Hybrid Classical-Quantum Framework for Sentiment Classification and Claim Check-Worthiness Identification in Bengali

Pritam Pal[⋆], **Dipankar Das**[⋆], **Anup Kumar Kolya**[†] **and Siddhartha Bhattacharyya**[⋄]
[⋆]Jadavpur University, Kolkata, India
[†]RCC Institute of Information Technology, Kolkata, India
[⋄]VSB Technical University of Ostrava, Ostrava, Czech Republic
{pritampal522, dipankar.dipnil2005, anup.kolya, dr.siddhartha.bhattacharyya}@gmail.com

Abstract

Traditional machine learning and deep learning models have demonstrated remarkable performance across various NLP tasks in multiple languages. However, these conventional models often struggle with languages with complex linguistic structures and nuanced contexts, such as Bengali. Recent advancements in quantum computing offer promising solutions for tackling complex, computationally challenging problems, providing faster, more efficient processing than classical systems. This research aims to address the challenges posed by the intricate linguistic structure of the less-resourced Bengali language by developing a quantum-enhanced framework for sentiment classification and claim-checkworthiness identification. We created a classical LSTM framework and proposed novel 2-qubit and 4-qubit classical-quantum frameworks, evaluating their effectiveness for sentiment classification and claim-checkworthiness identification tasks in Bengali. An entirely new dataset comprising ≈3K samples was developed by curating Bengali news headlines from prominent sources. We tagged these headlines with sentiment and claim checkworthy labels using state-of-the-art LLMs. Our findings indicate that the quantum-enhanced frameworks outperform the traditional models in both tasks. Notably, the 4-qubit-based framework achieved the highest F1-score in sentiment classification, while the 2-qubit-based framework demonstrated the best F1-score in claim checkworthiness identification.

1 Introduction

The rapid growth of information on the internet has intensified the challenges of processing and analyzing natural languages at a scale. Two critical tasks in this domain are sentiment analysis, which identifies the emotional tone intended in a sentence as positive, negative, or neutral, and claim checkworthiness identification, which determines whether a sentence constitutes a checkworthy claim or not, facilitating further fact-checking to mitigate misinformation and disinformation. While classical deep learning approaches have achieved remarkable performance in both tasks, their performance in a quantum-computing environment has not been broadly explored.

Quantum Computing, a trending and emerging topic in the computer science domain, is based on the fundamentals of quantum physics, such as superposition and entanglement (Gyongyosi and Imre, 2019). Due to the superposition, entanglement, and other unique characteristics, quantum computers can solve problems more efficiently than classical computers by speeding up computational time with less resource utilization (Gyongyosi and Imre, 2019; Pandey and Pakray, 2023). One of the best examples for assessing the power of quantum computing is the breaking of the famous Rivest–Shamir–Adleman (RSA) algorithm (Rivest et al., 1978). To break the RSA algorithm generally, a classical computer takes billions of years; however, a quantum computer takes only a few hours to break the RSA algorithm (Shor, 1997; Proos and Zalka, 2004).

One primary application of quantum computing is Quantum Machine Learning (QML). Where classical computers require a large amount of data and enormous computational resources, quantum computers could learn from less data, understand complex patterns in data, and handle noisy data in a better way than classical computers (Neumann et al., 2019). These advantages of quantum computing inspire us to analyze NLP tasks, such as sentiment classification and claim checkworthiness identification, using QML methods, particularly in less-resourced languages like Bengali.

The seventh most widely spoken language globally, Bengali represents over 272 million speakers, with a majority portion in India and Bangladesh, yet remains significantly underrepresented in the natural language processing research community

Proceedings of the QuantumNLP: Integrating Quantum Computing with Natural Language Processing, pages 10–19
November 24, 2025

compared to high-resource languages like English. This disparity becomes particularly pronounced when addressing sophisticated computational tasks such as sentiment classification or claim checkworthiness identification, where the linguistic complexity and contextual nuances of Bengali pose substantial challenges for traditional machine learning approaches.

Our research addresses the linguistic challenges of the Bengali language and develops a novel classical-quantum hybrid framework for sentiment classification and claim checkworthiness detection in Bengali texts. The contributions in this paper can be summarized as follows:

- We have developed an entirely new Bengali dataset for claim checkworthiness detection and sentiment classification, with a sample size of approximately 3,000, curating data from the prominent Bengali news portal 'Sangbad Pratidin', and annotating sentiment and claim labels using three state-of-the-art Large Language Models (LLMs): GPT-4o-mini (OpenAI et al., 2024), Llama-4 (Touvron et al., 2023), and GPT-4.1-mini, followed by majority voting.[1]
- We developed a classical LSTM framework and a classical-quantum hybrid framework using Variational Quantum Circuit (VQC) for sentiment classification and claim checkworthiness identification.
- We perform comparative analysis between classical LSTM and classical-quantum hybrid frameworks for both sentiment classification and claim checkworthiness detection, providing valuable insights into their performance.

The remainder of this paper is organized as follows: Section 2 presents the related work, providing an overview of recent studies in the field of quantum NLP. In Section 3, we discuss our data collection strategy, the process of dataset preparation using LLMs, and the analysis of inter-annotator agreement. Section 4 covers the methodologies for developing both classical LSTM models and classical-quantum hybrid frameworks utilizing VQCs, along with the training hyperparameters. Section 5 presents the results, discussing the outcomes of different frameworks for both tasks. Finally, Section 6 concludes the paper by outlining the valuable findings from this research and suggesting future directions for this work.

