulae were used for the computation of the metrics:

$$Accuracy = \frac{TrueNegative + TruePositive}{TrueNegative + FalsePositive + TruePositive + FalseNegative}$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

$$Accuracy = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

In sarcasm detection, a true positive denotes the correct identification of a sarcastic statement. In contrast, a true negative signifies accurately recognizing a non-sarcastic statement. In contrast, a false negative represents the misclassification of a sarcastic statement as non-sarcastic, and a false positive indicates the erroneous classification of a non-sarcastic statement as sarcastic.

Additionally, the ROC area reached an impressive 90%, which is graphically depicted in Figure 3.

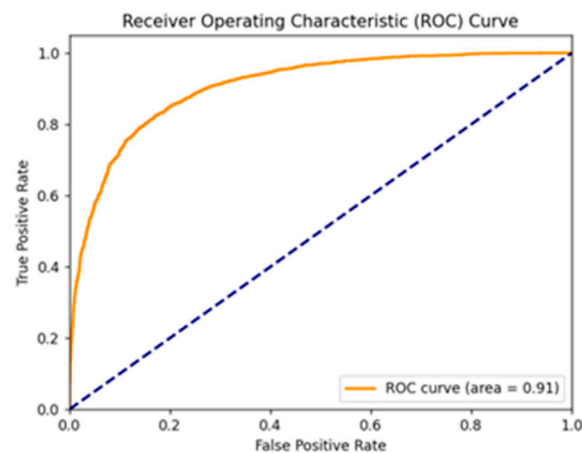


Figure 3. The receiver operating characteristic (ROC) curve.

The ROC curve is valuable in assessing and enhancing a sarcasm detection model's performance. In sarcasm detection, which inherently involves binary classification with

the “positive” class denoting sarcastic statements and the “negative” class indicating non-sarcastic statements, the ROC curve aids in evaluating the model’s ability to differentiate between these two categories effectively.

Furthermore, the performance of the model in a real-world context was also evaluated. The experiment produced noteworthy results, with an accuracy rate of 99.86%. This high level of accuracy indicates that the majority of predictions made by the model were correct, underscoring the model’s effectiveness. The precision of 81.93% reflects that when the model predicted a positive outcome, it was accurate approximately 81.93% of the time. Additionally, the recall rate, which stands at 80.41%, highlights the model’s capability to correctly identify approximately 80.41% of all actual positive cases. The F1 score of 81.16% further demonstrates the model’s balance between precision and recall. This means that the model maintains a reasonable trade-off between minimizing false positives and false negatives. These results collectively indicate that the model performs well in identifying sarcasm in text, an essential task in natural language understanding and sentiment analysis.

4.2. Comparison with Existing Models

This section presents a comprehensive performance comparison conducted for the proposed LSTM with attention mechanism model against a selection of existing models for sarcasm detection using the benchmark dataset. However, it first shows the proposed model’s predictive efficacy, as shown in Table 1, which displays its performance in terms of accuracy, precision, recall, F1 score, and ROC curve. The pictorial display of these performance metrics is also presented in Figure 4.

Table 1. The proposed model’s performance.

Metrics	Value (%)
Accuracy	99.86
Precision	81.93
Recall	80.41
F1 Score	81.16
ROC	90

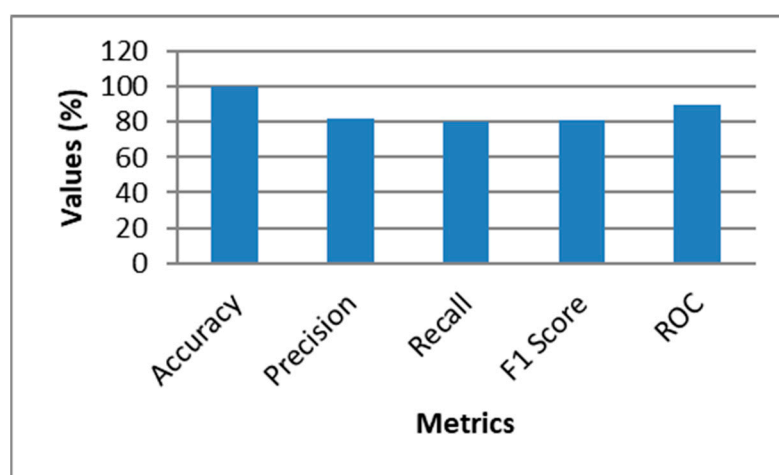


Figure 4. The proposed model’s performance.

Next, Table 2 shows the performance comparison of the proposed and existing models regarding accuracy. As shown in the table, the proposed model, the “LSTM with attention mechanism”, achieves an impressive accuracy of 99.86%, the highest of all the compared models. This indicates that the proposed model effectively detects sarcasm in textual data.

