UTM UNIVERSITI TEKNOLOGI MALAYSIA
INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

Enhancing Fake News Analysis in Bangla: A Hybrid Model with LDA, and LIME for Interpretable Classification

Gazi Tahsina Sharmin Jahin[1] & Nuren Nafisa[2]
Department of Computer Science and Engineering International Islamic University Chittagong Chattogram, 4318, Bangladesh
Email: gazi3jahin@gmail.com[1]; nurennafisa@gmail.com[2]

Firoze Maliha
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology
Chattogram, 4349, Bangladesh
Email: firozemaliha.radhika@gmail.com

Submitted: 9/3/2025. Revised edition: 16/8/2025. Accepted: 25/8/2025. Published online: 30/11/2025
DOI: https://doi.org/10.11113/ijic.v15n2.543

***Abstract*—The proliferation of fake news presents a critical challenge in ensuring information authenticity, especially in the context of Bangla text. This study addresses the issue by developing a robust pipeline for detecting and categorizing fake news using advanced machine learning (ML) and natural language processing (NLP) techniques. A dataset comprising 5,000 fake news articles was prepared, collected from 20 different newspapers in Bangladesh between 2020 and 2024. Topic modeling was performed using Latent Dirichlet Allocation (LDA), whose results in topic categorization (World, Business, Entertainment, Bangladesh, and Sports) were significantly better compared to other models. The proposed model, consisting of Bidirectional Encoder Representations from Transformers (BERT), Support Vector Machine (SVM), and Bidirectional Long Short-Term Memory (BiLSTM), outperforms the traditional models in all evaluation measures, yielding a mean accuracy of 97%. Moreover, the integration of Local Interpretable Model-agnostic Explanations (LIME) adds interpretability to the model by explaining individual predictions, thereby enhancing transparency. This holistic approach sets a new benchmark for Bangla fake news detection with high accuracy, interpretability, and reliability, opening avenues for further research in combating misinformation in Bangla.**

***Keywords*—Fake News Detection, Bangla NLP, Machine Learning, Topic Modeling, Model Interpretability**

INTRODUCTION

In today's digital age, spreading fake news has become a major contemporary societal challenge threatening common trust, economic stability, and social cohesion. Misinformtion can reach millions in no time on rapidly growing online platforms and may influence decision-making or cause widespread panic [1].That is very concerning, especially in Bangladesh, where more than half the population depends on digital news platforms and social media as the main sources of information. This underlines the importance of accurate classification to understand misinformation, which is often rooted in specific classes. For instance, in March 2024, the fake news about Bangladesh receiving a $10 billion economic bailout from the IMF spread like wildfire [2]. The fake news was full of fabricated government press releases and statements that had already caused public discussion and confusion. While the claim was debunked within days [3], its impact on public perception and confidence highlighted the dangers of unchecked misinformation. This incident, among many others, underscores the necessity for robust systems to detect and mitigate fake news. The ability to accurately classify fake news and identify which categories misinformation belongs to is crucial for understanding its spread and devising targeted countermeasures. Addressing this, a dataset has been created from scratch by web scraping data from 20 reputed online newspapers in Bangladesh [4]. After cleaning and preprocessing, the features are extracted using TF-IDF and LDA, yielding five important categories of fake news (World, Business, Entertainment, Bangladesh, and Sports). Then, a hybrid model is used combining BERT (for contextual word embeddings) with SVM (for the classification) and Bi-LSTM (to capture sequential dependencies), performing better than traditional classifiers based on extensive performance metrics. The LIME model is introduced for explainability in the decisions [5]. The contributions of this study are threefold:

Introduction to a comprehensive Bangla fake news analysis framework utilizing advanced feature extraction and classification techniques, demonstrating the superior performance of a hybrid BERT+BiLSTM+SVM model compared to traditional approaches, and enhancement of interpretability through explainable AI techniques like LIME.

Since the proposed work overcomes two major challenges prevalent in related works data imbalance and explainability-it would set a benchmark for detection and classification in Bangla fake news, thereby adding considerably to the fight against misinformation in the digital space.

LITERATURE REVIEW

Different studies have addressed Bangla fake news detection using various ML and DL techniques. Among them, the study of [4] introduces the BanFakeNews dataset and shows that SVM with linguistic features achieves an F1-score of 91%, while CNN, LSTM, and BERT perform poorly. [6] reports 96% accuracy using Bi-LSTM on 57,000 news articles, though the work would benefit from better preprocessing. [1] compares SVM and MNB, where SVM reaches 96.64% accuracy, but the dataset size is small. [7] tests several models and finds DNN performs best at 95.9%, but does not explore hybrid models or transformer fine tuning. [8] combines traditional ML and DL models for COVID-19 fake news detection, achieving 94.25% accuracy, but notes challenges such as generalization and computational cost. Researchers have also investigated the potential of hybrid models to enhance performance. [9] proposes a hybrid CNN-BiLSTM-FLN model, achieving 99.4% accuracy on English datasets but without Bangla evaluation. [10] uses Random Forest for document classification, obtaining 99.91% accuracy. [11] The DeepCnnLstm and DeepCnnBilstm models perform strongly on ISOT and moderately on FA-KES by capturing spatial and contextual features, but struggled to generalize across datasets and were limited to binary classification. [12] applies GRU for fake news detection with 94% accuracy, showing its potential for real-time applications. [13] proposes RoBERTa-GCN with 98.6% accuracy, noting that handling real-time data remains a challenge. Using XAI, [14] explores interpretability for news category classification with TF-IDF, word2vec, SBERT, BERT, LR, KNN, RF, and DT; RF achieves 91.48% accuracy but risks overfitting. [15] applies a hybrid CNN+LSTM with LIWC-based features outperformed other methods, but its success depends heavily on training data quality and may face issues with changing language patterns or new fake news styles. [16] also applies a hybrid CNN-LSTM with FastText embeddings, outperformed ML, DL, and transformer models, reaching up to 0.99 accuracy, but was limited to English datasets and had occasional errors. [17] uses BERT-mCNN-sBiGRU on FakeNewsNet, achieving over 93% accuracy by modeling semantic interactions between content, headlines, and user behavior, but mainly handles uniform inputs and requires high computational resources. [18]have found that SVM with TF-IDF (99.03%) and CNN with TF-IDF (98.99%, F1 = 99.02%) outperform other approaches, though results depended on embedding choice and did not include advanced models like BERT. [19] uses HHO-based feature selection with CNN-BiLSTM, achieving up to 98.89% accuracy on several datasets, though performance varied slightly between datasets and required manual hyperparameter tuning. The discussion indicates that deep learning and transformer-based models are powerful but still face issues such as class imbalance, lack of interpretability, and challenges in real-world applications.

