

Multi-Task BanglaBERT for Joint Sentiment and Fake News Detection in COVID-19 Discourse

Arshadul Hoque

Department of Computer Science and Engineering, University of Chittagong, Chittagong, 4331, Bangladesh

E-mail: ahsbd95@gmail.com

ORCID iD: <https://orcid.org/0009-0004-4636-7446>

Received: June 9, 2026; Revised: June 25, 2026; Accepted: July 13, 2026; Published: August 8, 2026

Abstract: The COVID-19 pandemic catalyzed an unprecedented surge of misinformation on social media, frequently intertwined with emotionally charged language. Understanding both the sentiment and truthfulness of this content is critical for public health monitoring and misinformation mitigation. However, Bangla—despite being a globally prominent language—remains severely underrepresented in joint sentiment and fake news detection research, with existing studies largely restricted to single-task settings. To bridge this gap, this paper proposes a novel multi-task BanglaBERT-based framework for the simultaneous classification of sentiment and truthfulness in COVID-19 discourse. Furthermore, we introduce the first publicly available, dual-annotated Bangla corpus for this domain, comprising 35,526 textual samples aggregated from social media and news sources. Our architecture employs a shared BanglaBERT encoder with dual task-specific heads, optimized using a task-prioritized loss function that combines modified Focal Loss and weighted cross-entropy to address inherent class imbalances. Extensive experiments demonstrate that the proposed model achieves 75.1% accuracy (Macro F1: 0.707) for sentiment classification and 88.0% accuracy (Macro F1: 0.851) for truthfulness detection. Ablation studies and error analyses confirm that our tailored loss strategies significantly enhance the recognition of underrepresented and semantically ambiguous classes, particularly neutral sentiments. By releasing our dataset, code, trained models, and a Gradio-based interactive demo, this work establishes a robust benchmark for multi-task learning in low-resource Bangla NLP and provides a practical tool for fact-checking during health crises.

IndexTerms: BanglaBERT, COVID-19 misinformation, Multi-Task learning, sentiment analysis, fake news detection, low-resource NLP

1. Introduction

The COVID-19 pandemic disrupted public health systems worldwide and simultaneously fueled an unprecedented surge of misinformation on social media. False claims about vaccines, treatments, and the virus itself circulated rapidly, often framed with emotionally charged language. Understanding both the sentiment (positive, neutral, negative) and the truthfulness (real, fake) of such content is critical for mitigating misinformation and for monitoring public perception during health crises.

While English and other high-resource languages have benefited from extensive research in sentiment analysis and fake news detection, Bangla remains severely underrepresented, despite being the seventh most spoken language globally. Existing Bangla NLP studies have typically focused on either sentiment classification [1–4] or misinformation detection [5–7], but rarely address both in a joint setting. Multi-task frameworks that could exploit shared linguistic representations across tasks remain largely unexplored. This gap limits our ability to capture the nuanced interplay between emotional framing and misinformation spread in Bangla discourse.

In this work, we propose a multi-task BanglaBERT-based framework for simultaneous sentiment and truthfulness classification on COVID-19 related textual content.

To the best of our knowledge, this is the first publicly available Bangla corpus that provides joint sentiment and truthfulness annotations on the same COVID-19 discourse samples. While prior datasets such as Ban-FakeNews [8], BanFakeNews-2.0 [9], BANGLABOOK [10], and MultiBanFakeDetect [7] address single tasks or single modalities, none provide paired sentiment–truthfulness labels on a unified COVID-19 corpus. We therefore position our contribution not as the first Bangla dataset, but as the first dual-annotated, multi-task Bangla resource for joint sentiment–truthfulness modeling in pandemic discourse.

To facilitate future research, we have released our code implementation on GitHub and the trained model on

Hugging Face.

Extensive experiments show that this model achieves 75% sentiment accuracy (macro F1 = 0.70) and 88% truth accuracy (macro F1 = 0.85). These results establish a strong benchmark for future multi-task research in Bangla NLP and demonstrate the feasibility and effectiveness of integrating sentiment and misinformation detection in the context of limited linguistic resources.

The remainder of this paper is organized as follows. Section 2 reviews related work on Bangla sentiment analysis and misinformation detection. Section 3 provides details about dataset construction, preprocessing, and model architecture. Section 4 presents experimental results, ablation studies, error analysis, and compares our approach with previous work, and Section 5 concludes with a discussion of future research directions.

2. Related Work

We review previous work most relevant to our multi-task framework for fake news detection (FND) and sentiment analysis (SA) in Bangla, focusing on linguistic challenges, modalities, and low-resource multi-task learning (MTL).

2.1. Fake News and Sentiment Analysis in Bangla

Fake news—disinformation designed to manipulate public opinion—has surged with digital platforms. Automated FND is critical, yet most research targets high-resource languages. In Bangla, early efforts relied on *unimodal* text-based approaches.

Hossain et al. [8] presented BanFakeNews, a dataset of 50,000 Bangla articles (48.7K genuine, 1.3K fake), establishing benchmarks using traditional features (n-gram TF-IDF, POS tags, punctuation) and neural models. SVM with TF-IDF at the character level achieved an F1 score of 91%, outperforming neural networks and highlighting the importance of linguistic cues. Majumdar Shibu et al. [9] extended this with BanFakeNews-2.0 (60,000 articles in 13 categories) and presented a transformer-based benchmark using mBERT and BLOOM with quantized fine-tuning, achieving an F1 score of 89%, surpassing handcrafted feature benchmarks and demonstrating the superiority of embeddings.

Bangla sentiment analysis faces similar resource constraints. Maisha et al. [1] achieved 99% accuracy in document-level SA using TF-IDF and Random Forest. Meanwhile, Mahmud et al. [11] proposed a hybrid model *BSPS* + *BanglaBERT* that improved accuracy from 79% to 89% by combining rule-based polarity estimation with pre-trained language model representations. These results highlight the value of hybrid architectures with language models.

Research specifically addressing COVID-19 in Bangla remains sparse. Ahmed et al. [12] compared RNN, LSTM, and BERT models, with BERT achieving 95% accuracy on a dataset of 10,000 articles. However, to our knowledge, no large-scale, publicly available dataset exists for COVID-19 misinformation specifically on social media in Bangla. To address general data scarcity, Kabir et al. [10] released *BANGLABOOK* (158,000 book reviews) and Rashid et al. [4] released the largest dataset of e-commerce reviews (1.74 million posts in Bangla, English, and Romanized Bangla), which includes images to enable multimodal analysis.

2.2. Multimodal and Context-Aware Approaches

While multimodal methods are gaining popularity [7], text remains the dominant modality for content analysis and provides a solid foundation for understanding task synergies. We focus on text-based multi-task learning to create a robust and interpretable framework for collaborative sentiment and truth detection.

Faria et al. [7] pioneered multimodal FND in Bangla using MultiBanFakeDetect (9.6k text-image pairs). Their model MultiFusionFake (mBERT + DenseNet-169, early fusion) achieved 79.7% accuracy (6.6% improvement over text-only baseline) and strong multiclass performance. Our textual approach isolates common linguistic cues across tasks and allows for a clearer analysis of task interactions, providing a baseline for future multimodal integration.

In Aspect Sentiment Analysis (ABSA), Ahmed et al. [12] proposed a hybrid BERT random forest model tRF-BERT that outperforms previous work on benchmark datasets (improving F1 from 0.85 to 0.89) by exploiting the contextuality of BERT. This reinforces the value of contextual embeddings for capturing nuances in Bangla text, a principle we leverage in our multi-task setup.

2.3. Multi-Task Learning and Low-Resource Transfer

Multi-task learning (MTL) improves generalization by learning shared representations across related tasks, making it especially valuable in low-resource settings [13–15]. Among MTL architectures, hard parameter sharing—using a shared encoder with task-specific output heads—has become a cornerstone for NLP due to its simplicity, data efficiency, and strong empirical performance [16].