[1]The dataset is publicly available at: https://github.com/pritampal98/quantum-sentiment-claim

2 Related Works

In this section, we discuss some related works on quantum computing in the natural language processing (NLP) domain. Although quantum computing has been a hot topic for the last decade, the applications of quantum computing in the NLP domain have not been extensively explored and are still in the early stages (Varmantchaonala et al., 2024).

Basile and Tamburini (2017) proposed quantum language models using quantum probability theory. The authors have shown that their proposed quantum language models outperform the state-of-the-art language models in terms of perplexity scores. Tamburini (2019) also used quantum probability theory for developing a word sense disambiguation algorithm.

A joint multi-modal multi-task learning framework for sentiment and sarcasm detection using quantum probability was proposed by Liu et al. (2021). The authors evaluated their proposed framework on two datasets, $\text{MUStARD}_{\text{ext}}$ (Chauhan et al., 2020) and Memotion (Sharma et al., 2020), demonstrating that its performance surpasses that of the state-of-the-art. Phukan and Ekbal (2023) proposed a multimodal framework for sentiment analysis using a variational quantum circuit (VQC) (Qi et al., 2021). The authors have also demonstrated that their framework outperforms other frameworks for the CMU-MOSEI dataset (Bagher Zadeh et al., 2018). A multimodal quantum-based framework for emotion detection was also explored in the study by Li et al. (2023).

One of the popular NLP tasks, part-of-speech (POS) tagging, was also explored by several researchers (Sipio et al., 2021; Pandey et al., 2022; Pandey and Pakray, 2023) utilizing QLSTM. While Sipio et al. (2021) and Pandey et al. (2022) worked with unidirectional QLSTM, Pandey and Pakray (2023) used bidirectional QLSTM in their study to identify POS tags in a text. In contrast, Pandey et al. (2022) used the Mizo language, which is a low-resourced Indian language, and Pandey and Pakray (2023) used codemixed texts in their experiments.

Quantum frameworks are also explored in the domain of text classification (Xu et al., 2024; Shi et al., 2023), sentiment analysis (Yan et al., 2024; Zhang et al., 2019), sarcasm detection, claim identification (Pal and Das, 2025), and metaphor detection (Qiao et al., 2024) tasks. While Xu et al. (2024) used quantum RNNs to develop their text classification framework and evaluated their models in the Rotten Tomatoes dataset (Pang and Lee, 2005), Shi et al. (2023) developed quantum-inspired convolution neural network-based models and evaluated their models on popular benchmark datasets such as SST, SUBJ, MPQA, etc.

Coecke et al. (2020) proposed 'DisCoCat', a quantum framework for NLP tasks that preserves the linguistic meaning and structure of a text and converts them into a quantum circuit. The applications of the DisCoCat framework are shown in the papers (Ruskanda et al., 2023, 2022; Ganguly et al., 2022) where the authors performed sentiment analysis using the DisCoCat framework with 'lambeq'(Kartsaklis et al., 2021) toolkit. 'Lambeq'[2] is the first open-source Python library for quantum natural language processing, which provides a vast range of modules and classes to develop quantum circuits for sentence representation, training of quantum circuits, and many others.

There are several survey papers (Wu et al., 2021; Guarasci et al., 2022; Varmantchaonala et al., 2024; Widdows et al., 2024) that discussed quantum natural language processing and its applications in a more elaborate and detailed way. Among them, one of the interesting articles proposed by Wu et al. (2021) discusses and categorizes different quantum algorithms and NLP tasks, showing that quantum NLP models produce better or equivalent results than classical NLP models.

3 Dataset

A completely new dataset was developed for this experiment with sentiment and claim checkworthy labels. The data was collected from news headlines from one of the popular and prominent Bengali news portals, 'Sangbad Pratidin' [3]. We utilized Python's BeautifulSoup web-scraping method to systematically scrape news headlines and store them in an Excel file. Following the collection of data, the entire crawled data was reviewed by the authors to check for inconsistent entries, such as HTML tags or undefined Unicode characters, and the texts were manually cleaned.

Upon collection of data, the news headlines were annotated with claim checkworthy and sentiment labels using Large Language Models (LLMs). Due to their state-of-the-art performance across various NLP tasks, including question answering, machine translation, and classification tasks, we employed three distinct LLMs for the data annotation task: GPT-4o-mini, GPT-4.1-mini, and Llama-4. It is evident that annotation data with professional human annotation is always of high quality and provides gold-standard annotated labels. However, the human annotation requires specialized training, significant annotation costs, and time. Also, in the context of resource-constrained languages, such as Bengali, finding quality professional data annotators is quite challenging. The following prompt was provided to each LLM model to annotate the claim and sentiment labels:

```
You are a language expert annotating Bengali news headlines.

Now classify the sentiment of news headline as:
- Positive: Expresses praise, hope, success, happiness, or celebration
- Negative: Expresses criticism, fear, conflict, danger, sadness, or loss
- Neutral: Factual or informational, without emotional tone

Then decide if the headline is check-worthy:
- Check-worthy: A verifiable claim with potential public impact
- Not check-worthy: Opinion-based, vague, or unverifiable

Output Format: ["<Positive|Negative|Neutral>", "<Check-worthy|Not check-worthy>", "<A brief justification in English enclosed with quotation>"]

Now annotate the headline: "{txt}"
```

All the LLM models were accessed through their corresponding APIs, and the temperature and top-p values were set to 0 and 0.95, respectively.