TABLE I: Summary of Literature Review

Ref	Year	Dataset	Method	Results	Contribution	Limitation
[4]	2020	50,000 Bangla news items	Traditional linguistic features and neural network-based methods	Linear classifiers with traditional linguistic features perform better than neural network-based models	Introduced BanFakeNews, a 50K Bangla news dataset, showing linear classifiers with linguistic features outperform neural models	Neural models lacked certain feature integrations, potentially understating their true potential
[6]	2021	57,000 Bangladeshi news articles	ML: DTC, RF, SVM, NB, GB, LR, KNN; DL: CNN, LSTM, BiLSTM	Bi-LSTM achieved up to 96% accuracy	Introduces a robust ML and DL pipeline on a large Bangla dataset	Requires further optimization and data handling
[1]	2022	2,500 articles	SVM and MNB classifiers	MNB: 93.32%, SVM: 96.64% accuracy	Introduces a 2,500-article Bangla fake news dataset, shows SVM outperforming MNB	Small dataset size, lacks advanced features or stemming, hybrid classifiers not explored
[7]	2022	Combined Bangla dataset: 4,678 fake + 1,754 real	ML: SVM, MNB, LR, KNN, DT, AdaBoost; DNN: CNN, LSTM, BiLSTM, CNN+LSTM, CNN+BiLSTM; Transformers: mBERT, Bangla-BERT	DNNs 95.9%, SVM 95.2%, Bangla-BERT 94.1%	Clean, balanced Bangla dataset, established baselines across ML, DNN, transformer models	Limited dataset size, transformer models not fully trained, hybrid models/added features could improve performance
[8]	2022	10,700 manually annotated English social media posts	ML: MNB, RF; DL: RNN, LSTM, Bi-LSTM, GRU; LIME applied to Bi-LSTM	BiLSTM + LIME: 94.25% accuracy	Proposes interpretable BiLSTM model using LIME and PMI	Weak alignment between LIME and global PMI, longer posts reduce interpretability, baseline comparison limited
[9]	2023	ISOT: 44,898, FA-KES: 804	CNN+BiLSTM features fused early, classified by FLN	High accuracy on both datasets, outperformed existing methods	Unified framework combining global, local, temporal features with early fusion	Tested only on English datasets
[10]	2023	130,307 Bangla documents	RF, DT, SVM, KNN, LR, MNB, Bernoulli NB; LIME and SHAP for interpretability	RF: 99.91% accuracy, 0.9991 F1-score	Large Bangla corpus, high performance, explainable NLP techniques applied	Focused on supervised ML; deep learning, transfer learning, advanced NLP remain future work
[11]	2023	FA-KES: 804, ISOT: 44,898	DeepCnnLstm, DeepCnnBilstm, LSTM/BiLSTM for feature extraction	DeepCnnBilstm: 68% FA-KES, near 100% ISOT	Hybrid CNN-RNN models extract spatial/contextual features using ELU and dropout	Limited generalization across datasets, binary classification only
[12]	2024	58,478 Bangla news articles	GRU-based sequential neural network	94% accuracy	GRU-based Bangla fake news model outperforms prior works	Slightly lower validation accuracy; needs broader language and real-time testing
[13]	2024	BanFakeNews	LR, KNN, RF, SVM, BERT-BiLSTM, XLM-RoBERTa	RoBERTa-GCN: 0.986 accuracy	Bangla-specific RoBERTa-GCN combining NLP and relational learning	May need retraining for other languages or real-time data; nuanced misinformation types remain challenging
[14]	2024	66,940 news headlines (Kaggle)	LR, RF, DT, KNN; LIME for interpretability	RF: 91.48% accuracy	Combined SBERT, ML models, and LIME for interpretable classification	Models may struggle with complex patterns
[15]	2024	7,796 news articles (Kaggle)	Classical ML: SVM, RF, Gradient Boosting, k-means, DBSCAN; DL: CNN, LSTM, Bi-LSTM, Bi-directional LSTM, CNN+LSTM	CNN+LSTM hybrid outperformed other models	Hybrid CNN+LSTM using LIWC-based linguistic features	Performance depends heavily on training data quality; less robust to evolving language patterns
[16]	2024	WELFake, FakeNewsNet, FakeNewsPrediction	ML: DT, RF, SVM, LR, XGBoost, CatBoost; DL: LSTM, BiLSTM, GRU, CNN-LSTM; Transformers: BERT, XLNet, RoBERTa	CNN-LSTM + FastText: accuracy/F1 up to 0.99	Hybrid CNN-LSTM effective with interpretable decisions via LDA/LIME	ML models inconsistent; CNN-LSTM had occasional errors; limited to English datasets
[17]	2024	FakeNewsNet (Politifact 600, GossipCop PG 18,782)	RST, LIWC, text-CNN, SAF, LR (N-gram), HAN, dEFEND, BERT, hybrid BERT-mCNN-sBiGRU (CMAFD)	CMAFD ¿93% accuracy	Models semantic interactions between news content, headlines, user behavior	Handles mainly homogeneous inputs, relies on computationally heavy BERT embeddings
[18]	2024	TruthSeeker (180,000 tweets)	Classical ML: NB, LR, KNN, SVM, MLP, DT, RF; DL: CNN (3 architectures) with TF-IDF, Word2Vec, FastText	SVM TF-IDF: 99.03%, CNN-3 TF-IDF: 98.99% accuracy, 99.02% F1	Identifies optimal model–embedding combinations	Model performance varies with embedding type; advanced embeddings like BERT not tested
[19]	2025	ISOT, Kaggle, ConFake, McIntire	ML: KNN, RF, GNB, LR, Bagging, AdaBoost, SVM; DL: CNN, LSTM, Word Embedding + CNN/LSTM; Proposed: HHO-CNN-BiLSTM	Accuracy up to 98.89% (ISOT)	Integrates HHO-based feature selection with hybrid CNN-BiLSTM for accurate multi-dataset detection	Slightly lower performance on some datasets, manual hyperparameter tuning required

METHODOLOGY

The following research adopts a systematic approach to identifying and analyzing fake news (Fig. 1). News articles are collected through web scraping. Rigorous cleaning and preprocessing are performed, followed by categorization into topics. The proposed hybrid model is then applied, interpreted, and evaluated for transparency.

