International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group| 9th ICAILLMBA (ALLAM) 2025-Conference Proceedings
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.ICAILLMBA-122

Original Article

Multiclass Text Classification Using Deep Learning

Seepana Sridhararao[1], Varaprasad K S B[2], Buridi Ganesh[3]
[1] Research Scholar, Department of CS & SE, Andhra University, Visakhapatnam, India.
[2]Department of Electrical and Electronics Engineering, Visakha Institute of Engineering and Technology, Narava, Visakhapatnam, India.
[3]Assistant Professor, Electrical and Electronics Engineering, Visakha Institute of Engineering and Technology, Visakhapatnam, India.

Abstract *- The exponential growth of digital news content has created a critical need for automated techniques that can efficiently organize and categorize textual information. Multiclass text classification plays a vital role in structuring large-scale news articles by assigning them to predefined thematic categories. This study proposes a deep learning–driven framework for the multiclass categorization of news articles using the BBC Text Dataset. The approach incorporates essential natural language processing steps, including text tokenization, removal of stop words, stemming, and lemmatization, to clean and standardize the input data. Textual features are represented using both TF-IDF and word embedding methods to effectively capture semantic information. An LSTM-based neural network is utilized to model the sequential nature and contextual relationships within the text. The model's performance is evaluated using widely accepted metrics such as accuracy, precision, recall, and F1-score. The experimental findings indicate that the proposed method attains an average classification accuracy of about 91%, demonstrating superior performance compared to conventional machine learning techniques. The study highlights the effectiveness of LSTM-based architectures for automated news categorization and demonstrates their potential for real-world content management and information retrieval systems.*

Keywords *- Deep Learning, Long Short-Term Memory (Lstm), Natural Language Processing, News Article Classification, Bbc Text Dataset, Tf-Idf, Word Embeddings, Sequence Modeling, Information.*

1. Introduction

The rapid expansion of digital communication and online news platforms has resulted in an enormous volume of textual data being generated and disseminated every day. News articles covering a wide range of domains including sports, business, politics, technology, and entertainment are published continuously, making manual organization and categorization increasingly impractical. Efficient management of this unstructured textual information is essential for timely information retrieval, content recommendation, and decision-making processes. Text classification is a fundamental task in natural language processing (NLP), addresses this challenge by automatically the process of allocating predefined labels to textual documents. In particular, multiclass text classification enables the categorization of news articles into multiple topic-based classes, allowing users to quickly access relevant information and improving the overall usability of digital news systems.

Conventional text classification approaches largely depend on handcrafted feature extraction techniques methods like Bag-of-Words and TF-IDF combined with traditional machine learning algorithms. Although these approaches are commonly used because they are simple and easy to interpret, they suffer from several inherent limitations. Such approaches often fail to capture semantic meaning, contextual relationships, and word order dependencies present in natural language text. Furthermore, as the scale and diversity of news content continue to grow, traditional models encounter challenges related to scalability, adaptability, and generalization across evolving topics. These shortcomings result in reduced classification accuracy, particularly when dealing with complex and sequential textual data such as news articles.

Recent advances in deep learning have significantly transformed text classification by enabling models to These models are capable of automatically learning informative representations from raw textual data, eliminating the need for extensive manual feature extraction. Recurrent neural networks, particularly Long Short-Term Memory (LSTM) architectures, are well suited for sequential data processing, as they preserve long-range contextual dependencies while addressing challenges such as the vanishing gradient problem. By integrating word embeddings with LSTM architectures, deep learning models can better understand both syntactic structure and semantic context within news articles. Motivated by these developments, this work focuses on designing a deep learning–based multiclass text classification framework that accurately categorizes news articles into predefined classes. The proposed approach aims to improve classification performance, enhance scalability, and provide a robust automated solution for organizing large-scale news content in real-world applications.

2. Literature Survey

McCallum et al. presented a comparative analysis of Naive Bayes event models emphasizing the multivariate Bernoulli and multinomial models. Their tests on real-world text datasets revealed that the multinomial Naive Bayes model consistently outperformed the Bernoulli model, especially with larger vocabularies, due to its ability to utilize word frequency information.

