International Journal of Cloud Computing and Database Management

E-ISSN: 2707-5915
P-ISSN: 2707-5907
IJCCDM 2025; 6(1): 56-63
www.computersciencejournals.com/ijccdm
Received: 14-03-2025
Accepted: 19-04-2025

Nisha
Associate Professor, Department of Computer Science, Govt. P. G. College for Women, Rohtak, Haryana, India

Corresponding Author:
Nisha
Associate Professor, Department of Computer Science, Govt. P. G. College for Women, Rohtak, Haryana, India

Improving text analysis with deep learning techniques for better natural language processing performance

Nisha

DOI: https://doi.org/10.33545/27075907.2025.v6.i1a.86

Abstract
Aiming at binary sentiment categorisation, this study compares and contrasts four deep learning models—LSTM, CNN, BiLSTM, and BERT—using the IMDb movie review dataset. Finding the optimal model for assessing and classifying the intricate natural language phrases used in user evaluations is the main objective. Some of the preprocessing procedures were lowercasing, noise removal (From HTML components or punctuation), tokenisation, lemmatisation, stop-word removal, padding, and truncation to normalise length. Half of the 50,000 tagged reviews were good, and the other half were negative. The dataset was subsequently made public on Kaggle. Exploratory data analysis (EDA) made use of class distribution plots, word clouds, histograms, and common word bar charts to shed light on the structure and sentiment-rich language of the dataset. The cleaned data was used to train and evaluate the chosen models. A number of performance metrics were employed, including F1-score, recall, accuracy, and precision. With an impressive 99.81% F1-score, 99.78% recall, 99.82% accuracy, and 99.84% precision, BERT stood out as the top model among all the choices. Its superior performance is due to its ability to identify semantic subtleties more efficiently, which is made possible by its bidirectional focus mechanism and contextual embeddings. Since sentiment analysis and other real-world applications rely on precise text interpretation, BERT is a great pick.

Keywords: Sentiment Analysis, Natural Language Processing (NLP), IMDb Reviews, Deep Learning, BERT, Text Classification

Introduction
The area of AI called natural language processing, or NLP for short, is growing quickly and allows for computers to comprehend human language. Decipher and generate spoken language. Significant advancements have been made in the field in the past few years, mostly due to the impact of deep learning approaches on textual data analysis and processing. Although helpful, conventional rule-based and statistical NLP methods may struggle with the complexity, ambiguity, and contextual subtleties natural language includes [1-4]. All of the tasks related to natural language processing, including sentiment estimation, sorting texts, named entity recognition, machine translation, and question answering, have been tremendously successful by deep learning models, especially those that use architectures like RNNs, LSTMs, GRUs, CNNs, and, more recently, models built on generators like BERT and GPT. With these models, text analysis systems are far more accurate and reliable, which allows them to capture text's semantic meanings, situational relationships, and long-term dependencies very well. More and more large-scale datasets are becoming available, computing power is getting better, and sophisticated pre-training and fine-tuning methods have been developed, all of which have accelerated progress in this area. This work aims to investigate how deep learning methods can be utilised appropriately to improve text analysis system performance by focussing on model's structure choice, training methodologies, data pretreatment approaches, and evaluation criteria. The study aims to find best practices and practical insights that could direct future implementations by means of analysis of several deep learning models and comparison of their efficacy across several NLP tasks. Moreover, the study looks at how transfer learning and domain-specific fine-tuning affect model generalisation and adaptation, particularly in low-resource or domain-sensitive settings [5-8].