While recent advances such as adapter-based modularity [17] and hypernetworks [18] offer fine-grained control, we adopt the well-established hard-sharing framework in our dual-head BanglaBERT architecture. This choice is motivated by its proven ability to facilitate robust cross-task knowledge transfer in low-resource regimes [16], while avoiding the complexity and additional hyperparameter tuning required by modular alternatives.

In the context of Bangla, where annotated data is scarce, transfer learning via models like BanglaBERT [19] has

shown strong results, outperforming from-scratch training [20, 21]. By combining strong parameter sharing with pre-trained contextual embeddings, our approach maximizes representation reuse, improving generalization and reducing overfitting.

2.4. Summary and Positioning

Research in Bangla FND and SA has evolved from unimodal, feature-engineered models to multimodal, transformer-based systems, supported by growing datasets (BanFakeNews-2.0, MultiBanFakeDetect, BanglishRev). However, most work remains single-task and modality-siloed.

While prior Bangla work has advanced single-task models and multimodal fusion, no research has jointly modeled sentiment and truthfulness detection. Our work introduces:

1. A multi-task framework leveraging hard parameter sharing for joint sentiment and fake news detection,
2. A comprehensive dual-annotated dataset of COVID-19 related Bangla content. To the best of our knowledge, this is the first publicly available Bangla corpus that provides joint sentiment and truthfulness annotations on the same samples. While prior datasets address single tasks, none provide paired sentiment–truthfulness labels on a unified COVID-19 corpus.

3. Materials and Methods

3.1. Dataset Construction

We compiled a dataset of 35,526 Bangla textual samples, comprising both social media posts and news articles, related to COVID-19. The dataset was annotated with two distinct labels: sentiment (positive, negative, neutral) and truthfulness (real vs. fake). Data were aggregated from Facebook posts, the Bangla Newspaper Dataset (ebD) [22], BanFakeNews-2.0 [9], and the Rumor Scanner¹ fact-checking portal.

To ensure dataset quality, the following filters were applied:

- Removal of exact and near-duplicate entries.
- Exclusion of non-Bangla textual data and irrelevant noise.
- Retention only of samples with both sentiment and truthfulness annotations.

The final sentiment distribution consisted of 15,453 negative, 14,939 positive, and 5,134 neutral samples, as illustrated in Figure 1. Regarding truthfulness, the dataset comprised 25,940 real samples and 9,586 fake samples, as shown in Figure 2.

Figure 3 shows the joint distribution of sentiment and truthfulness. Among negative samples, 4,827 were fake and 10,626 were genuine. Among neutral samples, 1,761 were fake and 3,373 were real. Among positive samples, 2,998 were fake and 11,941 were genuine. Authentic content predominated across all sentiment categories, with the largest proportion in positive sentiment.

We also analyzed text length by sentiment, as illustrated in Figure 4. Neutral samples tended to be shorter on average, while negative samples exhibited greater variability with longer outliers. Positive instances showed intermediate length distributions.

Sentiment labels were initially produced by a BanglaBERT model fine-tuned on the SentiGOLD dataset [23], then manually corrected. Critically, this auxiliary sentiment model was trained on external data and does not overlap with our COVID-19 corpus, preventing data leakage. The multi-task model evaluated in Section 4 represents a separate training run on our dual-annotated corpus. To enable independent verification and reproducibility, we release the fine-tuned sentiment model checkpoint on Hugging Face [21].

3.2. Preprocessing and Label Encoding

To handle noisy, code-mixed text, we designed a multi-step preprocessing pipeline:

¹<https://rumorscanner.com/>

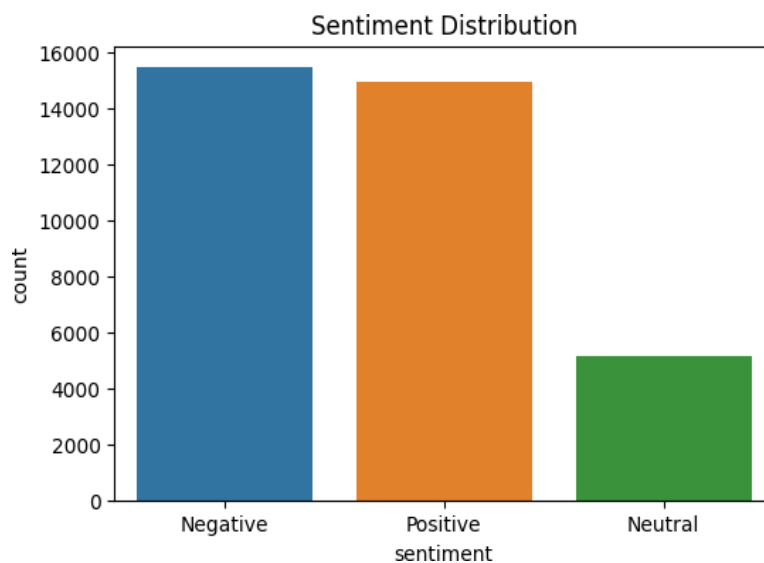


Fig. 1. Sentiment Distribution

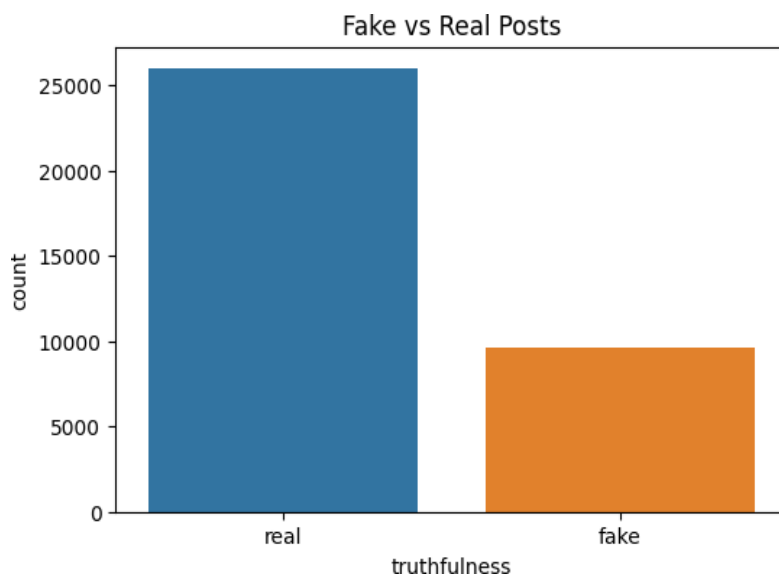


Fig. 2. Truthfulness Distribution

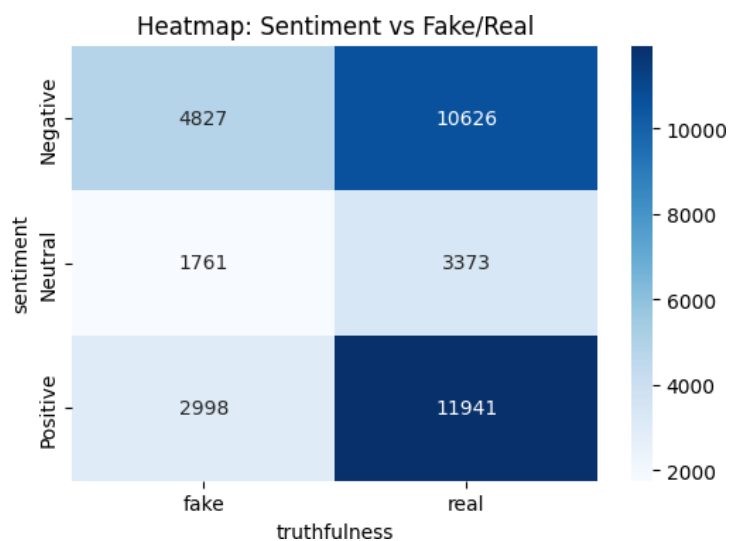


Fig. 3. Heatmap showing the distribution of sentiment (negative, neutral, positive) by truthfulness (fake vs. real), with color intensity indicating frequency.

