# Abstract

Text sentiment analysis is an important task in natural language processing and has always been a hot research topic. However, in low-resource regions such as South Asia, where languages like Bengali are widely used, the research interest is relatively low compared to high-resource regions due to limited computational resources, flexible word order, and high inflectional nature of the language.（研究背景、现状—A急需研究） With the development of quantum technology, quantum machine learning models leverage the superposition property of quantum bits to enhance model expressiveness and achieve faster computation compared to classical systems. （引出了研究对象—B发挥了重要作用）To promote the development of quantum machine learning in low-resource language domains, we propose a classical-hybrid quantum architecture.（研究目的）This architecture utilizes a pre-trained multilingual BERT model to obtain vector representations of words and combines the proposed Batched Upload Quantum Recurrent Neural Network (BUQRNN) and Parameter Non-Shared Batched Upload Quantum Recurrent Neural Network (PN-BUQRNN) as feature extraction models for sentiment analysis in Bengali.（研究方法和结果—B在A的作用 通过xx手段得到了什么） Our numerical results demonstrate that the proposed BUQRNN structure achieves a maximum accuracy improvement of 0.993% in Bengali text classification tasks while reducing average model complexity by 12%. The PN-BUQRNN structure surpasses the BUQRNN structure once again and outperforms classical architectures in certain tasks.（研究结果—B的作用）

# Introduction

As one of the classical subfields of machine learning [1]– [3], natural language processing (NLP) [4], [5] has been a hot research topic in recent years. Text sentiment analysis (SA) [6], as a subtask of NLP, aims to classify text into positive and negative sentiment categories by detecting the polarity of the text. SA has been applied in various domains, including lexicon-based SA [7], machine learning-based SA [8], and deep learning-based SA [9]. Remarkable results have been achieved in SA for high-resource languages such as English and Chinese [10], [11]. However, due to the complexity of language grammar, limited usage, and expensive computational resources, SA in low-resource languages has not been extensively explored. With the development of the internet, a large influx of textual comments has made SA in low-resource languages feasible. In general, effective SA tasks can be achieved by combining good word embedding models with efficient feature extraction models. In the case of studying word embeddings for Bengali texts, a significant challenge lies in capturing the rich expressions of sentiment present in the Bengali language. Due to the complexity of grammar rules, the extraction of features for sentiment classification becomes intricate as sentiment information may be expressed differently in sentences. The emergence of pretrained language models [12], [13] has improved feature extraction in Bengali sentiment classification tasks, as they can effectively capture sentiment information within Bengali sentences by learning rich language representations and contextual understanding.

自然语言处理(natural language processing, NLP)作为机器学习的经典子领域之一[1]-[3]，[4]，[5]是近年来的研究热点。文本情感分析(Text sentiment analysis, SA)[6]是NLP的一个子任务，其目的是通过检测文本的极性，将文本分为积极和消极情感两类。情景分析已经应用于各个领域，包括基于词典的情景分析[7]、基于机器学习的情景分析[8]和基于深度学习的情景分析[9]。对于英语和汉语等资源丰富的语言，SA已经取得了显著的成果[10]，[11]。然而，由于语言语法的复杂性、使用范围的有限性和计算资源的昂贵性，低资源语言中的情景识别还没有得到广泛的探索。随着互联网的发展，大量文本评论的涌入使得低资源语言的SA成为可能。通常，通过将好的词嵌入模型与高效的特征提取模型相结合，可以实现有效的情景分析任务。在研究孟加拉语文本的词嵌入的情况下，一个重大的挑战在于捕捉孟加拉语中存在的丰富的情感表达。由于语法规则的复杂性，情感信息在句子中的表达可能不同，因此情感分类特征的提取变得复杂。预训练语言模型[12]，[13]的出现改善了孟加拉语情感分类任务中的特征提取，因为它们可以通过学习丰富的语言表示和上下文理解来有效地捕获孟加拉语句子中的情感信息。

