International Journal of Computer Sciences and Engineering
Vol.13, Issue.7, pp.58-63, July 2025
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

IJCSE
ISSN: 2347-2693 (E)

Review Article

A Review on News Article Classification Using Different Machine Learning Algorithms

Dimple Choudhury[1]

[1]Faculty of Computer Technology, Assam down Town University, Guwahati, India

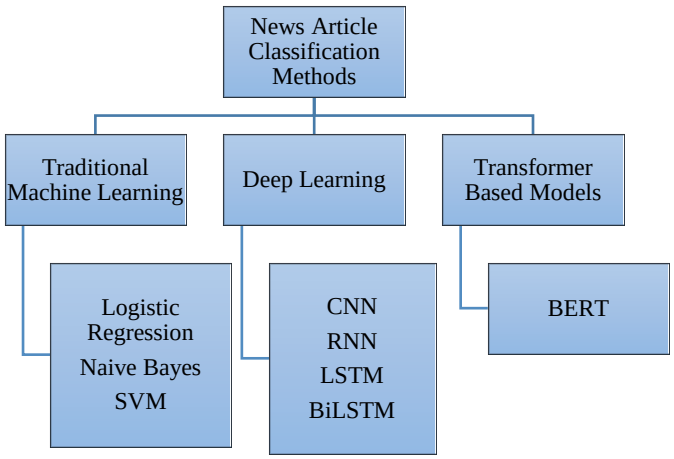
Corresponding Author:

Received: 22/May/2025; **Accepted**: 24/Jun/2025; **Published**: 31/Jul/2025. **DOI**: https://doi.org/10.26438/ijcse/v13i7.5863

Abstract: In today's digital landscape, the majority of textual data is unstructured, requiring automated classification techniques for better organization and usability. News article classification into distinct categories such as politics, technology, crime, education, and entertainment is a crucial task in Natural Language Processing (NLP), with applications in recommendation systems, fake news detection, and content personalization. This review paper presents a comprehensive overview of various machine learning approaches employed for multi-label news classification across multiple languages and domains. It synthesizes methodologies, comparative results, and key insights from numerous studies while identifying prevailing challenges, emerging trends, and future research directions. The work primarily focuses on contrasting traditional machine learning algorithms with advanced deep learning and transformer-based models to provide a broad perspective on the evolution of techniques in this domain.

Keywords: News Article, Classification, Machine Learning, Deep Learning, NLP, Text Classification

Graphical Abstract- This illustration provides an overview of the evolution of machine learning approaches used in multi-label news classification, beginning with traditional models, advancing through deep learning neural networks, and culminating in transformer-based architectures that now play a pivotal role in modern classification tasks.



1. Introduction

News articles are essential to public knowledge and play a key role in informing the masses. However, the explosion of online content demands systems that can automatically categorize news into relevant topics. Traditional methods such as Naïve Bayes and SVM have evolved into advanced deep learning models like CNNs, RNNs, LSTM, and transformer-based models like BERT, capable of handling semantic complexity.

Natural Language Processing (NLP) and Machine Learning (ML) techniques, ranging from traditional algorithms like Naïve Bayes to advanced deep learning models like LSTM and BiLSTM, are used to classify news articles, as manual categorization is impractical. These systems can analyze textual content and assign categories based on patterns in language and context. Applications include personalized news feeds, misinformation detection, content moderation, and event detection.

This study synthesizes and evaluates a wide array of approaches—from classic ML algorithms to cutting-edge deep learning and transformer-based techniques—used for multi-class and multi-label news classification across different languages and datasets. Unlike previous reviews that often focus on a specific algorithm or language, this paper provides a holistic comparison across multiple linguistic domains, architectures, and methodological choices. It identifies key challenges and outlining future directions to guide research in this evolving field.

1.1 Objective of the Study

The primary objective of this study is to systematically review and analyze existing machine learning and deep learning approaches used for multi-label news classification. This includes examining traditional algorithms, neural network-based methods, and transformer-based architectures in terms of their performance, scalability, and applicability to diverse news categories such as politics, crime, entertainment, education, and technology. The study aims to identify key trends, challenges, and gaps in the current research while proposing future directions for building more accurate and context-aware news classification systems.

1.1 Organization

The rest of the paper is structured as follows: Section 2 provides literature review. Section 3 outlines key evaluation measures. Section 4 describes common architectures. Section 4 discusses results and comparative analysis and Section 5 concludes the paper in reference to future directions.