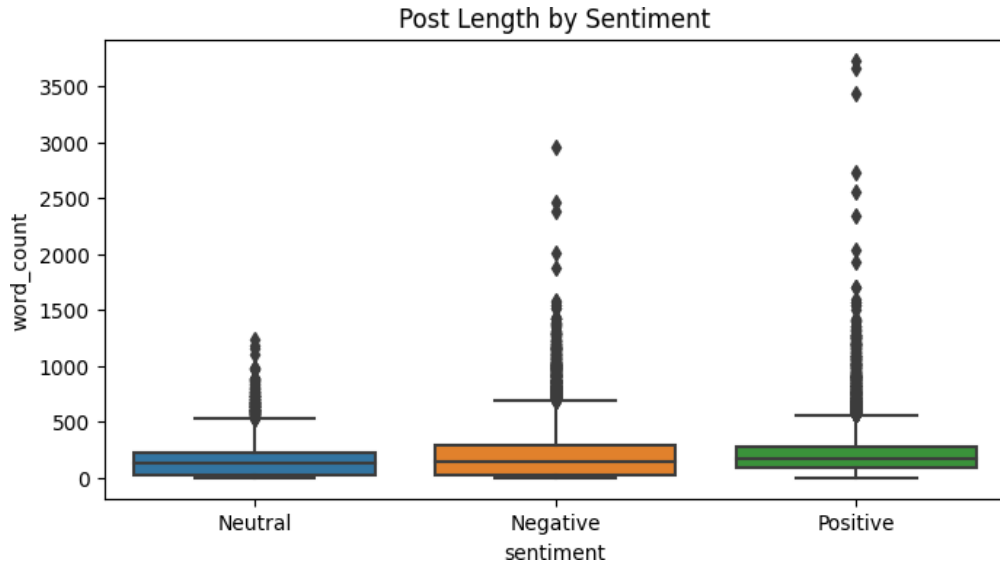


Fig. 4. Boxplot showing word count distribution by sentiment type: Neutral, Negative, and Positive, highlighting median values and outliers.

- **Normalize text using the CSEBUET NLP normalizer** [20] to fix Unicode inconsistencies, strip emojis, remove URLs, and standardize punctuation.
- **Stopword removal** via a manually curated Bangla stopwords list.
- **Whitespace** collapsing and token filtering to reduce artifacts.

Preprocessing was parallelized using multiprocessing, reducing runtime by nearly 70%. The BanglaBERT WordPiece tokenizer was used for tokenization, with a maximum sequence length of 512. We used batch tokenization with caching to increase efficiency by avoiding duplicate calculations between runs.

To maintain alignment between tokenized inputs, task-specific labels, and original row indices, we implemented a custom dataset wrapper. This enabled precise alignment of predictions with raw samples during error analysis, while a specialized collator ensured index preservation between training and evaluation batches.

Finally, alias dictionaries were used to encode labels (e.g., mapping ‘pos’ to ‘positive’, ‘fake news’ to ‘fake’). A label sanity check verified raw labels against these dictionaries; to guarantee zero mismatches between raw and encoded labels across all splits, any unmapped entries triggered explicit errors.

3.3. Data Splits

We split the dataset into training (80%), validation (10%), and test (10%) sets. Stratification was applied to the *sentiment* label to ensure a balanced representation of negative, neutral, and positive samples across all splits. The complete distribution of sentiment and authenticity (real/fake) across splits is shown in Table 1.

Table 1. Dataset Distribution

Split	Negative	Neutral	Positive	Real	Fake	Total
Train	12,362	4,107	11,951	20,752	7,668	28,420
Val	1,545	513	1,495	2,594	959	3,553
Test	1,546	514	1,493	2,594	959	3,553

3.4. Class Weighting and Loss Functions

Class imbalance was addressed with tailored loss functions for each task:

Sentiment task: A modified Focal Loss [24] with $\gamma = 2$ and class-specific α values ($\alpha_{\text{neutral}} = 1.5$; $\alpha_{\text{positive}} = 1.0$; $\alpha_{\text{negative}} = 1.0$). This penalized misclassification of the underrepresented neutral class while down-weighting easy examples.

Truthfulness task: Weighted cross-entropy loss, which up-weighted the minority false class using class weights calculated as the inverse of class frequencies.

The total multi-task loss was defined as:

$$L = \frac{1}{G} (\alpha L + (1 - \alpha) L_t) \quad (1)$$

where L_s is the sentiment loss, L_t the truthfulness loss, $\alpha = 0.90$ (prioritizing sentiment), and G is the gradient accumulation factor. This formulation ensured stable optimization under gradient accumulation.

We adopt the asymmetric weighting ($\alpha = 0.90$) based on the inherent complexity of the tasks: sentiment analysis is a more fine-grained, three-class objective requiring a stronger gradient signal to resolve semantic ambiguity (particularly for the neutral class), whereas truthfulness detection benefits from stronger lexical and stylistic cues. This prioritization ensures the model does not neglect the more challenging sentiment task during joint optimization. The results are shown in Table 2.

Table 2. Performance metrics for different values of (sentiment weight). The proposed value of $\alpha = 0.90$ achieves the best balance between sentiment and truth detection tasks.

α (sentiment weight)	Sentiment F1	Truth F1
0.50	0.661	0.863
0.70	0.688	0.857
0.90 (proposed)	0.707	0.851
1.00	0.712	0.753

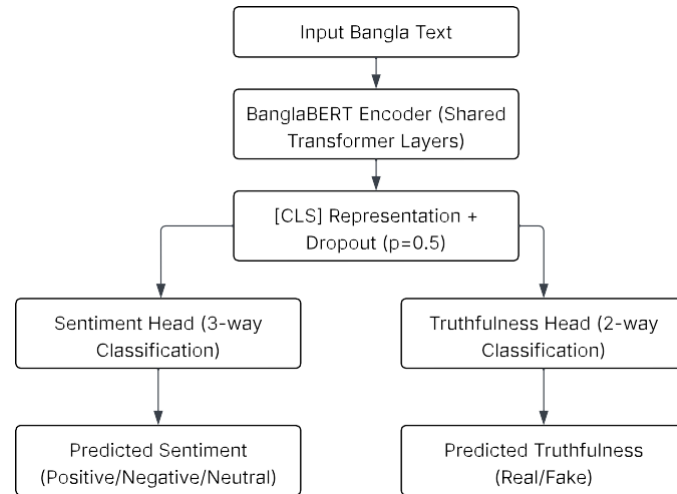


Fig. 5. BanglaBERT Multi-Task Architecture (Sentiment + Truthfulness)

3.5. Model Architecture

We implemented a multi-task learning system built on BanglaBERT [19], a BERT-based model pre-trained on 27.5 GB of Bangla text. The shared encoder generates context-aware representations. The architecture consists of a shared BanglaBERT encoder followed by two task-specific classification heads: a sentiment head (3-way classifier) and a truth head (2-way classifier), with dropout ($p=0.5$) applied to the [CLS] token representation to prevent overfitting. This dual-head configuration (Figure 5) enables parameter sharing across tasks.

