Preprints.org

Article

Not peer-reviewed version

Enhanced Quantum-Inspired Deep Learning with Multi-Head Attention and Contrastive Learning for Multimodal Emotion Recognition in Human-Computer Interaction

Fumin Zou , Lei Zou , Feng Guo * , Xunhuang Wang , Jianqing Weng , Tao Fang , Haocai Jiang , Xueming Wu

Posted Date: 1 April 2026

doi: 10.20944/preprints202603.2525.v1

Keywords: quantum-inspired computing; deep learning; multi-head attention; contrastive learning; emotion recognition

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Enhanced Quantum-Inspired Deep Learning with Multi-Head Attention and Contrastive Learning for Multimodal Emotion Recognition in Human-Computer Interaction

Fumin Zou [1,2], Lei Zou [1,2], Feng Guo [1,2,*], Xunhuang Wang [1,2], Jianqing Weng [1,2], Tao Fang [1,2], Haocai Jiang [1,2] and Xueming Wu [1,2]

[1] Fujian Key Laboratory of Automotive Electronics and Electric Drive, Fujian University of Technology, Fujian 350118, China

[2] Renewable Energy Technology Research Institute of Fujian University of Technology, Fujian University of Technology, Ningde 352101, China

* Correspondence: mapli@fjut.edu.cn

Abstract

This paper proposes an enhanced quantum-inspired sentiment analysis model incorporating a self-embedding mechanism for sentiment feature extraction and classification tasks. The method integrates phase-pre-trained self-embedding, bidirectional GRUs, a multi-head attention mechanism, and a multi-layer Transformer structure, effectively capturing semantic and emotional features in texts. Simultaneously, the model introduces contrastive learning and an enhanced feature interaction module, further improving feature discriminability. Extensive experiments on the RECCON dataset demonstrate that the proposed model significantly outperforms mainstream baseline methods (KEC, MPEG, Window Transformer) on key metrics such as macro-F1, positive-class F1, and negative-class F1. The experimental results show that the method not only improves overall accuracy and recall but also effectively mitigates challenges arising from class imbalance, achieving a macro-F1 of 0.95, positive-class F1 of 0.93, and negative-class F1 of 0.97 on the test set. The findings suggest that the combination of quantum-inspired structures and self-embedding mechanisms holds broad application prospects for complex sentiment analysis tasks.

Keywords: quantum-inspired computing; deep learning; multi-head attention; contrastive learning; emotion recognition

1. Introduction

1.1. Research Background

Natural Language Processing (NLP), a core branch of artificial intelligence, aims to enable computers to understand, interpret, and generate human language. It has long been a key focus of research. This technology has been widely applied in various downstream tasks, such as text classification [1,2], word sense disambiguation (WSD) [3,4], knowledge graphs [5], and question-answering systems [6]. In recent years, the NLP field has seen significant progress, primarily driven by the widespread adoption and successful application of deep learning methods, including convolutional neural networks (CNN) [7,8] and recurrent neural networks (RNN) [9].

Since 2018, the emergence of models such as ELMo, GPT, and BERT has marked the dawn of a new era in NLP. These models learn deep representations of language by pre-training on vast amounts of unlabeled text and are adapted to specific tasks through fine-tuning [10]. For example,

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

the BERT (Bidirectional Encoder Representations from Transformers) model generates deep bidirectional representations by jointly training on both left and right contexts [11].

Sentiment Classification: Sentiment analysis aims to assess and identify the emotion or sentiment conveyed by textual data [12]. Transformer-based models and their variants have demonstrated excellent performance in this task [13]. For example, a hybrid architecture combining BERT, BiLSTM, and CNN layers has been employed for sentiment classification of student reviews in MOOCs, enhancing classification performance by fusing BERT's context-aware capability, BiLSTM's ability to capture long- and short-term dependencies, and CNN's local feature extraction [14]. Other research has integrated BERT with Support Vector Machines (SVM) for sentiment analysis, where the two methods complement each other: BERT provides powerful language understanding, while SVM excels at classification [15].