2. Literature Review

Several research studies have explored the classification of news articles using different machine learning and deep learning techniques. In [1], a comparative analysis was conducted using three machine learning models such as support vector machine (SVM) linear support vector classification and random forest alongside two deep learning models namely LSTM and GRU to classify Bengali News. The study found that among the ML models, SVM (with Singular Value Decomposition) performed best, while GRU achieved the highest accuracy among the DL models. To improve final classification accuracy, the authors proposed a majority voting technique, where the predicted class from each model was counted and the most frequently predicted class was selected as the final output. Using this approach, the proposed method achieved a highest overall accuracy of 95.45%, surpassing all individual models. Similarly, the study in [2] used a variety of ML models such as SVC, Naïve Bayes, Standard gradient descent, Decision trees, Random Forest, K-nearest neighbors CNN, and LSTM to classify Arabic news articles by country, with deep learning models outperforming classical approaches.

For the Urdu language, [3] evaluated three ML classifiers such as Decision tree, SVN and K- nearest neighbor on a large corpus, achieving highest accuracy with an SVM. In a study focusing on the Swahili language, [4] employed classical ML models, DL models and Transformer based models. Among traditional machine learning algorithms, Support Vector Machine (SVM) demonstrates superior performance. In the domain of deep learning, Convolutional Neural Networks (CNN) and transformer-based models, particularly RoBerta Weschel sw have shown the most promising results. Their transformer-based model sets a new state of art with 93% test accuracy. For Bangla news classification, [5] found that a Multi-layer Dense Neural Network outperformed traditional ML models, reaching 95.50% accuracy.

The study in [6] employed both traditional models like Logistic Regression and advanced models like BERT for English news classification, with BERT achieving up to 98% accuracy. Similarly, [7] demonstrated the effectiveness of BiLSTM on the Kaggle dataset, showing superior results over CNN and LSTM.

The work in [8], categorized news articles based solely on titles using models like SVM, Naïve Bayes, Random Forest, and RNN, showing that high accuracy can still be achieved with limited input. In [9], a hybrid model combining CNN, GRU and attention mechanism was applied to three different Chinese datasets, including the THUCNews dataset, and outperformed previous models in terms of classification accuracy.

Machine learning models such as Naïve Bayes, LR, SVM, KNN discussed in [10] and Naïve Bayes works better in terms of accuracy with 93%. Another study [11] experimented with two different datasets, BBC datasets and five groups of 20Newsgroup datasets and obtained accuracy 97.84% and 94.93% respectively.

The study in [12] emphasized multi-label classification using 5 ML models, NB, RF, DT, SVM, and KNN where SVM achieved highest accuracy of 89.35%. Similar work is carried out in [13], where Naïve Bayes with TFIDF produces satisfactory result.

The classification of crime news titles using SVM, Linear SVM, LR, XGBoost, Multinomial Niave Bayes, KNN, Decision Trees and Random Forest models was explored in [14] , achieving up to 99% accuracy with Decision Tree and Random Forest classifiers.

In [15], researchers focused on classifying Mizo news into local, national, and international categories, where SVM outperformed other classifiers such as KNN, decision tree, Naïve Bayes. A similar study in [16] used event categorization from news articles, where Random Forest achieved the highest accuracy among compared models.

A comprehensive comparison of classification models for Myanmar news is presented in [17], demonstrating that LSTM models outperformed SVM for local and international news. Another work [18] introduced transformer- based BERT model for news article classification and compare its performance with CNN, finding that BERT achieved the highest accuracy.

The work in [19] experimented with 3 ML models, including Linear Support Vector Machine (SVM) and (SGD) training & multinomial Naïve Bayes (NB) using two feature extraction methods - BoW and TF-IDF. The result showed that multinomial Naïve Bayes with BoW achieved highest accuracy of 92.83%. In [20], word2vec and its two variants, CBOW and skipgram were applied in conjunction with CNN on both news and tweet datasets. The results indicated that CBOW performed better on news data, while Skip-gram yielded superior results on tweet data.

