An Ensemble Approach using RNN, LSTM, BiLSTM and DistilBERT for Sarcasm Detection in News Headlines

Aditya Shinde[1], Dr. Soumya Jana[1]

***Abstract*— One of the most important parts of natural language processing is sarcasm detection, which looks for ironic or mocking meaning in written texts. The complexity and ambiguity of language make the task very difficult. The importance of effective sarcasm detection is growing as a result of the increasing prevalence of social media sites like Facebook, Instagram, and Twitter, where people often share their thoughts on entertainment, politics, and products. The study looks into using broad sequence models to find sarcasm. These include Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Bidirectional Encoder Representations from Transformers (BERT). Additionally, the study thoroughly explores the ensemble approach for sarcasm detection using various combinations of these models. The performance of these models is then evaluated through experimentation and by using the different performance measures, like accuracy, precision, recall, and F1-score. The results demonstrate that the transformer models outperform the sequence models in sarcasm detection by giving the highest testing accuracy of 92.58%.**

I. Introduction

An important challenge in Natural Language Processing (NLP) especially while being used to find sarcasm in text data specifically news headlines [1][2]. Often, sarcasm communicates the opposite meaning of intended meaning, making it hard for their corresponding mechanisms of sentiment analysis and text classification algorithms to correctly interpret meaning [3][4]. The complexity of identifying sarcasm also due to contextual and subtle linguistic clues that help pinpoint sarcasm [5]. Findings from deep learning and natural language processing research led to the construction of many models for the improvement of sarcasm detection [6][7]. RNN, LSTM networks, and BERT proved rather successful. RNNs and LSTMs respectively well function in extracting a better sense of a sequence's temporal dependency and contextual context, which is key for the understanding of sarcasm [1][6]. BERT uses the transformer architecture that delivers full contextual understanding by allowing the bidirectional context of words in the sentence [3][8]. The benefits of the varied independent models are combined to increase the effectiveness of the system for sarcasm detection; thus, various ensemble models have been proposed. The RNN, LSTM, and BERT are to be integrated so that an ensemble model can avail the different strengths each method can offer to achieve better and accurate detection of sarcasm [9]. For instance, RNNs and LSTMs work well with sequential features but BERT provides deep contextual and semantic understanding [7]. This paper proposes an ensemble model that combines RNN, LSTM, and BERT in detecting sarcasm in news headlines. Hopefully, this methodology will get rid of the shortcomings of one model with the advantage of the other through their combination. The proposed model is tried with a dataset of news headlines that capture quite problematic sets wherein issues get intricate and sophisticated because of the nature of sarcasm within those texts [7]. Using such an ensemble approach, we aim to apply them to improve the resilience accuracy of detecting sarcasm while advancing NLP applications like sentiment analysis and fake news detection [8]. The ensemble model combining RNN, LSTM, and BERT together for the fight against sarcasm in news headlines presents an attractive solution to such tricky problems. This study aims to demonstrate the effectiveness of this method and the improvement in efficiency of detection in sarcasm.

II. Neural Networks Models for Natural Language Processing (NLP)

The models are categorized in following two categories:

A. Sequence models for Text Data

1) Recurrent Neural Networks: Recurrent Neural Networks (RNNs) have been widely applied to the challenge of sarcasm detection. Since they are perfectly suited for effective processing of sequential data, which may be critical when dealing with this type of temporal dependency and contextual information necessary in order to determine sarcasm in news headlines. For instance, a study used deep learning that included an RNN-based model as given in Figure 1 for the detection of sarcasm in Telugu news headlines, which ascertained the capability of the model regarding the nuance of sarcasm in a non-English language [9]. Another research stated that it applied neural networks, like RNNs, for the classification of news headlines between sarcastic and not sarcastic, which performed reliablely in the detection of sarcasm [10]. These works establish the capability of RNNs to capture and analyze the sequential aspect of news headlines and interpret them for sarcasm effectively.