However, the approach was limited by the assumption of word independence, which restricts contextual and semantic understanding. The authors concluded that despite these limitations, multinomial Naive Bayes serves as an efficient and effective baseline for text classification and influenced subsequent research in this domain [1].

Cortes and Vapnik introduced Support Vector Machines (SVMs) for text classification by maximizing the margin between classes using linear and kernel-based hyperplanes. Their results demonstrated strong generalization performance and robustness to high-dimensional text data. However, Support Vector Machines require appropriate kernel choice and meticulous parameter optimization, which can lead to high computational costs for large datasets. The study concluded that SVMs provide a powerful and theoretically sound approach for text classification tasks [2]. Collobert et al. proposed an integrated deep learning framework designed to address multiple NLP tasks, such as text classification, using word embeddings and convolutional layers. Their model achieved superior performance by automatically learning hierarchical features without manual feature engineering. However, the approach required large training datasets and high computational resources. The authors concluded that deep learning models significantly advance NLP performance and enable end-to-end learning [3].

Graves et al. introduced Bi-LSTM networks to capture contextual information from both past and future sequences. Their experiments showed improved accuracy over unidirectional models in sequential classification tasks. However, Bi-LSTMs increase computational complexity and training time. The study concluded that bidirectional architectures are highly effective for context-sensitive sequence modeling [4]. Yang and Pedersen investigated feature selection techniques for text classification. Their results showed that appropriate feature selection significantly improves classification accuracy and reduces dimensionality. However, selecting optimal features can be computationally intensive for large vocabularies. The authors concluded that feature selection is crucial for efficient and accurate text classification [5].

Zhang and Wallace analyzed the sensitivity of Convolutional Neural Networks (CNNs) to hyperparameters in sentence classification tasks. Their experiments demonstrated that CNNs effectively capture local and hierarchical textual features with proper configuration. However, CNN performance is sensitive to filter size and pooling strategies. The study concluded that CNNs are competitive models for sentence-level text classification [6]. Zhang, Zhao, and Liu proposed a hybrid CNN–RNN model to capture both local features and long-term dependencies in text data. Their results showed improved classification accuracy relative to single-architecture CNN or RNN approaches. However, the hybrid architecture increased model complexity and training cost. The authors concluded that combining neural architectures enhances feature representation for text classification [7].

Kim proposed a CNN-based architecture for sentence classification using multiple filter sizes and max pooling. The model achieved state-of-the-art results across multiple benchmark datasets. However, CNNs have capacity to capture long-range contextual dependencies. The study concluded that CNNs are highly effective for extracting local textual patterns in classification tasks [8]. Devlin et al. introduced BERT, a transformer-based pre-trained language model that captures deep bidirectional context. The model demonstrated leading results across multiple natural language processing tasks, including text classification. However, BERT requires significant computational resources for training and inference. The authors concluded that pre-trained transformer models represent a major breakthrough in natural language understanding [9]. Johnson and Zhang proposed a CNN-based model that preserves word order information for text classification. Their experiments showed improved performance over models that ignore word sequence. However, the approach still struggles with very long documents. The study concluded that incorporating word order is essential for accurate text classification [10].

Yu and Hu introduced a multi-channel attention-based CNN model to enhance feature extraction in text classification. Their results demonstrated superior performance compared to traditional CNNs. However, the attention mechanism increases model complexity. The authors concluded that attention improves model focus on informative textual features [11]. Lin, Yang, and Socher proposed a structured self-attentive sentence embedding model to capture both local and global dependencies. Their approach achieved strong performance on benchmark datasets. However, self-attention models require careful tuning and higher computational cost. The study concluded that self-attention significantly improves sentence representation quality [12]. Sun, Xu, and Yang introduced ERNIE, a knowledge-enhanced pre-trained language model that integrates external structured knowledge. The model achieved leading results in text classification benchmarks. However, integrating knowledge graphs increases model complexity. The authors concluded that knowledge integration enhances semantic understanding in language models [13].

Wang and Jiang conducted a comparison of deep learning models, including CNNs and RNNs, for text classification. Their results highlighted that no single model is optimal for all tasks and datasets. However, model selection requires extensive experimentation. The study concluded that architecture choice plays a critical role in classification performance [14]. Howard and Ruder proposed ULMFiT, a transfer learning approach for text classification using fine-tuned language models. Their experiments showed significant accuracy improvements with limited labeled data. However, fine-tuning requires careful training strategies. The authors concluded that transfer learning is highly effective for improving text classification performance [15].

