JIRBDAI

Journal of Institutional Research, Big Data Analytics and Innovation

A Hybrid LSTM-CNN Model for Text-based Sarcasm Detection

[1]Adenike Azeezat ADAM, [2]Sakirat Adedayo AKEE, [3]Temitayo MARTINS, [4]Obarinde Adetomiwa ABAYOMI-RUFAI

[1,4] Department of Computer Science, Crescent University, Abeokuta, Nigeria
[2]Department of Cybersecurity, Crescent University, Abeokuta, Nigeria
Department of Computing and Engineering, University of Plymouth, Plymouth, United Kingdom
[3]Department of Computer Science, Crescent University, Abeokuta, Nigeria
adenike.adam@cuab.edu.ng, adedayo.akee@cuab.edu.ng,
temitayo.martins@postgrad.plymouth.ac.uk, abayomiobarinde@gmail.com

ABSTRACT

Sarcasm detection is still a difficult task in a Natural Language Processing (NLP) system as the meaning of sarcastic expression is not always the same as the literal meaning. While deep learning methods have proven successful for sarcasm detection, many existing models are ineffective at handling both contextual and local semantic information to correctly classify the sarcasm. This work introduces a hybrid LSTM-CNN model for sarcasm detection in the text domain and compares its accuracy with the accuracy of models such as Simple Neural Network (SNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), BERT and DistilBERT. The News Headlines Dataset for Sarcasm Detection, which contains 28,619 news headlines labeled as sarcastic or not, was used for the experiments. To prepare the text data for the model, it was first preprocessed using a combination of cleaning, tokenization, lower-casing, and padding sequences. The performance of models was assessed through the use of Accuracy, Precision, Recall, F1-score and Binary Cross-Entropy Loss. The experimental results revealed that the proposed sequential CNN-LSTM model achieved better performance than the CNN and LSTM models alone, highlighting the advantage of the sequential CNN-LSTM model in combining sequential contextual learning with CNN feature extraction. Transformer models (BERT and DistilBERT) however had the best overall classification performance. The results show transformer architectures outperform them in terms of predictive accuracy, but the proposed LSTM–CNN model is a lightweight model that is also appropriate for text classification tasks with limited resources.

KEYWORDS: Sarcasm Detection, Convolutional Neural Network, Long Short-Term Memory, Deep Learning, Natural Language Processing, Sentiment Analysis

I. INTRODUCTION

Sarcasm is one of the most developed examples of figurative language used in human communication as the meaning of an expression is often different than what appears in the literal meaning. Sarcastic expressions often require background information, prior discourse, tone and situational context, and shared human experiences to interpret the intended meaning. In principle, humans know how to interpret such context as part of the communication process, but in a computational system, the text is usually taken at its face value, meaning that sarcasm detection is a hard problem in NLP (Sarsam et al., 2020). Such challenges have a profound impact on language understanding applications downstream, such as sentiment analysis, opinion mining, conversational AI, recommendation systems, and social media analytics, where sarcasm can result in wrong sentiment classification and flawed decision-making. With the growing popularity of social media sites, online discussion forums, digital news sites and product review websites, there is an even greater need for reliable automatic sarcasm detection technology. Sarcasm is frequently used to convey criticism, humor, irony, or dissatisfaction with something; millions of user-generated textual contents are generated across platforms like X (formerly Twitter), Reddit, Facebook and news

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.21381960

comment sections online every day. Organizations are increasingly relying on automated text analysis techniques to gain insights into customers' opinions, public perception, political discourse, and intelligent conversational systems (Bharti et al., 2022; Razali et al., 2021). In contrast with oral expression, where tone of voice, facial expression and body language are cues that give additional context, written sarcasm offers relatively few explicit cues making computational detection more difficult.

In the field of sarcasm detection, traditional machine learning methods such as Support Vector Machines (SVM), Naïve Bayes, Logistic Regression and Random Forests have shown good results, but they rely on manually designed lexical and syntactic features (Afiyati et al., 2020). In recent years, deep learning has made significant advancements in feature extraction using techniques like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. CNNs have been shown to be effective at capturing local semantic information from textual data while LSTM networks are better at capturing long-range contextual dependencies within sequential text (Razali et al., 2021). However, stand-alone architectures often do not extract local semantic information and global contextual relationships at the same time, which reduces the performance of the overall classification. More recently, transformer-based language models, such as BERT and its variants, have enjoyed new state-of-the-art performance on a wide range of NLP tasks, by making use of contextualized word representations learned from large-scale corpora. Transformer models, while being powerful in their predictive ability, tend to be far more expensive in terms of computational resources needed for training and deployment, making them impractical in settings with limited computational resources. As a result, lightweight hybrid deep learning architectures are still very appealing for applications where memory efficiency and speed of inference are key practical considerations.