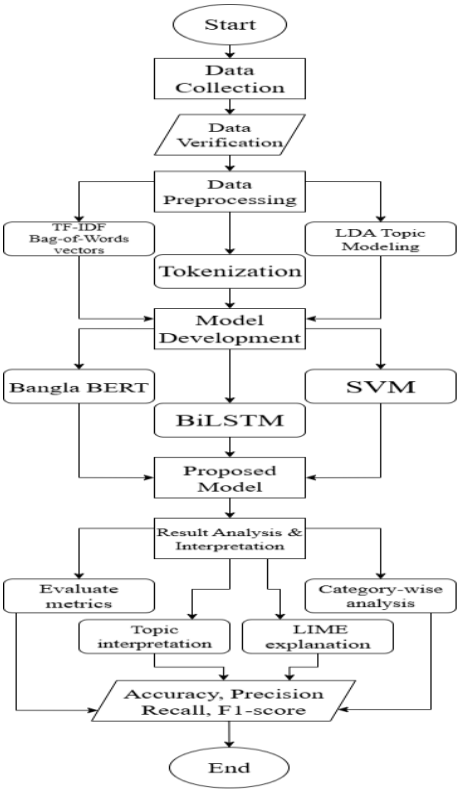

Fig.1. Fake News Detection Pipeline

A. Dataset Collection

The dataset involves news articles crawled from 20 reputed online newspapers in Bangladesh, covering the years 2020 to 2024 (Table II).

TABLE II: Domain-wise Distribution of Fake News (Total = 5000)

Domain	Website	Number of Fake Data
Prothom Alo	prothomalo.com	500
Daily Star Bangla	bangla.thedailystar.net	383
Pratidin	pratidin.com	457
Bdnews24 Bangla	bangla.bdnews24.com	393
Jagonews24	jagonews24.com	500
Kaler Kantho	kalerkontho.com	478
Samakal	samakal.com	309
Ittefaq	ittefaq.com.bd	319
Bangla Tribune	banglatribune.com	447
Naya Diganta	dailynayadiganta.com	351
Channel 24	channel24bd.tv	330
Somoy TV	somoynews.tv	340
Independent TV	independent24.com	298
Amader Shomoy	amadershomoy.com	287
Shomoyer Alo	shomoyeralo.com	234
Desh Rupantor	deshrupantor.com	223
Ajker Patrika	ajkerpatrika.com	245
Risingbd	risingbd.com	319
Bonik Barta	bonikbarta.net	213
Bhorer Kagoj	bhorerkagoj.com	202
Total		**5000**

The features examined include, but are not limited to, clickbait, sensational headlines, exaggerated claims, and politically biased narratives. A manual verification process is used, involving cross-checking the content against trusted sources to ensure factuality, realism of claims, mismatch against other news, and reliability of the source. Fact-checking tools such as Boom Live, Jachai.com, and FactWatch BD are integrated into this verification process (Fig. 2). Finally, the results of the pipeline feed into the creation of the Fake News Dataset, which forms the foundation for further analysis and model training (Fig. 5).

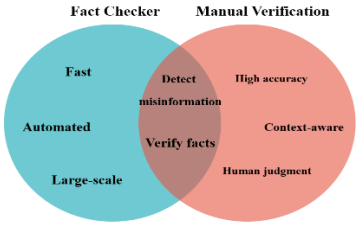

Fig. 2. Fact Checker and Manual Verification Venn Diagram

B. Preprocessing

The raw text data are cleaned to handle inconsistencies such as redundant whitespaces, irrelevant non-Bengali characters, advertisements, and unrelated text. Additional preprocessing includes stemming/lemmatization, removing stop words, duplicates, punctuation, and stripping HTML tags and hyperlinks (Fig. 3). This ensures that the data are in an appropriate format for tokenization and further processing.

C. Feature Extraction

Feature extraction methods convert the Bangla fake news text into numerical representations. TF-IDF and Count Vectorizer with Bag-of-Words (BoW) quantify word importance, while tokenization splits text into individual tokens. LDA is applied to identify thematic structures, grouping words based on co-occurrence patterns into topics such as Bangladesh News, World News, Business News, Entertainment News, and Sports News.

1) TF-IDF (Term Frequency–Inverse Document Frequency): The TF-IDF score for a word ω in a document d is given by:

$$TF - IDF(\omega, d) = TF(\omega, d) \times IDF(\omega)$$

where:

$$TF(\omega, d) = \frac{Count\ of\ \omega\ in\ d}{Total\ number\ of\ words\ in\ d}$$

and:

$$IDF(\omega) = log\frac{N}{DF(\omega)}$$

Here, N is the total number of documents, and DF(ω) is the number of documents containing the word ω .

2) LDA (Latent Dirichlet Allocation): The likelihood of a word w in a document d for a topic t is given by:

$$P(\omega|d) = \sum_{t=1}^{K} P(t|d)P(\omega|t)$$

where:

- K = total number of latent topics in the topic model
- P (t|d) is the topic distribution for document d, representing the proportion of each topic in the document.
- P (ω |t) is the word distribution for topic t, representing the probability of ω occurring in t.

LDA introduces two hyperparameters:
• α: controls the document–topic distribution.
• β: controls the word–topic distribution.

While both TF-IDF and LDA were tested, LDA performed better as it captures contextual relevance and semantic relationships between words, making it more effective for categorizing Bangla text. In contrast, TF-IDF focuses on unique word importance, which is less effective for this dataset.