Table 1. Comparative Analysis of on News Article Classification

Ref. No	Language/Region	Dataset Used	Algorithms Used	Feature Engineering	Accuracy/F1/Remarks
[1]	Bengali	Custom Dataset	SVM, Linear SVM, RF, LSTM, GRU	BoW	Good performance with SVM, GRU
[2]	Arabic	Arabic news articles (Hajj, Brexit)	SVC, NB, SGD, Dicision Tree, RF, KNN, CNN, LSTM,BiLSTM, GRU, CNN+LSTM, CNN+BiLSTM,CNN+GRU	N-gram, TF-IDF,	DL methods outperform ML
[3]	Urdu	Urdu News Dataset	SVM, KNN, Dicision Tree	TF-IDF	SVM reached highest accuracy
[4]	Swahili	Swahili Dataset	SVM, LR, RF, XGBoost BiLSTM,CNN, BiLSTM+CNN AfriBERTa, XLM-RoBERTa, RoBERTa Wechsel sw	TF-IDF, Attention	RoBERTa Wechsel sw outperforms all
[5]	Bangla	Prothom Alo, Kaggle	NB, KNN, SVM, MLP	TF-IDF	MLP reached 95.5% accuracy
[6]	English	Kaggle News	Linear SVC, BERT	BoW, TF-IDF	BERT with TF-UIDF achieved 98% accuracy
[7]	English	Kaggle News Dataset	CNN, LSTM, BiLSTM	Word-level and contextual	BiLSTM outperformed all
[8]	English	News Aggregator Data Set	SVM, RF, NB, RNN	BoW, Word2Vec	High accuracy with short input
[9]	Chinese	IMDB film review dataset, Fudan University Chinese dataset and THUCNews dataset	CNN + GRU	Word2Vec, Attention	Their model outperforms other state of art.
[10]	English	Custom Corpus	Naive Bayes, LR, SVM, KNN	Naïve Bays showed highest accuracy with 93%	Outperformed traditional models
[11]	English	BBC datasets 20Newsgroup datasets	SVM	TF-IDF	97.48% accuracy for BBC datasets and 94.93% accuracy for 20Newsgroup datasets
[12]	English	AG news classification dataset	SVM, NB, DT, RF, KNN	Text Labeling, TF-IDF	SVM reached 89.35%
[13]	English	Source - Google	Naïve Bayes	TF-IDF	NB gives satisfactory result
[14]	English	Source – Online	SVM, Linear SVM, Multinomial NB, DT, RF, KNN, XGBoost, LR	TF-IDF, Count Vectorizer	RF and DT achieved highest accuracy up to 99%
[15]	Mizo	Collected from Zonet website	NB, KNN, SVM, Decision Tree	TF-IDF	SVM gives best result
[16]	English	-	LR, Gaussian NB, RF, Multinomial NB, KNN, SVM, DT	TF-IDF, Count Vectorizer	RF outperformed all other classifiers
[17]	Myanmar	Myanmar News	SVM, LSTM	TF-IDF	High accuracy in LSTM
[18]	English	Kaggle dataset	BERT	BERT	BERT gives 91% accuracy.
[19]	English	-	Linear Support Vector Machine (SVM) and (SGD) training & multinomial Naïve Bayes (NB)	BoW, TF-IDF	Multinomial Naïve Bayes with BoW achieved highest accuracy of 92.83%
[20]	English	News dataset and Tweet dataset	CNN	Word2vec	CBOW performed better on news data, while Skip-gram yielded superior results on tweet data.

Table 1 summarizes language scope, datasets used, algorithms applied, feature engineering techniques, and key outcomes (accuracy/F1-score or observations) from the above studies related to news classification using machine learning and deep learning models.

3. Evaluation Measures

In the field of news classification, evaluating the effectiveness of machine learning and deep learning models is essential for benchmarking and comparison. Different studies have used various evaluation metrics based on whether it is a binary, multi-class, or multi-label classification challenge. The most commonly used evaluation measures include:

3.1. Accuracy

This measure, which is the simplest to understand, shows the proportion of accurately predicted observations to all observations. In multi-class news classification, accuracy is one of the most commonly used evaluation metrics. However, it's important to interpret it carefully, especially when the dataset is imbalanced.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions}$$

3.2. Precision

Precision is the ratio of true positive predictions to the total number of predicted positive instances. It indicates the model's ability to avoid labeling negative instances as

positive. In multi-class news classification, Precision answers the question: "Of all the news articles the model labeled as a particular category (e.g., 'Crime'), how many actually belong to that category?"

$$Precision = \frac{TP}{TP + FP}$$

Where:
TP (True Positives): Number of news articles correctly classified as a specific category.
FP (False Positives): Number of articles wrongly predicted as that category (they actually belong to another class).