Word Sense Disambiguation (WSD): WSD is a crucial research topic in NLP, aimed at determining the correct meaning of a word within a specific context [16]. It has broad applications in areas such as text classification, machine translation, and information retrieval [17]. Early studies focused on context-sensitive and statistical methods [18], while modern approaches leverage deep learning models. For instance, Graph Convolutional Networks (GCN) have been applied to WSD tasks, improving disambiguation accuracy by extracting discriminative features—such as words, part-of-speech tags, and semantic categories—from the context of ambiguous words.

Advances in NLP have also benefited the biomedical field, where researchers have proposed various pre-trained language models trained on biomedical datasets (e.g., text, electronic health records, protein and DNA sequences) for a range of biomedical tasks [19]. Furthermore, NLP techniques are being applied to AI-assisted programming tasks such as code generation, code completion, and code translation. As a widely researched area in NLP, text classification has seen remarkable achievements with deep learning models, including large pre-trained models like BERT and DistilBERT.

The advent of Large Language Models (LLMs) represents a revolutionary breakthrough in artificial intelligence. With unprecedented training scales and model parameters, they have significantly advanced capabilities in language understanding, synthesis, and commonsense reasoning, achieving performance levels close to human proficiency[20]. Pre-trained language models like BERT, GPT, and ERNIE are representative examples of LLMs. The evolution of NLP has progressed from Word2Vec and GloVe word embeddings in 2013, to the introduction of the attention mechanism and Transformer in 2017, culminating in the development of large multimodal models like GPT-4 in 2023 and Gemini in 2024.

Despite these significant advancements, NLP still faces challenges. The first is the "black box" nature of neural network models [21]. The millions, or even billions, of parameters in these models are difficult to interpret, limiting their application in critical areas such as medical diagnosis [22] or financial decision-making [23]. Secondly, the inherent uncertainty and ambiguity of natural language are not fully embedded in the models, which is inconsistent with how humans perceive language and may result in inadequate modeling [24]. To address these challenges, the emergence of quantum-inspired models offers a new research direction for NLP. These models, which construct neural networks based on quantum theory [25], hold the potential to enhance model interpretability, thereby improving the understanding and handling of the complexities of natural language.

1.2. Development of Quantum-Inspired Models

In recent years, researchers have extensively investigated the construction of quantum-inspired models and their application in various NLP scenarios. In 2018, Zhang et al. [26] introduced an end-to-end quantum-inspired language model for question-answering tasks, representing sentences as density matrices and proposing a joint representation for question answering. Building on this, Li et al. [27] developed a complex-valued word-embedding neural network, which defines semantic units in Hilbert space and uses complex vectors to represent words. In 2019, Li et al. proposed a novel matching model termed Complex-Valued Network (CNM), achieving performance comparable to

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

traditional CNN and RNN baselines. In the same year, Tamburini [28] introduced a quantum-like Word Sense Disambiguation (QWSD) model based on quantum probability theory, representing words and sentences in complex domains. In 2024, Shi et al. [29] proposed QPFE and QPFE-ERNIE, which enhance quantum-like models with gated recurrent units (GRU) and incorporate attention mechanisms and CNNs, yielding improved experimental results in text classification tasks.

These studies have undoubtedly made significant contributions to the application of quantum theory in NLP, advancing the development of quantum-inspired models in the field. However, these improvements have largely overlooked a critical issue: prior knowledge is indispensable for constructing high-performance quantum-inspired models. Even though models based on quantum probability theory have been theoretically demonstrated to be suitable for natural language modeling, ELMo, BERT, GPT, and other language models have achieved great success by learning textual knowledge through pre-training tasks on large corpora. Currently, few works integrate pre-trained textual feature embeddings into quantum-inspired models, which may limit the performance enhancement of such models.

To improve the performance of quantum-inspired models, especially for emotion recognition in dialogues for sentiment classification, we adopt a self-embedding mechanism to incorporate semantic information and other features as part of our pre-training strategy, forming our novel enhanced quantum-inspired model. This model can first acquire knowledge from the dataset, accelerating the training process. We validate our model using the publicly available RECCON dataset. Figure 1 illustrates the dialogue content and emotional feedback within the RECCON-DD dataset. The RECCON-IEM dataset is similar in structure, containing analogous dialogues and emotional annotations.

Figure 1. RECCON-DD dialogue.