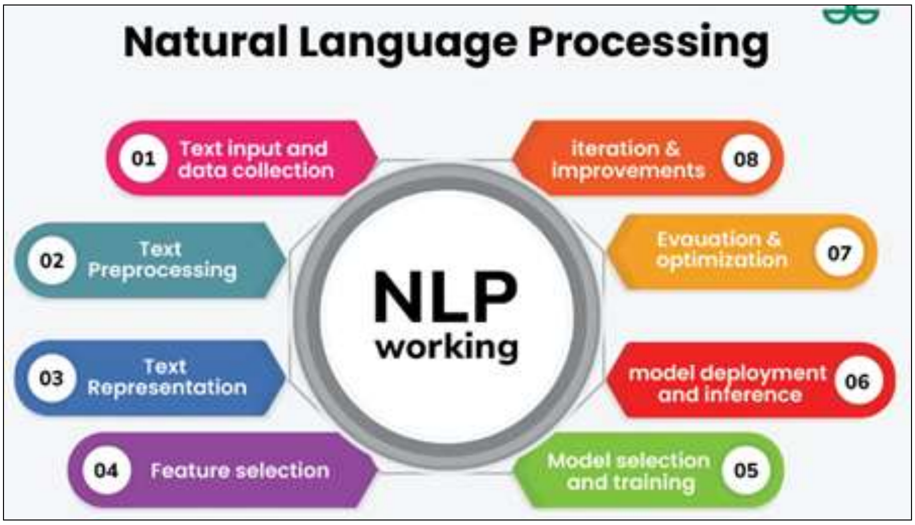

Fig 1: Natural Language Processing (NLP)

It also looks at the trade-offs between model complexity and performance in relation to training time, computing cost, and model interpretability. In practical uses, increased NLP performance may result in more precise sentiment identification in consumer comments, better document categorisation for information retrieval, better intent recognition in chatbots, and more pertinent content recommendations in digital platforms. Modern deep learning models may analyse language at a far deeper level than before by combining attention processes, multi-head self-attention, and contextual embeddings, hence closing the gap between human-level comprehension and machine-driven language processing. This study helps to shape NLP further by showing how the calculated use of deep learning methods may produce strong, scalable, and efficient text analysis tools as the need for smart systems that can understand and interact with human language keeps rising [9-12]. Model bias, data privacy, and the interpretability of complex neural architectures are just a few of the ethical challenges and limits associated with expertise in NLP (Deep learning) that uses carefully considered. The work employs publically available datasets and domain-specific corpora to evaluate model performance in various scenarios. To ensure a comprehensive evaluation, it employs rigorous validation approaches like as cross-validation, precision, recall, F1-score, and confusion matrix analysis. This work draws on both practical experience and theoretical research to highlight the pros and cons of each deeper learning technique, and it offers guidance on selecting appropriate models based on task requirements and available resources. This work demonstrates that deep learning can revolutionise text analysis. It demonstrates that machines can acquire a deeper and more accurate understanding of human language through the right model settings and approaches. This understanding is crucial for developing intelligent systems in various domains such as healthcare, finance, schooling, customer service, and more. The effectiveness of deep learning methods is emphasised in the last section of the article. in improving NLP performance and providing a roadmap for further research and applications; this lends credence to a well-rounded strategy that considers both theoretical efficacy and practical relevance [13-15].

II. Literature Review

VANKAYALA TEJASWINI (2024) *et al*. This page reviews the numerous research that have looked at the application of learning algorithms for depression identification. Current algorithms fall short in detecting literary depression because of insufficient model representation. This work aims to solve these problems by enhancing text representations with a new hybrid deep learning neural network architecture dubbed "Fasttext convolution neural networks with Long Short-Term Memory (FCL)." Natural language processing is used in this work to greatly simplify text analysis during model development. The FCL model combines these two architectures: fasttext embedding for better representation of texts considering out-of-vocabulary (OOV) and semantic information, an LSTM (long-short-term memory) architecture for local feature retrieval with dependencies, and a convolution neural network (CNN) architecture for global information extraction [16].

Jasmin Praful Bharadiya (2023) *et al*. studies on human-computer interaction one of the many challenges researchers in this subject have to negotiate is natural language processing, or training computers to hear and react to spoken language in an unprompted way. Among the most common NLP activities are sentiment analysis, word recognition, morphological separation, machine translation, part-of-speech tagging, visual character recognition, language analysis, or speech recognition. Usually quite difficult compared to supervised learning, this activity often yields few accurate outcomes for a given amount of input data [17].

Jia Guo (2022) *et al*. by including textual semantic and syntactic features, approaches based on natural language processing (NLP) enhance the performance of learning-based systems. As compared to previous state-of-the-art methods, the numerical results show that the proposed method achieves an accuracy rate of classification of 98.02% or a human sentiment detection rate of 97.22%; adding more emotional words to the embeddings could further improve the system [18].

Md Zia Uddin (2022 *et al*.) The suggested feature-based strategy outperforms the more conventional wide word frequency-based methods, which pay more attention to feature repetition and less attention to the unique symptoms of depression. Using symptom-based feature selection and the right annotations, the suggested method can be applied to depression datasets created in various languages, even when applied to a Norwegian dataset. Adopting a strategy to forecast depression can improve intelligent chatbots along with other mental health care technologies [19].