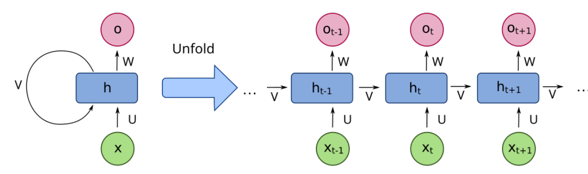


Fig. 1: RNN Architecture

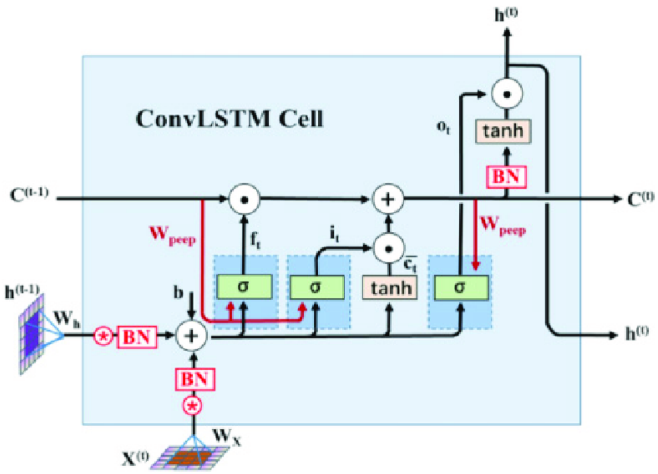


Fig. 2: LSTM Architecture

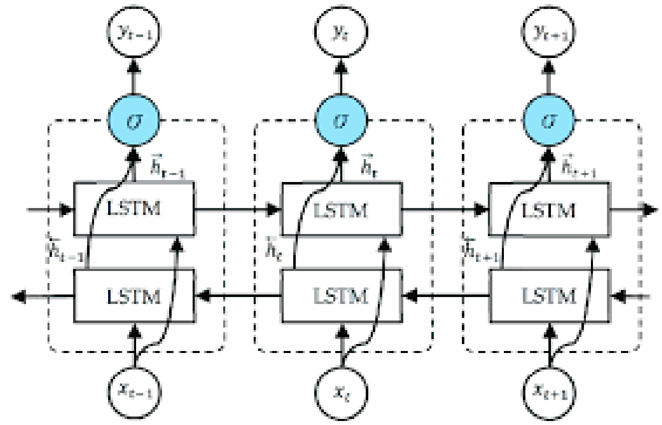


Fig. 3: BiLSTM

2) Long Short-Term Memory: Long Short-Term Memory (LSTM) networks-a variant of RNNs, have especially been apt for sarcasm detection because of maintaining long-term dependencies in text and its architecture is as given in Figure 2. Several studies have applied LSTMs to enhance sarcasm detection in news headlines. For example, one suggested an architecture of deep learning comprising an LSTM along with the GlobalMaxPool1D layer to get the temporal dependencies as well as feature extraction. It reported a high accuracy in sensing sarcasm [11]. Another has combined LSTM with CNN to seize the features of both the time and spatial axis and coming up with an effective model for classifying news headlines to be either sarcastic or genuine [12]. The following are instances that illustrate, in actual fact, LSTMs may enhance the prediction of sarcasm significantly by handling the complexity of language from the news headlines.

3) Bidirectional Long Short-Term Memory: The Bidirectional Long Short-Term Memory (BiLSTM) networks are an improved version of RNN that effectively analyze the sequential data by incorporating both past (backward) and future (forward) contextual information as shown in Figure 3. This bidirectional feature is especially beneficial for tasks such as sarcasm detection, where the comprehension of a word or phrase in a news headline often depends on its wider context. BiLSTMs are built on the LSTM framework, which minimises the vanishing gradient issue present in conventional RNNs, hence boosting the model's capacity to learn long-term dependencies efficiently [10]. BiLSTMs improve the comprehension of nuanced linguistic elements, such as irony and contrast, characteristic of sarcastic remarks by analysing text sequences bidirectionally [13]. The capacity to extract nuanced contextual information renders BiLSTM an effective option for identifying sarcasm in textual data such as news headlines.

B. Transformer Models

1) Bidirectional Encoder Representations from Transformers: Bidirectional Encoder Representations from Transformers (BERT) has emerged as a highly powerful model in many NLP tasks, including sarcasm detection [14]. BERT can understand context from both directions, making it more suitable for detecting sarcasm in the news headlines as shown in Figure 4. It was demonstrated that the BERT-based models significantly outperform classical approaches in tasks related to sarcasm detection. For instance, BERT was applied on a news headline dataset and showed higher accuracy compared to any state-of-the-art model so far, achieving proper results. Another research combined BERT with gradient boosting decision trees (GBDT) towards optimizing efficiency in the detection of sarcasm, which it demonstrated to work better than other strategies [15]. The results above demonstrate that BERT is able to successfully capture complex contextual cues necessary for identifying sarcasm in news headlines.