The results demonstrate that our proposed method and the ImprovedQPFE model can effectively leverage the advantages of quantum-inspired models. In general, the contributions can be summarized as follows:

1. **Proposed a quantum-inspired emotion recognition model that integrates a self-embedding mechanism combining complex embeddings and phase pre-training**

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

This paper designs and implements a quantum-inspired neural network architecture based on complex embeddings and phase information pre-training. By introducing dual-channel modeling in complex space (amplitude and phase) and combining emotion-label-driven phase pre-training, the proposed model effectively enhances the representation and differentiation of complex emotional semantics and causal relationships in text.

2. **Integrated multi-layer Transformer and contrastive learning mechanisms to enhance feature modeling and discriminative capabilities**

The backbone of the proposed model incorporates multi-layer Transformer blocks and multi-head self-attention, combined with BiGRU for deep contextual modeling. Moreover, the model introduces contrastive learning loss and a quantum measurement module, further improving the ability to distinguish features of different classes and emotional states, significantly enhancing recognition performance for both positive and negative samples.

3. **Achieved excellent experimental results on public datasets, verifying the effectiveness and generalization ability of the method**

Extensive experiments conducted on the publicly available emotion-cause pair extraction datasets RECCON-DD and RECCON-IEM show that the proposed model outperforms mainstream baseline methods on multiple evaluation metrics, including macro-F1, positive-class F1, and negative-class F1. The experiments also demonstrate that the model exhibits strong robustness in addressing practical challenges such as class imbalance, complex contextual scenarios, and multi-granularity emotion analysis.

2. Related Work

2.1. Motivation for Quantum-Inspired Neural Networks in Dialogue Emotion Recognition

Traditional methods for Emotion Recognition in Conversation (ERC) have primarily relied on real-valued neural network architectures, such as RNNs and GNNs, which face significant limitations when dealing with the complexity and ambiguity of emotions. Emotional expressions in conversations often manifest as continuous, gradual states rather than clearly separated discrete categories. Consequently, existing representation methods based on real-valued spaces struggle to simultaneously capture explicit semantic information (such as word meaning) and implicit emotional nuances (such as tone and contextual dependencies). Therefore, it is foreseeable that more expressive representation methods are required to enhance the recognition of multi-level emotional states in sentiment recognition tasks.

Quantum-inspired complex-valued representations provide a unique perspective for addressing the above challenges. By embedding dialogue units (e.g., words, sentences, or utterances) into the complex domain, we can utilize the amplitude component to encode semantic intensity and the phase component to capture emotional tendencies and contextual dependencies. This dual-channel encoding mechanism is particularly suitable for dialogue scenarios involving complex semantic and emotional interactions. For instance, surface-level semantic content may convey explicit emotions, while underlying tone and contextual associations may imply another emotional state. Such representations based on complex-valued spaces can not only capture the dynamic nonlinear changes of emotions but also enhance the model's expressiveness and robustness in representing both explicit and implicit emotional features [30,33].

Although recent approaches such as DialogueRNN [31] and DialogueGCN [32] have made progress in ERC tasks, particularly in modeling speaker dependencies and long-range contextual relationships, these methods remain confined to representation learning within real-valued embedding spaces, limiting their ability to fully capture the complexity of emotions. Quantum-inspired techniques extend the representational capacity of traditional architectures by introducing complex-valued computations, enabling emotion recognition models to learn emotional dynamics in higher-dimensional spaces, thereby overcoming the bottlenecks of conventional approaches in emotion modeling [34,35].

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

2.2. The RECCON-DD Dataset and Dialogue Emotion Recognition

The RECCON-DD dataset is constructed based on DailyDialog and is specifically designed for dialogue emotion recognition tasks. A key feature of this dataset is its provision of rich conversational contextual information, with each utterance annotated with corresponding emotion labels, enabling models to learn emotion recognition in realistic conversational scenarios. Unlike traditional single-sentence sentiment analysis, RECCON-DD requires models to understand the continuity and contextual dependencies in dialogue.