Based on these observations, this study introduces a hybrid LSTM–CNN model which integrates the sequential learning ability of LSTM networks with the local feature extraction power of CNN networks for automatic sarcasm detection. To assess the proposed architecture from a broader perspective, this study introduces BERT and DistilBERT as transformer-based benchmark models, whereas the previous studies only compared the proposed architecture with conventional deep learning models. In addition, multiple trials of the experiments are conducted and the results are statistically verified to enhance the results' reliability and repeatability. This extended assessment examines the pros and cons of employing lightweight hybrid deep-learning models compared to the latest LLM models for sarcasm detection.

II. RELATED WORKS

Sarcasm is a well-documented issue in Natural Language Processing (NLP) as it is a form of expression that does not necessarily convey its literal meaning. In contrast to explicit sentiment expressions, sarcastic statements rely heavily on contextual knowledge, pragmatic reasoning, semantic incongruity and/or commonsense understanding and thus are hard for computational systems to accurately interpret. Joshi et al. (2017) noted that to detect sarcasm, models should be able to pick up on implicit contextual information when they do not rely on lexical information. Likewise, Sarsam et al. (2020) claimed that sarcasm often flips the sentiment of a given statement, making it difficult for traditional sentiment analysis systems to capture the sentiment correctly. Hence, the automatic sarcasm detection is an essential research field due to its impact on further NLP applications like sentiment analysis, opinion mining, conversational agents, fake news detection, and recommender systems (Bharti et al., 2022; Li et al., 2024). Most of the sarcasm detection algorithms generated the initial results used traditional machine learning algorithms (e.g. Support Vector Machines (SVM), Naïve Bayes, Decision Trees (DT) and Logistic Regression (LR)) with hand-crafted lexical, syntactic and sentiment-based features. These strategies performed well on comparatively small datasets, and relied heavily on manual feature engineering and on the domain knowledge of the programmer. Joshi et al. (2017) found that hand-crafted features tend to fail to produce any contextual context for identifying sarcastic vs. literal expressions, nor do they provide any pragmatic information. Similarly, Li et al. (2024) reported that conventional machine learning approaches suffer from poor generalization due to their inability to capture semantic relationships and contextual information

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.21381960

in sarcastic texts. This led researchers to increasingly focus on deep learning techniques that can automatically learn the discriminative representations from raw data in the text. The advent of deep learning revolutionized the field of sarcasm detection, allowing the detection of sarcasm without the need for manual feature engineering. Deep neural networks learn hierarchical semantic representations directly from the text without making any assumption about the handcrafted linguistic features used. Kim (2014) illustrated that Convolutional Neural Networks (CNNs) have been proved to be very effective for sentence classification as they can capture informative local semantic patterns and n-gram features. CNN-based architectures have since been widely used for sarcasm detection, due to their ability to efficiently pick up on salient lexical cues with the aid of pooling operations to reduce the number of features. CNNs, however, tend to only capture local context, neglecting to capture long-range context that is important for understanding sarcastic expressions, especially when these are dependent on information that is also present elsewhere in the sentence/discourse (Sarsam et al., 2020; Li et al., 2024).

However, with these limitations, the use of recurrent neural networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, began to rise in the field of sarcasm detection. LSTM networks were introduced by Hochreiter and Schmidhuber (1997), where memory cells and gates are added to allow the relevant contextual information to be stored over long sequences of text and reduce the vanishing gradient problem that is common in RNNs. In the context of sarcasm detection, this ability is especially useful, as sarcastic meaning is often based on interactions between words at long distances in sentences. However, LSTM networks are good at capturing sequential information, but they may fail to capture important local semantic patterns that may be good indicators for sarcasm. However, stand-alone LSTM models may fail to provide optimal results in certain cases where both the contextual and lexical information needs to be taken into account (Bharti et al., 2022; Li et al., 2024). To combine the advantages of convolutional architecture and recurrent architecture, hybrid deep learning models based on convolutional network (CNN) and long short-term memory (LSTM) network have been studied in recent years. In these architectures, convolutional layers are first used to capture salient local semantic features of textual input followed by LSTM layers to model the contextual relationships between the extracted features. Such a combination allows hybrid architectures to more effectively learn discriminative lexical patterns as well as long-range contextual dependencies than either architecture alone. CNN-LSTM hybrid models have been found consistently to achieve superior results on various sarcasm detection datasets, with only moderate computational overhead, compared to CNN and LSTM models. Hybrid models are therefore found to be an efficient approach to compromise on the computational efficiency and predictive performance, especially for the applications which may not be possible to implement using large transformer models (Sarsam et al., 2020).