Stage	Text Representation
Original Bengali Text	নাটোর স্বেচ্ছাসেবক লীগের কমিটিকে 'অবৈধ' বললেন আগের নেতারা, "প্রায় ২৪ বছর পর কেন্দ্র অনুমোদিত নাটোর জেলা আওয়ামী স্বেচ্ছাসেবক লীগের নতুন কমিটিকে 'অবৈধ' ও 'পকেট কমিটি' হিসেবে দাবি করেছেন সদ্য বিলুপ্ত হওয়া জেলা কমিটির নেতারা।...
After Tokenization	[নাটোর, স্বেচ্ছাসেবক, লীগের, কমিটিকে, অবৈধ, বললেন, আগের, নেতারা, প্রায়, ২৪, বছর, পর, কেন্দ্র, অনুমোদিত, জেলা, আওয়ামী, নতুন, পকেট, কমিটি, হিসেবে, দাবি, করেছেন, সদ্য, বিলুপ্ত, হওয়া]
Bag-of-Words Representation	"নাটোর": 2, "স্বেচ্ছাসেবক": 2, "লীগের": 2, "কমিটিকে": 2, "অবৈধ": 2, "বললেন": 1, "আগের": 1, "নেতারা": 2, "প্রায়": 1, "২৪": 1, "বছর": 1, "পর": 1, "কেন্দ্র": 1, "অনুমোদিত": 1, "জেলা": 2, "আওয়ামী": 1, "নতুন": 1, "পকেট": 1, "কমিটি": 1, "হিসেবে": 1, "দাবি": 1, "করেছেন": 1, "সদ্য": 1, "বিলুপ্ত": 1, "হওয়া": 1

Fig. 3: Transformation of Bengali text using Count Vectorizer with Bag-of-Words (BoW)

D. Model Description

The proposed model combines BERT for contextual embeddings, BiLSTM for capturing sequential dependencies, and SVM for robust classification (Fig. 4). BERT (Bidirectional Encoder Representations from Transformers) serves as the first layer, generating deep contextual embeddings for each token based on its surrounding words. Its multi-layer transformer architecture, equipped with self-attention mechanisms, captures complex semantic relationships and long-range dependencies, providing rich representations for downstream processing. These embeddings are then fed into a BiLSTM (Bidirectional Long Short-Term Memory) layer, which processes sequences in both forward and backward directions. The BiLSTM uses input, forget, and output gates to control the flow of information, enabling the model to learn sequential patterns and preserve long-term dependencies essential for understanding sentence structure and news context. Finally, the outputs from BiLSTM are passed to a Support Vector Machine (SVM) classifier. SVM constructs optimal hyperplanes to separate classes in high-dimensional feature space and can employ kernel functions for non-linear decision boundaries, ensuring precise classification. By combining deep contextual and sequential representations with a stable, traditional classifier, the model achieves both rich language understanding and robust predictive power (Fig. 6). The hybrid model:

• BERT: contextual word representations.
• BiLSTM: sequential pattern recognition.
• SVM: robust decision boundaries.

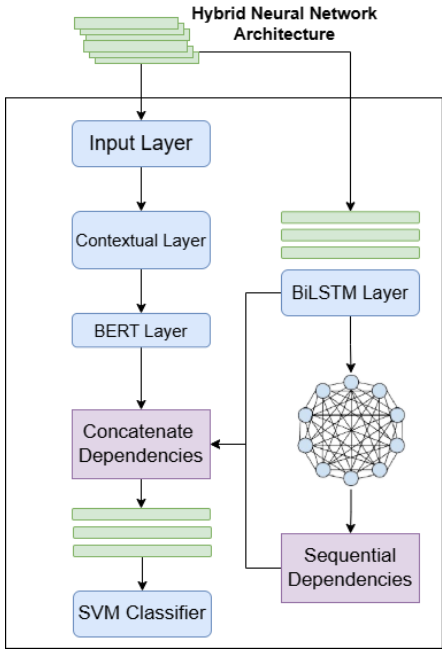

Fig. 4. Hybrid Model Architecture(BERT+BiLSTM+SVM)

E. Proposed Pipeline Algorithm

Algorithm 1: Bangla Fake News Detection Pipeline

Require: *Raw articles A (2020–2024)*
Ensure: *Label ∈ {Fake, Real} + explanation*
1: Collect → Verify
2: Clean: ads, HTML, non-Bengali, stopwords
3: Preprocess: stemming / lemmatization
4: Features: Tokenize, TF–IDF, BoW, LDA
5: Model: BanglaBERT → BiLSTM → SVM
6: Evaluate: Acc, Prec, Rec, F1
7: Interpret: LIME
8: return Label + explanation

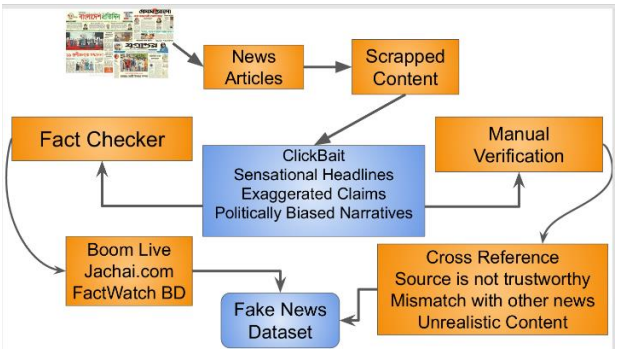

Fig. 5. Data collection Overview

F. *Interpretability with LIME*

LIME (Local Interpretable Model-agnostic Explanations) is applied to explain the predictions of the hybrid model. LIME identifies key features or words influencing each decision, thereby enhancing transparency. This ensures that predictions are not only accurate but also interpretable, which is essential for deploying fake news detection systems in real-world applications.

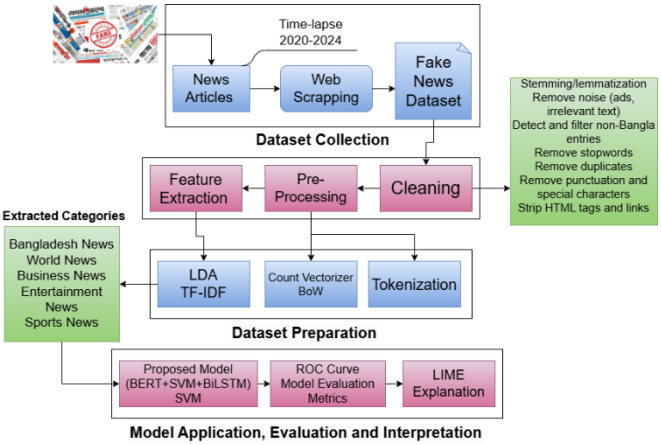

Fig. 6. Structural Overview of Framework

RESULTS ANALYSIS AND DISCUSSION

A. *Experimental Setup*

The proposed BERT+SVM+BiLSTM hybrid model was implemented using Python and executed on Google Colab Pro. Additional experiments were run on a PC with a Windows 11 operating system, an 11th-generation Intel Core i5-1135G7 processor (2.40–2.42 GHz), 8 GB RAM, and a 64-bit operating system. The dataset was split 80%–20% for training and testing. Dropout and batch normalization were applied to reduce overfitting.

