Revolutionizing Fake News Detection: A Quantum Machine Learning Paradigm

Ainaz Jamshidi
Department of Information Systems
University of Maryland Baltimore County
Baltimore, USA
ainazj1@umbc.edu

Muhammad Arif
Department of Information Systems
Colorado State University
Pueblo, USA
muhammad.arif@csupueblo.edu

***Abstract*—Today's digital world facilitates the rapid growth of social media platforms and online news sources and consequently the spread of true and false information. While the detection of fake news is extensively studied using classical machine learning (ML)-based approaches, quantum ML approaches remain largely unexplored. In this study, we aim to investigate the effectiveness of a hybrid quantum-classical neural network on three open access datasets. This study proposes a Hybrid Quantum Neural Network (HQNN) model for classifying news articles as either "fake" or "real". The approach combines classical machine learning techniques with quantum computing to leverage the strengths of both paradigms. It incorporates a quantum layer, implemented via a parameterized quantum circuit that utilizes angle embedding and entanglement to extract complex features, followed by a classical fully connected layer for final decision-making. Our results demonstrate the effectiveness of the HQNN model, achieving high accuracy across multiple datasets, with values reaching up to 90.71%. These findings highlight the potential of hybrid quantum-classical models to improve fake news detection and pave the way for further research into quantum-enhanced ML applications.**

***Index Terms*—quantum computing, fake news detection, quantum artificial neural network**

I. Introduction

The rapid growth of social media platforms and online news outlets has transformed how information is shared, making it easier to spread both true and false content at an unprecedented speed. Fake news—deliberately created to mislead readers—has become a major issue, influencing public opinion, democratic processes, and global stability. Addressing this challenge is critical to maintaining trust in digital information. Traditional methods for detecting fake news primarily rely on machine learning (ML) and deep learning (DL) techniques, which analyze patterns in text, user behavior, and social network activity to identify fake content [1], [2].

Although traditional methods have shown promising results, the growing sophistication of fake news generation demands new and innovative solutions that go beyond the limits of current models. Techniques like advanced feature extraction using natural language processing and DL have paved the way for improvements, but the challenges of modern misinformation, including its complexity and high computational demands, require a stronger and more efficient approach [2], [3].

Quantum Machine Learning (QML) has emerged as a promising paradigm to address these challenges. By leveraging the principles of quantum computing—such as superposition and entanglement—QML enables the efficient processing of large and complex datasets [4]. Recent studies highlight the potential of quantum-enhanced algorithms, such as quantum k-nearest neighbors, to outperform traditional ML approaches in accuracy [3]. The unique capabilities of QML when combined with classical ML can be a power tool for advancing fake news detection systems.

This paper introduces a novel Hybrid Quantum-Classical Neural Network (HQNN) for fake news detection. The proposed model combines the strengths of classical machine learning techniques with a quantum layer. We demonstrate the effectiveness of HQNN in accurately identifying fake news by evaluating the approach three open-access datasets.

In summary, the **contributions** of the paper are as follows:

- We introduce a hybrid quantum-classical model for fake news detection, combining quantum computing's strengths with established ML techniques.
- We demonstrate the efficacy of our approach in detecting fake news, utilizing a robust evaluation of our method.

The remainder of this paper is organized as follows: Section 2 details the proposed methodology, Section 3 presents the experimental results, and Section 4 concludes the study with potential future directions.

II. Overview of the proposed system

A high-level overview of our study is illustrated in Figure 1. The process can be broken down into the following steps:

A. DataSets

To explore the application of our proposed quantum-based approach in fake news detection, we test our model using different publicly available datasets. Two of these datasets are sourced from the Kaggle platform, and the third dataset is obtained from Hugging Face. The first Kaggle dataset [5] (referred to as KG1) contains 23,481 fake and 21,417 real news articles. The second Kaggle dataset [6] (referred to as KG2) contains 3,164 fake and 3,171 real news articles. The Hugging Face dataset [7], referred to as HF, includes 15,478