A significant development in NLP was the introduction of the Transformer architecture by Vaswani et al. (2017), which abandoned the use of recurrent computations in favor of the self-attention mechanism that can be used to model the relationship between all of the words in a sentence in parallel. To this architecture, Devlin et al. (2019) added Bidirectional Encoder Representations from Transformers (BERT) that learns contextualized word representations by taking into account not only the next word but also the previous word in the training set. With its ability to represent context, BERT quickly became the state-of-the-art solution for many NLP tasks, such as sentiment analysis, question answering and sarcasm detection. Unlike the traditional recurrent models, which process the text in a sequence, BERT is more powerful in modeling the bidirectional context information in the text, which in turn enhances its semantic understanding and classification accuracy (Devlin et al., 2019; Helal et al., 2024). BERT has an excellent predictive performance, but has too many parameters to be trained or utilized during inference, which entails high computation and memory costs. To increase computational efficiency, Sanh et al. (2019) proposed DistilBERT, a shrinkage of BERT by knowledge distillation, that preserves most of the capability of BERT in understanding language. A key advantage of DistilBERT is that it offers more memory efficient and faster inference speeds, allowing it to be used for real-time NLP applications while maintaining almost negligible changes in classification accuracy. Comparative studies have shown that

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.21381960

DistilBERT performs well on a variety of text classification tasks and consumes much less computational resources than BERT (Sanh et al., 2019; Helal et al., 2024).

Sarcasm detection research has elevated attention to the transformer-based architecture in the recent years, owing to its ability to account for contextual information that goes beyond the sentence level lexical patterns. Incorporating contextual information with the News Headlines and MUSTARD datasets, Helal et al. (2024) added RoBERTa and DistilBERT models, and they found that the models achieved a considerable improvement in sarcasm detection accuracy over the traditional deep learning approaches. In the same vein, by carefully engineering fine-tuning strategies, Shu (2024) has shown that transformer-based representations of context, such as BERT and RoBERTa, work well for sarcasm classification. Later, a thorough review by Li et al. (2024) found that the transformer models have become the preferred approach for sarcasm detection studies due to their ability to capture more sophisticated semantic representations compared to traditional deep learning models, but also due to their significantly higher computational requirements for deployment. Despite these advances, recent literature continues to identify several unresolved challenges. Most transformer-based sarcasm detection models prioritize predictive accuracy while paying comparatively less attention to computational complexity, inference time and deployment on resource-constrained devices. Conversely, lightweight deep learning architectures offer greater computational efficiency but often sacrifice contextual representation capability. Recent reviews therefore recommend further investigation into computationally efficient hybrid architectures that balance predictive performance with practical deployment requirements (Li et al., 2024; Helal et al., 2024). Motivated by these observations, the present study develops a hybrid LSTM–CNN architecture and evaluates its performance against both conventional deep learning models and transformer-based language models, including BERT and DistilBERT, using the News Headlines Dataset for Sarcasm Detection. This broader comparative evaluation provides a more comprehensive assessment of the trade-off between classification performance and computational efficiency than previous studies.

Nevertheless, there are still some issues that have not been solved, as described in recent literature. The majority of models that have been developed for sarcasm detection focused on maximizing prediction accuracy with comparatively little attention to factors such as computational complexity, inference time and deployment on resource constrained devices. In contrast to the heavy weight deep learning architectures, the lightweight deep learning architectures are more efficient in computation but might lack capability of representing contextual information. This has led to a number of recent reviews that call for the continued exploration of computationally efficient hybrid architectures that meet both computational and deployment demands (Li et al., 2024; Helal et al., 2024). To address these issues, the present study proposes a hybrid LSTM–CNN model and compares its results with those of traditional deep learning models and the transformer-based language models such as BERT and DistilBERT on the News Headlines Dataset for Sarcasm Detection. This more general comparative analysis offers a more complete comparison between classification accuracy and computational efficiency than past analyses.