While hard-parameter-sharing dual-head transformers are standard, our method incorporates three specific design choices: (i) a task-prioritized loss formulation (Eq. 1) with asymmetric weighting α that reflects the semantic primacy of sentiment framing in misinformation analysis; (ii) the integration of focal loss and weighted cross-entropy within a single MTL objective, which to our knowledge has not been combined for low-resource Bangla sentiment-truthfulness modeling; and (iii) a gradient-accumulation-stable optimization scheme tailored for constrained data scenarios. Our contribution is not in the encoder architecture itself, but in the training recipe and loss design adapted for low-resource Bangla. This targeted approach is essential because sentiment framing and emotional language are frequently weaponized in COVID-19 disinformation, and our loss design explicitly captures this task interdependency.

3.6. Training Setup

Experiments were conducted on two NVIDIA T4 GPUs (16GB each) using mixed-precision (FP16) for efficiency. Training details:

- **Optimizer:** AdamW with learning rate $lr = 2 \times 10^{-5}$ and weight decay = 0.01.
- **Scheduler:** Cosine decay with 10% warm-up.
- **Batch size:** 16, with gradient accumulation over 4 steps (effective batch size = 64).
- **Training duration:** Up to 15 epochs, with early stopping triggered if the sentiment macro-F1 score failed to improve by at least 0.005 over 2 consecutive epochs.

In practice, early stopping was triggered at epoch 4, yielding the best balance between convergence and generalization. Table 3 lists the key hyperparameters used during training, including the learning rate, batch size, and task weighting factor α .

Table 3. Hyperparameters

Parameter	Value	Purpose
Learning rate	2×10^{-5}	Optimization stability
Batch size	16	Memory constraints
α (sentiment weight)	0.90	Task prioritization
Dropout rate	0.5	Regularization
Gradient accumulation steps	4	Effective batch size
Warm-up percentage	10%	Scheduler warm-up

3.7. Evaluation Metrics

Models were evaluated using accuracy, macro-average precision, recall, and F1, as well as confusion matrices for detailed analysis of errors per class. Macro-averaging emphasized balanced performance across imbalanced classes.

To enhance reproducibility, we exported a complete error analysis table containing raw ground truth labels, predicted labels, mismatch indicators, and correctness indicators for each instance. This artifact ensured transparent reporting of model limitations and sources of error.

Accuracy was computed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Precision, recall, and F1 were defined for each class j :

$$Precision_j = \frac{TP_j}{TP_j + FP_j} \quad (3)$$

$$Recall_j = \frac{TP_j}{TP_j + FN_j} \quad (4)$$

where TP_j , FP_j , and FN_j denote the true positives, false positives, and false negatives for class j , respectively.

The macro-averaged F1 score was computed as:

$$F1_{macro} = \frac{1}{C} \sum_{j=1}^C \frac{2 \cdot Precision_j \cdot Recall_j}{Precision_j + Recall_j} \quad (5)$$

where C is the number of classes.

To ensure statistical transparency, 95% confidence intervals (CIs) for all reported accuracy and macro-F1 scores were computed via non-parametric bootstrapping. Specifically, we drew 1,000 resamples with replacement from the test set predictions and calculated the 2.5th and 97.5th percentiles of the resulting metric distributions. This approach avoids distributional assumptions, which is critical given the class imbalance in our evaluation set.

4. Results and Discussion

4.1. Overall Performance

Our multi-task BanglaBERT model, trained with class-weighted Focal Loss for sentiment classification and weighted cross-entropy for truthfulness classification, demonstrated strong performance on the held-out test set. The joint optimization framework, with sentiment prioritized at $\alpha = 0.90$, balances the demands of both tasks without significant degradation in either. Sentiment classification achieved 75.1% accuracy (95% CI: 73.8–76.4%) and macro F1 of 0.707 (95% CI: 0.692–0.722). Truthfulness detection reached 88.0% accuracy

(95% CI: 86.9–89.1%) and macro F1 of 0.851 (95% CI: 0.838–0.864).

Key Results

Sentiment Classification.

- **Macro F1:** 0.707, **Accuracy:** 75.1%
- **Negative:** F1 = 0.77, **Positive:** F1 = 0.78
- **Neutral:** Precision = 0.69, Recall = 0.48 (F1 = 0.57)

Performance on the majority classes (Negative, Positive) is consistently strong, while the underrepresented Neutral class achieves a precision of 0.69 – a substantial improvement compared to baseline configurations. The main limitation is neutral recall (0.48), reflecting both the semantic ambiguity of neutrality and its underrepresentation in the training set (14.4% of samples).

Truthfulness Detection.

- **Macro F1:** 0.851, **Accuracy:** 88.0%
- **Real:** F1 = 0.92, **Fake:** F1 = 0.79

The model distinguishes real from fake news effectively. The majority *real* class achieves near-perfect precision and recall, while the minority *fake* class still obtains a competitive F1 of 0.79, aided by inverse-frequency weighting that mitigates imbalance.

Interpretation

These findings validate our core hypothesis: task-specific loss weighting and prioritization significantly improve performance on semantically ambiguous and underrepresented categories. In particular:

- **Focal Loss with $\alpha_{\text{neutral}} = 1.5$** enhances sensitivity to difficult Neutral cases.
- **Joint-task prioritization ($\alpha = 0.90$)** sustains sentiment as the primary objective while preserving robust truthfulness classification.

Consequently, this approach improves the accuracy of the minority class without sacrificing the accuracy of the majority class or the secondary task. Additionally, the early stopping mechanism triggered at epoch 4 demonstrates fast convergence and stable optimization supported by mixed precision learning, cosine scheduling, gradient accumulation, and balanced loss design.

4.2. Ablation Study on Loss Strategies

We systematically evaluated five loss functions for sentiment classification (Table 4). The proposed Focal Loss with fixed class weights delivers the best performance (Accuracy: 0.75, Macro-F1: 0.71), outperforming plain Focal Loss (0.74, 0.69) and weighted cross-entropy variants. Standard cross-entropy performs worst (0.52, 0.43), though label smoothing and inverse-frequency weights improve it substantially (0.62, 0.58).

Notably, combining Focal Loss with inverse-frequency weights reduces performance slightly (0.70, 0.66), suggesting that focal modulation and frequency reweighting may interfere. Crucially, the proposed method achieves the highest Neutral F1 (0.57), underscoring its advantage in handling ambiguous, minority sentiment.

Table 4. Effect of Different Loss Strategies on Sentiment Classification

Loss Strategy	Accuracy	Macro-F1	Neg F1	Neu F1	Pos F1
Cross-Entropy (CE)	0.52	0.43	0.66	0.00	0.62
CE + Inverse Frequency Weights + Label Smoothing	0.62	0.58	0.66	0.41	0.67
Focal Loss	0.74	0.69	0.76	0.54	0.78
Focal Loss + Inverse Frequency Weights	0.70	0.66	0.72	0.48	0.78
Focal Loss + Fixed Class Weights*	0.75	0.71	0.77	0.57	0.78

* Proposed method: fixed class weights tuned to balance rare-class sensitivity.

4.3. Confusion Matrix Insights

The confusion matrices (Figures 6, 7) reveal class-specific error patterns.

Sentiment Classification. The model excels at recognizing Negative (82% accuracy) and Positive (77%) sentiments. The Neutral class, however, shows frequent misclassification: many neutral instances are labeled as Negative (167) or Positive (100), confirming its semantic ambiguity. Additionally, confusion between Negative and Positive (213 vs. 297 instances) suggests overlapping linguistic markers in polarized discourse.