3.3. Recall

Recall is the ratio of true positive predictions to the total number of actual positive instances. It reflects the model's ability to identify all relevant instances. In scenarios such as disaster or political news classification, where missing critical items could have major implications, high recall is preferred. So while Precision focuses on how correct the model's predictions are, Recall focuses on how complete they are.

$$Recall = \frac{TP}{TP + FN}$$

Where:
TP (True Positives): Articles correctly predicted to be in a category.

FN (False Negatives): Articles that actually belong to that category but were missed by the model (classified as something else).

3.4. F1-Score

The F1-Score is the harmonic mean of precision and recall. It balances the trade-off between these two metrics, making it suitable for datasets with imbalanced class distributions. This metric is frequently used in multi-label classification tasks where both false positives and false negatives are critical. Basically, it gives a single score that balances both the correctness and completeness of the predictions.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

3.5. AUC-ROC (Area Under the Receiver Operating Characteristics Curve)

This metric measures a model's ability to distinguish between classes and is widely used for binary and multi-class classification problems. AUC scores closer to 1 indicate high classification ability. For instance, studies on Swahili and multilingual datasets have shown CNN-BiLSTM models with attention mechanisms achieving AUC scores above 0.95 [4].

3.6. Hamming Loss (for multi-label classification)

Especially relevant in multi-label classification tasks, Hamming Loss measures the fraction of wrong labels to the total number of labels. A lower Hamming Loss indicates better performance in scenarios where each article can belong to multiple classes like politics, technology, and sports simultaneously [19], [25].

3.7. Macro and Micro Averages

Macro-average computes the metric independently for each class and then takes the average. It treats all classes equally and is useful when classes are of equal importance. Micro-average aggregates the contributions of all classes to compute the average metric. It is better suited when there is class imbalance, as it weights each instance equally regardless of class [6], [11].

Different models and datasets require different evaluation priorities. For example, while some applications prioritize precision (e.g., fake news detection), others may prioritize recall (e.g., crime reporting). Most reviewed papers use a combination of metrics to ensure a comprehensive evaluation.

4. Common Architectures

In the domain of news classification, the choice of model architecture significantly influences the performance of the classification system. Common architectures include:

4.1. Traditional Machine Learning

ModelsNaïve Bayes, Logistic Regression, Decision Trees, K-Nearest Neighbors, and Support Vector Machines are widely used due to their simplicity and interpretability. These models often rely on bag-of-words, TF-IDF, or other statistical features [12].

4.2. Convolutional Neural Networks (CNNs)

CNNs are effective in capturing local features and patterns in text, such as n-gram semantics. They are often used with embeddings like Word2Vec or GloVe for news classification tasks [13]. Their strength lies in their ability to detect key phrases and hierarchical patterns within limited text contexts.

4.3. Recurrent Neural Networks (RNNs) and LSTMs

RNNs, especially LSTMs and BiLSTMs, are suitable for sequence modeling. They can capture long-range dependencies and are effective for understanding the contextual flow of sentences [1]. BiLSTMs improve upon standard LSTMs by analyzing data in both forward and backward directions, offering a more comprehensive understanding of the text.

4.4. Hybrid Models

Combining CNNs and LSTMs allows models to capture both spatial and temporal features of text. Attention mechanisms are often added to improve focus on significant words [9]. These models are effective in handling complex text patterns and improving performance in multilingual and multi-label settings.

4.5. Transformer-based Models

Pre-trained models like BERT utilizes self-attention mechanisms to understand context at a deeper level. These models achieve state-of-the-art performance in news classification by encoding rich semantic information [18]. Sentence-level embeddings enhance performance in short-text scenarios such as headlines or summaries. Their transfer

learning capabilities allow quick adaptation to low-resource languages or new domains.

5. Research Methodology

This study was conducted through a systematic review of existing literature on news classification. Research papers were collected from reputable academic databases such as Google Scholar, Semantic Scholar, IEEE Xplore, and others. The initial pool of papers included a wide range of publications, which were then screened to remove duplicates and irrelevant studies. After a thorough evaluation based on relevance, methodology, and contribution to the field, a total of 45 high-quality papers were selected for in-depth analysis and review.