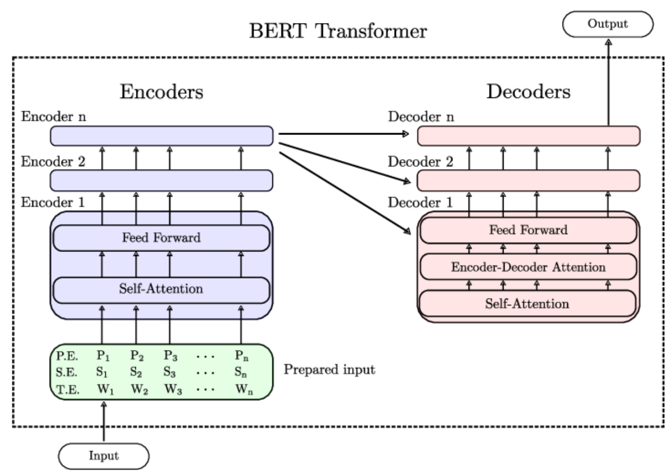


Fig. 4: BERT Architecture

2) DistilBERT: DistilBERT is a concise and efficient variant of the Bidirectional Encoder Representations from Transformers (BERT) Encoder, intended for optimising performance while maintaining computational efficiency. It maintains 97% of BERT's efficiency while being 40% lighter and 60% faster in inference, making it an ideal selection for resource-limited settings. DistilBERT achieves this reduction by knowledge distillation, wherein the smaller model acquires knowledge from the bigger pre-trained BERT model by simulating its behaviour [16]. Its architecture is as shown in Figure 5. Despite its diminished size, DistilBERT effectively generates contextual embeddings, making it especially suitable for natural language understanding tasks like as text classification, sentiment analysis, and, in this instance, sarcasm detection. The small design facilitates rapid training and inference, needed for applications requiring large datasets or real-time processing.

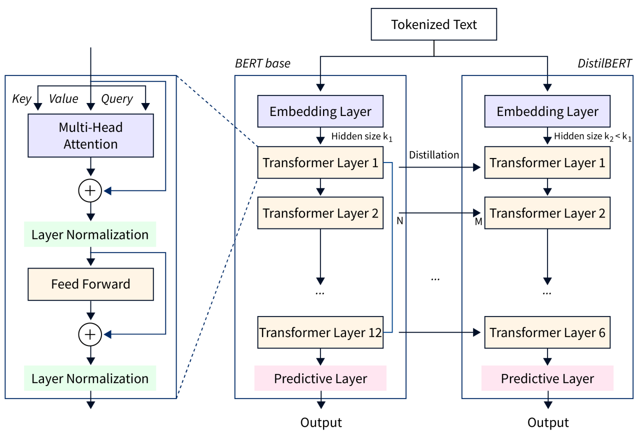


Fig. 5: DistilBERT Architecture

C. Ensemble Variations for Sarcasm detection in News Headlines

In this paper, following variations of ensemble models built from RNN, LSTM and BiLSTM, and BERT are explored and evaluated.

- Ensemble of RNN+LSTM+BiLSTM
- Ensemble of LSTM-BiLSTM
- Stacked Ensemble of Conv-Dense Layer
- Ensemble DistilBERT Models

The proposed architecture of ensemble model for sarcasm detection is as shown in Figure 6. The Section on Experimental Results discusses the performance of each of these ensemble models.

III. Dataset Description

The challenge with some of the more recent Twitter-based datasets about sarcasm was to be overcome in curating this dataset for the News Headlines Dataset. While some noisy datasets rely on hashtag-based supervision, this dataset is actually from two of the most reputable platforms: high-quality sarcastic content from The Onion and authentic, non-sarcastic news by HuffPost. Each record in the data set consists of three attributes: a binary label$is_sarcastic$, which says whether a remark in the headline is sarcastic, the headline of the news, and an article link for further reference [17]. This data set offers several benefits. Professional writing of the headlines eliminates the linguistic inconsistencies of informal platforms like Twitter, eliminating noise, and making

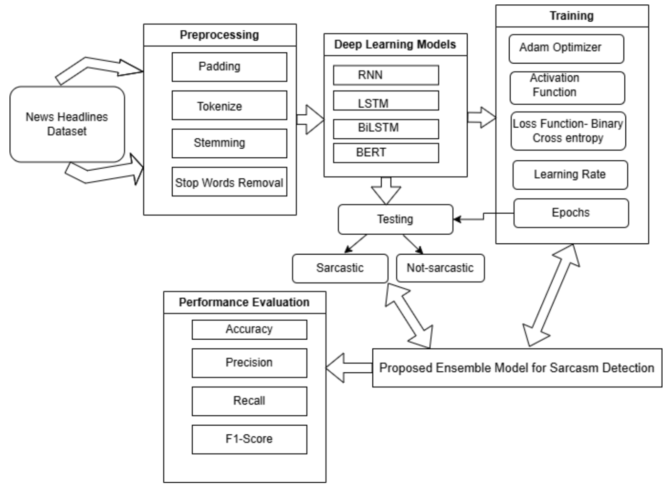


Fig. 6: Ensemble Architecture For Sarcasm Detection

this dataset compatible with pre-trained word embeddings. Furthermore, labels are of extremely high quality and contain minimal ambiguity, as The Onion specializes in sarcasm. Compared to the Twitter reply model, the headlines are self-contained, so researchers can isolate and analyze sarcasm without the need for further contextual information. With these features, the News Headlines Dataset presents a strong and significant resource for advancing research in sarcasm detection. The details are as given in Table I.