3. Methodology

3.1. About Dataset

This study utilizes the BBC Text Dataset, a well-established benchmark commonly employed in text classification research. The dataset comprises news articles sourced from various sections of BBC News, covering a diverse range of topics. The articles are categorized into five predefined classes: Sports, Business, Politics, Technology, and Entertainment. The dataset contains approximately 2,225 text samples, with each sample representing a single news article assigned to one of the above categories. The variety of topics and writing styles provides a rich and realistic text corpus for training and evaluating machine learning and deep learning approaches. To ensure fair and unbiased classification performance, the dataset was handled carefully so that all news categories were adequately represented, minimizing the impact of class imbalance. This balanced representation makes the BBC Text Dataset suitable for evaluating the effectiveness of different text classification techniques..

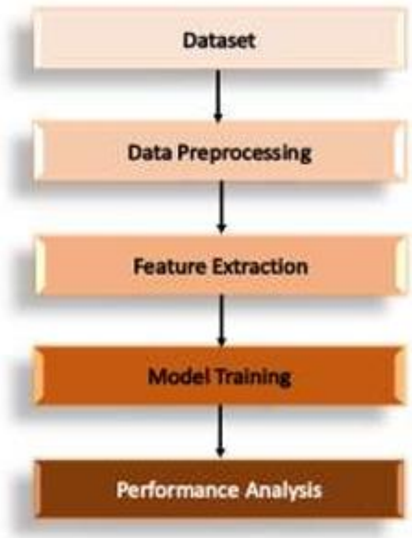

Fig1: Block Diagram of the Text Classification Process

Figure 1 illustrates the end-to-end workflow of the proposed multiclass news article classification system. It begins with the dataset collection, followed by data preprocessing is performed to clean and structure the text data. Next, relevant features are obtained from the processed data and used for model training. Finally, the trained model is evaluated through performance analysis to assess its effectiveness in accurately classifying news articles into different categories..

3.2. Proposed LSTM Model for Multiclass Text Classification

This study employs a Long Short-Term Memory (LSTM)–based deep learning approach for multiclass classification of news articles from the BBC Text Dataset. LSTM networks are particularly effective for text-based tasks because they can capture both sequential order and contextual relationships between words. The internal architecture of the LSTM cell used in this work, including its gate mechanisms and information flow, is illustrated in Figure 2.

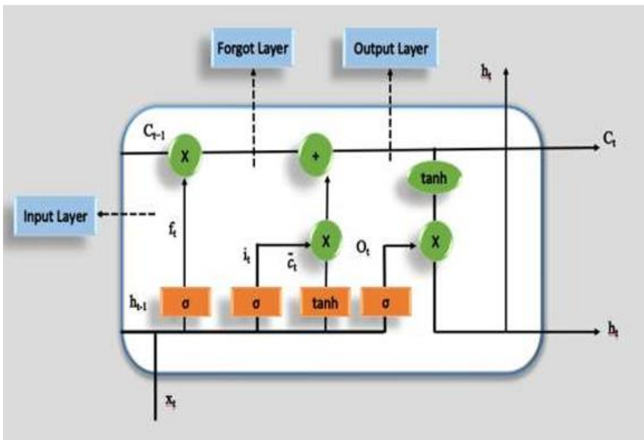

Fig 2: Structure of Lstm

Initially, each news article is preprocessed and transformed into a sequence of numerical vectors, which are fed into the LSTM network word by word. At the start of sequence processing, the hidden state and cell state are initialized, allowing the model to begin learning dependencies across the text. As the sequence progresses, the LSTM processes one word at a time while maintaining and updating these states to preserve meaningful information from earlier words. The forget gate is responsible for determining which information from the prior cell state should be preserved and which should be eliminated. This decision is made by considering the current input token along with the preceding hidden state, allowing the network to discard unnecessary or outdated information while retaining relevant context. This mechanism helps the model focus on important contextual features within the news article.