Upon annotating claim and sentiment labels with three distinct LLMs, the final annotation was carried out through a majority voting scheme. For both sentiment and claim checkworthy labels, the label with the most frequent outcome was selected as the final label. The annotations, where no majority was found, were further annotated manually by the annotators. The inter-annotation agreement score between different LLMs was calculated using Fleiss' Kappa (Fleiss, 1971) and Gwet AC1 (Gwet, 2006) metrics. For the sentiment label annotation, the Fleiss' Kappa score was 0.7751, and the Gwet AC1 score was 0.8209. In case of claim checkworthiness, the Fleiss' Kappa and Gwet AC1 scores were 0.3554 and 0.6516, respectively.

However, instead of fully relying on LLM anno-

[2]https://docs.quantinuum.com/lambeq/
[3]https://www.sangbadpratidin.in/

tating data, all the final annotations (after majority voting selection) were further reviewed through a rigorous review process by three undergraduate computer science interns. If any inconsistencies were found, those are marked by the interns and further reviewed and resolved by the authors. A complete flow diagram of the overall data annotation process is provided in Figure 1. The distribution of sentiment and claim labels is provided in Table 1.

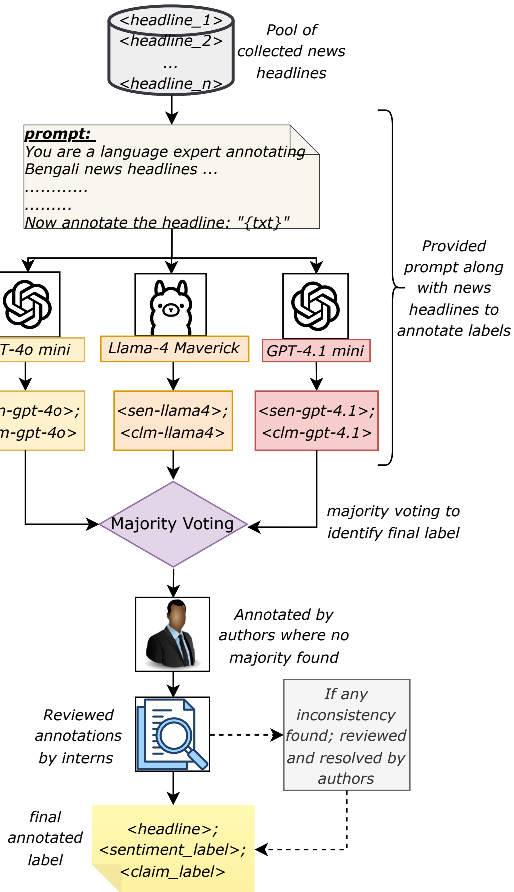

Figure 1: Overall flow diagram of the data sentiment and claim-checkworthy data annotation process utilizing three state-of-the-art LLMs followed by majority voting and manual intervention.

	Label	#Train	#Test
Sentiment	Negative	1640	396
	Neutral	463	118
	Positive	665	179
Claim	Check-worthy	2164	529
	Not Check-worthy	604	164

Table 1: Distribution of sentiment label and claim-checkworthy label for training and testing set.

4 Methodology

This section provides a brief overview of the methodology for sentiment classification and claim-checkworthiness detection, with a classical-quantum LSTM framework.

4.1 Task Definition

Given a tokenized sequence $S = [t_1, t_2, t_3, ..., t_n]$ where S is the sentence or text and t_i's are text tokens or words. For sentiment classification, each text was annotated with either positive, negative, or neutral labels. For claim identification, each text was annotated with either claim-checkworthy or not claim-checkworthy labels. Our objective is to predict appropriate labels using quantum machine learning algorithms.

4.2 Framework Description

We developed three frameworks for both sentiment analysis and claim checkworthiness identification: **1)** a classical LSTM framework where no quantum modules are used, **2)** a 2-qubit-based classical-quantum framework where we used a 2-qubit-based VQC layer, and **3)** a 4-qubit-based classical-quantum framework where a 4-qubit-based VQC layer was utilized. A flow diagram of classical LSTM and 4-qubit-based classical-quantum framework is provided in Figure 2.

As depicted in Figure 2, for both classical and classical-quantum frameworks, the tokenized sequence was first provided through an embedding layer of 128 dimensions to get a vector representation of each t_i in S, let's say X of dimension $n \times 128$. Next, the embedding matrix $[X]_{n\times128}$ was provided as input to an LSTM layer with 128 hidden units and a $tanh$ activation function.

Following the LSTM layer, the last hidden state representation from the LSTM, with a dimension of 1×128, was further passed through a fully connected layer (FC layer) with 32 hidden units and a sigmoid activation function.

$$Z_{fc} = sigmoid(LSTM_{out})$$

Here, Z_{fc} represents the output of the FC layer and $LSTM_{out}$ is the last hidden state output from the LSTM layer.