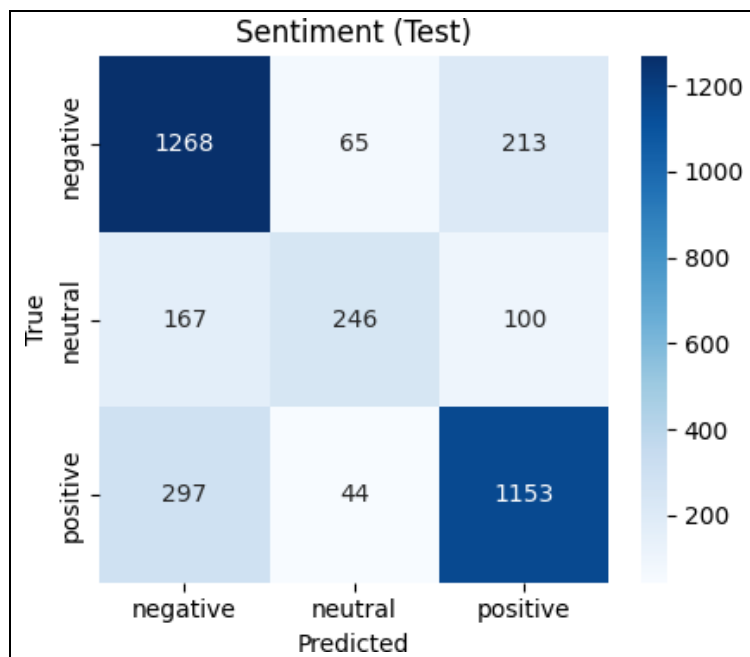


Fig. 6. Confusion matrix for the three-class sentiment-analysis model.

Truthfulness Classification. Real samples are identified with high reliability (90%), while Fake samples reach 82%. Errors primarily occur when genuine news employing critical or sensational language is flagged as fake (251 cases). In contrast, Fake → Real misclassifications (177) are less common, suggesting a conservative bias against false-positive fake predictions.

Together, these matrices highlight that the binary truthfulness classification task is handled more effectively than the ternary sentiment task, where neutrality remains the primary bottleneck.

4.4. Error Analysis

Manual inspection confirmed systematic weaknesses. Neutral instances with factual tones were frequently mislabeled as polarized. For example, "বাংলাদেশে আগামী ১ অক্টোবর থেকে ইংলিশ মিডিয়াম স্কুলের 'ও' এবং 'এ' লেভেলের পরীক্ষা নেওয়ার অনুমতি দিয়েছে শিক্ষা মন্ত্রণালয়" ("The Ministry of Education has permitted English-medium schools in Bangladesh to hold O- and A-level exams from October 1") was classified as positive.

Such errors highlight the inherent difficulty of classifying neutral content, which is often informational and lacks overt evaluative language, resulting in sparse discriminative cues (consistent with the *feature-sparsity hypothesis*). Furthermore, interpretation is highly context-dependent; stakeholders may perceive factual updates—such as the exam permission mentioned above—variably based on health or logistical concerns. This gradient polarity suggests that models trained solely on surface features miss nuances, while annotator subjectivity introduces noise (consistent with the *annotation noise hypothesis*). Because neutrality lies between positive and negative poles, subtle contextual shifts or subjective biases can easily subvert labels. Addressing this may require incorporating non-textual context or adopting a graded sentiment scale rather than a strict three-factor classification.

Similarly, descriptive statements such as "ফাইজার ভ্যাকসিনটি চীনে তৈরি করা হয়েছে" were predicted as negative, reflecting the blurred boundary between neutral and polarized discourse.

For truthfulness, real instances employing sensationalist language (e.g., official COVID-19 death counts in alarming press reports) were mislabeled as fake. Conversely, some fake claims with measured or factual tones were predicted as real, such as "এই বিশেষ মহামারী এমন এক যেখানে আমি মনে করি না যে দেশব্যাপী ২৫ বছরের কম বয়সী এক জনের মৃত্যু হয়েছে" (fake → real). These asymmetries suggest a stylistic bias: real news with dramatic framing is mistaken for fabricated content, while sophisticated disinformation mimics factual reporting. Two persistent challenges emerge: (1) separating neutral from weakly polarized sentiment, and (2) distinguishing factual but sensational real news from fabricated claims. Remedies may involve discourse-aware encoders or stylistic features that capture tone beyond factuality.

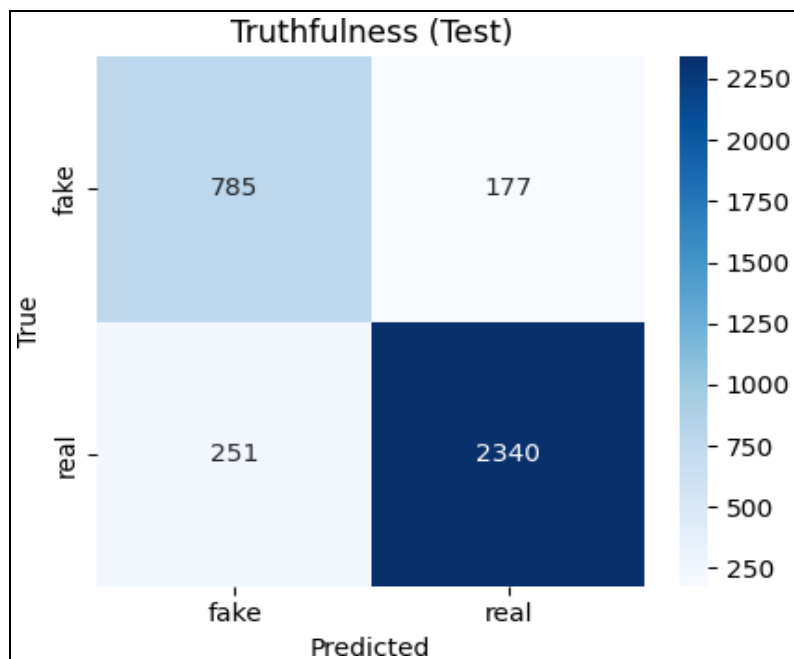


Fig. 7. Confusion matrix for the binary truthfulness classification model (fake vs. real).

4.5. Deployment via Gradio

To demonstrate practical utility, we developed a Gradio interface (Figure 8) for real-time inference. Users input raw Bangla text, which is preprocessed and analyzed to produce simultaneous predictions for sentiment and truthfulness classification. The interactive demo illustrates how our system can support monitoring of public opinion and misinformation during health crises.

4.6. Comparison with Previous Work

We compare our multi-task BanglaBERT-based framework with earlier research in Bangla sentiment analysis and fake news detection to contextualize our findings. Hossain et al. [8] introduced BanFakeNews, a foundational dataset where linear classifiers like SVM achieved strong baselines (91% F1 for fake news). Later multimodal approaches like MultiBanFakeDetect [7] further improved performance (79.69% accuracy) by fusing text and image modalities. More recently, Shibu et al. [9] demonstrated the potential of transformer-based techniques by using LLMs to improve low-resource fake news identification.

In terms of sentiment analysis, Kabir et al. [10] introduced BANGLABOOK, the largest Bangla sentiment dataset to date, while Maisha et al. [1] used conventional ML approaches on Bangla news articles and reported strong performance using Random Forest (99% accuracy). Mahmud [11] combined lexicon-based scoring with BanglaBERT to achieve 89% accuracy, surpassing BanglaBERT alone (79%), demonstrating the promise of hybrid approaches. Aspect-based sentiment analysis (ABSA) has been investigated in parallel work, such as TRF-BERT [12], which outperformed previous benchmarks in domain-specific ABSA tasks.

Direct numerical comparison with prior work is not directly applicable because of differences in datasets, label schemas, and evaluation protocols; Table 5 is therefore provided for contextual orientation rather than as a direct benchmark. Our framework offers three key advantages: First, the framework captures the interaction between sentiment framing and false information by simultaneously addressing sentiment and truthfulness classification inside a single model, in contrast to unimodal or single-task techniques. Second, it improves minority class recognition by directly addressing the ongoing issue of class imbalance through the use of inverse-frequency class weights and focal loss [24]. Finally, we fill a crucial gap left by previous resources like BanFakeNews [8] and BANGLABOOK [10], which tackled each task separately. We offer a curated COVID-19 dataset of over 35,000 Bangla social media posts and news articles annotated for both sentiment and truthfulness classification.