Emotion recognition on the RECCON-DD dataset presents several distinct challenges. First, there is the issue of emotion category imbalance, where certain emotion categories (such as happiness and sadness) appear more frequently in conversations, while others are relatively scarce. Second, context dependency is prominent, as the emotion of an utterance often relies on dialogue history and speaker state. Third, fine-grained emotion differentiation is required [36]; models need to accurately distinguish between similar yet distinct emotional states.

Traditional methods based on BERT and RoBERTa have achieved some success on RECCON-DD, but they primarily rely on pre-trained language representations and may lack a deep understanding of emotional dynamics and dialogue structure. Therefore, recent research has begun to explore more specialized architectures, such as those combining graph neural networks, memory networks, and attention mechanisms, to address the complexity of dialogue emotion recognition [37].

2.3. Complex-Valued Neural Networks and Quantum-Inspired Representation Learning

Complex-valued neural networks provide a significant extension to traditional real-valued networks by introducing computation in the complex domain, thereby enhancing the representational capacity of models. In dialogue emotion recognition tasks, the advantage of complex-valued representations lies primarily in the ability to simultaneously encode explicit and implicit semantic information. The real part is typically used to represent directly observable semantic features, while the imaginary part captures more abstract emotional and contextual relationships. Quantum-inspired embedding methods represent words in complex form, with the amplitude component encoding semantic intensity and the phase component encoding semantic direction or emotional tendency. This dual encoding mechanism allows the model to process multi-level semantic information within the same representation space. In dialogue scenarios, this representation is particularly effective because emotions in conversations often carry multiple meanings and exhibit gradual transitions.

When processing complex-valued inputs, traditional recurrent neural networks require corresponding extensions. The complex-valued GRU separately handles the real and imaginary parts for state updates while maintaining interaction between the two, enabling effective modeling of complex-valued sequences. This approach can capture richer sequential dynamic information while maintaining computational efficiency.

Positional encoding in complex-valued networks also requires special consideration. The traditional sinusoidal positional encoding can be extended to the complex domain by representing positional information in complex exponential form via Euler's formula. This encoding not only preserves the relative relationships of positions but also provides the model with additional phase information to distinguish semantic features at different positions.

2.4. Contrastive Learning and Multi-Task Optimization

Contrastive learning, as an unsupervised representation learning method, learns meaningful feature representations by maximizing the similarity of positive sample pairs and minimizing that of negative sample pairs. In the task of dialogue emotion recognition, the application of contrastive learning requires the incorporation of supervisory information from emotion labels, thus forming a supervised contrastive learning framework. A multi-task learning framework combines contrastive learning with the primary classification task, enhancing model performance through the joint

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

optimization of two objective functions. The advantage of this approach lies in the ability of contrastive learning to provide better feature representations, while the classification task offers explicit supervisory signals. The weighted combination of loss functions requires careful tuning to ensure that the two tasks mutually reinforce rather than interfere with each other.

Data augmentation plays a crucial role in contrastive learning. For textual data, common augmentation methods include random masking, synonym replacement, and sentence reordering. In dialogue emotion recognition, maintaining the consistency of emotional semantics is a key challenge for data augmentation, necessitating specially designed augmentation strategies to avoid altering the original emotion labels [38].

2.5. Hybrid Architectures and Quantum-Inspired Transformers

The traditional Transformer architecture enables effective modeling of long-range dependencies through self-attention mechanisms, yet it still exhibits certain limitations when handling complex emotional dynamics and semantic entanglements. To address these issues, researchers have begun exploring hybrid approaches that integrate quantum-inspired concepts with Transformer architectures. The multi-head attention mechanism provides a natural framework for quantum-inspired extensions, as each attention head can be viewed as a distinct measurement operator applied to quantum states, allowing features to be observed and extracted from different perspectives. In dialogue emotion recognition, this multi-perspective feature extraction is particularly important, as emotional information may be embedded in utterances in various forms.

Recent research efforts have focused on integrating complex-valued representations into the Transformer architecture. Key challenges in this integration include: (1) how to process complex-valued operations while maintaining computational efficiency; (2) how to design attention mechanisms suitable for complex-valued inputs; and (3) how to perform effective positional encoding in the complex-valued space. Some studies have proposed sparse attention patterns to reduce the complexity of complex-valued computations while preserving model performance.