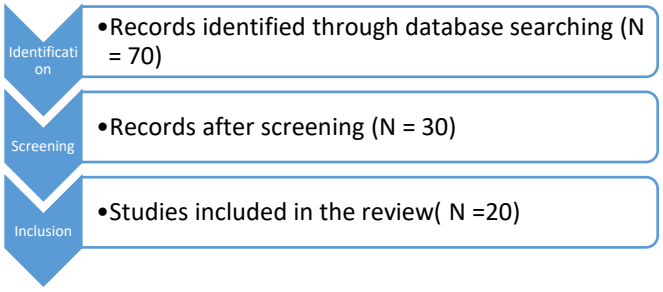


Figure 1 Review procedure of the research

6. Results and Discussion

Results from various studies highlight the strengths and limitations of different models and techniques. While traditional ML methods provide speed and simplicity, they often lag behind in performance compared to deep learning models. The integration of word embeddings significantly boosts accuracy across models.

The classification performance varies significantly with language and resource availability. For instance, in resource-constrained languages like Mizo [15] and Myanmar [17], Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) respectively achieved high accuracy, demonstrating their robustness even with limited data. In Swahili, the integration of transformer-based models such as RoBERTa with Wechsel embeddings outperformed all other methods [4], underscoring the benefit of leveraging multilingual pre-trained models for low-resource languages. Similarly, GRU yielded competitive results for Bengali news classification [1].

Overall, the comparative analysis demonstrates that the performance of news article classification models is influenced by a combination of factors including the language of the dataset, the choice of algorithms, feature extraction techniques, and the size and nature of the corpus. Traditional machine learning models such as SVM, Random Forest, and Decision Trees remain highly effective when combined with TF-IDF or BoW features, particularly for structured and moderately sized datasets. Deep learning models, especially LSTM and BiLSTM, show significant improvements in capturing contextual dependencies, while transformer-based architectures like BERT and RoBERTa consistently outperform others due to their ability to model rich contextual relationships. Additionally, studies involving underrepresented or low-resource languages are increasingly leveraging multilingual and pre-trained models, showing promising results. These findings suggest a growing trend towards the adoption of deep and contextual models, although traditional approaches still hold value, especially in resource-constrained settings.

7. Conclusion

This review has explored a wide spectrum of approaches in the domain of news article classification, ranging from traditional machine learning algorithms to advanced deep learning models. Across the studies reviewed, it is evident that while traditional methods like Naïve Bayes and SVM remain relevant for their simplicity and efficiency, deep learning architectures, particularly those utilizing contextual embeddings and transformer-based models like BERT, consistently outperform in accuracy, contextual comprehension, and generalizability.

A major takeaway is the importance of data preprocessing and the use of semantic word embeddings, which significantly enhance classification performance. Furthermore, hybrid models that combine CNNs and LSTMs with attention mechanisms offer robust solutions for handling sequential and semantic dependencies in text

Despite notable advancements, challenges remain in scaling models to diverse and multilingual datasets, handling class imbalance, and ensuring interpretability. Future research may benefit from integrating Explainable AI (XAI) techniques, addressing low-resource languages, and expanding the use of transfer learning for domain adaptation. By synthesizing existing methodologies and highlighting gaps, it provides valuable insights that can guide future innovations and experimental designs in the field of automated text classification.

Author's statements

Acknowledgements- The authors would like to sincerely thank the Assam down town University, for offering financial support through seed money grant. We also extend our appreciation to the peer reviewers and editorial team for their valuable comments and insightful suggestions.

Funding Source- This work was supported by the Assam down town University through seed money grant project.

Authors' Contributions-

Dimple Choudhury: researched literature and conceived the study, wrote the manuscript, reviewed and edited the manuscript.

Conflict of Interest- The authors affirm that there are no conflicts of interest associated with the publication of this manuscript.

References

[1] T. A. Mahmud, S. Sultana and A. Mondal, "A new technique to classification of bengali news grounded on ml and dl models," International Journal of Computer Applications, Vol.**185**, No.**18**, pp.**8887**, **2023**.

[2] N. Zamzami, H. Himdi and S. F. Sabbeh, "Arabic News Classification Based on the Country of Origin Using Machine Learning and Deep Learning Techniques," Applied Sciences, Vol.**13**, No.**12**, pp.**70-74, 2023.**

[3] I. Rasheed, V. Gupta, H. Banka and C. Kumar, "Urdu Text Classification: A comparative study using machine learning techniques," in 2018 Thirteenth International Conference on Digital Information Management (ICDIM), Berlin, Germany, **2018**.