TABLE I: Dataset Details

Dataset Semeval	**Headlines**
No. of Records 3,000	28,619
No. of Sarcastic Records 2,396	13,635
No. of Non-sarcastic Records 604	14,984
% Pre-trained Word Embeddings Not Available 35.53%	23.35%

IV. Experimental Results

A. RNN Model

The experimental results of RNN are as given in Table II and in Figure 7. The findings indicate that the model attains perfect training accuracy by the fourth epoch with no training loss, indicating successful optimisation. Validation accuracy is consistent between 82% and 85%, however validation loss escalates from 0.35 to 0.92, indicating overfitting. This underscores the necessity for measures such as regularisation or early stopping to enhance generalisation.

B. LSTM Model

The findings indicate that the model attains 100% training accuracy by the sixth epoch, along with negligible training loss, indicating effective learning from the training data. The validation accuracy stabilises between 83-85%, however

TABLE II: RNN-Training and Validation Results

Epoch	Train-Acc	Train-Loss	Val-Acc	Val-Loss
1	0.68	0.56	0.85	0.35
2	0.94	0.15	0.83	0.44
3	0.99	0.03	0.82	0.53
4	1.00	0.01	0.82	0.65
5	1.00	0.00	0.83	0.72
6	1.00	0.00	0.83	0.78
7	1.00	0.00	0.83	0.81
8	1.00	0.00	0.84	0.85
9	1.00	0.00	0.84	0.88
10	1.00	0.00	0.83	0.92

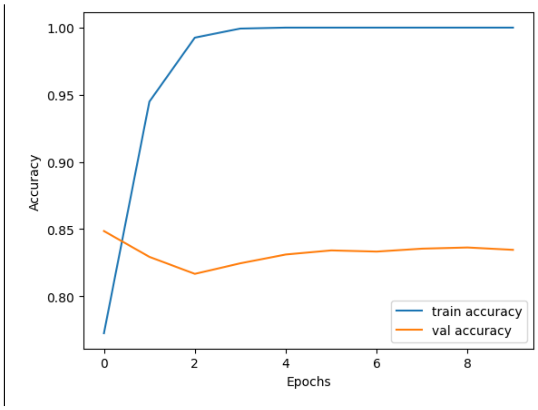


Fig. 7: RNN

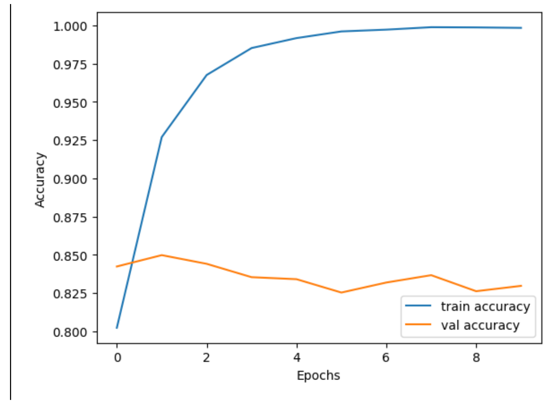


Fig. 8: LSTM

TABLE IV: BiLSTM-Training and Validation Results

Epoch	Train-Acc	Train-Loss	Val-Acc	Val-Loss
1	0.79	0.42	0.35	0.85
2	0.93	0.18	0.38	0.85
3	0.97	0.08	0.50	0.83
4	0.99	0.04	0.60	0.84
5	0.99	0.02	0.69	0.83
6	1.00	0.01	0.90	0.83
7	0.99	0.02	0.78	0.83
8	1.00	0.01	0.84	0.82
9	1.00	0.01	0.87	0.84
10	1.00	0.00	1.00	0.84

the validation loss progressively rises from 0.35 to 0.92, indicating overfitting and necessitating regularisation or early halting to enhance generalisation. These results are given in Table III and Figure 8.