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.21381960

Table 1: A Summary of the Major Studies Reviewed

S/N	Year	Authors	Research Method	Strengths	Weaknesses
1.	2021	Saadat etal;	2 datasets were used; Covid19 twitter data and Expo2020 twitter data to train the hybrid model that used the support vector machine (SVM), Naïve Bayes (NB), and Augmented Naïve Bayes (AUGNB).	Expo Dataset performed outstandingly well over (BFTAN)	Low efficiency rate.
2.	2022	Bharti etal;	A text model for textual features, audio model for audio features, and a hybrid that combines both textual and audio features.	The hybrid model performed better than the other models	Multilingual text was not considered.
3.	2022	Lara I Novic	Support vector Machine (SVM) and Logical Regression (LR)	High Processing time.	Small dataset was used.
4.	2018	Rajeswari & Shantibala	Multi-Nominal Naïve Bayes for sarcastic emotions of individuals, and Support Vector Machine (SVM) to sarcasm type	Sarcasm type	No real time tweet.
5.	2021	Razalin etal;	Convolutional Neural Network (CNN) and handcrafted feature sets.	Excellent detection with crafted feastyure sets	Limited dataset
6.	2022	Barhoom etal;	Recurrent Neural network (RNN)	Faster learning rate	Dataset was limited
7.	2021	Bhardwaj etal;	CNN and facial feature	Both performed well for the crafted data	Performance on real time data was lagging
8.	2020	Bingol & Yildirim	Four machine learning algorithms namely; J48 algorithm a WEKA-edited version of the C4.5 decision tree algorithm, Naïve Bayes, Logical Regression & Random Forest	Random forest performed better than other models.	Fewer algorithms and limited feature extraction technique.
9.	2020	Afiyaati etal;	Sarcasm classifiers included Nave Bayes (NB), support vector machine (SVM), K-nearest neighbors (K-NN), LSTM, Convolutional neural network, and Deep neural network.	LSTM, and Deep Neural Network performed better than the other models	Low efficiency rate
10.	2020	Sarsam etal;	Support vector machine and Convolutional neural	High prediction accuracy	Different text processing and classification features does not work swell for the model

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.21381960

III. METHODOLOGY

This study uses a quantitative experimental research approach to create and test a hybrid deep-learning model designed to detect sarcasm in text. The proposed framework is based on the sequential learning ability of a Long Short-Term Memory (LSTM) network combined with the local feature extraction capability of a Convolutional Neural Network (CNN) to classify textual statements as sarcastic and non-sarcastic. The methodological workflow includes: acquisition of datasets, data preprocessing, vectorisation of the texts, development of models, training of models, and performance evaluation and comparison. Four benchmark models were implemented in addition to the proposed hybrid model and they were Simple Neural Network (SNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), BERT and DistilBERT.To give a comprehensive performance comparison, four benchmark models were then implemented: Simple Neural Network (SNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), BERT, and DistilBERT. In this study, the publicly available News Headlines Dataset for Sarcasm Detection from Kaggle was used. This data set consists of 28,619 English-language news headlines (13,634 sarcastic, 14,985 non-sarcastic), each of which is accompanied by a headline, a class label, and an article link. Duplicate records and missing data were