4.2.1 VQC Layer

In the case of the classical-quantum hybrid framework (Figure 2 (b)), the Z_{fc} was further split into equal chunks to serve as an input to the VQCs, i.e.,

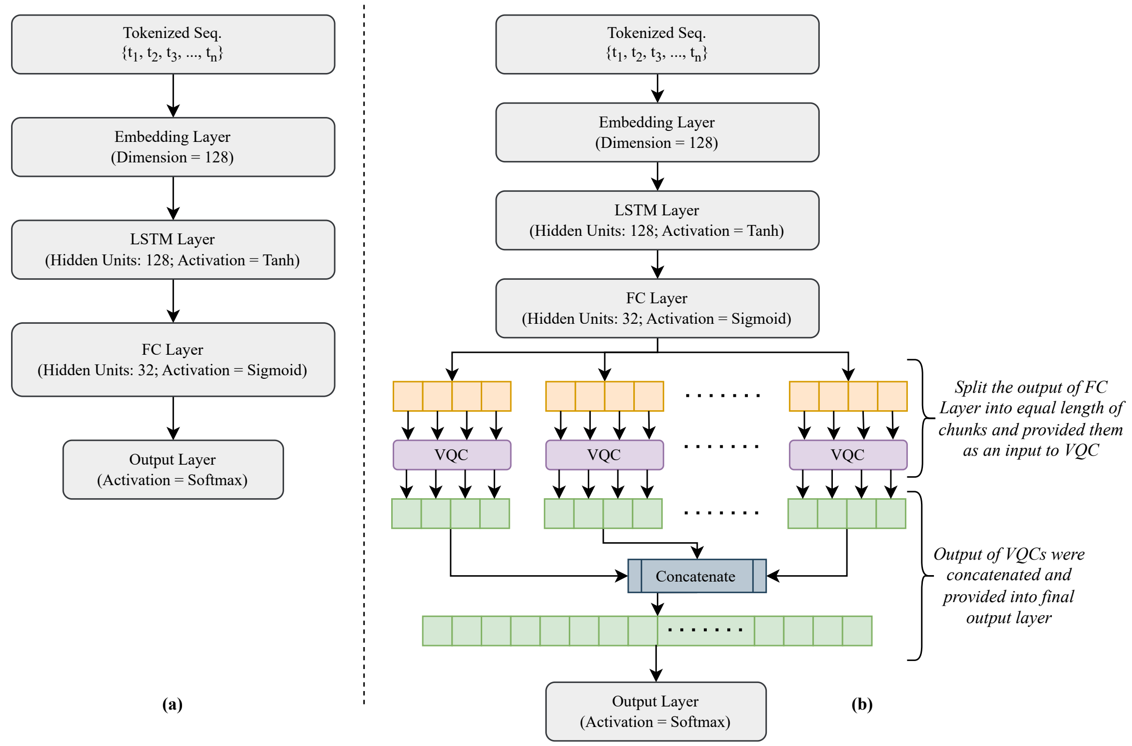

Figure 2: System framework for sentiment classification and claim-chekworthy identification: (a) classical LSTM framework, (b) hybrid classical-quantum framework utilizing classical LSTM followed by a layer of VQCs

for the 2-qubit-based framework, the Z_{fc} was divided into 16 equal chunks (each chunk with vector length 2), and for the 4-qubit-based framework, the Z_{fc} was split into eight equal chunks (each chunk with vector length 4).

VQCs or Variational Quantum Circuits are a special type of quantum circuit that has tunable parameters, and the parameters are updated iteratively by the gradient descent method. A typical VQC consists of three blocks: First, a data encoding block ($U(x)$) where the classical data is encoded into a quantum state, Second, a variational block ($V(\theta)$) where the encoded quantum state representation of classical bits gets a parameterized rotation with learnable parameter weights followed by several CNOT gates, and a quantum measurement block which measure the output for every qubit in the Pauli-Z basis. The diagrammatic representation of 2-qubit and 4-qubit-based VQCs used for developing the classical-quantum hybrid framework is provided in Figure 3.

As depicted in Figure 3, the first block is the data encoding block ($U(x)$), where the H gate or Hadamard gate first transforms each qubit state $|0\rangle$ to a superposition state $(|0\rangle + |1\rangle)/2$. Followed by H gate, for each classical input x_i, the R_y gate is used as an angle to rotate a qubit around the Y-axis

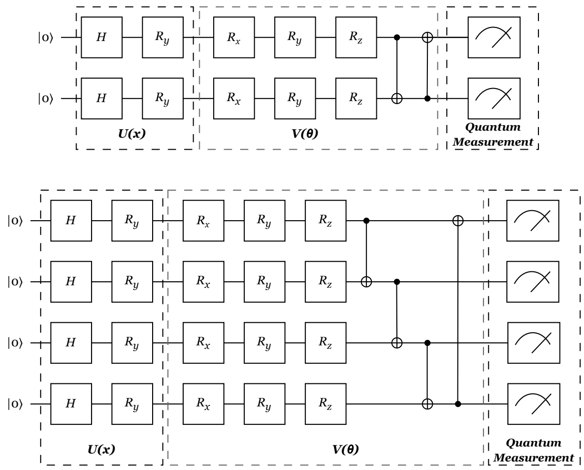

Figure 3: Diagramatic representation of VQCs developed for the classical-quantum hybrid: the top figure is the VQC with 2-qubit, and the bottom figure is the VQC with 4-qubit. H represents the Hadamard gate, R_x, R_y, R_z are the rotation gates around the X-axis, Y-axis, and Z-axis of the Bloch sphere, respectively.

of the Bloch sphere.