Despite the positive role of pre-trained language models in Bengali sentiment classification tasks, traditional feature extraction models still face efficiency challenges. Quantum deep learning [14], [15] combines the concepts of quantum computing and deep learning, leveraging the parallelism advantage [16] of quantum computing to accelerate the training and inference processes of models. By utilizing quantum neural networks and quantum gate operations, quantum deep learning models can handle complex sentiment classification tasks more efficiently [17]. A classical-hybrid quantum recurrent neural network model (QRNN) based on the quantum variational circuit (VQC) core was proposed in literature [18]. Such networks have been successfully applied as feature extractors in text classification tasks for high-resource languages, demonstrating better performance compared to their classical counterparts. Considering the characteristics of low-resource languages, this sparks the idea of using quantum algorithms to improve low-resource text SA tasks. However, previous studies have shown that QRNNs may struggle to effectively capture semantic information and may even result in information loss when dealing with longer sequences. The current challenges can be summarized as follows:

尽管预训练语言模型在孟加拉语情感分类任务中发挥了积极的作用，但传统的特征提取模型仍然面临着效率方面的挑战。量子深度学习[14]，[15]结合了量子计算和深度学习的概念，利用量子计算的并行性优势[16]来加速模型的训练和推理过程。通过利用量子神经网络和量子门操作，量子深度学习模型可以更有效地处理复杂的情感分类任务[17]。文献[18]提出了一种基于量子变分电路(VQC)核的经典-混合量子递归神经网络模型(QRNN)。这种网络已经成功地作为特征提取器应用于高资源语言的文本分类任务中，与传统网络相比表现出更好的性能。考虑到低资源语言的特点，这激发了使用量子算法来改进低资源文本SA任务的想法。然而，先前的研究表明，qrnn在处理较长的序列时可能难以有效捕获语义信息，甚至可能导致信息丢失。当前的挑战可以概括如下:

* Due to the limitations of current Noisy IntermediateScale Quantum (NISQ) devices [19], it is necessary to match the dimensionality of the input sequence with the number of quantum bits. Previous QRNN models employed parameter-sharing linear layers to reduce the dimensionality of the input data, which may potentially result in the loss of semantic information to some extent.
* 由于当前有噪声的中间尺度量子(Noisy Intermediate Scale Quantum, NISQ)器件的局限性[19]，需要将输入序列的维数与量子比特数相匹配。以往的QRNN模型采用参数共享线性层来降低输入数据的维数，这可能会在一定程度上造成语义信息的丢失。
* Previous QRNN models did not optimize the Quantum Variational Circuit (VQC) specifically but instead utilized parameter-sharing linear layers for optimization across all VQCs.

先前的QRNN模型没有专门优化量子变分电路(VQC)，而是利用参数共享线性层对所有VQC进行优化。

In response to the first situation mentioned above, we designed and utilized a Batched Uploading Quantum Neural Network (BUQNN), which is essentially a structure that incorporates VQC circuits. The BUQNN divides the input feature sequence into batches and loads them into the circuit to obtain the complete semantic information. We refer to the QRNN model that utilizes this BUQNN as BUQRNN. By adopting this approach, we can alleviate the semantic information loss caused by previous methods with a S number of quantum bits.

Regarding the second situation mentioned, in reference [20], a non-parameter-sharing linear layer was applied after the VQC to enhance the expressiveness of the circuit. We followed this idea and made improvements by using non-parameter-sharing linear layers both before and after each VQC. This allows for independent optimization of each VQC, which is advantageous for the model.

Our contributions are as follows:

* We proposed a Batched Uploading Quantum Recurrent Neural Network (BUQRNN) specifically designed for sequential data. This method requires only a S number of quantum bits and mitigates the loss of semantic information caused by previous approaches.
* We introduced a Parameter Non-sharing BUQRNN (PN-BUQRNN) that employs independent linear layers for each VQC, enabling independent optimization of each VQC in the model structure.
* For the first time, we applied quantum algorithms to address sentiment classification tasks in low-resource languages such as Bengali, which holds significant importance for advancing the development of quantum in low-resource languages.

The rest of the paper is organized as follows. Section 2 focuses on the text classification process and discusses word embedding techniques and the QRNN network for lowresource language text classification. Section 3 presents the specific implementation approach to address the aforementioned issues. Section 4 describes the numerical simulation results, and Section 5 concludes the paper, providing insights into future directions and prospects.