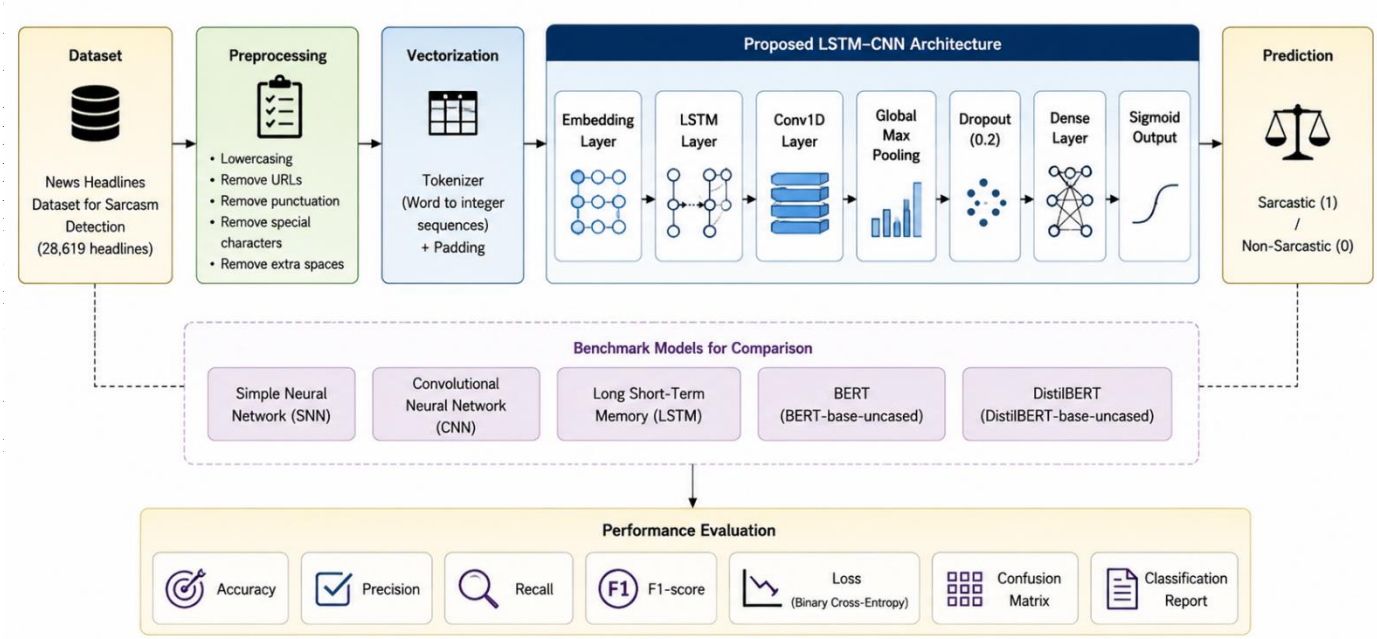

Figure 1: Overall Framework of the Proposed LSTM-CNN Model for Text-Based Sarcasm Detection

The proposed hybrid architecture is the fusion of the complementary strengths of LSTM and CNN networks. An LSTM layer with 64 memory units was then used to model long-range contextual dependencies among textual sequences, after the embedding layer. The LSTM layer-generated contextual feature representations were then passed to a one-dimensional convolutional layer with 128 filters, kernel size 5 and Rectified Linear Unit (ReLU) activation function, for automatic extraction of discriminative local textual features. The extracted features were then fed into a Global Max Pooling layer to reduce the number of features while keeping the most informative features. To prevent overfitting, a Dropout layer with a dropout rate of 0.2 was added before the features were passed to a fully connected dense layer with a sigmoid activation function which was used to classify them into two classes: sarcastic and non-sarcastic. It is presented in Figure 1 the architectural framework of the proposed model. In addition, three classical deep learning models and two transformer-based language models were also developed for comparison. The Simple Neural Network (SNN) consisted of an embedding layer, dropout layers and a fully connected dense output layer. The CNN model consisted of an embedding layer, a 1-D convolutional layer, a Global Max Pooling layer and a dense classifier. The standalone LSTM model consisted of an embedding layer,

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.21381960

an LSTM layer and a dense output layer. In addition, the Sarcasm dataset was used to fine-tune the pretrained BERT-base-uncased and DistilBERT-base-uncased transformer models through the technique of transfer learning. Both transformer models used contextual embeddings, which are created using a self-attention mechanism, and were fine-tuned using the Hugging Face Transformers library.

The conventional deep learning models were trained to the Adam optimization algorithm with an error function of Binary Cross-Entropy, batch size of 128 and six epochs of training. Model learning was tracked with the validation set during the learning process. The optimizer for the transformer models was the AdamW, with a learning rate of 2×10^{-5}, a batch size of 16, weight decay of 0.01, early stopping and 5 training epochs. To enhance the generalization of the model, the best performing model checkpoint was automatically selected based on the model validation F1-score. The performance of all models was evaluated using Accuracy, Precision, Recall, F1-score and Binary Cross-Entropy Loss. Furthermore, a confusion matrix and classification report were created for each model to give a detailed view of each model's classification accuracy. The proposed hybrid LSTM–CNN model was trained and tested in five different experimental runs with different random seeds to ensure the reliability and reproducibility of experimental results. The average and standard deviation of the evaluation metrics were then calculated to evaluate the consistency and statistical stability of the proposed model. Finally, the ability of the hybrid LSTM–CNN model was evaluated for automatic sarcasm detection by comparing it with the SNN, CNN, LSTM, BERT and DistilBERT models.