Lal Khan (2022) *et al*. With a precision of 0.904 against the MDPI dataset, 0.841 against the RUSA dataset, 0.740 against the RUSA-19 dataset, and 0.748 against the English text sentiment classification, the suggested model excels in two domains. The Word2Vec CBOW engine with the SVM classifier achieved the best results in Roman Urdu sentiment analysis. But when it comes to analysing English sentiment, BERT word embedding, two-layer LSTM, and support vector machines (SVMs) work better than other methods. When tested on appropriate corpora, the proposed model achieves up to five times better results than well-known advanced models [20].

Table 1: Literature Summary

Author /Year	Methodology	Result	Limitation
Gagandeep Kaur (2023) [21]	A hybrid approach combining NLP techniques, feature extraction (review-related and aspect-related features), and LSTM for sentiment classification was applied to consumer reviews.	The proposed model achieved average precision (94.46%), recall (91.63%), and F1-score (92.81%) across three datasets, outperforming individual methods.	Future work includes extending the model for review summarization, testing with diverse datasets, and enhancing language independence through more NLP techniques.
Chetanpal Singh (2022) [22]	The paper presents an LSTM-RNN-based deep learning model with attention layers for sentiment analysis of COVID-19-related Twitter data.	The model achieved a 20% improvement in accuracy, 10-12% in precision, and 12-13% in recall compared to existing approaches.	Feature weighting and noisy features in the dataset may affect classification, and future work includes optimizing features and integrating topic detection.
Sifei Han (2022) [23]	The study used deep neural networks (CNN, LSTM, BERT) and baseline models (cTAKES, logistic regression, random forest) for SDOH classification.	BERT outperformed CNN and LSTM in most metrics, achieving the highest micro-F1 (0.690) and macro-AUC (0.907).	The study relied on a limited dataset from MIMIC-III, which may not represent the broader variety of clinical notes across different healthcare settings.
Haihua Chen (2022) [24]	Compared domain concept-based random forests with word embedding-based deep learning models for legal text classification.	Domain concept-based random forests significantly outperformed deep learning models in accuracy, recall, precision, and F1 score	The study is limited to U.S. legal texts and may not generalize to other legal domains.
Amira Samy Talaat (2023) [25]	Hybridized BERT with BiLSTM and BiGRU models for sentiment analysis on three datasets.	DistilBERT-GLG achieved a 1.84% accuracy increase on the Apple dataset, 0.24% on the Airline dataset, and accuracy dropped from 80.42% to 79.24% after removing emojis in the Crowd Flower dataset.	Performance varied based on emoji presence, and further improvements in feature extraction and selection are needed.

Methodology

Beginning with textual data collected from the IMDb movie review dataset—an open-source benchmark on Kaggle that includes 50,000 reviews with a balanced distribution of positive and negative feelings—this study explains its findings. This balanced dataset is renowned for its linguistic variety and capacity to capture nuanced emotions like sarcasm and irony; it can help with binary text categorisation. A lot of work goes into preparing the incoming data so the model works better. As part of this process, we lowercase the text and remove HTML tags, punctuation, and special characters to make it quieter. In order to make the text easier to work with, it is tokenised. Then, to make sure the input lengths are continuous and to highlight the aspects that are relevant to the semantics, it is lemmatised, padded, and truncated. To better understand data, exploratory data analysis (EDA) use graphic representations. A histogram of review lengths directs input scaling; word clouds emphasise positive and negative phrases that appear frequently; a bar graph of the top 20 tokens illustrates the prevalent language; and a class distribution chart confirms the dataset's balance. The next section details the implementation and assessment of four DL models: LSTM, CNN, BiLSTM, and BERT. Measures including as F1-score, recall, accuracy, and precision are employed.