Next, the input gate regulates the incorporation of new information into the cell state. A candidate memory vector is first produced using a tanh activation function, and the input gate then determines the extent to which this newly generated information is written to the cell state. Together, the forget and input gates update the cell state, enabling the LSTM to maintain long-term dependencies while incorporating relevant new content from the text. The output gate controls which information from the updated cell state is propagated as the hidden state. This hidden state serves as the LSTM’s output at the current time step and is passed to subsequent layers and carries contextual information to the next step in the sequence. Through this process, the LSTM effectively learns relationships between words across the entire news article.

After processing all words in a news article, the final hidden state summarizes the overall contextual and sequential information. This representation is forwarded to a fully connected output layer, which predicts the appropriate news category such as Sports, Business, Politics, Technology, or Entertainment. The model is trained using Backpropagation Through Time (BPTT), which updates the network weights while the gate mechanisms help prevent vanishing and exploding gradient problems, ensuring stable and efficient learning.

4. Results

The proposed LSTM-based model was trained to perform multiclass classification of news articles using an embedding layer, LSTM layers, and dropout for regularization. The embedding layer transformed input words into dense vector embeddings, allowing the model to capture semantic relationships, while the LSTM layers effectively learn long-range dependencies within the textual sequences. Dropout was applied during training t o minimize overfitting and enhance generalization, the model was trained for roughly 25 epochs using a batch size of 50. The Adam optimizer was adopted to ensure efficient gradient-based learning, while categorical cross-entropy was selected as the loss function due to its suitability for multiclass classification problems. These training settings ensured stable convergence and effective learning across all news categories.

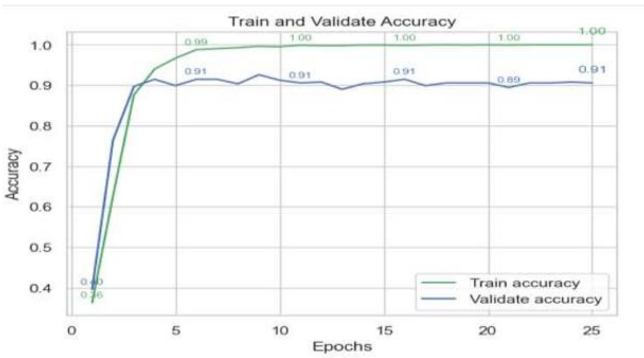

Fig 3: Train and validate accuracy

Figure 3 illustrates the learning behavior of the model over training epochs. The training accuracy steadily increases and reaches nearly 91%, showing that the model successfully learns patterns from the training data. At the same time, the validation accuracy also improves but later stabilizes and shows slight fluctuations, suggesting the onset of overfitting, in which the model achieves high performance on the training set but exhibits limited generalization to unseen validation data.

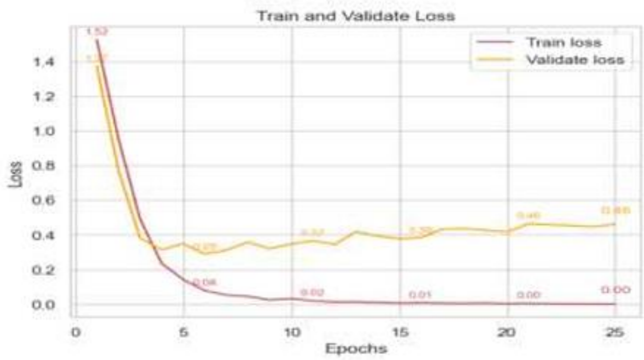

Fig 4: Train and Validate Loss

Figure 4 presents the variation of training and validation loss over successive epochs. The training loss, shown by the red curve, declines sharply during the early epochs, indicating that the model efficiently learns patterns from the training data. The validation loss, shown by the yellow line, also decreases at first, reflecting improved performance on unseen data; however, it later stabilizes and shows slight fluctuations, suggesting that the model has reached a learning plateau and may be approaching overfitting.