B. *Evaluation Metrics*

Evaluation metrics are essential for assessing models, guiding parameter tuning for optimal performance, and ensuring reliable deployment on unseen datasets. This study emphasizes NLP metrics like precision, recall, F1-score, and accuracy, derived from the confusion matrix using true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

Accuracy: Accuracy is the percentage of correct predictions. It measures how frequently the model predicts the correct outputs and the ratio of all correctly anticipated cases to all cases in the data.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Precision: Precision represents the ratio of correctly predicted positive cases compared to all predicted positive cases.

$$\Pr e\,cision = \frac{TP}{TP+FP}$$

Recall (Sensitivity): Recall shows the percentage of positive cases correctly identified by the model.

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

F1-Score: The F1-score balances precision and recall, providing a better measure of model performance when false positives and false negatives have different costs.

$$F1 = 2 \times \frac{\Pr ecision \times Recall}{\Pr ecision + Recall}$$

C. *Dataset Overview*

Data Distribution Across Sources: (Fig. 7) illustrates the number of articles collected from each newspaper, highlighting variations in contribution from different sources. Category-wise Distribution: Category-wise data distribution of the dataset (Fig. 8), indicating a moderate class imbalance across news categories such as Bangladesh news, sports news, entertainment news, world news, and business news. The pie chart demonstrates the diversity of content in the dataset, which is crucial for training models capable of handling multiple news domains.

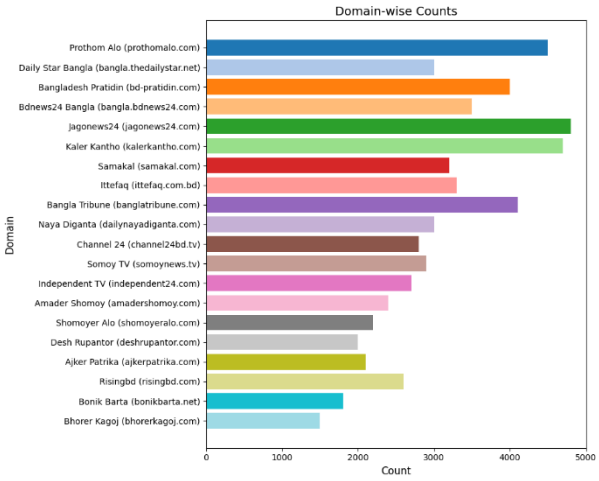

Fig. 7. Number of Articles Collected per Newspaper

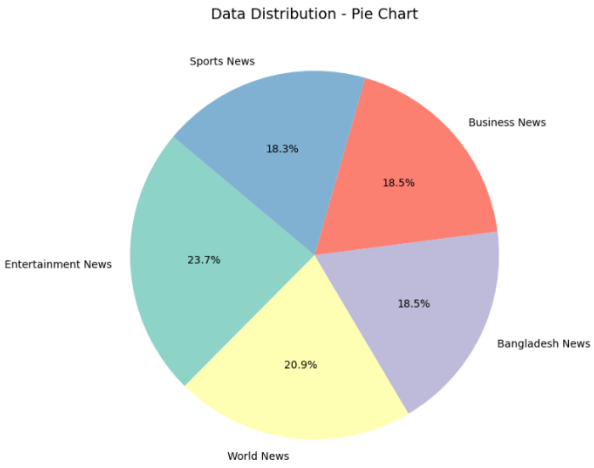

Fig. 8. Distribution of Articles by News Category

D. Polar Bar Chart Analysis for LDA and TF-IDF

Polar bar charts are used to visualize the frequency of top words and thematic relevance across topics for both LDA and TF-IDF methods (Fig. 9, 10).

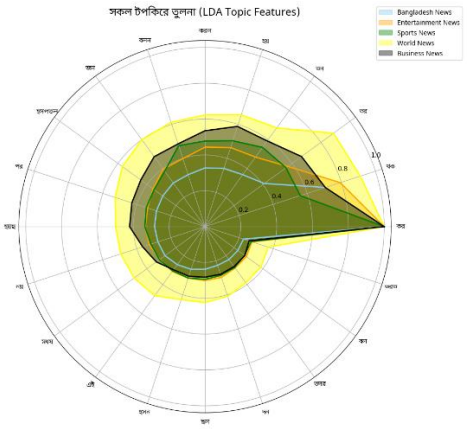

Fig. 9. Polar bar chart visualization for LDA.

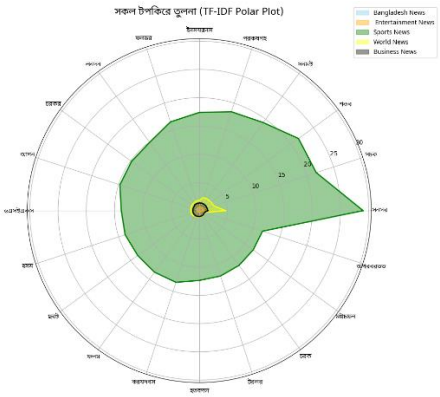

Fig. 10. Polar bar chart visualization for TF-IDF

These visualizations highlight the strengths of the models in identifying and focusing on key linguistic patterns across categories. First, TF-IDF was applied to determine word importance across the corpus. While it is effective for keyword extraction, TF-IDF underperforms in capturing underlying topics because it relies solely on term frequency without considering contextual relationships. Consequently, TF-IDF was less effective compared to LDA in identifying hidden topics in the text. In contrast, LDA produced more meaningful and coherent topics, such as Bangladesh, Business, Entertainment, World, and Sports. This improvement is due to LDA's ability to capture word co-occurrence and contextual relevance, making it more suitable for topic modeling tasks.