After data encoding, the next step is the variational block ($V(\theta)$), where every qubit gets a trainable Euler rotation (RX, RY, RZ), followed by an entangling ring of CNOTs, which enables the model to learn interactions between features. And, the final block is the quantum measurement block, which measures the expectations $\langle Z \rangle$ on each qubit ($\langle Z \rangle \in [-1, 1]$).

4.2.2 Classification

For classification, in the classical LSTM framework, Z_{fc} was passed to a final output layer with a softmax activation function. For sentiment classification, the output layer consists of three hidden units; for claim checkworthiness detection, it consists of two hidden units.

$$\mathcal{P} = softmax(Z_{fc})$$

On the other hand, for the classical-quantum hybrid framework, the output from each VQC unit was concatenated and further passed through the final output layer.

$$Z_{VQC} = Concatenate(z^1_{VQC}, z^2_{VQC}, ..., z^k_{VQC})$$
$$\mathcal{P} = softmax(Z_{VQC})$$
$$\hat{Y} = \underset{\mathbf{j}}{argmax}(\mathcal{P})$$

Here, z^i_{VQC} is the output of each VQC unit ($i = 1, 2, ..., k$), Z_{VQC} is the concatenated output, $\mathcal{P}$ represents the probability value for each class and $\hat{Y}$ represents the predicted class label and j represents the number of classes.

4.3 Training

In order to accomplish the training process, the training dataset was divided into a 90-10 ratio, where 90% of the data was used for training the framework and 10% of the data was reserved as a validation set. The `CrossEntropy` loss was used with a learning rate of 0.0025 to train all the frameworks. The optimizer selected was Adam (Kingma and Ba, 2017), and the number of epochs chosen for training the frameworks was 10 with a batch size of 64.

5 Experiment and Results

5.1 Experimental Setup

All experiments were performed using the PyTorch and Pennylane libraries [4] with an NVIDIA RTX-5000 GPU. PennyLane is a robust and open-source framework for quantum computing and quantum machine learning. It enables us to execute and train quantum circuits with a variety of backends, including real quantum computers and quantum simulators.

The VQC modules were trained and executed using the Pennylane quantum computing simulator with the '`default.qubit`' backend. For evaluation, the precision, recall, and macro-F1 score metrics were computed for the test dataset for both sentiment classification and claim checkworthiness identification.

To ensure a fair comparison between classical and 2-qubit and 4-qubit-based classical-quantum hybrid frameworks, all the frameworks were trained on the same training data with the same hyperparameters as reported in Section 4.3 and evaluated on the same test datasets as mentioned in Table 1.

5.2 Result

The performance for sentiment classification and claim checkworthiness detection is provided in Table 2. For sentiment classification, the best precision score of 62.14 is provided by the classical LSTM framework. Conversely, the best recall and F1-score of 52.64 and 52.67 is provided by a 4-qubit-based classical-quantum framework, which is a performance improvement of 0.25% in terms of F1-score when compared with the classical LSTM framework. Notably, the performance in the 2-qubit-based classical-quantum hybrid framework is surprisingly decreased to 47.62 F1-score, which is a performance dropout (F1-score) of 9.36% and 9.59% compared to the classical LSTM framework and classical-quantum hybrid framework, respectively.

	Framework	Precision	Recall	F1
Sentiment	classical	**62.14**	51.45	52.54
	2-qb	54.69	46.82	47.62
	4-qb	56.88	**52.64**	**52.67**
Claim	classical	65.95	**72.02**	64.27
	2-qb	**68.35**	65.58	**66.63**
	4-qb	66.20	66.47	66.33

Table 2: Result of sentiment classification and claim-chekworthy identification for test dataset. 'classical' represents the classical LSTM framework, '2qb' and '4qb' represent the 2-qubit-based and 4-qubit-based classical-quantum hybrid frameworks, respectively.

One possible reason for the low F1-score in the 2-qubit-based classical-quantum hybrid framework for sentiment classification is the division of the output of the FC layer into small chunks (vector length of 2), which loses the overall contextual relationship in the text, resulting in a low F1-score.

[4] https://pennylane.ai/

In the case of claim checkworthiness detection, both the 2-qubit and the 4-qubit-based classical-quantum hybrid framework outperform the classical LSTM framework in terms of F1-score. The classical LSTM framework only provides the best recall score of 72.02. The 2-qubit and 4-qubit-based classical-quantum hybrid frameworks achieved F1-scores of 66.63 and 66.33, representing a performance improvement of 3.67% and 3.1%, respectively, compared to the classical LSTM framework.

5.3 Error Analysis

Error analysis was performed using confusion matrices for sentiment analysis and claim checkworthiness identification in both the classical LSTM framework and two quantum-enhanced frameworks: the 2-qubit and 4-qubit-based classical-quantum hybrid frameworks. The confusion matrix for sentiment classification and claim checkworthiness identification is provided in Figure 4 and Figure 5, respectively.

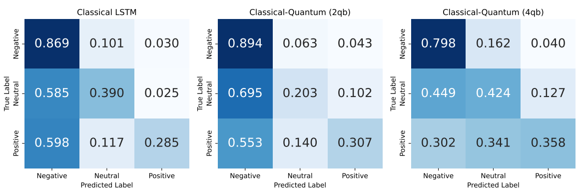

Figure 4: Confusion matrix for sentiment classification. The left confusion matrix is for the classical LSTM, the middle and right confusion matrices are for the 2-qubit and 4-qubit-based classical-quantum hybrid frameworks, respectively.