TABLE III: LSTM-Training and Validation Results

Epoch	Train-Acc	Train-Loss	Val-Acc	Val-Loss
1	0.73	0.52	0.84	0.35
2	0.93	0.19	0.85	0.38
3	0.97	0.08	0.84	0.44
4	0.99	0.04	0.84	0.52
5	0.99	0.03	0.83	0.67
6	1.00	0.01	0.83	0.75
7	1.00	0.01	0.83	0.81
8	1.00	0.01	0.84	0.89
9	1.00	0.00	0.83	0.98
10	1.00	0.01	0.83	0.92

C. BiLSTM Model

The findings demonstrate that the BiLSTM model attains a training accuracy of 100% by the sixth epoch, along with a consistently decreasing training loss. Validation accuracy, however, stabilises between 82% and 85%, whereas validation loss consistently rises from 0.35 in the initial epoch to 1.00 in the tenth epoch. This means that the model fits the training data too well and can't generalize well. To fix this, the model needs to be regularized or stopped early. These results are given in Table IV and Figure 9.

The comparative results of all these models are given in Table V and in Figure 10 indicate that RNN surpasses both LSTM and BiLSTM, achieving highest accuracy (0.85), precision (0.85), recall (0.84), and F1 score (0.84). LSTM and BiLSTM exhibit comparable performance, with BiLSTM demonstrating marginally superior recall (0.84). In summary, the RNN is the most efficacious model for this task. The bar plot clearly demonstrates that RNN consistently attains the greatest scores across all parameters, whilst LSTM and BiLSTM exhibit comparable performance, with BiLSTM marginally superior in recall.

TABLE V: Performance Metrics for RNN, LSTM, and BiLSTM Models

Model	Accuracy	Precision	Recall	F1 Score
RNN	0.85	0.85	0.84	0.84
LSTM	0.83	0.83	0.82	0.83
BiLSTM	0.84	0.83	0.84	0.83

D. DistilBERT Model

TABLE VI: DistilBERT-Training and Testing Accuracy

Metric	Accuracy (%)
Training Accuracy	95.24
Testing Accuracy	91.91

Table VI illustrates that DistilBERT attains a training accuracy of 95.24% and a testing accuracy of 91.91%, indicating that it is effective in generalizing to novel data despite slight overfitting.

Table VII illustrates DistilBERT's equitable performance across both categories, exhibiting uniform precision (0.92), recall (0.92), and F1-score (0.92). The macro and weighted

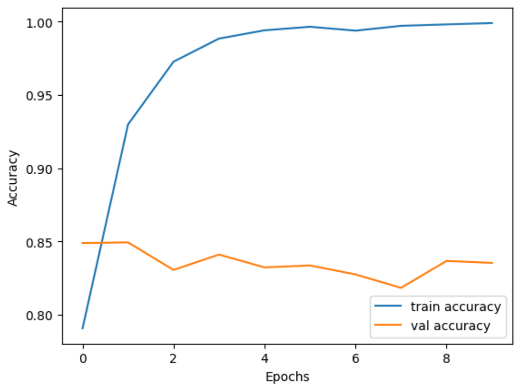


Fig. 9: BiLSTM

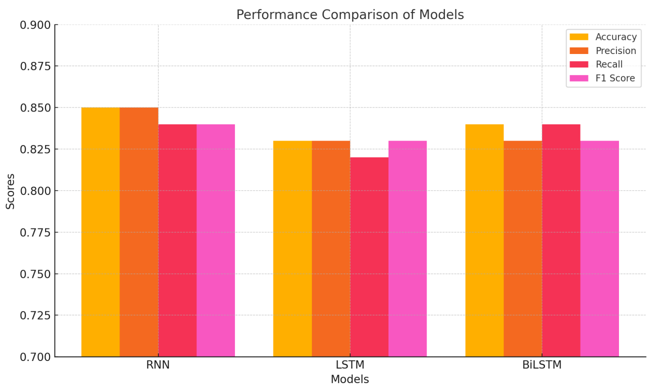


Fig. 10: Plot for RNN, LSTM, and BiLSTM Models

averages validate its strength and dependability in sarcasm detection throughout the sample, providing consistently high classification metrics.

E. Ensemble Model

- The following subsections provide the experimental results of the various ensemble methods mentioned above.

1) Ensemble of RNN+LSTM+BiLSTM: The performance of RNN, LSTM, BiLSTM, and the Ensemble method is given in Table VIII. Although all models attain comparable accuracy, from 85.06% (RNN) to 86.01% (Ensemble), the Ensemble model exhibits superior performance across all parameters, achieving the highest precision (85.26%), recall (85.42%), and F1 score (85.34%). The BiLSTM model exhibits competitive performance, achieving an accuracy of 85.52% together with balanced precision and recall, hence illustrating its proficiency in efficiently capturing bidirectional context.

Table IX depicts the confusion matrix of the ensemble performance. The Ensemble model attains superior classification performance, having the most true negatives (2592) and true positives (2331), and the lowest false positives (403) compared to all other models. This illustrates its enhanced capability to accurately identify both sarcastic and non-sarcastic headlines in comparison to RNN, LSTM, and BiLSTM.