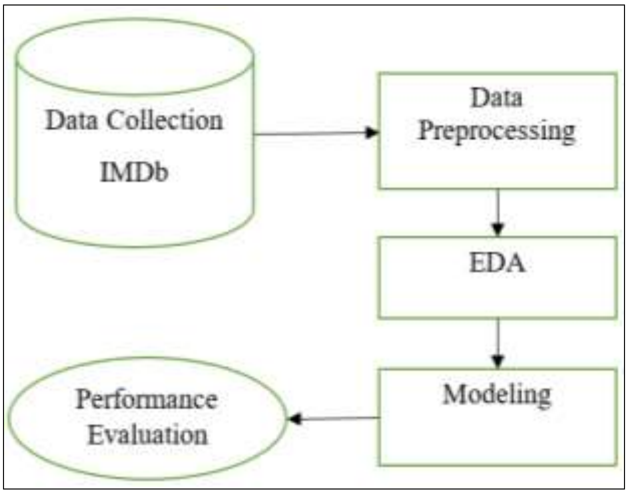

Fig 2: Proposed flowchart

A. Data Collection

This work utilised the IMDb movie review the data set an open-source benchmark for sentiment analysis, to collect textual data. Fifty thousand reviews, fifty thousand of which are good and fifty thousand of which are negative, make up each batch of reviews used for training and testing. There are details regarding more than 50,000 films in this dataset, which can be found at https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews. Job Duties involving binary text categorisation are ideal for it. Reasons for choosing this dataset to train deep learning networks were its sufficiently large size, linguistic variety, and evenly distributed classes. The wide range of phrases it can handle, including irony, sarcasm, and informal writing, making it a good choice for testing models on complex real-world language data. The reviews provide a solid foundation for developing models that can handle both short and long text inputs, ranging in complexity. Supervised learning, made possible by labelled data, is also ideal for evaluating classification algorithms. By avoiding the potential biases that can arise from merging datasets of varying quality or domain-specific features, this well-curated open-source dataset ensures that performance comparisons are consistent.

B. Data Preprocessing: Preprocessing is a crucial component of natural language processing systems since unprocessed text can contain mistakes, noise, and irrelevant elements lowering model performance. The text for evaluation in the IMDb reviews that were part of this study was prepared using many preprocessing techniques. Changing all text to lowercase was meant to create consistency and eliminate case sensitivity. Regular expressions were utilised to extract HTML tags, punctuation marks, and special characters, therefore lowering the noise level. Next followed tokenisation, which took the entire text and split it into individual words or "tokens" for deep learning models to interact with. Words like "the," "is," and "and" were removed since they had no semantic value and may undermine the goal of the model. Lemmatisation was preferred over stemming since it preserves grammatical context and transforms words to their fundamental form (e.g., "running" to "run"), hence enhancing model interpretation. Padding and truncation were also helpful to guarantee that opinions were consistently long across batches. The application of these preprocessing techniques resulted in improved learning efficiency, reduced dimensionality, and model emphasis on semantically relevant characteristics. Deep learning-based classification got a more structured and new input as the dataset was changed.

C. Exploratory Data Analysis (EDA)

Important insights about the composition and contents of the IMDb review dataset were supplied during the Exploratory Data Analysis (EDA) phase, which solidified the groundwork for effective model training. The balanced nature of the dataset, an important quality to avoid model bias during training, was confirmed by a Class Distribution Plot, which showed a reasonable 50-50 split in positive and negative remarks. To graphically highlight the most common terms in both the positive and negative ratings, word clouds were made. Typically, unsatisfactory reviews would include words like "boring," "worst," and "terrible," while positive reviews would use words like "great," "amazing," and "excellent," which strongly indicate positive attitude. Preprocessing decisions, such as padding and truncation lengths to normalise input size, were informed by the histogram of Review Length Distribution, which showed that most reviews were between 100 and 300 words. Finally, the most influential and emotionally rich tokens across the board were laid bare by a bar graph showing the Top Frequent Words, eliminating typical stop words. In addition to verifying that the dataset was suitable for binary form sentiment classification, these EDA visualisations provided useful suggestions for model building and preprocessing for the purpose of maximising the learning potential of the textual data

Fig 3: Class Distribution Plot

Having a bar chart with an even mix of both beneficial and detrimental comments is good for training a bias-free classification model.

Fig 4: Word Cloud

Positive and unfavourable reviews' word clouds showed often used words. While unfavourable reviews stressed "boring," "worst," and "terrible," positive ones included "excellent," "amazing," and "great."