From Figure 4, for sentiment analysis, the neutral class shows the majority of error cases. In the 2-qubit-based classical-quantum hybrid framework, only 20.3% instances are correctly classified, followed by the classical LSTM framework, where 39% neutral instances are appropriately classified. The 4-qubit-based classical-quantum hybrid framework achieves 42.4% accuracy in identifying neutral classes, which is the highest neutral class classification accuracy among the three frameworks. The majority of misclassification for the neutral class is observed towards the negative class, where 58.5%, 69.5%, and 44.9% of neutral instances are misclassified as the negative class for the classical LSTM, 2-qubit-based classical-quantum, and 4-qubit-based classical-quantum hybrid frameworks, respectively.

This misclassification trend is also observed for positive classes, where 59.8% and 55.8% positive instances are misclassified as the negative class in classical LSTM and 2-qubit-based classical-quantum hybrid frameworks, respectively. The misclassification rate for the negative class is reduced to 30.2% for the 4-qubit-based classical-quantum framework; however, 34.1% positive instances are misclassified as the neutral class for the 4-qubit-based classical-quantum hybrid framework.

One possible reason for the majority of misclassification as the negative class is that the distribution of sentiment labels is highly imbalanced, with more than 50% of instances tagged as the negative class, which makes the model slightly biased towards the negative sentiment. As a result, neutral and positive instances are misclassified as negative sentiment.

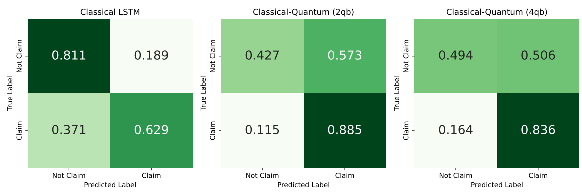

Figure 5: Confusion matrix for claim checkworthiness identification. The left confusion matrix is for the classical LSTM, the middle and right confusion matrices are for the 2-qubit and 4-qubit-based classical-quantum hybrid frameworks, respectively.

On the other hand, in the case of claim checkworthiness identification, 37.1% claim-checkworthy instances are misclassified in the classical LSTM framework. Notably, this misclassification rate is overcome in the classical-quantum hybrid frameworks, with misclassification rates of 11.5% and 16.4% in the 2-qubit-based classical-quantum hybrid framework and the 4-qubit-based classical-quantum hybrid framework, respectively.

However, while quantum-enhanced frameworks demonstrate strong performance in classifying checkworthy claims, their performance in identifying non-checkworthy claims deteriorates, with misclassification rates of 57.3% and 50.6% in the 2-qubit and 4-qubit-based classical-quantum hybrid frameworks, respectively. In contrast, the classical LSTM framework achieves a lower error rate of 18.9% in classifying non-checkworthy claims.

Along with analyzing confusion matrices, a few examples of error cases for sentiment classification and claim checkworthiness identification in different frameworks are provided in Table 3.

ID	Text	True Label	Predicted Label		
			classical	2qb	4qb
S_1	দর্শকসংখ্যায় সর্বকালের রেকর্ড, নতুন উচ্চতা ছুঁল আইসিসি চ্যাম্পিয়ন্স ট্রফি (**T:** ICC Champions Trophy reaches new heights, sets all-time record in viewership)	positive	negative	positive	positive
S_2	শেষ ম্যাচে পাঞ্জাবকে হারাতে মরিয়া মহামেডান, দলে একাধিক পরিবর্তনের ইঙ্গিত (**T:** Mohammedan desperate to beat Punjab in the last match, hints at multiple changes in the team)	neutral	neutral	positive	positive
S_3	‘এত সিলিন্ডার বিস্ফোরণ হলে ৫০ হাজার মানুষ মারা যাবে’, পার্ক স্ট্রিটের রেস্তরাঁ দেখে উদ্বেগে মমতা (**T:** 'If so many cylinders explode, 50,000 people will die', Mamata expresses concern after seeing Park Street restaurant)	negative	neutral	negative	negative
S_4	ভয়াবহভাবে বিশ্বে বাড়ছে দুর্ভিক্ষ ও অনাহার, উদ্বেগ বাড়াল রাষ্ট্রসংঘের রিপোর্ট (**T:** Famine and starvation are increasing alarmingly in the world, UN report raises concerns)	negative	positive	neutral	neutral
C_1	ভারত-বাংলাদেশ সীমান্ত থেকে বালুরঘাটের দুই কৃষককে অপহরণ! রাতভর চলল ফ্ল্যাগ মিটিং (**T:** Two farmers from Balurghat kidnapped from India-Bangladesh border! Flag meeting continues all night)	claim	n-claim	claim	claim
C_2	চিড়িয়াখানার নতুন আকর্ষণ, চেন্নাই থেকে কলকাতায় আসছে সবুজ অ্যানাকোন্ডা (**T:** New attraction at the zoo, green anaconda coming to Kolkata from Chennai)	claim	claim	n-claim	claim
C_3	হারের হ্যাটট্রিকের ভ্রুকুটি নিয়ে গোয়ায় ইস্টবেঙ্গল, চোট-আঘাতে দল সাজানোই চ্যালেঞ্জ অস্কারের (**T:** East Bengal face a hat-trick of defeats in Goa, Oscar's challenge is to shape the team due to injuries)	n-claim	n-claim	claim	claim
C_4	ইডেনে নাইটদের ম্যাচে বৃষ্টি? আজ কেমন থাকবে রাজ্যের আবহাওয়া (**T:** Rain in Knights match at Eden? What will the weather be like in the state today?)	n-claim	claim	n-claim	n-claim