[4] P. Manas, K. S. Avinash and S. Arpit, "Swahili News Classification: Performance, Challenges, and Explainability Across ML, DL, and Transformers," in Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025, Vienna, Austria, Association for Computational Linguistics, pp.**203-209, 2025.**

[5] Y. Sharmin, K. Ratnadip, R. ,. ,. H. R. A, U. Ashraf, Q. ,. ,. U. ,. A. and R. Hasnat, "Multi-category Bangla News Classification using Machine Learning Classifiers and Multi-layer Dense Neural Network," International Journal of Advanced Computer Science and Applications, Vol.**12, 2021.**

[6] S. Bihi, K. Amal, E. ,. Bouchra and R. Maryem, "Analyzing BERT's Performance Compared to Traditional Text Classification Models," in 25th International Conference on Enterprise Information Systems, **2023**.

[7] A. Shreyas and P. P, "Deep Learning-based Text News Classification using Bi-directional LSTM Model," in Conference: 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT), VIT Vellore, **2024**.

[8] L. Yongwei, L. Kejun and Z. T. M. Y. Ziyu Liu, "News categorization based on titles with SVM, Naïve Bayesian, random forest, and RNN algorithms," in 2nd International Conference on Artificial Intelligence, Automation, and High-Performance Computing (AIAHPC 2022), Zhuhai, China, **2022**.

[9] M. Yuqun, C. Hailong, W. Qing and Z. Xin, "Text classification model based on CNN and BiGRU fusion attention mechanism," in ITM Web of Conferences, **2022**.

[10] A. Jeelani and A. Muqeem, "ONLINE NEWS CLASSIFICATION USING MACHINE LEARNING TECHNIQUES," IIUM Engineering Journal, Vol.**22**, No.**2**, pp.**210-225, 2021.**

[11] M. H. D. Seyyed, S. ,. Mohammad and M. ,. Morteza, "A novel text mining approach based on TF-IDF and Support Vector Machine for news classification," in 2016 IEEE International Conference on Engineering and Technology (ICETECH), Coimbatore, **2016**.

[12] N. Beebi, P. C. Nagendra, D. Asutosh, C. S and S. N. Niranjan, "Machine Learning Models for News Article Classification," in 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, **2023**.

[13] D. M. Shweta, "News Classification Using Machine Learning," International Journal on Recent and Innovation Trends in Computing and Communication, Vol.**9**, No.**5**, pp.**23-27, 2021.**

[14] R. R. Mapitsi and J. S. Tshephisho, "Classification of Crime-related News Titles using Machine Learning Methods," in 2024 International Conference on Electrical, Communication and Computer Engineering (ICECCE), Kuala Lumpur, **2024**.

[15] L. Mercy, D. Ranjita and L. Robert, "Mizo News Classification Using Machine Learning Techniques," in Evolution in Computational Intelligence , **2023**.

[16] "Kogilavani, Shanmugavadivel; Malliga, Subramanian; Vasantharan, Kumarakurubaran; G., A., Prethish; S., Sankar;," in Speech and Language Technologies for Low-Resource Languages. SPELLL 2023., Springer, Cham, pp.**255-267, 2024.**

[17] Z. Y. Kyaw, M. N. Ye, N. Ye, P. N. Khant and Z. L. Thant, "Implementation of Burmese Language News Classification System by Using SVM and LSTM Machine Learning Algorithm," in 2023 IEEE 6th International Conference on Computer and Communication Engineering Technology (CCET), Beijing, **2023**.

[18] S. Parvathavarthini, M. Shreekanth, K. S. Vignesh and N. S. Santhosh, "News Category Classification using Natural Language Processing Transformer," in 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, **2023**.

[19] A. Feroz and G. Shabina, "Linear Support Vector Machine (SVM) with Stochastic Gradient Descent (SGD) training & multinomial Naïve Bayes (NB) in News Classification," International Journal of Computer Sciences and Engineering, Vol.**7**, No.**4**, pp.**360-363, 2019.**

[20] J. I. K. Beakcheol and W. K. Jong, "Word2vec convolutional neural networks for classification of news articles and tweets," PLoS ONE , Vol.**14**, No.**8**, **2022**.

[21] L. Bingyao, C. Jiajing, W. Rui, H. Junming, L. Yuanshuai and W. Jianjun, Optimizing News Text Classification with Bi-LSTM and Attention Mechanism for Efficient Data Processing, **2024**.

AUTHORS PROFILE

Dimple Choudhury earned her B. Tech., in Computer Science from Central Institute of Technology and M. Tech., in Information Technology from Gauhati University in 2021 and 2023, respectively. She is currently working as Assistant Professor in Department of Computer Technology from Assam down town University since 2023. Her primary research interests lie in Artificial Intelligence, Machine Learning, and Natural Language Processing.