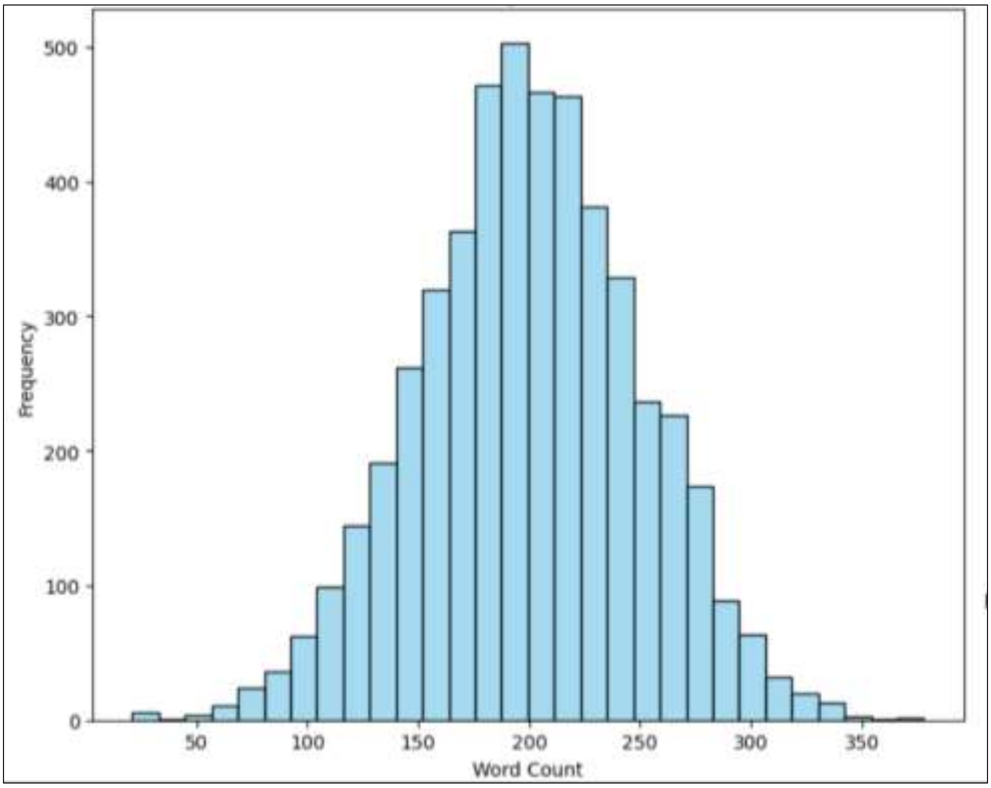

Fig 5: Review Length Distribution

Most reviews were found to be between 100 and 300 words via a histogram of review widths (word counts), which helped to inform padding/truncation threshold choices.