Table 3: Some examples of error cases in the test dataset. S_1 to S_4 are the error cases for sentiment classification and C_1 to C_4 are the error cases for claim checkworthiness identification. The red-coloured texts represent the misclassified labels, and the blue-coloured texts represent the correctly classified labels. (‘T:’ represents the translation of the Bengali text)

6 Conclusion

This paper represents a novel classical-quantum hybrid framework for sentiment classification and claim checkworthiness identification for the less-resourced Bengali language. We developed an entirely new dataset for sentiment classification and claim-checkworthiness identification, comprising approximately 3,000 samples, and experimented with a classical LSTM framework and two quantum-classical hybrid frameworks based on 2-qubit and 4-qubit VQC. Our experiments and findings show that the classical-quantum hybrid framework outperforms the classical LSTM framework for both sentiment classification and claim checkworthiness identification.

Furthermore, to more accurately and robustly justify our findings and observations, we'll evaluate the proposed frameworks with other languages, such as English, Hindi, Assamese, and Odia, among others. In addition, we will experiment with quantum-enhanced Bi-LSTM, GRU, or Bi-GRU models in our future work.

Limitations

Our proposed work also has some potential limitations. First, all the classical-quantum hybrid frameworks were trained and evaluated on a quantum simulator, which somewhat limits the actual potential of real quantum hardware.

Second, we experimented with only the 2-qubit-based and 4-qubit-based VQCs in the development of the classical-quantum hybrid framework due to resource limitations. Chunking the output of the FC layer into 2 and 4 chunks and providing it through 2-qubit-based and 4-qubit-based VQCs sometimes loses the original contextual relationship between the words in a sentence. In our future work, we will aim to develop advanced techniques to preserve the contextual relationships between words while chunking. Also, we'll experiment with higher qubit VQCs, such as 8-qubit or 16-qubit, in our future work.

Third, due to the lack of trained professional annotators, time constraints, and economic reasons, we have to annotate the sentiment and claim checkworthy labels with the help of LLMs. Although we used three LLMs, followed by majority voting and manual verification, instead of relying on a single LLM model, there might still be some incorrectly annotated samples, as no LLM is 100% accurate. However, we'll aim to develop a fully human-annotated dataset and evaluate our proposed framework with that dataset in our future work.

Fourth, there is a high level of imbalance in the claim-checkworthy labels and sentiment labels in the dataset, which sometimes makes the frameworks biased towards the majority labels. However, in our future work, we'll incorporate more samples into the existing dataset (especially those with minority labels) to make the dataset more balanced.

Lastly, the dataset is limited to news headlines, which restricts our ability to assess the framework's performance in a broader scope, such as with data from Twitter or Reddit.

Acknowledgment

This work was supported by the Defence Research and Development Organisation (DRDO), New Delhi, under the project "Claim Detection and Verification using Deep NLP: an Indian perspective".

References

AmirAli Bagher Zadeh, Paul Pu Liang, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. 2018. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2236–2246, Melbourne, Australia. Association for Computational Linguistics.

Ivano Basile and Fabio Tamburini. 2017. Towards quantum language models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1840–1849, Copenhagen, Denmark. Association for Computational Linguistics.

Dushyant Singh Chauhan, Dhanush S R, Asif Ekbal, and Pushpak Bhattacharyya. 2020. Sentiment and emotion help sarcasm? a multi-task learning framework for multi-modal sarcasm, sentiment and emotion analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4351–4360, Online. Association for Computational Linguistics.

Bob Coecke, Giovanni de Felice, Konstantinos Meichanetzidis, and Alexis Toumi. 2020. Foundations for near-term quantum natural language processing. *Preprint*, arXiv:2012.03755.

Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Srinjoy Ganguly, Sai Nandan Morapakula, and Luis Miguel Pozo Coronado. 2022. Quantum natural language processing based sentiment analysis using lambeq toolkit. In *2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T)*, pages 1–6.

Raffaele Guarasci, Giuseppe De Pietro, and Massimo Esposito. 2022. Quantum natural language processing: Challenges and opportunities. *Applied Sciences*, 12(11).

Kilem Li Gwet. 2006. Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, 61(1):29–48.

Laszlo Gyongyosi and Sandor Imre. 2019. A survey on quantum computing technology. *Comput. Sci. Rev.*, 31(C):51–71.

Dimitri Kartsaklis, Ian Fan, Richie Yeung, Anna Pearson, Robin Lorenz, Alexis Toumi, Giovanni de Felice, Konstantinos Meichanetzidis, Stephen Clark, and Bob Coecke. 2021. lambeq: An efficient high-level python library for quantum nlp. *Preprint*, arXiv:2110.04236.

Diederik P. Kingma and Jimmy Ba. 2017. Adam: A method for stochastic optimization. *Preprint*, arXiv:1412.6980.

Ziming Li, Yan Zhou, Yaxin Liu, Fuqing Zhu, Chuanpeng Yang, and Songlin Hu. 2023. QAP: A quantum-inspired adaptive-priority-learning model for multi-modal emotion recognition. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 12191–12204, Toronto, Canada. Association for Computational Linguistics.