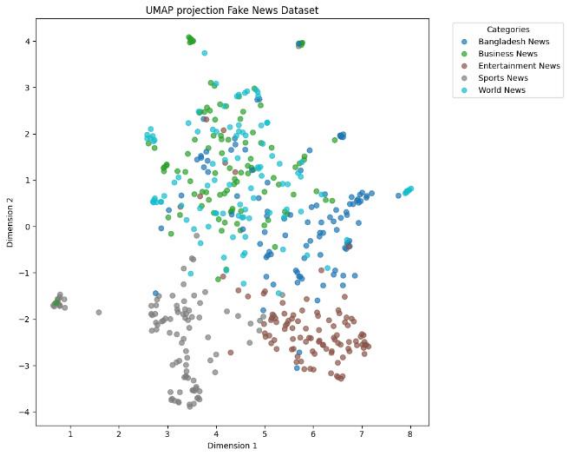

Fig. 11. UMAP Projection of news embeddings

E. Classification Performance of the Proposed Hybrid Model

To better understand how the hybrid model represents news articles, we applied UMAP to project high-dimensional embeddings into 2D space (Fig. 11). The projection reveals clear clustering of categories such as Entertainment, Sports, and Bangladesh News, indicating that the model captures distinct semantic patterns. Some overlap occurs between Business and World News, consistent with the confusion matrix results. Table III presents the classification report for the proposed BERT + BiLSTM + SVM hybrid model across five news categories. The hybrid approach achieved an overall accuracy of 97%. The hybrid model demonstrated high performance across all categories, with F1-scores reaching 0.97 for Entertainment News and Sports News, indicating its strong ability to capture complex linguistic patterns in the Bangla fake news dataset. Fig 12 shows the classification performance of the BERT+BiLSTM+SVM model across six categories: Bangladesh News, Business News, Entertainment News, Sports News, World News, and overall accuracy. Precision, recall, and F1-score are plotted for each class. The model performs consistently well across all categories, with scores mostly above 0.9. Sports News achieves the highest performance across all metrics, while Bangladesh News shows slightly lower but still strong scores. Overall, the model demonstrates robust classification ability with high precision, recall, and F1-score, indicating effective feature learning and balanced performance across different news categories.

TABLE III: CLASSIFICATION REPORT FOR NEWS CATEGORIES (BERT + BILSTM + SVM)

Category	Accuracy	Recall	Precision	F1-Score
Bangladesh News	0.96	0.95	0.95	0.96
Business News	0.96	0.96	0.95	0.96
Entertainment News	0.97	0.96	0.97	0.97
Sports News	0.98	0.97	0.97	0.97
World News	0.96	0.96	0.95	0.96
Overall Accuracy	**0.97**			

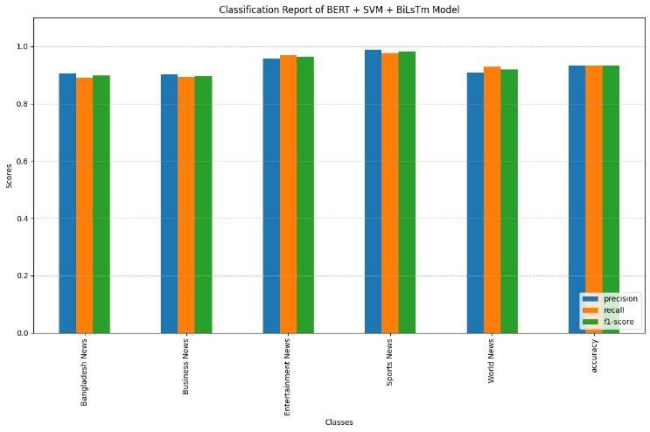

Fig. 12. Classification Report(BERT+SVM+BiLSTM)

F. Confusion Matrix Analysis

The category-wise confusion matrix for the hybrid model (Fig. 13) illustrates its high classification accuracy across all categories. The diagonal dominance indicates correct predictions, with minimal misclassifications between categories. Notably, World News and Bangladesh News achieved the highest correct classification counts, while occasional confusion occurred between closely related topics such as Business and World News.

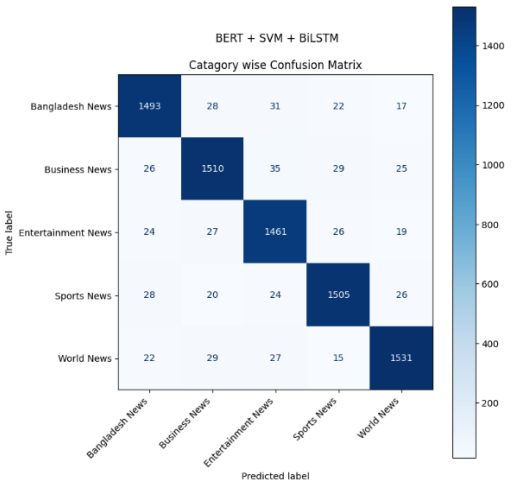

Fig. 13. Category-wise Confusion Matrix

G. Training and Validation Curves

The training history of the hybrid model (Fig. 14) shows a steady improvement in accuracy for both training and validation sets, reaching near convergence around epoch 50. Similarly, the loss curves indicate a consistent reduction in training and validation loss, with no significant signs of overfitting, confirming stable model learning (Fig. 15).