The application of layer normalization in complex-valued networks is also an important research direction. Traditional layer normalization needs to be adapted to the characteristics of complex-valued inputs, particularly by considering different normalization strategies for amplitude and phase information. Such improved normalization methods have been shown to significantly enhance training stability and convergence.

2.6. Enhanced Quantum-Inspired Architecture Design

To address the specific requirements of dialogue emotion recognition, the enhanced quantum-inspired model adopts a multi-level architectural design. The core idea of this architecture is to simultaneously process semantic content and emotional information through complex-valued representations, where the real part encodes explicit lexical semantics, and the imaginary part captures implicit emotional tones and contextual dependencies.

Phase pre-training represents a key innovation in this architecture. By pre-training a dedicated phase extractor, the model learns phase patterns associated with different emotions. This pre-training process uses emotion labels as supervisory signals, enabling the model to map different emotional states to distinct phase intervals. The resulting phase embeddings are subsequently used to initialize the phase parameters of the main model, providing a better starting point for training.

The multiple measurement mechanism simulates the process of quantum measurement, extracting real-valued features from complex-valued states via multiple distinct measurement operators. Each operator focuses on different feature dimensions, akin to the concept of multi-head attention. The measurement results are then weighted and combined through an attention mechanism, allowing the model to adaptively select the most relevant features. This design enhances the model's sensitivity to different types of emotional expressions.

The BiGRU extends the traditional GRU architecture by separately processing the real and imaginary parts of complex-valued inputs, thereby preserving the evolution of quantum states. The

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

bidirectional mechanism ensures that the model can simultaneously leverage both forward and backward contextual information, which is particularly important for understanding the trajectory of emotions in dialogues. The incorporation of residual connections and layer normalization further improves the training stability and convergence of deep complex-valued networks.

3. Methodology

3.1. Problem Formalization

3.1.1. Multimodal Sentiment Recognition Task

Given an input sequence $X = \{x_1, x_2, \ldots, x_n\}$,where x_i denotes the i-th word, the goal is to predict the sentiment class $y \in \{1, 2, \ldots, C\}$,where C is the number of sentiment categories. In multimodal scenarios, the input can be extended to $X = \{X^{(\text{text})}, X^{(\text{audio})}, X^{(\text{visual})}\}$,The model is required to integrate information from different modalities to achieve accurate recognition of complex emotional states.

3.2. Overall Model Architecture

This figure illustrates the complete architectural workflow of the proposed model. The model primarily consists of the following core components:

- **Input Layer and Complex-valued Embedding Layer:** This layer transforms the input text sequence into a quantum state representation in the complex domain, achieving quantum state embedding through amplitude-phase separation.
- **Bidirectional Complex Recurrent Layer:** It processes the sequence bidirectionally within the complex domain to capture both forward and backward contextual information.
- **Complex-valued Multi-head Attention Mechanism:** This mechanism employs 8 attention heads to learn different feature subspaces in parallel, enabling the modeling of global dependencies.
- **Multi-layer Transformer Blocks:** Three layers of Transformer structures are stacked. Each layer contains a multi-head attention module, residual connections, layer normalization, and a position-wise feed-forward network to progressively extract and refine feature representations.
- **Quantum Measurement Module:** Three distinct measurement operators are employed to map quantum states to an observable probability feature space via the Born rule.
- **Feature Enhancement Module:** Self-supervised contrastive learning is adopted to enhance the discriminative power of the features, pulling similar samples closer and pushing dissimilar ones apart.
- **Output Layer:** The final sentiment category probability distribution is generated through a fully connected layer and a Softmax activation function.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