Yaochen Liu, Yazhou Zhang, Qiuchi Li, Benyou Wang, and Dawei Song. 2021. What does your smile mean? jointly detecting multi-modal sarcasm and sentiment using quantum probability. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 871–880, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Niels Neumann, Frank Phillipson, and Richard Versluis. 2019. Machine learning in the quantum era. *Digitale Welt*, 3(2):24–29.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, and 9 others. 2024. Gpt-4 technical report. *Preprint*, arXiv:2303.08774.

Pritam Pal and Dipankar Das. 2025. Toward quantum-enhanced natural language understanding: Sarcasm and claim detection with qlstm. In *Proceedings of the 15th International Conference on Recent Advances in Natural Language Processing - Natural Language Processing in the Generative AI era*, pages 854–859, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Shyambabu Pandey, Pankaj Dadure, Morrel V. L. Nunsanga, and Partha Pakray. 2022. Parts of speech tagging towards classical to quantum computing. In *2022 IEEE Silchar Subsection Conference (SILCON)*, pages 1–6.

Shyambabu Pandey and Partha Pakray. 2023. Bi-quantum long short-term memory for part-of-speech tagging. In *Proceedings of the 20th International*

Conference on Natural Language Processing (ICON), pages 301–307, Goa University, Goa, India. NLP Association of India (NLPAI).

Bo Pang and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the ACL*.

Arpan Phukan and Asif Ekbal. 2023. QeMMA: Quantum-enhanced multi-modal sentiment analysis. In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, pages 815–821, Goa University, Goa, India. NLP Association of India (NLPAI).

John Proos and Christof Zalka. 2004. Shor's discrete logarithm quantum algorithm for elliptic curves. *Preprint*, arXiv:quant-ph/0301141.

Jun Qi, Chao-Han Huck Yang, and Pin-Yu Chen. 2021. Qtn-vqc: An end-to-end learning framework for quantum neural networks. *Preprint*, arXiv:2110.03861.

Wenbo Qiao, Peng Zhang, and ZengLai Ma. 2024. A quantum-inspired matching network with linguistic theories for metaphor detection. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 1435–1445, Torino, Italia. ELRA and ICCL.

R. L. Rivest, A. Shamir, and L. Adleman. 1978. A method for obtaining digital signatures and public-key cryptosystems. *Commun. ACM*, 21(2):120–126.

Fariska Z. Ruskanda, Muhammad Rifat Abiwardani, Muhammad Akram Al Bari, Kinantan Arya Bagaspati, Rahmat Mulyawan, Infall Syafalni, and Harashta Tatimma Larasati. 2022. Quantum representation for sentiment classification. In *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 67–78.

Fariska Zakhralativa Ruskanda, Muhammad Rifat Abiwardani, Infall Syafalni, Harashta Tatimma Larasati, and Rahmat Mulyawan. 2023. Simple sentiment analysis ansatz for sentiment classification in quantum natural language processing. *IEEE Access*, 11:120612–120627.

Chhavi Sharma, Deepesh Bhageria, William Scott, Srinivas PYKL, Amitava Das, Tanmoy Chakraborty, Viswanath Pulabaigari, and Björn Gambäck. 2020. SemEval-2020 task 8: Memotion analysis- the visuo-lingual metaphor! In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 759–773, Barcelona (online). International Committee for Computational Linguistics.

Jinjing Shi, Zhenhuan Li, Wei Lai, Fangfang Li, Ronghua Shi, Yanyan Feng, and Shichao Zhang. 2023. Two End-to-End Quantum-Inspired Deep Neural Networks for Text Classification . *IEEE Transactions on Knowledge & Data Engineering*, 35(04):4335–4345.

Peter W. Shor. 1997. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509.

Riccardo Di Sipio, Jia-Hong Huang, Samuel Yen-Chi Chen, Stefano Mangini, and Marcel Worring. 2021. The dawn of quantum natural language processing. *Preprint*, arXiv:2110.06510.

Fabio Tamburini. 2019. A quantum-like approach to word sense disambiguation. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 1176–1185, Varna, Bulgaria. INCOMA Ltd.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. *Preprint*, arXiv:2302.13971.

Charles M. Varmantchaonala, Jean Louis K. E. Fendji, Julius Schöning, and Marcellin Atemkeng. 2024. Quantum natural language processing: A comprehensive survey. *IEEE Access*, 12:99578–99598.

Dominic Widdows, Willie Aboumrad, Dohun Kim, Sayonee Ray, and Jonathan Mei. 2024. Quantum natural language processing. *KI - Künstliche Intelligenz*.

Sixuan Wu, Jian Li, Peng Zhang, and Yue Zhang. 2021. Natural language processing meets quantum physics: A survey and categorization. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3172–3182, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Wenduan Xu, Stephen Clark, Douglas Brown, Gabriel Matos, and Konstantinos Meichanetzidis. 2024. Quantum recurrent architectures for text classification. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 18020–18027, Miami, Florida, USA. Association for Computational Linguistics.

Kehuan Yan, Peichao Lai, and Yilei Wang. 2024. Quantum-inspired language model with lindblad master equation and interference measurement for sentiment analysis. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2112–2121, Mexico City, Mexico. Association for Computational Linguistics.

Yazhou Zhang, Qiuchi Li, Dawei Song, Peng Zhang, and Panpan Wang. 2019. Quantum-Inspired Interactive Networks for Conversational Sentiment Analysis. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5436–5442.