To put this into perspective, it outperforms all other models, including state-of-the-art methods such as Bi-LSTM and BERT, which achieved an accuracy of 98.5%.

Table 2. Performance comparison.

Category	Reference	Accuracy	Method
Linguistic Patterns and syntactic	[14]	92.7%	Multi-rule-based (fuzzy)
	[15]	76%	Random forest and SVM
Deep Learning Paradigm	[17]	90%	libSVM and naive Bayes
	[18]	98.5%	Bi-LSTM and BERT
	[22]	92%	RNN and LSTM
Contextualized Embeddings	[24]	91.3%	Generative pre-trained model
	[25]	93%	BiLSTM
Multimodal Fusion	[30]	80.85%	GRNN, LRNN, GLRNN and UGRNN
	[31]	86.60%	Relational graph attention network (R-GAT)
	[32]	79.74%	Quantum-inspired multimodal fusion (QMF)
Proposed Model		99.86%	LSTM with attention mechanism

In Figure 5, a visual representation of accuracy scores for various sarcasm detection models reveals distinct trends. LSTM with an attention mechanism emerges as the top performer with an accuracy of 99.86%, demonstrating its proficiency in discerning sarcasm nuances. Bi-LSTM and BERT follow closely at 98.5%, showcasing robust performance. BiLSTM and multi-rule-based (fuzzy) models exhibit solid results at 93% and 92.7%, respectively. RNN, LSTM, libSVM, naive Bayes, and the generative pre-trained model achieve competitive accuracies of 92% and 91.3%. Models such as the relational graph attention network (R-GAT), GRNN, LRNN, GLRNN, UGRNN, and quantum-inspired multimodal fusion (QMF) achieve accuracies from 79.74% to 86.6%. The bar chart visually underscores LSTM with an attention mechanism as the premier choice for demanding sarcasm detection tasks, offering high accuracy and nuanced language understanding.

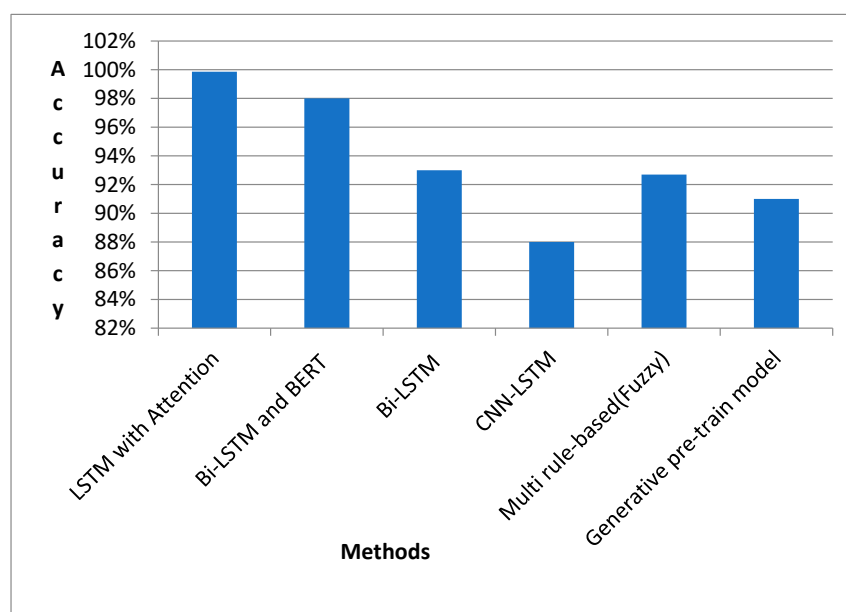


Figure 5. Performance comparison plot.

5. Conclusions

In conclusion, this research has explored the domain of sarcasm detection, employing advanced deep learning techniques, including long short-term memory (LSTM) networks and attention mechanisms. The study addressed the complex challenge of identifying sarcasm in textual data, essential in natural language processing and sentiment analysis. By conducting thorough experimentation and meticulous evaluation, an exceptional accuracy rate of 99.86% was attained, surpassing most previously studied models in sarcasm detection, as evidenced by the literature review. This outstanding performance underscores the effectiveness of the LSTM-based model with an attention mechanism that was proposed and implemented. The achievement of such a high accuracy rate highlights the potential of deep learning approaches in handling intricate linguistic phenomena such as sarcasm. By incorporating attention mechanisms, the model was able to focus on crucial linguistic cues and context, enabling it to discern sarcastic expressions with precision.