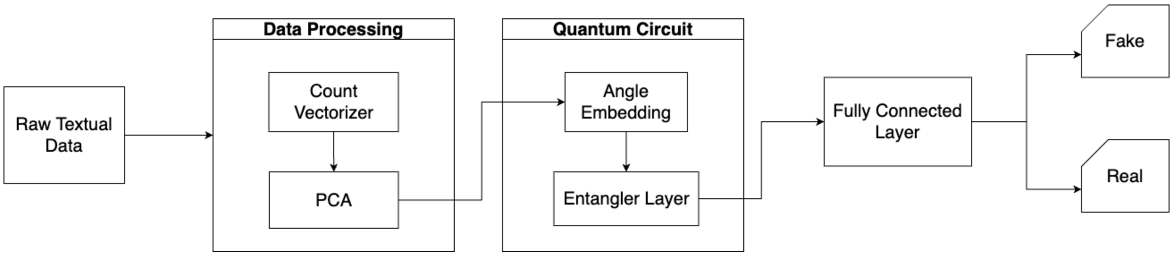


Fig. 1: An overview of the proposed quantum-based network for fake news classification

fake and 14,522 real news articles. All three datasets have a balanced distribution of classes between real and fake news.

B. Feature Engineering

- **Term Frequency-Inverse Document Frequency (TF-IDF)**: The texts are transformed into a numerical representation using TF-IDF. This technique assigns a score to each word based on how frequently it appears in a document compared to how often it appears across all documents.

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \log\left(\frac{N}{\text{DF}(t)}\right)$$

  where:
  - TF(t, d) is the term frequency of term t in document d,
  - N is the total number of documents,
  - DF(t) is the number of documents containing term t.
- **Principal Component Analysis (PCA)**: PCA is applied to reduce the dimensionality of the TF-IDF matrix. It projects the data into a lower-dimensional space while preserving the maximum variance.

The classical data, reduced using PCA, is embedded into quantum states. The details of this process are explained in the following section.

We develop a hybrid quantum-classical neural network model for binary classification. The model integrates classical preprocessing techniques, such as TF-IDF and PCA, with quantum layers to extract features from text data, leveraging the principles of quantum computing.

The proposed architecture is composed of two key components: a quantum layer and a classical fully connected layer. The input data is first processed through the quantum circuit, which extracts complex features. The resulting output is then passed to the classical fully connected layer, where a weighted sum of the quantum circuit outputs is computed. Finally, a softmax activation function is applied to predict the class labels.

- **Quantum Layer.** The quantum layer in the model is implemented using a quantum circuit and serves as the first layer of the artificial neural network. This quantum circuit processes input data by performing a series of transformations to extract meaningful features. The quantum device is initialized with 6 qubits. Classical input features, reduced to a lower dimension using PCA, are embedded into quantum states through angle encoding. This method uses the input features to set the rotation angles of the qubits, effectively mapping the classical data into the quantum state space. For classical data $x \in \mathbb{R}^n$, each feature x_i is encoded into a quantum state by rotating a qubit along the Y-axis. The encoding is represented mathematically as:

$$U(\theta) = R_Y(x_i) = \exp\left(-i\frac{x_i}{2}\sigma_y\right)$$

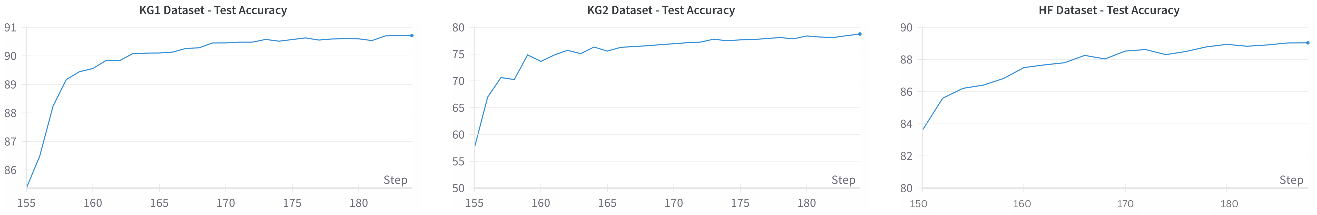
  where σ_y is the Pauli-Y matrix, and $R_Y(x_i)$ represents a rotation by angle x_i on the qubit.
  After embedding the input features into quantum states, the quantum circuit applies entangling operations using trainable parameters (weights). These operations ensure interaction between qubits, creating complex correlations that capture meaningful information from the data. The quantum circuit includes entanglement layers that use controlled operations, such as CNOT gates, to couple pairs of qubits. Mathematically, the entangling operation between qubits i and j is represented as:

$$\text{Entangler}(i, j) = \text{CNOT}(i, j),$$

  where the state of qubit j is flipped if qubit i is in the $|1\rangle$ state. These entangling layers enable the circuit to correlate information across all input features effectively.
  At the end of the quantum circuit, the expectation values of the Pauli-Z operator $\langle Z_i \rangle$ are computed for each qubit i. These expectation values are real numbers between -1 and 1 and are given by:

$$\langle Z_i \rangle = \langle \psi | Z_i | \psi \rangle,$$

  where $|\psi\rangle$ is the quantum state after applying the quantum gates. The expectation values represent the transformed features and serve as the outputs of the quantum layer. The quantum layer's outputs (expectation values for 6-qubits) are passed as inputs to the fully connected classical layer. This classical layer uses these transformed features to perform further processing and label prediction.
- **Fully Connected Layer.** The second layer of the HQNN is a fully connected classical layer, implemented as a



(a) KG1 Dataset

(b) KG2 Dataset

(c) HF Dataset

Fig. 2: Average test accuracy trends across three datasets (KG1, KG2, and HF) during 5-fold cross-validation. The results demonstrate the effectiveness of the HQNN model in achieving high accuracy, with KG1 and HF datasets performing better than KG2.

simple linear fully connected. The input to this layer is the output from the quantum circuit, which has a dimensionality equal to the number of qubits (6-qubits). The output of this layer has two dimensions, representing the two possible classes (real news or fake news).

III. Experimental Results and Discussion

We employ a 5-fold cross-validation approach to evaluate the model's performance. This strategy ensures that the model is trained and validated on diverse data splits. The evaluation metrics include accuracy, recall, precision, and F1-score. In Table I, we report the average performance of the model across five folds for each dataset. Our HQNN model is trained for 20 epochs with a batch size of 32 using the Adam optimizer with a learning rate of 0.001, and CrossEntropyLoss as the loss function.

The results in Table I demonstrate the effectiveness of the HQNN model in detecting fake news across diverse datasets. The model achieves high accuracy, precision, recall, and F1-scores, particularly on the KG1 and HF datasets. The slight variation in performance metrics for KG2 reflects the increased complexity and variability of this dataset. To illustrate this, Fig. 2 presents the average trend in test accuracy during training. The figure highlights the model's ability to converge to stable and high accuracy levels for fake news detection across all datasets, with particularly strong performance on KG1 and HF.

These findings indicate that the integration of quantum layers with classical machine learning techniques provides a robust and efficient framework for fake news detection.

TABLE I: Performance metrics of HQNN model on different datasets. (average of 5-fold cross-validation ± std).

Data	**Accuracy**	**Precision**	**Recall**	**F1-score**
KG1	90.71 ± 0.2	89.30 ± 0.00	91.49 ± 0.00	90.37 ± 0.00
KG2	78.75 ± 1.65	80.24 ± 0.01	76.36 ± 0.02	78.24 ± 0.01
HF	89.03 ± 0.49	87.59 ± 0.01	90.16 ± 0.01	88.83 ± 0.00

IV. Conclusion and Future Work

In this research, we explored and validated the feasibility of utilizing a quantum-based neural network model for fake news detection. The proposed quantum-enhanced fake news detection system demonstrated high accuracy across all three datasets used in our study.

For future work, we plan to extend our approach by incorporating additional datasets and conducting comparative evaluations with state-of-the-art classical methods. Moreover, we will explore the efficiency of HQNN in detecting AI-generated fake news, as these introduce unique challenges to the domain due to their impressive ability to mimic legitimate content.

References

[1] Muhammad Arif, Atnafu Lambebo Tonja, Iqra Ameer, Olga Kolesnikova, Grigori Sidorov, and Abdul Gafar Manuel Meque. Cic at checkthat!-2022: Multi-class and cross-lingual fake news detection. 2022.

[2] Jawaher Alghamdi, Yuqing Lin, and Suhuai Luo. A comparative study of machine learning and deep learning techniques for fake news detection. *Information*, 13(12):576, 2022.

[3] Ziyan Tian and Sanjeev Baskiyar. Fake news detection: An application of quantum k-nearest neighbors. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–6. IEEE, 2021.

[4] Ziyan Tian. Fake news detection with quantum machine learning. 2023.

[5] Mohsin Chaudhary. Detecting fake news dataset, 2021. Accessed: 2024-12-06.

[6] Clément Bisaillon. Fake and real news dataset. https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset/data, n.d.

[7] Erfan Moosavi Monazzah. Fake news detection dataset (english). https://huggingface.co/datasets/ErfanMoosaviMonazzah/fake-news-detection-dataset-English.