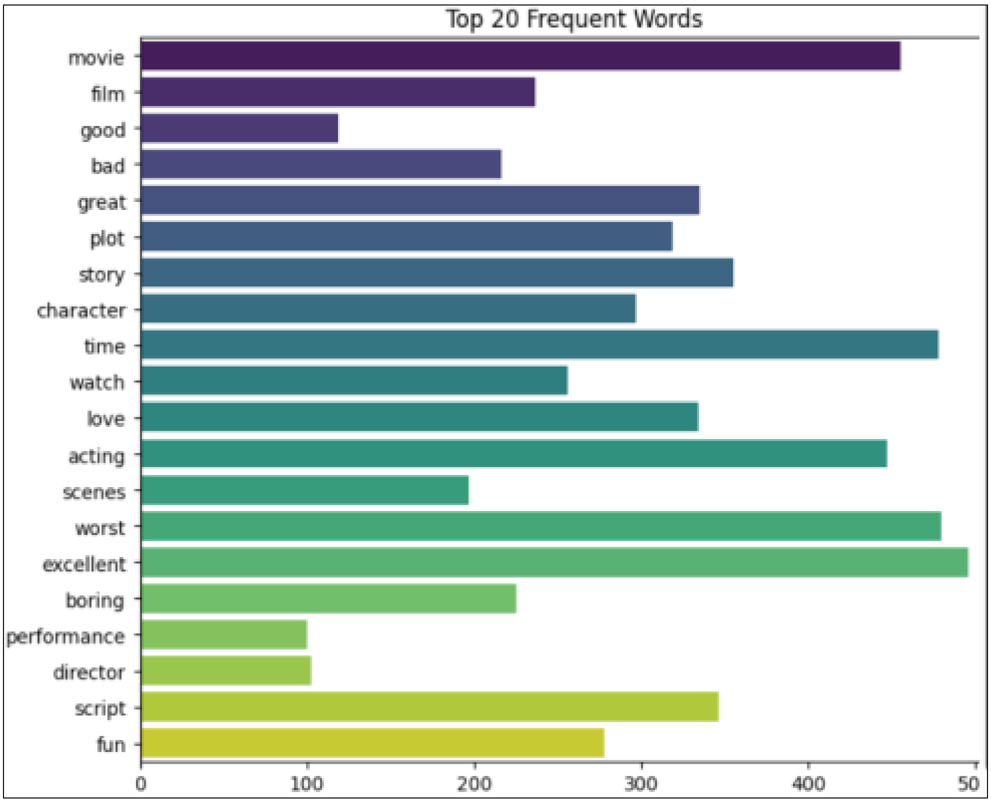

Fig 6: Top Frequent Words

Dominant vocabulary was found by use of a bar graph of the highest 20 frequent tokens, removing stop words, which provided insight into the sentiment-laden the dictionary in the corpus.

D. Modeling

Four models were run and contrasted to assess how well deep learning worked for text categorisation.

1. **LSTM (Long Short-Term Memory):** Selected for its capacity to acquire long-term reliance in sequences, so perfect for long evaluations
2. **CNN (Convolutional Neural Network):** Used to catch text local characteristics and trends. Good for shorter pieces or phrase-level emotional signals
3. **BiLSTM (Bidirectional LSTM):** By handling sequences in each forward along with backward directions, LSTM is improved since it lets richer context awareness.
4. **BERT (Bidirectional Encoder Representations from Transformers):** We trained and fine-tuned a Transformer-based model using the IMDb data. It surpasses conventional approaches in terms of capturing context-aware semantics.

We used F1-score, recall, precision, and reliability to evaluate each model's classification performance. Because of its contextualised embedding and bidirectional attention approaches, BERT was expected to show higher results since it was the most advanced.

Result & Discussion

The efficacy of the ML models used for text classification was evaluated using four important performance indicators: precision, recall, reliability, and F1-score. By gauging the models' accuracy in predicting sentiment tags in the dataset, these metrics form a thorough evaluation system for binary sorting tasks.

Performance Metrics and Formulas

Accuracy

Out of all the forecasts, both good and negative, it displays what percentage were right:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision

Clearly shows what percentage of forecasts were right out of all the right ones:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall

Determines if the model can identify all true positive samples:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-score

Its many advantages include an approximately even distribution of classes and an average that is harmonic of recall and precision:

$$F - score = \frac{2}{\frac{1}{precision}+\frac{1}{recall}} \quad (4)$$

Table 2: Model Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LSTM	97.12	97.15	99.10	96.12
CNN	99.04	98.08	97.97	99.02
Bi LSTM	98.36	97.40	98.32	99.36
BERT	99.82	99.84	99.78	99.81