BD Bangla Sentiment and Fake News Detector for COVID-19 Related Content

Model: Fine-tuned BanglaBERT for Multi-Task Learning. Enter a piece of Bengali text, click **Get Prediction**, and get instant predictions for:

- Sentiment** (Positive, Negative, Neutral)
- Truthfulness** (Real, Fake)

Enter Bengali Text

ভিটামিন সি, জিংক করোনা সারাতে সাহায্য করে।

Predicted Sentiment: POSITIVE

Predicted Truthfulness: FAKE

Flag

Clear **Get Prediction**

Examples

করোনা ভাইরাস নিয়ে সরকারের পদক্ষেপ অত্যন্ত প্রশংসনীয়। এই ভ্যাকসিনে মানুষ মারা যাচ্ছে, সরকার লুকছে তথ্য। লকডাউনের কারণে অনেকের চাকরি চলে গেছে, কিন্তু এটা প্রয়োজনীয় ছিল।

Use via API - Built with Gradio - Settings

Fig. 8. Gradio demo of Bangla Sentiment and Fake News Detector for COVID-19 Related Content.

Our model achieves 75% accuracy (macro F1: 0.70) for sentiment classification and 88% accuracy (macro F1: 0.85) for truthfulness classification, as summarized in Table 5, setting a strong benchmark that outperforms most previous unimodal baselines and demonstrates the usefulness of multi-task transformer models for Bangla NLP.

4.7. Contribution Summary

This work establishes a new benchmark for multi-task learning and fake news detection in Bangla COVID-19 textual content, spanning both social media and news domains. By jointly modeling sentiment nuances and truthfulness, the system provides a foundation for deeper error analysis and practical deployment.

5. Conclusion

The COVID-19 pandemic highlighted the dual challenges of understanding public opinion and mitigating the rapid spread of misinformation, particularly in under-resourced languages like Bangla. In this study, we addressed both issues simultaneously by proposing a multi-task BanglaBERT framework to perform joint sentiment and truthfulness classification of COVID-19-related Bangla content. Our approach introduces a new annotated dataset of over 35,000 textual samples, incorporates class-weighted Focal Loss and weighted cross-entropy to address imbalance, and demonstrates the feasibility of deployment via a Gradio interface. We release our code on GitHub and the trained model on Hugging Face to support reproducibility and the Bangla NLP community. This work aims to facilitate further research in low-resource language processing.

Experimental results confirm that the proposed model achieves 75% accuracy in sentiment classification (Macro F1: 0.70) and 88% accuracy in truthfulness classification (Macro F1: 0.85), outperforming or matching state-of-the-art baselines in Bangla NLP. However, detailed analysis, including ablation studies, confusion matrices, and error analysis, reveals persistent challenges. Specifically, the semantic ambiguity of neutral sentiment and the stylistic overlap between real sensational news and fabricated content remain significant challenges. These findings provide clear directions for future research, such as incorporating situational sentiment modeling, discourse-level features, or multimodal integration to capture nuances missed by text-only models.

Despite these strengths, certain limitations must be acknowledged. Although the dataset is large by Bangla standards, it remains modest compared to high-resource languages, which may limit generalizability. Furthermore, the current model does not utilize large-context windows or multimodal features (e.g., images or videos).

Table 5. Comparison of Performance with Previous Work in Bangla Sentiment Analysis and Fake News Detection

Study	Task	Dataset	Model/Approach	Accuracy	F1-score
Hossain et al. (2020) [8]	Fake News Detection	BanFakeNews (48k fake real, 1.3k)	SVM with linguistic + metadata features	–	0.91 (fake)
Shibu et al. (2025) [9]	Fake News Detection	BanFakeNews-2.0 (60k news)	LLMs with Quantized LoRA (BLOOM, mBERT-unc)	–	0.89
Faria et al. (2025) [7]	Multimodal Fake News Detection	MultiBanFakeDetect (9.6k text-image pairs)	MultiFusionFake (DenseNet-169 + mBERT, Early Fusion)	0.7969	–
Maisha et al. [1] (2021)	Sentiment Analysis	Bangla newspaper articles (2-class)	Traditional ML (Random Forest, SVM)	0.99 (RF)	0.80 (SVM)
Kabir et al. [10] (2023)	Sentiment Analysis	BANGLABOOK (158k book reviews)	BoW, n-grams + ML classifiers	–	–
Mahmud & Mahmud (2024) [11]	Sentiment Analysis	15k reviews, 9 sentiment classes	BSPS (lexicon) + BanglaBERT (hybrid)	0.89	–
Ahmed et al. [12] (2024)	Aspect-Based Sentiment Analysis	Bangla Cricket + datasets Restaurant	TRF-BERT (BERT + Random Forest)	0.89–0.92	0.85–0.89
This Work (2025)	Multi-Task Sentiment + Truthfulness Classification	35k+ COVID-19 social media posts and news articles	Dual-head BanglaBERT + class weighting + focal loss	0.75 (sentiment), 0.88 (truthfulness)	0.70 (sentiment), 0.85 (truthfulness)

Finally, our assessment is limited to the discourse surrounding COVID-19. To ensure broader resilience, future work should focus on extending this multi-task framework to other domains and crisis situations.

This study establishes a strong benchmark for future research by jointly identifying sentiment and misinformation in Bangla, a globally spoken but under-resourced language. The Gradio interface highlights the system’s practical utility for fact-checking and monitoring public discourse during crises.

We acknowledge several limitations in this study that present opportunities for future research. First, the semi-automated annotation process and the focus on a single domain (COVID-19) may limit generalization to external or independent benchmark datasets; future work should evaluate this framework on diverse, cross-domain Bangla corpora. Second, while we addressed class imbalance via loss weighting, more advanced techniques such as synthetic minority oversampling or threshold-moving could be explored. Third, our ablation study focused primarily on loss functions; future iterations should include comprehensive ablations of preprocessing pipelines, single-task baselines, and encoder variants. Fourth, our error analysis remains largely qualitative; systematic quantitative categorization of failure modes (e.g., stylistic vs. factual errors) is warranted. Finally, while the Gradio interface demonstrates practical utility, future deployment studies must rigorously evaluate inference latency, scalability, and robustness under domain shift. Addressing these limitations will be critical for advancing multi-task NLP in low-resource settings.

All the Declarations and Statements

Author Contributions Statement

Arshadul Hoque – The author is the sole contributor to this work. He was responsible for all aspects of the research, including study conception and design, methodology, software development, data collection and curation, formal analysis, investigation, and manuscript preparation.

Conflict of Interest Statement

The author has no competing interests to disclose.

Funding Declaration

This work was self-funded and received no external financial support.

Data Availability Statement

The datasets used during the current study are available in GitHub.

Ethical Declarations

The researcher affirms that this work adhered to all ethical standards. The study involved no direct participation from humans or animals; any datasets employed were fully anonymized and used strictly for academic purposes. Additionally, the author verifies that the manuscript is an original piece of work, has not been published before, and is not currently being considered for publication elsewhere.