TABLE VII: DistilBERT- Classification Report (1 epoch)

Metric	Class 0	Class 1	Macro Average	Weighted Average
Precision	0.93	0.91	0.92	0.92
Recall	0.92	0.92	0.92	0.92
F1-Score	0.92	0.92	0.92	0.92
Support	2995	2729	-	-

TABLE VIII: Detailed Metrics for RNN, LSTM, BiLSTM, and Ensemble Model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
RNN	85.06	87.84	79.70	83.57
LSTM	85.45	82.94	87.47	85.14
BiLSTM	85.52	83.59	86.63	85.08
Ensemble	86.01	85.26	85.42	85.34

TABLE IX: Confusion Matrix Interpretation for RNN, LSTM, BiLSTM, and Ensemble Model

Model	True Negatives	False Positives	False Negatives	True Positives
RNN	2694	301	554	2175
LSTM	2504	491	342	2387
BiLSTM	2531	464	365	2364
Ensemble	2592	403	398	2331

2) Ensemble of LSTM+BiLSTM: The LSTM achieves a test accuracy of 84.26%, while BiLSTM slightly outperforms it with 84.42%, benefiting from bidirectional context. The ensemble combined the predictions of LSTM and BiLSTM by a simple weighted average:

$P_{ensemble} = 0.5 \times P_{LSTM} + 0.5 \times P_{BiLSTM}$

Both models equally contributed to the final ensemble prediction. The ensemble model further improves test accuracy to 84.77%, with balanced precision (82.25%), recall (86.77%), and F1-score (84.45%), showcasing the effectiveness of combining both models for improved performance. The performance plots of each of these models is given in Figure 11.

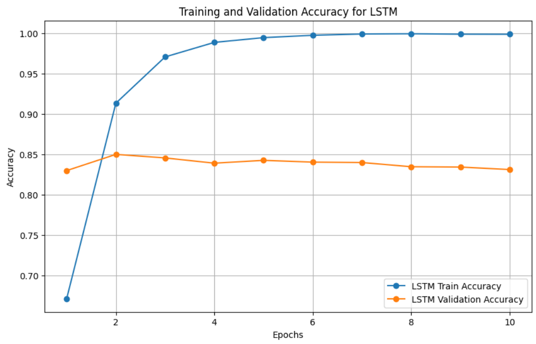
3) Stacked Ensemble of Conv-Dense Layer: The stacked ensemble approach creates multiple base models and uses a meta-learner to combine their predictions, potentially capturing more nuanced patterns in the data. Each model has:

- Embedding layer
- Convolutional layer (Conv1D)
- Global max pooling
- Dropout layers
- Dense layers
- Sigmoid output layer

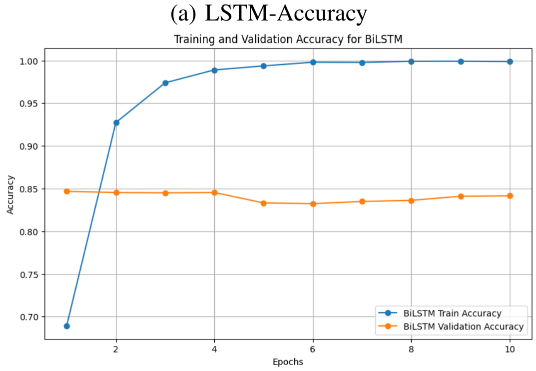
Also to avoid overfitting, following features are implemented-

- Introduced dropout to prevent overfitting
- Early stopping
- Low learning rate

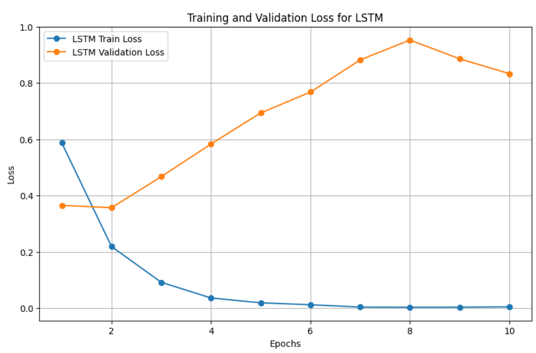
The performances by each of these models are as given Table X and Table XI. Table X presents the performance metrics of various models within the stacked ensemble, with Model2 attaining the highest validation accuracy (84.58%) and the lowest validation loss (0.61). This signifies that Model2 substantially enhances the ensemble's overall effi-