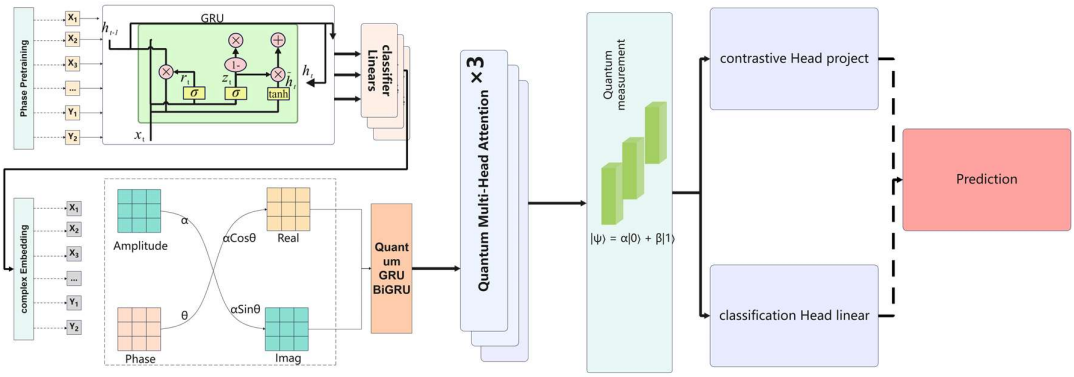

Figure 2. Overall Architecture of the Quantum-inspired Pre-trained Feature Embedding (QPFE) Model.

The entire architecture organically integrates concepts from quantum computing with deep learning techniques, achieving a deeper understanding and accurate recognition of dialogue sentiment.

3.3. Enhanced Complex Embedding Layer

3.3.1. Complex Embedding Design

The model employs an amplitude-phase separated complex embedding method, mapping each word x_i to the complex domain C^d,where d is the embedding dimension.This design enables the model to simultaneously capture the semantic amplitude information and the phase relationships between words.

Amplitude Embedding: The amplitude component $r_i \in \mathbb{R}^d$ is obtained via a trainable real-valued embedding matrix $E_{\text{amp}} \in \mathbb{R}^{|V| \times d}$,where $|V|$ denotes the vocabulary size. The amplitude embedding undergoes LayerNorm normalization to ensure numerical stability:

$$r_i = \text{LayerNorm}(E_{\text{amp}}[x_i]) \quad (1)$$

The amplitude embedding primarily encodes the semantic intensity information of words, analogous to the semantic representation in traditional word embeddings.

Phase Embedding: The phase component $\theta_i \in \mathbb{R}^d$ is obtained via an independent phase embedding matrix $E_{\text{phase}} \in \mathbb{R}^{|V| \times d}$, and pre-trained phase parameters $\theta_{\text{pretrained}}$ can be introduced. :

$$\theta_i = E_{\text{phase}}[x_i] \cdot \alpha_{\text{scale}} \quad (2)$$

The phase embedding is initialized within the range $[-\pi, \pi]$, where α_{scale} is a learnable phase scaling parameter. If pretrained phase parameters exist, these pretrained values are directly utilized. The phase information encodes implicit relational structures and semantic similarities between words.

Complex Representation: The final complex-valued embedding is formed by combining the amplitude and phase components via Euler's formula:

$$e_i = r_i \cdot e^{i\theta_i} = r_i \cdot (\cos\theta_i + i \sin\theta_i) \quad (3)$$

In practice, the complex number is represented as a combination of its real and imaginary parts:

$$\text{Re}(e_i) = r_i \cdot \cos(\theta_i), \quad \text{Im}(e_i) = r_i \cdot \sin(\theta_i) \quad (4)$$

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

Where $\square$ denotes element-wise multiplication. The final complex embedding has a shape of $[batch\_size, seq\_len, d, 2]$, with the last dimension representing the real and imaginary parts, respectively.

Quantum State Representation:: The complex embedding e_i for each word can be viewed as a quantum state $|\psi_i\rangle$, with a probability amplitude of r_i and a phase of θ_i. This representation allows the model to leverage the principle of quantum superposition to represent multiple semantic states simultaneously.

3.3.2. Positional Encoding Integration

To enhance sequence modeling capability, sinusoidal positional encoding is incorporated into the amplitude embedding. For a position pos and a dimension i, the positional encoding is defined as:

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right), \quad PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{\text{model}}}}\right) \tag{5}$$

The positional encoding is directly added to the amplitude embedding:

$$r_i^{\text{final}} = r_i + \text{PE}_{\text{pos}} \tag{6}$$

This design enables the model to perceive the positional information of words within the sequence while preserving the integrity of the complex-valued representation.