IV. RESULTS AND DISCUSSION

In this section, the proposed hybrid LSTM–CNN model is tested by experimental analysis and the performance of the proposed model is compared with five benchmark models, namely, Simple Neural Network (SNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), BERT and DistilBERT. To make comparison fair, all models are trained and evaluated on the same News Headlines Dataset for Sarcasm Detection and the same train-test partition is used for all the models. The accuracy, precision, recall and F1-score were used to determine the model's performance. Repeated experimental runs were also performed to check the consistency of the proposed model as well. The performance comparison of all six models is shown in Table 1 and the graphical comparison of the evaluation metrics is shown in Figure 2. The performance of the models evaluated is compared in Table 1.

Table 1: Performance of the Models Evaluated

Model	Accuracy	Precision	Recall	F1-score
BERT	0.929	0.926	0.925	0.925
DistilBERT	0.925	0.923	0.920	0.921
Simple Neural Network (SNN)	0.871	0.865	0.863	0.864
LSTM–CNN	0.835	0.817	0.840	0.829
CNN	0.829	0.835	0.795	0.813
LSTM	0.735	0.659	0.925	0.769

Here is the performance comparison of the evaluated models.

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.21381960

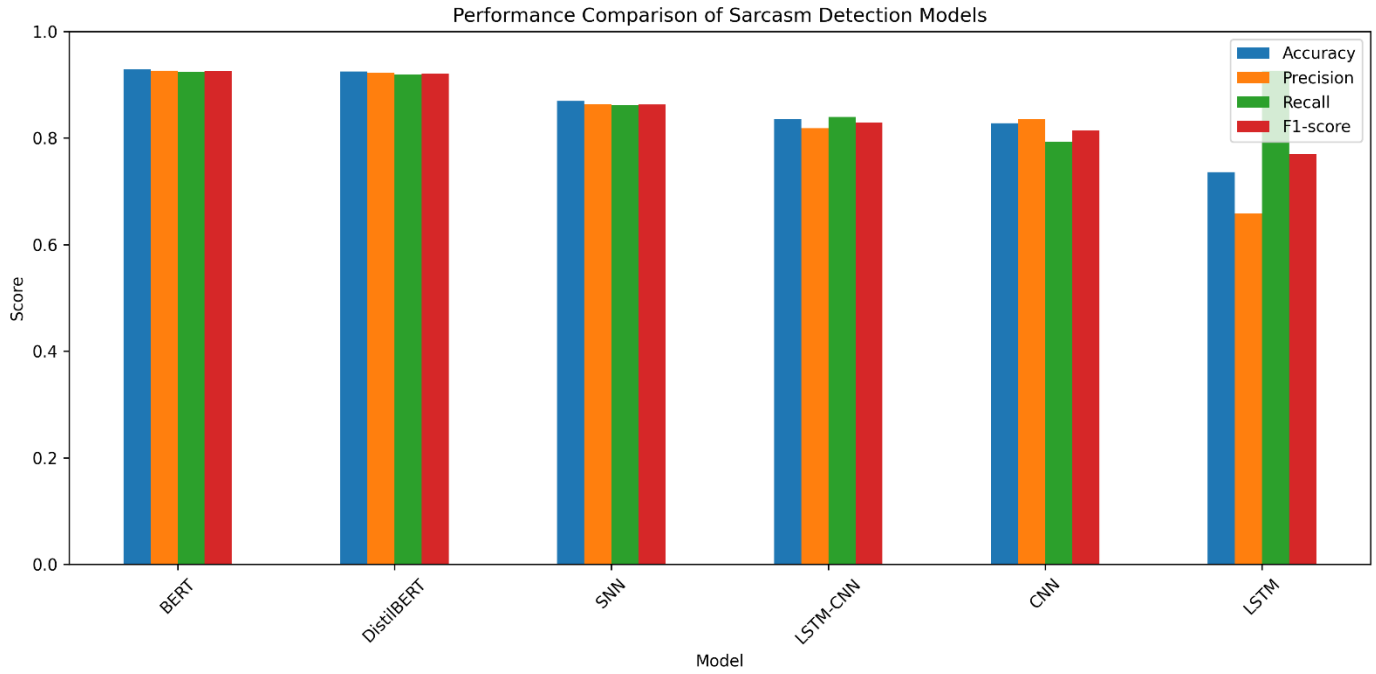

Figure 2: Performance Comparison of the Evaluated Models

As can be seen in Table 1, among the approaches tested, the transformer models showed the best classification performance. In terms of overall performance, BERT achieved the best accuracy (92.9%), precision (92.6%), recall (92.5%) and F1-score (92.5%) scores, demonstrating its superior performance at learning the contextual semantic representations with its self-attention mechanism. The accuracy of DistilBERT was also very high at 92.5%, which shows that knowledge distillation retains the majority of BERT's predictive power with a more efficient architecture. It is interesting to note that among the conventional deep learning models, the proposed LSTM–CNN, CNN and standalone LSTM models, the Simple Neural Network (SNN) happened to be the highest classification accuracy (87.1%). One potential reason for this is that the length and structure of news headlines are relatively short, making it easier to capture textual representations that are sufficiently discriminative to be learnt by simpler neural architectures, which do not need to model sequences as much. However, the classification performance of the proposed LSTM–CNN was slightly better than that of the standalone CNN and significantly better than that of the standalone LSTM model with regard to the overall accuracy and F1-score. This observation indicates that the combination of sequential contextual learning and convolutional feature extraction is a more balanced representation of textual information compared to using either method alone. The proposed LSTM–CNN model achieved an accuracy of 83.5%, precision of 81.7%, recall of 84.0% and an F1-score of 82.9%. The hybrid architecture achieved an improvement in recall and F1-score, better at correctly identifying sarcastic headlines with balanced classification performance, relative to the standalone CNN. Likewise, the standalone LSTM model achieved maximum recall (92.5%) with respect to the other conventional deep learning models, but with a relatively low precision (65.9%), this model misclassified a significant proportion of non-sarcastic headlines as sarcastic, which resulted in a relatively low value of accuracy (73.5%). As a result, CNN+LSTM was able to counterbalance the respective drawbacks of these two networks.