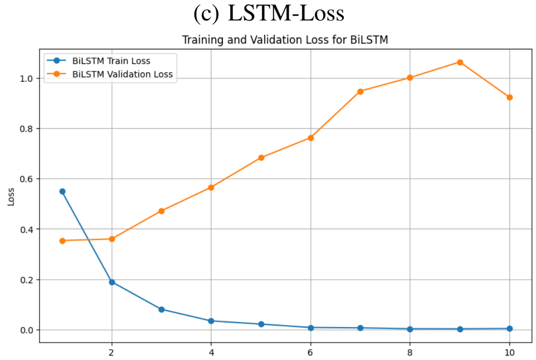
(a) LSTM-Accuracy



(b) BiLSTM Accuracy



(c) LSTM-Loss



(d) BiLSTM-Loss

Fig. 11: LSTM and BILSTM Performances in an Ensemble Model

cacy. Table XI encapsulates the overall performance of the stacked ensemble, attaining an accuracy of 84.62%, precision of 84.51%, recall of 83.20%, and an F1-score of 83.85%, illustrating the ensemble's ability to use the advantages of individual models for enhanced generalisation.

TABLE X: Stacked Ensemble - Individual Model Performance

Model	Train Acc (%)	Val Acc (%)	Train Loss	Val Loss
Model1	98.67	84.15	0.15	0.63
Model2	98.76	84.58	0.14	0.61
Model3	98.22	84.17	0.17	0.62

TABLE XI: Stacked Ensemble-Collective Performance

Accuracy	Precision	Recall	F1 Score
0.8462	0.8451	0.8320	0.8385

The Figure 12 shows that all three models attain elevated training accuracy throughout epochs, with validation accuracy stabilising at comparable levels, signifying equivalent performance among the models in the ensemble.

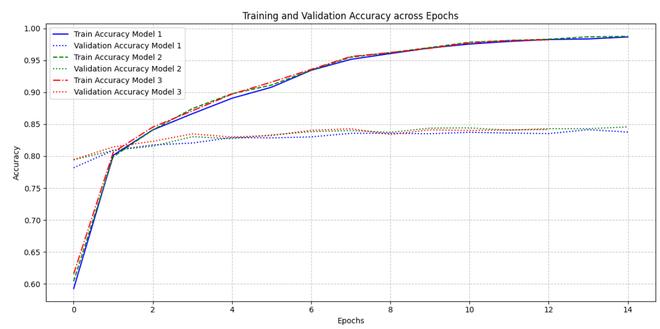


Fig. 12: Plot for RNN, LSTM, and BiLSTM Models

4) Ensemble DistilBERT Models: The ensemble of three DistilBERT models was constructed by training each model with minor variations, including different training parameters. The final predictions are obtained by averaging the outputs of these individually trained models, so assuring diversity and robustness within the ensemble.

The performance of this ensemble is given in Table XII and Confusion matrix of ensemble output is given in Table XIII.

Each model (Model 1, Model 2, and Model 3) attains consistent validation accuracies (92.45%–92.75%) and minimal validation losses (0.34–0.35). The ensemble model attains a maximum test accuracy of 92.58%, utilising the combined advantages of the three models. The confusion matrix shows that the ensemble model exhibits strong performance, by accurately classifying a majority of instances. It attains improved precision and recall for both non-sarcastic (2997 accurate) and sarcastic (2727 correct) categories, with little misclassification (180 and 240 cases, respectively). In addition, the results highlight the ensemble model's enhanced generalisation and its capacity to integrate predictions for superior accuracy and equitable performance across both classes.

TABLE XII: Performance Metrics of Three DistilBERT Models

Model	Acc (%)	Val Acc (%)	Train Loss	Val Loss
Model 1	90.58	92.75	0.0884	0.3443
Model 2	91.07	92.45	0.0839	0.3523
Model 3	91.07	92.50	0.0839	0.3523
Ensemble	92.58 (TestAcc)			

TABLE XIII: Confusion Matrix for Ensemble Model

	Pred Non-Sarc	Pred Sarc
Actual Non-Sarc	2997	180
Actual Sarc	240	2727

V. Conclusion

This study explored sarcasm detection in news headlines using both sequence models (RNN, LSTM, BiLSTM) and transformer-based models (DistilBERT). Multiple ensemble techniques were employed, including an LSTM+BiLSTM ensemble, a stacked ensemble including three models and a meta-learner, and an ensemble of three DistilBERT models. The sequence models and their ensembles exhibited indications of overfitting, attaining a test accuracy of roughly 84%, however the DistilBERT model surpassed all other methods with a test accuracy of 92.58%. This highlights the effectiveness of DistilBERT in identifying subtle patterns in the data while preserving its minimalist design, making it an optimal model for sarcasm detection in resource-limited settings.