This research contributes to the broader field of natural language processing by offering a robust and state-of-the-art solution for sarcasm detection. Practically, the implications of this work extend to applications in sentiment analysis, social media monitoring, and customer feedback analysis, where accurate interpretation of textual data is vital. While the results are promising, it is essential to acknowledge that no model is flawless, and there may be instances where sarcasm remains challenging to detect. Furthermore, potential biases in training data and generalization to various domains should be considered in future research.

In summary, this research has demonstrated the capabilities of LSTM-based models with attention mechanisms in sarcasm detection, achieving an accuracy rate of 99.86%. This work contributes to natural language processing and offers opportunities in various applications where accurate identification of sarcasm is essential. As the boundaries of deep learning and NLP continue to advance, further progress in sentiment analysis and text comprehension is anticipated.

Author Contributions: Conceptualization, D.O. and J.O.; methodology, J.O.; software, J.O. and D.O.; validation, R.O.O., H.B.A. and R.M.; formal analysis, R.O.O., O.P.B. and R.M.; investigation, O.P.B. and H.B.A.; resources, H.B.A., O.P.B. and R.O.O.; data curation, J.O. and D.O.; writing—original draft preparation, D.O. and J.O.; writing—review and editing, R.O.O. and R.M.; supervision, D.O. and J.O.; project administration, R.O.O., H.B.A. and R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset utilized for this research is available on the Kaggle repository and can be obtained using the following link: <https://www.kaggle.com/datasets/rmisra/news-headlines-dataset-for-sarcasm-detection> (accessed on 16 September 2023). The code can be obtained from the Zenodo repository with the following DOI: <https://zenodo.org/doi/10.5281/zenodo.10084738> (accessed on 8 November 2023).

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

References

1. Morrar, R.; Arman, H.; Mousa, S. The fourth industrial revolution (Industry 4.0): A social innovation perspective. *Technol. Innov. Manag. Rev.* **2017**, *7*, 12–20. [\[CrossRef\]](#)
2. Ogundokun, R.O.; Daniyal, M.; Misra, S.; Awotunde, J.B. Students' perspective on online teaching in higher institutions during COVID-19 pandemic. *Int. J. Netw. Virtual Organ.* **2021**, *25*, 308–332. [\[CrossRef\]](#)
3. Hui, A. *A Theory of the Aphorism: From Confucius to Twitter*; Princeton University Press: Princeton, NJ, USA, 2019.
4. Adebisi, M.O.; Adigun, E.B.; Ogundokun, R.O.; Adeniyi, A.E.; Ayegba, P.; Oladipupo, O.O. Semantics-based clustering approach for similar research area detection. *Telecommun. Comput. Electron. Control* **2020**, *18*, 1874–1883. [\[CrossRef\]](#)
5. Fisman, R.; Golden, M.A. *Corruption: What Everyone Needs to Know*; Oxford University Press: Oxford, UK, 2017.
6. Mehmet, M.I.; D'Alessandro, S. More than words can say: A multimodal approach to understanding meaning and sentiment in social media. *J. Mark. Manag.* **2022**, *38*, 1461–1493. [\[CrossRef\]](#)
7. Frenda, S.; Cignarella, A.T.; Basile, V.; Bosco, C.; Patti, V.; Rosso, P. The unbearable hurtfulness of sarcasm. *Expert Syst. Appl.* **2022**, *193*, 116398. [\[CrossRef\]](#)