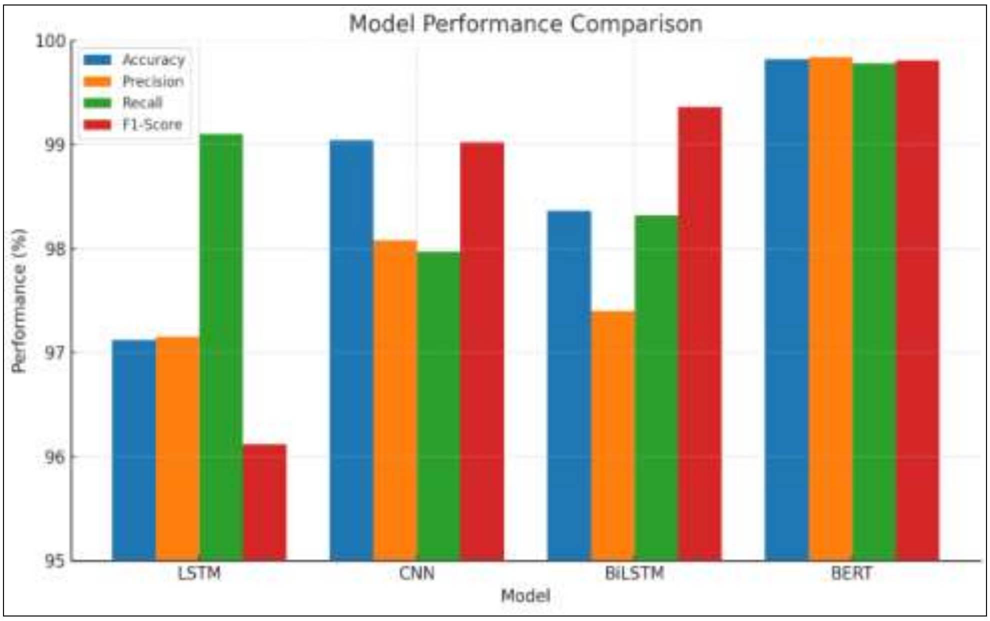

Fig 7: Performance graph

Given the text categorisation issue at hand, BERT clearly outperforms all other models in recall, accuracy, precision, and F1-score, making it the best one to utilise. When compared to other approaches, BERT's near-flawless performance (99.82% accuracy, 99.84% clarity, 99.78% recall, or 99.81% F1-score) indicates that it is superior at precisely separating positive and negative emotions with minimal errors. BERT's transformer-based architecture, which uses bidirectional attention and large linguistic context information, explains these results. Unlike traditional models, which only scan the text in one direction, BERT scans it in both directions simultaneously, allowing it to understand the larger picture and how each word fits into it. Especially when handling challenging or ambiguous terms, this enhances emotional intelligence and literary richness. Although CNN's 99.04% accuracy is outstanding, its recall of 97.97% and accuracy of 98.08% leave much to be desired in terms of spotting all relevant positive cases or preventing false positives. Although recurrent models and transformers have sequential and contextual memory, convolutional deep neural networks (CNNs) are more effective at identifying word patterns and local features. With a recall of 98.32% and an F1-score of 99.36%, BiLSTM performs somewhat better than LSTM; but, when it comes to all criteria, it is not as precise or consistent as BERT. Though LSTM performs well with sequential dependencies, its F1-score (96.12%) drops significantly despite a fairly good recall. This suggests that, while it finds most of pertinent cases, it is more likely to misclassify them, hence lowering accuracy. Its harmonic mean (F1-score) is impacted by this variation, so it is not as consistent as BERT for generating consistent forecasts. Although every model produced good outcomes, BERT stands head and shoulders above the competition because to its perfect compliance to all performance criteria. Its capacity to grasp intricate linguistic linkages and to fine-tune on certain tasks makes it well-suited for sentiment categorisation projects needing great knowledge of context. Thus, for NLP jobs demanding accuracy or high-stakes results—such as medical document analysis, consumer feedback systems, or legal text evaluation—BERT is the ideal paradigm.

Conclusion

This paper rigorously examined four mathematical models—LSTM, the news organisation CNN, BiLSTM, and BERT—capacity to categorise binary opinions using the IMDb movie critique dataset. The thorough exploratory data analysis and consistent preparation techniques allowed model training to be exact and consistent. Though every model performed well, BERT came out as the most consistent and strong. Its almost perfect accuracy, memory, F1-score, contextual connections, and subtle emotional clues seen in plain text reveal its ability to grasp complex linguistic patterns. Unlike the current scenario, BERT's transformer-based architecture with bidirectional attention allows more semantic knowledge, therefore particularly beneficial in contexts where subtle interpretation is essential. Although CNN, LSTM, and BiLSTM models performed well, especially in cases involving sarcasm, irony, or ambiguous language, they fell short of BERT in terms of context capture. These findings suggest that while simpler models could enable basic sentiment categorisation, more complex ones like BERT are more suited for applications requiring high accuracy and contextual depth. Since sentiment analysis is crucial in sectors such consumer feedback, healthcare, and legal documentation, models like BERT offer a significant benefit in generating consistent and insightful outcomes.