This performance improvement of BERT and DistilBERT is in line with the recent findings of transformer-based language models achieving outstanding results for sarcasm detection due to its capabilities of producing contextualized word representations by using self-attention mechanisms (Devlin et al., 2019; Helal et al., 2024). In contrast to traditional recurrent models that process text sequences sequentially, transformer models can understand the semantic meaning of distant words in the same

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.21381960

sentence, allowing them to detect subtle contradictions that are typical of sarcasm. This boosts the predictive performance, but at the cost of significantly higher computational complexity, memory usage and training time than traditional deep learning models. The proposed LSTM–CNN network, on the other hand, did not achieve the best classification accuracy as compared to the transformer models, but it could still achieve competitive classification accuracy with a much simpler network structure. This is especially appealing for applications with limited resources, where computational efficiency, reduced memory usage, and quick inference are crucial. The proposed hybrid architecture thus provides a reasonable balance between accuracy of prediction and computational complexity, particularly when using a large transformer model is not an option for a specific application. The proposed model was validated using statistical methods. The proposed model was then statistically validated. Five separate experimental runs were performed in order to assess the reliability and stability of the proposed LSTM–CNN model, where the random initialization seeds were varied each time. The results from these the repeated experiments are given in the table2 as descriptive statistics. The statistical validation of the proposed LSTM–CNN model was performed as shown in Table 2.

Table 2: Statistical Validation of the Proposed LSTM–CNN Model

Performance Metric	Mean	Standard Deviation
Accuracy	0.8314	0.0042
Precision	0.8417	0.0220
Recall	0.7965	0.0329
F1-score	0.8178	0.0084

These statistical results show that the proposed LSTM–CNN model achieved stable classification performance for the five independent experimental runs. The relatively small standard deviations of Accuracy (0.0042) and F1-score (0.0084) indicate that the model is relatively stable and not overly sensitive to random initialization of weights. There was a little more variation for Precision (0.0220) and Recall (0.0329), but it is still small enough to show consistent and stable results. The results confirm the validity of the hybrid architecture proposed, and they are also good indication that the reported performance is not the result of a single lucky experimental run. In summary, the experimental results show that although transformer-based language models are currently the most predictive models for sarcasm detection, the proposed hybrid LSTM–CNN model is also a competitive lightweight option that is able to achieve high predictive performance with significantly reduced computational budget while retaining the advantage of being able to capture context while learning local features.