References

[1] Ali, R., Iqbal, M. A., Abdullah, S., Akram, S., Farhat, T., Alhajlah, M., Mahmood, A. (2023). Deep Learning for Sarcasm Identification in News Headlines. Applied Sciences, 13(9), 5586. https://doi.org/10.3390/app13095586

[2] Najafabadi, M. K., Ko, T. Z. C., Chaeikar, S. S., Shabani, N. (2024). A Multi-Level Embedding Framework for Decoding Sarcasm Using Context, Emotion, and Sentiment Feature. Electronics, 13(22), 4429. https://doi.org/10.3390/electronics13224429.

[3] Hao, S., Xu, S., Shi, C., Yao, J., Cheng, Y., Li, D., Zhou, Y. (2023). Enhanced Semantic Representation Learning for Sarcasm Detection by Integrating Context-Aware Attention and Fusion Network. Entropy (Basel, Switzerland), 25(6), 878. https://doi.org/10.3390/e25060878

[4] Sharma, D. K., Singh, B., Alhussan, A. A., Pachauri, N., Abdallah, H. A., Agarwal, S. (2023). Sarcasm Detection over Social Media Platforms Using Hybrid Ensemble Model with Fuzzy Logic. Electronics, 12(4), 937. https://doi.org/10.3390/electronics12040937

[5] Olaniyan, D., Akande, H. B., Ogundokun, R. O., Maskeliūnas, R., Olaniyan, J., Bernard, O. P. (2023). Utilizing an Attention-Based LSTM Model for Detecting Sarcasm and Irony in Social Media. Computers, 12(11), 231. https://doi.org/10.3390/computers12110231

[6] Jamil, R., Ashraf, I., Rustam, F., Saad, E., Mehmood, A., & Choi, G. Detecting sarcasm in multi-domain datasets using convolutional neural networks and long short term memory network model. PeerJ Computer Science. 2021; 7. https://doi.org/10.7717/peerj-cs.645.

[7] Jayaraman, A. K., Trueman, T. E., Ananthakrishnan, G., Mitra, S., Liu, Q., Cambria, E. (2022, December). Sarcasm Detection in News Headlines using Supervised Learning. In 2022 International Conference on Artificial Intelligence and Data Engineering (AIDE) (pp. 288-294). IEEE.

[8] Osei-Brefo, E., & Liang, H. UoR-NCL at SemEval-2022 Task 6: Using ensemble loss with BERT for intended sarcasm detection. 2022. https://doi.org/10.18653/v1/2022.semeval-1.121.

[9] Bharti, S., Naidu, R. Babu, K. (2021). Hyperbolic Feature-based Sarcasm Detection in Telugu Conversation Sentences. Journal of Intelligent Systems, 30(1), 73-89. https://doi.org/10.1515/jisys-2018-0475.

[10] Krishna, M. M., Vankara, J. (2023). Detection of Sarcasm Using Bi-Directional RNN Based Deep Learning Model in Sentiment Analysis. Journal of Advanced Research in Applied Sciences and Engineering Technology, 31(2), 352-362.

[11] Eke, Christopher Norman, Azah Shuib, Liyana. (2021). Context-Based Feature Technique for Sarcasm Identification in Benchmark Datasets Using Deep Learning and BERT Model. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3068323.

[12] A. Kumar, V. T. Narapareddy, V. Aditya Srikanth, A. Malapati and L. B. M. Neti, "Sarcasm Detection Using Multi-Head Attention Based Bidirectional LSTM," in IEEE Access, vol. 8, pp. 6388-6397, 2020, doi: 10.1109/ACCESS.2019.2963630.

[13] L. H. Son, A. Kumar, S. R. Sangwan, A. Arora, A. Nayyar and M. Abdel-Basset, "Sarcasm Detection Using Soft Attention-Based Bidirectional Long Short-Term Memory Model With Convolution Network," in IEEE Access, vol. 7, pp. 23319-23328, 2019, doi: 10.1109/ACCESS.2019.2899260.

[14] Potamias, R.A., Siolas, G. Stafylopatis, A . A transformer-based approach to irony and sarcasm detection. Neural Comput Applic 32, 17309–17320 (2020). https://doi.org/10.1007/s00521-020-05102-3

[15] Mandal, P., & Mahto, R. Deep CNN-LSTM with Word Embeddings for News Headline Sarcasm Detection. 16th International Conference on Information Technology-New Generations (ITNG 2019). 2019. https://doi.org/10.1007/978-3-030-14070-069.

[16] Thenmozhi, D. (2023). Sarcasm Detection in Dravidian Languages using Transformer Models.

[17] Data Set link- https://www.kaggle.com/datasets/rmisra/news headlines-dataset-for-sarcasm-detection