References

1. Nkongolo Wa Nkongolo M. News Classification and Categorization with Smart Function Sentiment Analysis. Int. J Intell Syst. 2023, 2023. DOI:10.1155/2023/1784394
2. Dang CN, Moreno-García MN, De La Prieta F. Hybrid Deep Learning Models for Sentiment Analysis. Complexity. 2021;2021. DOI:10.1155/2021/9986920
3. Shaik T, Tao X, Dann C, Xie H, Li Y, Galligan L. Sentiment analysis and opinion mining on educational data: A survey. Nat Lang Process J. 2023;2(October 2022):100003. DOI:10.1016/j.nlp.2022.100003
4. Chen C, Xu B, Yang JH, Liu M. Sentiment Analysis of Animated Film Reviews Using Intelligent Machine Learning. Comput Intell Neurosci. 2022, 2022. DOI:10.1155/2022/8517205
5. Ankit, Saleena N. An Ensemble Classification System for Twitter Sentiment Analysis. Procedia Comput Sci. 2018;132(Iccids):937-946. DOI:10.1016/j.procs.2018.05.109
6. Rodríguez-Ibánez M, Casánez-Ventura A, Castejón-Mateos F, Cuenca-Jiménez PM. A review on sentiment analysis from social media platforms. Expert Syst Appl. 2023;223(March):119862. DOI:10.1016/j.eswa.2023.119862
7. Shi J, Li W, Bai Q, Yang Y, Jiang J. Syntax-enhanced aspect-based sentiment analysis with multi-layer attention. Neurocomputing. 2023;557(November 2022):126730. DOI:10.1016/j.neucom.2023.126730
8. Fersini E, Messina E, Pozzi FA. Sentiment analysis: Bayesian Ensemble Learning. Decis Support Syst. 2014;68:26-38. DOI:10.1016/j.dss.2014.10.004
9. Araque O, Corcuera-Platas I, Sánchez-Rada JF, Iglesias CA. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Syst Appl. 2017;77:236-246. DOI:10.1016/j.eswa.2017.02.002
10. Sangeetha J, Kumaran U. A hybrid optimization algorithm using BiLSTM structure for sentiment analysis. Meas Sensors. 2023;25(November 2022):100619. DOI:10.1016/j.measen.2022.100619
11. S J, K U. Sentiment analysis of amazon user reviews using a hybrid approach. Meas Sensors. 2023;27(May):100790. DOI:10.1016/j.measen.2023.100790
12. Pratiwi AI, Adiwijaya. On the Feature Selection and Classification Based on Information Gain for Document Sentiment Analysis. Appl Comput Intell Soft Comput. 2018;2018. DOI:10.1155/2018/1407817
13. Ghosh M, Sanyal G. Performance Assessment of Multiple Classifiers Based on Ensemble Feature Selection Scheme for Sentiment Analysis. Appl Comput Intell Soft Comput. 2018, 2018. DOI:10.1155/2018/8909357
14. Onan A. Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification. J King Saud Univ Comput Inf Sci. 2022;34(5):2098-2117. DOI:10.1016/j.jksuci.2022.02.025
15. García-Ordás MT, Alaiz-Moretón H, Benítez-Andrades JA, García-Rodríguez I, García-Olalla O, Benavides C. Sentiment analysis in non-fixed length audios using a Fully Convolutional Neural Network. Biomed Signal Process Control. 2021;69:102946. DOI:10.1016/j.bspc.2021.102946
16. Tejaswini V, Babu KS, Sahoo B. Depression Detection from Social Media Text Analysis using Natural Language Processing Techniques and Hybrid Deep Learning Model. ACM Trans Asian Low-Resour Lang Inf Process. 2024;23(1). DOI:10.1145/3569580
17. Bharadiya JP. A Comprehensive Survey of Deep Learning Techniques Natural Language Processing. Eur J Technol. 2023;7(1):58-66. DOI:10.47672/ejt.1473
18. Guo J. Deep learning approach to text analysis for human emotion detection from big data. J Intell Syst. 2022;31(1):113-126. DOI:10.1515/jisys-2022-0001
19. Uddin MZ, Dysthe KK, Følstad A, Brandtzaeg PB. Deep learning for prediction of depressive symptoms in a large textual dataset. Neural Comput Appl. 2022;34(1):721-744. DOI:10.1007/s00521-021-06426-4
20. Khan L, Amjad A, Afaq KM, Chang HT. Deep Sentiment Analysis Using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media. Appl Sci. 2022;12(5). DOI:10.3390/app12052694
21. Kaur G, Sharma A. A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. J Big Data. 2023, 10(1). DOI:10.1186/s40537-022-00680-6
22. Singh C, Imam T, Wibowo S. A Deep Learning Approach for Sentiment Analysis of Reviews. Appl Sci. 2022.
23. Han S, *et al*. Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing. J Biomed Inform. 2022;127(September 2021):103984. DOI:10.1016/j.jbi.2021.103984
24. Chen H, Wu L, Chen J, Lu W, Ding J. A comparative study of automated legal text classification using random forests and deep learning. Inf Process Manag. 2022;59(2). DOI:10.1016/j.ipm.2021.102798
25. Talaat AS. Sentiment analysis classification system using hybrid BERT models. J Big Data. 2023, 10(1). DOI:10.1186/s40537-023-00781-w