V. CONCLUSION

This study developed a hybrid deep learning framework that combines Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures for automatic text-based sarcasm detection. The proposed model was designed to improve sarcasm classification by integrating the contextual learning capability of LSTM with the local feature extraction strength of CNN. Prior to model development, the textual data were preprocessed through text cleaning, tokenization, lower-case conversion and sequence padding to generate suitable numerical representations for model training. To provide a comprehensive evaluation, the proposed LSTM–CNN model was compared with five benchmark models, namely Simple Neural Network (SNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), BERT and DistilBERT. The experimental results demonstrate that the proposed LSTM–CNN model achieved better overall classification performance than the standalone CNN and LSTM models, confirming that combining sequential contextual learning with convolutional feature extraction provides a more balanced approach to sarcasm detection than employing either architecture independently. However, the transformer-based models, particularly BERT and DistilBERT, achieved the highest

Volume 2, Issue 3, 2026
https://doi.org/10.5281/zenodo.21381960

classification performance among all evaluated models, highlighting the effectiveness of contextualized language representations learned through self-attention mechanisms. Nevertheless, the proposed LSTM–CNN architecture remains a computationally efficient alternative that offers competitive predictive performance with considerably lower model complexity, making it suitable for deployment in resource-constrained environments. Furthermore, the repeated experimental runs demonstrated that the proposed model produces stable and reproducible results, thereby strengthening the reliability of the reported findings. Although the proposed framework demonstrated satisfactory performance, the study was limited to an English-language news headlines dataset obtained from Kaggle. Future research should investigate multilingual and domain-specific sarcasm datasets, explore larger and more diverse benchmark corpora, and evaluate recent lightweight transformer architectures and large language models. In addition, incorporating contextual, multimodal and conversational information may further improve sarcasm detection performance and enhance the applicability of automatic sarcasm detection systems in real-world Natural Language Processing applications.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., … Zheng, X. (2016). *TensorFlow: A system for large-scale machine learning*. Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI), 265–283.

Afiyati, A., Azhari, A., Sari, A. K., & Karim, A. (2020). Challenges of sarcasm detection for social network: A literature review. *JUITA: Jurnal Informatika, 8*(2), 169–176. https://doi.org/10.30595/juita.v8i2.8709

Bharti, S. K., Gupta, R. K., Shukla, P. K., Hatamleh, W. A., Tarazi, H., & Nuagah, S. J. (2022). Multimodal sarcasm detection: A deep learning approach. *Wireless Communications and Mobile Computing, 2022*, Article 1653696. https://doi.org/10.1155/2022/1653696

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 4171–4186. https://doi.org/10.18653/v1/N19-1423

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

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Joshi, A., Bhattacharyya, P., & Carman, M. J. (2017). Automatic sarcasm detection: A survey. *ACM Computing Surveys, 50*(5), Article 73. https://doi.org/10.1145/3124420

Kim, Y. (2014). Convolutional neural networks for sentence classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746–1751. https://doi.org/10.3115/v1/D14-1181

Li, Y., Wang, X., Zhang, Z., & Liu, H. (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

Misra, R. (2022). *News headlines dataset for sarcasm detection* [Data set]. Kaggle. https://www.kaggle.com/datasets/rmisra/news-headlines-dataset-for-sarcasm-detection

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Razali, S., Halin, A. A., Ye, L., Doraisamy, S., & Norowi, N. M. (2021). Sarcasm detection using deep learning with contextual features. *IEEE Access, 9*, 68609–68618. https://doi.org/10.1109/ACCESS.2021.3076789

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *arXiv*. https://arxiv.org/abs/1910.01108

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), 578–598. https://doi.org/10.1177/1470785320921779

Shu, X. (2024). *BERT and RoBERTa for sarcasm detection: Optimizing performance through advanced fine-tuning*. *Applied and Computational Engineering, 97*, 1–11. https://doi.org/10.54254/2755-2721/97/20241354

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Drame, M., & Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6