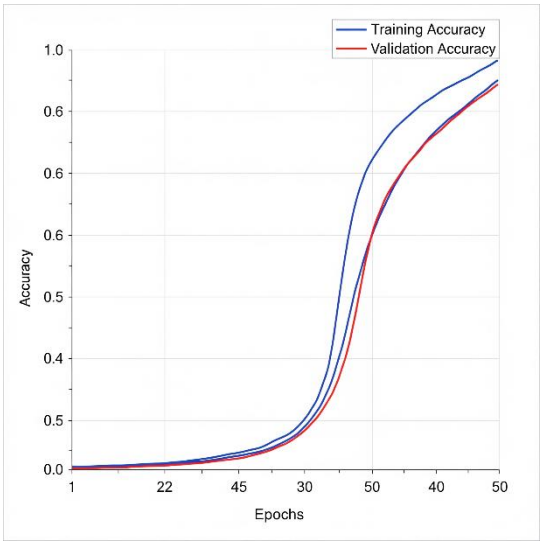

Fig. 14. Training Accuracy over Epochs

H. Model Accuracy Comparison

Figures 16–18 compare the hybrid model's accuracy against other machine learning, deep learning, and transformer-based approaches. These comparisons in table IV highlight the hybrid architecture's ability to leverage the strengths of contextual embeddings, sequential modeling, and robust classification to deliver superior performance. Specifically, the hybrid model achieved an absolute accuracy gain of 12 percentage points over the best-performing machine learning model (Random Forest, 85%), 15 points over the best deep learning baseline (CNN, 85%), and 10 points over the strongest transformer-based model (XLM-RoBERTa, 87%).

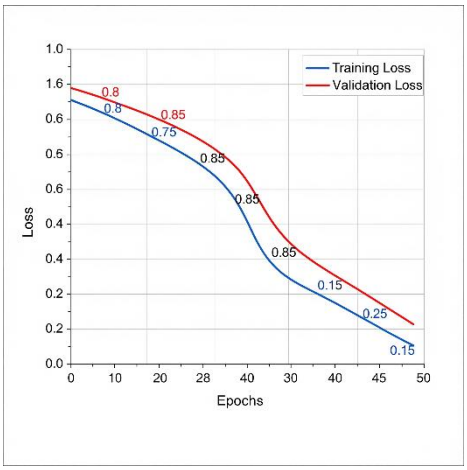

Fig. 15. Training Loss over Epochs

TABLE IV. PERFORMANCE COMPARISON OF MODELS

Category	Model	Accuracy(%)
Machine Learning	Logistic Regression	78
	AdaBoost	80
	Random Forest	85
Deep Learning	CNN	85
	BiLSTM	82
	GRU	75
Transformer Based	XLM-RoBERTa	87
	GPT-Neo	85
	DistilBERT	83
Hybrid Model	**BERT+BiLSTM+SVM**	**97**

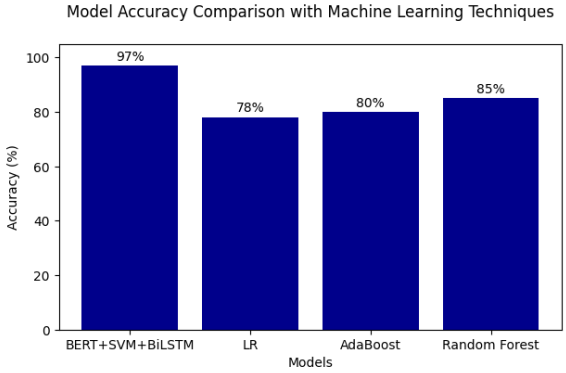

Fig. 16: Comparison of Model Accuracy with ML Baselines

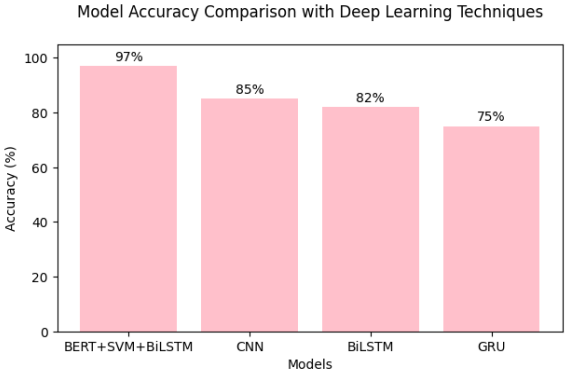

Fig. 17. Comparison of Model Accuracy with DL Baselines

I. ROC and AUC Analysis

The ROC curves (Fig 19) for the hybrid model show AUC values between 0.99 and 1 across all categories, demonstrating exceptional discrimination capability. The model performed particularly well in distinguishing Entertainment News and Sports News, further confirming its robustness.

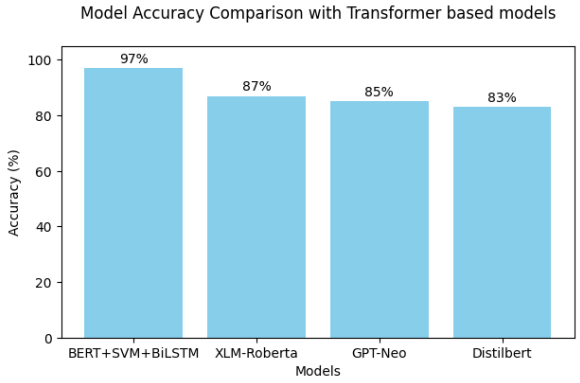

Fig. 18. Comparison of Model Accuracy with TBM Baselines

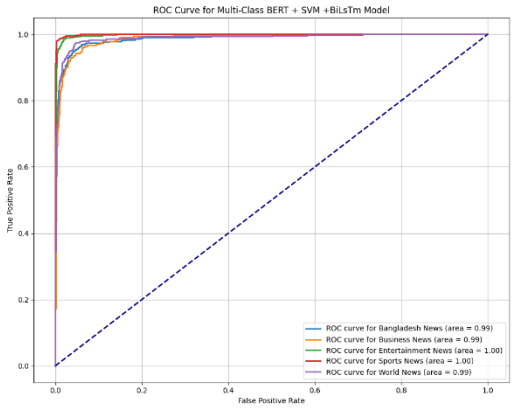

Fig. 19. ROC Curve: Hybrid Model

J. Model Interpretability with LIME

After training the model, LIME (Local Interpretable Model agnostic Explanations) is applied to understand and interpret its predictions. LIME works by slightly perturbing the input data and observing how the model's outputs change, allowing it to identify the words or features that have the most significant impact on each prediction. This analysis is crucial because deep learning models such as CNNs and LSTMs are often considered black boxes, providing accurate results without revealing the reasoning behind them. In our work, LIME produces visual explanations, including bar charts showing the top features contributing to a prediction, highlighted text within articles indicating words that strongly influence the model toward a "true" or "fake" label, and visual comparisons of true versus predicted labels (Fig. 20). These explanations not only reveal which elements of the text drive the model's decisions but also help understand why misclassifications occur. By providing this level of transparency, LIME makes the model's outputs more interpretable and trustworthy, which is particularly important in real-world applications where users need to understand and trust the decisions made by the system.