Acknowledgments

None

Declaration of Generative AI in Scholarly Writing

The author used GLM 4.5 (Zhipu AI) for language editing to improve clarity, grammar, and consistency. The AI tool was not involved in conceptualization, study design, data collection, data analysis, interpretation of results, or the decision to publish. All intellectual contributions, technical decisions, analyses, and final revisions were performed solely by the author.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation - Full Form

NLP – Natural Language Processing

FND – Fake News Detection

SA – Sentiment Analysis

GPU - Graphics Processing Unit

ML - Machine Learning

TP - True Positive

TN - True Negative

FP - False Positive

FN - False Negative

Appendix A/B/C..., with appendix title

Appendix A. Experimental Computing Environment

Parameter	Specification
Implementation Platform	Google Colab
GPU Used	NVIDIA T4 GPU
GPU Memory	16 GB
Programming Language	Python
Deep Learning Framework	Pytorch
Optimization Algorithm	AdamW

References

- [1] Sabrina Jahan Maisha, Nuren nafisa, and Abdul Kadar Muhammad Masum. Supervised machine learning algorithms for sentiment analysis of bangla newspaper. International Journal of Innovative Computing, 2021. <https://doi.org/10.11113/ijic.v11n2.321>
- [2] Md. Atikur Rahman and Emon Kumar Dey. Datasets for aspect-based sentiment analysis in bangla and its baseline evaluation. Data, 3(2), 2018. <https://doi.org/10.3390/data3020015>
- [3] Salim Sazzed and Sampath Jayarathna. A sentiment classification in bengali and machine translated english corpus. In 2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), pages 107–114, 2019. <https://doi.org/10.1109/iri.2019.00029>
- [4] Mohammad Nazmush Shamael, Sabila Nawshin, Swakkhar Shatabda, and Salekul Islam. Banglishrev: A large-scale bangla-english and code-mixed dataset of product reviews in e-commerce. arXiv preprint arXiv:2412.13161, 2024. <https://doi.org/10.48550/arXiv.2412.13161>
- [5] Md. Zahin Hossain George, Naimul Hossain, Md. Rafiuzzaman Bhuiyan, Abu Kaisar Mohammad Masum, and Sheikh Abujar. Bangla fake news detection based on multichannel combined cnn-lstm. International Conference on Computing Communication and Networking Technologies, 2021. <https://doi.org/10.2139/ssrn.5190043>

- [6] Md. Tanvir Rouf Shawon, G. M. Shahariar, Faisal Muhammad Shah, Mohammad Shafiul Alam, and Md. Shahriar Mahub. Bengali fake review detection using semi-supervised generative adversarial networks. *ICON*, 2023. <https://doi.org/10.1109/iconlp58431.2023.00011>
- [7] Fatema Tuj Johora Faria, Mukaffi Bin Moin, Zayeed Hasan, Md. Arafat Alam Khandaker, Niful Islam, Khan Md. Hasib, and M. F. Mridha. Multibanfakedetect: Integrating advanced fusion techniques for multimodal detection of bangla fake news in under-resourced contexts. *Int. J. Inf. Manag. Data Insights*, 2025. <https://doi.org/10.1016/j.jjime.2025.100347>
- [8] Zobaer Hossain, Ashrafur Rahman, Saiful Islam, and Sudipta Kar. Banfakenews: A dataset for detecting fake news in bangla. *arXiv: Computation and Language*, 2020. <https://doi.org/10.48550/arXiv.2004.08789>
- [9] Hrithik Majumdar Shibu, Shrestha Datta, Md. Sumon Miah, Nasrullah Sami, Mahruba Sharmin Chowdhury, and Md Saiful Islam. From scarcity to capability: Empowering fake news detection in low-resource languages with LLMs. In Ruwan Weerasinghe, Isuri Anuradha, and Deshan Sumanathilaka, editors, *Proceedings of the First Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages*, pages 100–107, Abu Dhabi, January 2025. Association for Computational Linguistics. <https://doi.org/10.48550/arXiv.2501.09604>
- [10] Mohsinul Kabir, Obayed Bin Mahfuz, Hasan Mahmud, and Md. Kamrul Hasan. Banglabook: A large-scale bangla dataset for sentiment analysis from book reviews. *Annual Meeting of the Association for Computational Linguistics*, 2023. <https://doi.org/10.18653/v1/2023.findings-acl.80>
- [11] Hemal Mahmud and Hasan Mahmud. Enhancing sentiment analysis in bengali texts: A hybrid approach using lexicon-based algorithm and pretrained language model bangla-bert. *arXiv.org*, 2024. <https://doi.org/10.48550/arXiv.2411.19584>
- [12] Shihab Ahmed, Moythry Manir Samia, Maksuda Haider Sayma, M. Kabir, and M. F. Mridha. Trf-bert: A transformative approach to aspect-based sentiment analysis in the bengali language. *PLoS ONE*, 2024. <https://doi.org/10.1371/journal.pone.0308050>
- [13] Shijie Chen, Yu Zhang, and Qiang Yang. Multi-task learning in natural language processing: An overview. *ACM Computing Surveys*, 2021. <https://doi.org/10.48550/arXiv.2109.09138>
- [14] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017. <https://doi.org/10.48550/arXiv.1706.05098>
- [15] Lovre Torbarina, Tin Ferkovic, Łukasz Roguski, Velimir Mihelčić, Bruno Šarlija, and Željko Kraljević. Challenges and opportunities of using transformer-based multi-task learning in nlp through ml lifecycle: A survey. *arXiv.org*, 2023. <https://doi.org/10.48550/arXiv.2308.08234>
- [16] Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. Multi-task deep neural networks for natural language understanding. *Annual Meeting of the Association for Computational Linguistics*, 2019. <https://doi.org/10.48550/arXiv.1901.11504>
- [17] Yukang Xie, Chengyu Wang, Junbing Yan, Jiyong Zhou, Feiqi Deng, and Jun Huang. Making small language models better multi-task learners with mixture-of-task-adapters. *Web Search and Data Mining*, Pages 1094 – 1097, 2023. <https://doi.org/10.1145/3616855.3635690>
- [18] Jonathan Pilault, Amine Elhattami, Christopher Pal, and Chris Pal. Conditionally adaptive multi-task learning: Improving transfer learning in NLP using fewer parameters & less data. *arXiv: Learning*, 2020. <https://doi.org/10.48550/arXiv.2009.09139>
- [19] Abhik Bhattacharjee, Tahmid Hasan, Wasi Ahmad, Kazi Samin Mubasshir, Md Saiful Islam, Anindya Iqbal, M. Sohel Rahman, and Rifat Shahriyar. BanglaBERT: Language model pretraining and benchmarks for low-resource language understanding evaluation in Bangla. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1318–1327, Seattle, United States, July 2022. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.findings-naacl.98>
- [20] Tahmid Hasan, Abhik Bhattacharjee, Kazi Samin, Masum Hasan, Madhusudan Basak, M. Sohel Rahman, and Rifat Shahriyar. Not low-resource anymore: Aligner ensembling, batch filtering, and new datasets for Bengali-English machine translation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2612–2623, Online, November 2020. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-main.207>
- [21] Arshadul Hoque. Fine-tuned banglabert for bengali sentiment analysis. <https://huggingface.co/ahs95/banglabert-sentiment-analysis>, 2025
- [22] Enam Biswas. Bangla newspaper dataset - ebd, 2021. <https://doi.org/10.34740/kaggle/dsv/3563095>
- [23] Md Ekramul Islam, Labib Chowdhury, Faisal Ahamed Khan, Shazzad Hossain, Md Sourave Hossain, Mohammad Mamun or Rashid, Nabeel Mohammed, and Mohammad Ruhul Amin. Sentigold: A large bangla gold standard multidomain sentiment analysis dataset and its evaluation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4207–4218, 2023. <https://doi.org/10.1145/3580305.3599904>
- [24] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–327, 2020. <https://doi.org/10.1109/iccv.2017.324>
- [25] Fahmida Khanam, Anik Chakraborty, Md. Ahsan Habib, and Md. Sadiq Iqbal. Bangla sentiment analysis on highly imbalanced data using hybrid cnn-lstm & bangla bert. *2024 3rd International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, 2024. <https://doi.org/10.1109/icaeee62219.2024.10561678>
- [26] Matteo Bodini and Matteo Bodini. Opinion mining from machine translated Bangla reviews with stacked contractive autoencoders. *Journal of Ambient Intelligence and Humanized Computing*, 2022. <https://doi.org/10.1007/s12652-022-03760-w>
- [27] Sheikh Sadi Bandan, Md Sharuf Hossain., MD. Samiul Islam Sabbir, and Khadiza Tul Kobra. Deep learning for bengali fake news detection: Innovative approaches for accurate classification. *International journal of research and innovation in applied science*, 393-403, 2024. <https://doi.org/10.51584/ijrias.2024.908034>
- [28] Md. Yasmi Tohabar, Nahiyah Nasrah, and Asif Mohammed Samir. Bengali fake news detection using machine learning and effectiveness of sentiment as a feature. *International Conference on Informatics, Electronics and Vision*, 2021. <https://doi.org/10.1109/icievicivpr52578.2021.9564138>
- [29] S. Majumder and Dipankar Das. Studies towards language independent fake news detection. *ICON*, 439-446, 2021.
- [30] Tanjina Sultana Camelia, Faizur Rahman Fahim, and M. Anwar. A regularized lstm method for detecting fake news articles. *2024 IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON)*, 2024. <https://doi.org/10.1109/spiccon64195.2024.10941441>