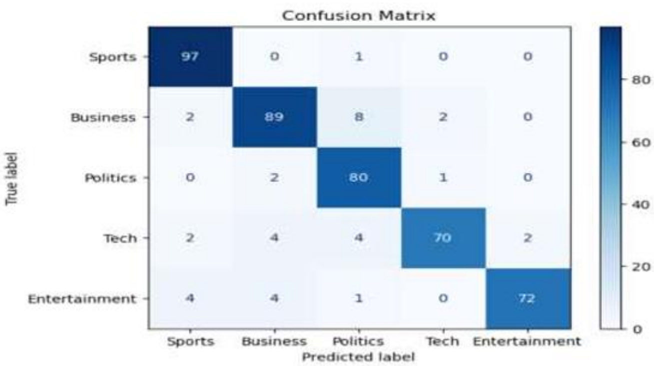

Fig 5: Confusion Matrix

Figure 5 presents the confusion matrix for the proposed LSTM-based multiclass news classification model, showing the relationship between true labels and predicted labels across the five categories: Sports, Business, Politics, Technology, and Entertainment. The majority of values appear along the diagonal, indicating a high number of correct predictions for each class, while only a few instances are misclassified into other categories. This demonstrates that the model effectively distinguishes between different news topics, achieving strong classification performance with minimal confusion among classes.

5. Conclusion

In this study, multiclass text classification was explored as an effective approach for organizing and managing large volumes of news articles across predefined categories. Various techniques, including Count Vectorizer, TF-IDF–based approaches and deep learning models were employed for the news classification task, among which Long Short-Term Memory (LSTM) networks achieved the best performance owing to their effectiveness in modeling contextual information and sequential dependencies in text. The experimental results confirm that LSTM-based models provide accurate and reliable classification, making them well suited for practical applications such as news curation, content recommendation, and information retrieval. Moreover, the proposed framework offers a scalable solution for handling continuously expanding news datasets. With future enhancements such as attention mechanisms, multilingual capabilities, and periodic model retraining, the system can be further strengthened to adapt to evolving news patterns and real-world deployment requirements.

References

[1] A. McCallum and K. Nigam, “A comparison of event models for naive Bayes text classification,” Proc. AAAI Workshop on Learning for Text Categorization, Madison, WI, USA, 1998, pp. 41–48.

[2] Cortes, C.,& Vapnak, V.(1995). Support Vector Networks. Machine Learning, 20(3), 273-297

[3] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, “Natural language processing (almost) from scratch,” *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, 2011.

[4] A. Graves, S. Fernández, J. Schmidhuber, and P. Frasconi, "Bidirectional LSTM networks for improved phoneme classification and recognition," in *Proc. Int. Conf. Artificial Neural Networks (ICANN)*, 2013, pp. 277–284.

[5] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text classification," in *Proc. 14th Int. Conf. Machine Learning (ICML)*, 1997, pp. 412–420.

[6] X. Zhang and B. C. Wallace, "A sensitivity analysis of (and practitioner's guide to) convolutional neural networks for sentence classification," in *Proc. 2015 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2015, pp. 253–263.

[7] Y. Zhang, X. Zhao, and H. Liu, "Hybrid convolutional and recurrent neural networks for text classification," in *Proc. 33rd Conf. Neural Information Processing Systems (NeurIPS)*, 2019, pp. 4541–4551.

[8] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1746–1751.

[9] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. 2018 Conf. North American Chapter of the Assoc. for Computational Linguistics (NAACL)*, 2018, pp. 4171–4186.

[10] R. Johnson and T. Zhang, "Effective use of word order for text classification with convolutional neural networks," in *Proc. 2015 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2015, pp. 103–112.

[11] M. Yu and X. Hu, "Multi-channel attention-based convolutional neural networks for text classification," in *Proc. 2020 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 4062–4070.

[12] Z. Lin, A. Yang, and R. Socher, "A structured self-attentive sentence embedding," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2017, pp. 1–10.

[13] C. Sun, H. Xu, and Z. Yang, "ERNIE: Enhanced representation through knowledge integration," in *Proc. 57th Annu. Meeting Assoc. for Computational Linguistics (ACL)*, 2019, pp. 1444–1454.

[14] S. Wang and J. Jiang, "A compare-and-contrast study of deep learning models for text classification," in *Proc. 2016 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2016, pp. 1115–1123.

[15] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in *Proc. 56th Annu. Meeting Assoc. for Computational Linguistics (ACL)*, 2018, pp. 328–339.