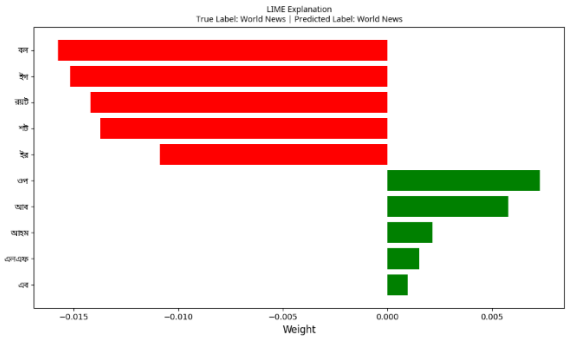

Fig. 20. Top features for LIME explanation

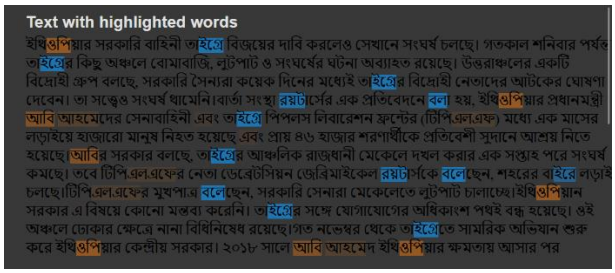

Fig. 21. Highlighted text of the original document

Fig. 22. True vs predicted label

CONCLUSION

Advanced feature extraction, a robust hybrid model, and interpretability tools like LIME together set up a pipeline for the analysis of fake news in Bangla text. A dataset of 5,000 fake news articles is scraped from 20 Bangladeshi newspapers spanning the years 2020 to 2024. The categorization of topics using LDA outperforms other techniques, providing clear and interpretable clusters for topic modeling. The proposed BERT+BiLSTM+SVM hybrid model outperforms the traditional models comprehensively, with high accuracy and precision in all categories. The results highlight how transparently the model captures features with high accuracy, setting a strong foundation for future research and real-world applications In combating fake news in Bangla.

FUTURE WORK

Future research can focus on expanding the dataset to include more sources and longer time spans, which may improve the model's generalizability. Incorporating additional multilingual transformer models could enhance detection performance, especially for mixed-language news articles. Furthermore, extending the framework to handle multi-class classification, such as partially true or misleading news, would make the system more versatile in practical scenarios.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to our supervisor, Nuren Nafisa, for her valuable guidance and support throughout the research process. Her expertise and insights were invaluable in shaping our research and helping us to overcome challenges.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCE

[1] Hussain, M. G. H. M. R. R. M. P. J., & Hossain, S. A. (2020). *Detection of Bangla fake news using MNB and SVM classifier*. In iCCECE 2020: Electronics & Communications Engineering Conference.

[2] International Monetary Fund. (2024, May 7). *IMF News*. https://www.imf.org/en/Home.

[3] Reuters. (2024, June 24). *Bangladesh economic and business news*. https://www.reuters.com/.

[4] Hossain, M. Z. R., M. A., I., M. S., & Hossain, S. (2020). *BanFakeNews: A dataset for detecting fake news in Bangla*. In Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), Marseille.

[5] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *Why should I trust you?* In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco.

[6] Hossain, E., Karim, K., M., N. J., M., J. U. R., & M. M. (2022). A study towards Bangla fake news detection using machine learning and deep learning. In *Advances in Intelligent Systems and Computing* (Vol. 1408, pp. 79–95).

[7] Rasel, R. I., Z., A. H., S. N., & H. (2022). *Bangla fake news detection using machine learning, deep learning and transformer models*. In 2022 25th International Conference on Computer and Information Technology (ICCIT).

[8] Ahmed, M. H., M. S. I., R. U., & A. (2022). Explainable text classification model for COVID-19 fake news detection. *Journal of Internet Services and Information Security, 12*(2), 51–69.

[9] Almarashy, A. H., J., F.-D., M.-R., & S. (2023). Enhancing fake news detection by multi-feature classification. *IEEE Access, 11*, 139601–139613.

[10] Habibullah, M. I., M. S., J. F., T., & (2023). *Bangla document classification based on machine learning and explainable NLP*. In 2023 6th International Conference on Electrical Information and Communication Technology (EICT).

[11] Tokpa, F. W., R. K., B. H., M. V., & O. (2023). Fake news detection in social media: Hybrid deep learning approaches. *Journal of Advances in Information Technology, 14*(3), 606–615.

[12] Mondal, P. K., K., S. S. R. M. M., R. S. S. S. A., & R. (2024). *Breaking the fake news barrier: Deep learning*

approaches in Bangla language. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT).

[13] Ahammad, M. A., S. K., R. K., I. M., T. M., & M. M. R. (2024). RoBERTa-GCN: A novel approach for combating fake news in Bangla using advanced language processing and graph convolutional networks. *IEEE Access, 12*, 132644–132663.

[14] Tabassoum, N., & A. M. (2024). Interpretability of machine learning algorithms for news category classification using XAI. In 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT).

[15] Dev, D. B. V. B. B. S. G. M., & Goyal, N. (2024). LSTMCNN: A hybrid machine learning model to unmask fake news. *Heliyon, 10*(3), e25244.

[16] Hashmi, E. Y., S. Y., Y. M., M. A., S., & (2024). Advancing fake news detection: Hybrid deep learning with FastText and explainable AI. *IEEE Access, 12*, 44462–44480.

[17] Alghamdi, J. L. Y., & L. (2024). Unveiling the hidden patterns: A novel semantic deep learning approach to fake news detection on social media. *Engineering Applications of Artificial Intelligence, 137*(Part B), 109240.

[18] Al-Tarawneh, M. A. B., A.-i. O., A.-M. K., S. K., H., & (2024). Enhancing fake news detection with word embedding: A machine learning and deep learning approach. *Computers, 13*(9), 239.

[19] Jain, M. K., G. D., & M. (2025). Hybrid CNN-BiLSTM model with HHO feature selection for enhanced fake news detection. *Social Network Analysis and Mining, 15*, Article 43.