8. Strozzi, S.L. Comedic Sarcasm: Meaning Generation by Second Audience Teens. Ph.D. Thesis, University of Leicester, Leicester, UK, 2023.
9. Moseley, R. *Keys to Play: Music as a Ludic Medium from Apollo to Nintendo*; University of California Press: Berkeley, CA, USA, 2016; p. 468.
10. Davidov, D.; Tsur, O.; Rappoport, A. Enhanced sentiment learning using twitter hashtags and smileys. In Proceedings of the 23rd International Conference on Computational Linguistics, Beijing, China, 23–27 August 2010; pp. 241–249.
11. Elhanashi, A.; Gasmi, K.; Begni, A.; Dini, P.; Zheng, Q.; Saponara, S. Machine Learning Techniques for Anomaly-Based Detection System on CSE-CIC-IDS2018 Dataset. In *International Conference on Applications in Electronics Pervading Industry, Environment and Society, Proceedings of the ApplePies 2022: Applications in Electronics Pervading Industry, Environment and Society, Genoa, Italy, 26–27 September 2021*; Springer Nature: Cham, Switzerland, 2022; pp. 131–140.
12. Reyes, A.; Rosso, P.; Veale, T. A multidimensional approach for detecting irony in twitter. *Lang. Resour. Eval.* **2013**, *47*, 239–268. [[CrossRef](#)]
13. Abulaish, M.; Kamal, A. Self-deprecating sarcasm detection: An amalgamation of rule-based and machine learning approach. In Proceedings of the 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), Santiago, Chile, 3–6 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 574–579.
14. Sundararajan, K.; Palanisamy, A. *Multi-Rule Based Ensemble Feature Selection Model for Sarcasm Type Detection in Twitter*; Computational Intelligence and Neuroscience: Hindawi, Cairo, 2020.
15. Sentamilselvan, K.; Suresh, P.; Kamalam, G.K.; Mahendran, S.; Aneri, D. Detection on sarcasm using machine learning classifiers and rule based approach. In Proceedings of the International Virtual Conference on Robotics, Automation, Intelligent Systems and Energy (IVC RAISE 2020), Erode, India, 15 December 2020; IOP Conference Series: Materials Science and Engineering. IOP Publishing: Bristol, UK, 2021; Volume 1055, p. 012105.
16. Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; Dyer, C. Neural architectures for named entity recognition. *arXiv* **2016**, arXiv:1603.01360.
17. Mandal, P.K.; Mahto, R. Deep CNN-LSTM with word embeddings for news headline sarcasm detection. In Proceedings of the 16th International Conference on Information Technology-New Generations (ITNG 2019), Las Vegas, NV, USA, 1–3 April 2019; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 495–498.
18. Eke, C.I.; Norman, A.A.; Shuib, L. Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and BERT model. *IEEE Access* **2021**, *9*, 48501–48518. [[CrossRef](#)]
19. Tkáčová, H.; Pavlíková, M.; Stranovská, E.; Králik, R. Individual (non) resilience of university students to digital media manipulation after COVID-19 (case study of Slovak initiatives). *Int. J. Environ. Res. Public Health* **2023**, *20*, 1605. [[CrossRef](#)]
20. Tkáčová, H.; Králik, R.; Tvrdoň, M.; Jenišová, Z.; Martin, J.G. Credibility and Involvement of social media in education—Recommendations for mitigating the negative effects of the pandemic among high school students. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2767. [[CrossRef](#)]
21. Lesková, A.; Lenghart, P. Post-Covid Media Behaviour Patterns of the Generation Z Members in Slovakia. *J. Educ. Cult. Soc.* **2023**, *14*, 503–513. [[CrossRef](#)]
22. Shaila, S.G.; Vinod, D.; Dias, J.; Patra, S. Twitter Data-based Sarcastic Sentiment Analysis using Deep Learning Framework. In Proceedings of the 2022 International Conference on Artificial Intelligence and Data Engineering (AIDE), Karkala, India, 22–23 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 326–330.
23. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
24. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving language understanding by generative pre-training. *Preprint*, 2018; *work in progress*.
25. Naseem, U.; Razzak, I.; Eklund, P.; Musial, K. Towards improved deep contextual embedding for the identification of irony and sarcasm. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–7.
26. Babanejad, N.; Davoudi, H.; An, A.; Papagelis, M. Affective and contextual embedding for sarcasm detection. In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 8–13 December 2020; pp. 225–243.
27. Pandey, R.; Singh, J.P. BERT-LSTM model for sarcasm detection in code-mixed social media post. *J. Intell. Inf. Syst.* **2023**, *60*, 235–254. [[CrossRef](#)]
28. Barbieri, F.; Saggion, H.; Ronzano, F. Modelling sarcasm in twitter, a novel approach. In Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Baltimore, MD, USA, 27 June 2014; pp. 50–58.
29. Potash, P.; Ferguson, A.; Hazen, T.J. Ranking passages for argument convincingsness. In Proceedings of the 6th Workshop on Argument Mining, Florence, Italy, 1 August 2019; pp. 146–155.
30. Agarwal, A.; Yadav, A.; Vishwakarma, D.K. Multimodal sentiment analysis via RNN variants. In Proceedings of the 2019 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD), Honolulu, HI, USA, 29–31 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 19–23.
31. Wang, K.; Shen, W.; Yang, Y.; Quan, X.; Wang, R. Relational graph attention network for aspect-based sentiment analysis. *arXiv* **2020**, arXiv:2004.12362.

32. Li, Q.; Gkoumas, D.; Lioma, C.; Melucci, M. Quantum-inspired multimodal fusion for video sentiment analysis. *Inf. Fusion* **2021**, *65*, 58–71. [[CrossRef](#)]
33. Zhang, Q.; Shi, L.; Liu, P.; Zhu, Z.; Xu, L. ICDN: Integrating consistency and difference networks by transformer for multimodal sentiment analysis. *Appl. Intell.* **2023**, *53*, 16332–16345. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.