- [31] Pronoy K. Mondal, Sadman Sadik Khan, Md. Masud Rana, Shahriar Sultan Ramit, A. Sattar, and Md. Sadekur Rahman. Breaking the fake news barrier: Deep learning approaches in bangla language. International Conference on Computing Communication and Networking Technologies, 2024. <https://doi.org/10.1109/iccncnt61001.2024.10725195>
- [32] Sheetal Harris, Hassan Jalil Hadi, Naveed Ahmad, and M. Alshara. Fake news detection revisited: An extensive review of theoretical frameworks, dataset assessments, model constraints, and forward-looking research agendas. Technologies, 2024. <https://doi.org/10.3390/technologies12110222>
- [33] Quazi Adibur Rahman Adib, Md. Humaion Kabir Mehedi, Md. Sadman Sakib, Kabbya Kantam Patwary, Sabbir Hossain, and Annajiat Alim Rasel. A deep hybrid learning approach to detect bangla fake news. 2021 5th International Symposium on Multi-disciplinary Studies and Innovative Technologies (ISMSIT), 2021. <https://doi.org/10.1109/ismsit52890.2021.9604712>
- [34] Rich Caruana. Multitask learning. Machine learning, 28(1):41–75, 1997. https://doi.org/10.1007/978-1-4615-5529-2_5
- [35] Jinlan Fu, See-Kiong Ng, and Pengfei Liu. Polyglot prompt: Multilingual multitask prompt training. Conference on Empirical Methods in Natural Language Processing, 2022. <https://doi.org/10.18653/v1/2022.emnlp-main.674>
- [36] Yiren Wang, ChengXiang Zhai, and Hany Hassan Awadalla. Multi-task learning for multilingual neural machine translation. arXiv: Computation and Language, 2020. <https://doi.org/10.18653/v1/2020.emnlp-main.75>
- [37] Jing Lü, Jiasen Lu, Vedanuj Goswami, Marcus Rohrbach, Devi Parikh, and Stefan Lee. 12-in-1: Multi-task vision and language representation learning. arXiv: Computer Vision and Pattern Recognition, 2019. <https://doi.org/10.48550/arXiv.1912.02315>
- [38] Chae-Gyun Lim, Young-Seob Jeong, and Ho-Jin Choi. Multi-task learning approach for utilizing temporal relations in natural language understanding tasks. Scientific Reports, 2023. <https://doi.org/10.1038/s41598-023-35009-7>
- [39] Amirhossein Farzam, Shashank Shekhar, Isaac D. Mehlhaff, and Marco Morucci. Multi-task learning improves performance in deep argument mining models. Workshop on Argument Mining, 2023. <https://doi.org/10.18653/v1/2024.argmining-1.5>
- [40] Clara Vania, Yova Kementchedjhieva, Anders Søgaard, and Adam Lopez. Edinburgh research explorer a systematic comparison of methods for low-resource dependency parsing on genuinely low-resource languages. 1105–1116, 2019. <https://doi.org/10.18653/v1/d19-1102>
- [41] Iker García-Ferrero. Cross-lingual transfer for low-resource natural language processing. arXiv.org, 2025. <https://doi.org/10.48550/arXiv.2502.02722>
- [42] Xiaobo Liang, Robert Mao, Lijun Wu, Juntao Li, Min Zhang, and Qing Li. Enhancing low-resource nlp by consistency training with data and model permutations. IEEE/ACM Transactions on Audio Speech and Language Processing, 189 – 199, 2024. <https://doi.org/10.1109/taslp.2023.3325970>
- [43] Jeffrey Otoihi, Oduguwa Damilola, and Okpare David. Sabiyarn: Advancing low resource languages with multitask nlp pre-training. Proceedings of the Sixth Workshop on African Natural Language Processing (AfricaNLP 2025), 95–107, 2025. <https://doi.org/10.18653/v1/2025.africanlp-1.14>
- [44] Michael A. Hedderich, Lukas Lange, Heike Adel, Jannik Strötgen, and Dietrich Klakow. A survey on recent approaches for natural language processing in low-resource scenarios. arXiv: Computation and Language, 2545–2568, 2020. <https://doi.org/10.18653/v1/2021.naacl-main.201>
- [45] Emily Chang and Nada Basit. How many words does it take to understand a low-resource language? North American Chapter of the Association for Computational Linguistics, 207–224, 2025. <https://doi.org/10.18653/v1/2025.naacl-srw.21>
- [46] Kumar Saunack, Saurav Kumar, and Pushpak Bhattacharyya. How low is too low? a monolingual take on lemmatisation in indian languages. North American Chapter of the Association for Computational Linguistics, 4088–4094, 2021. <https://doi.org/10.18653/v1/2021.naacl-main.322>
- [47] Yudong Wang, Chao Ma, Qingxiu Dong, Lingpeng Kong, and Jun Xu. A challenging benchmark for low-resource learning. arXiv.org, 2057–2080, 2023. <https://doi.org/10.18653/v1/2024.findings-acl.123>
- [48] Ping Yan and Jiangxu Wu. Low-resource named entity recognition based on multi-hop dependency trigger. China National Conference on Chinese Computational Linguistics, 325–334, 2022. https://doi.org/10.1007/978-3-031-18315-7_21

Authors’ Profiles



Arshadul Hoque is a graduate in Computer Science and Engineering from the University of Chittagong. His research focuses on natural language processing, machine learning, and low-resource language modeling, with an emphasis on Bangla text classification. He investigates transformer-based architectures, such as Bangla BERT, to tackle challenges like class imbalance, linguistic ambiguity, and the scarcity of annotated data in Bangla language processing. In addition to the current publication, another research article is pending acceptance in a Q-ranked journal.

Professionally, he works as a Business Analyst and Prompt Engineer at Cobalt Partners, a Ghana-based advisory firm, where he contributes to the design and refinement of AI-assisted prompts for business process specifications, user stories, and clarification logs. His work has led to measurable improvements in specification accuracy and stakeholder alignment.

He actively participates in data science competitions on Kaggle to further strengthen his applied machine learning expertise and is preparing to pursue graduate studies in Data Science and Artificial Intelligence, aiming to advance research at the intersection of NLP and real-world, impact-driven applications.

How to cite this paper

Arshadul Hoque, “Multi-Task BanglaBERT for Joint Sentiment and Fake News Detection in COVID-19 Discourse”, International Journal of Education and Management Engineering (IJEME), Vol.16, No.4, pp. 29-44, 2026. DOI:10.5815/ijeme.2026.04.03.