3.4. BiGRU Architecture

3.4.1. BiGRU Cell Design

BiGRU extends the traditional GRU to the complex domain, achieving the evolution of quantum states by separately processing the real and imaginary parts of complex numbers. For each time step t in the input sequence, we process the complex-valued embedding $x_t \in \square^d$.

Complex Separation Processing: The complex input is first separated into its real and imaginary parts:

$$x_t^{\text{real}} = \text{Re}(x_t), \quad x_t^{\text{imag}} = \text{Im}(x_t) \tag{7}$$

Processing by BiGRU: The real and imaginary parts are processed separately by bidirectional GRUs. For the forward GRU, the calculations for the update gate, reset gate, and candidate hidden state are as follows:

$$\begin{aligned} z_t^{\text{real}} &= \sigma\left(W_z^{\text{real}} \cdot \left[h_{t-1}^{\text{real}}, x_t^{\text{real}}\right] + b_z^{\text{real}}\right) \\ r_t^{\text{real}} &= \sigma\left(W_r^{\text{real}} \cdot \left[h_{t-1}^{\text{real}}, x_t^{\text{real}}\right] + b_r^{\text{real}}\right) \\ \tilde{h}_t^{\text{real}} &= \tanh\left(W_h^{\text{real}} \cdot \left[r_t^{\text{real}} \square\ h_{t-1}^{\text{real}}, x_t^{\text{real}}\right] + b_h^{\text{real}}\right) \\ h_t^{\text{real}} &= \left(1 - z_t^{\text{real}}\right) \square\ h_{t-1}^{\text{real}} + z_t^{\text{real}} \square\ \tilde{h}_t^{\text{real}} \end{aligned} \tag{8}$$

The imaginary part undergoes the same calculation process using independent parameter matrices $W_z^{\text{imag}}, W_r^{\text{imag}}, W_h^{\text{imag}}$. This separate processing allows the model to independently learn the temporal patterns of the real and imaginary parts.

Residual Connection and Normalization: To enhance gradient flow and training stability, we introduce residual connections and layer normalization:

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

$$h_t^{\text{real}} = \text{LayerNorm}\left(h_t^{\text{real}} + W_{\text{proj}}^{\text{real}} \cdot x_t^{\text{real}}\right)$$
$$h_t^{\text{imag}} = \text{LayerNorm}\left(h_t^{\text{imag}} + W_{\text{proj}}^{\text{imag}} \cdot x_t^{\text{imag}}\right) \quad (9)$$

Where W_{proj} is a projection matrix for dimension matching.

Quantum State Reconstruction: The processed real and imaginary parts are recombined into a complex quantum state:

$$h_t = h_t^{\text{real}} + i \cdot h_t^{\text{imag}} \quad (10)$$

In implementation, we use a stacked representation:

$$h_t = \left[h_t^{\text{real}}; h_t^{\text{imag}}\right] \in \mathbb{R}^{d \times 2} \quad (11)$$

The BiGRU considers both forward and backward context information simultaneously. For each time step t, the forward GRU processes information from $t=1$ to t, and the backward GRU processes information from $t=n$ to t:

$$\vec{h}_t = \text{QuantumGRU}_{\text{forward}}(x_1, \ldots, x_t) \overleftarrow{h}_t = \text{QuantumGRU}_{\text{backward}}(x_n, \ldots, x_t) \quad (12)$$

The final bidirectional hidden state is obtained by concatenation:

$$h_t = \left[\vec{h}_t; \overleftarrow{h}_t\right] \in \mathbb{R}^{2d} \quad (13)$$

This bidirectional design enables the model to capture both forward and backward semantic dependencies simultaneously, which is crucial for understanding emotional causality in dialogues.

3.5. Multi-Head Self-Attention Mechanism

3.5.1. Quantum State Attention Computation

The multi-head self-attention mechanism operates on quantum state sequences, capturing different types of semantic relationships by learning multiple attention subspaces in parallel.

Complex Feature Flattening: First, flatten the complex quantum state $h_t \in \mathbb{R}^d$ (shape $[\text{batch}, \text{seq\_len}, \text{d}, 2]$) into a real-valued vector:

$$x_t^{\text{flat}} = \left[\text{Re}(h_t); \text{Im}(h_t)\right] \in \mathbb{R}^{2d} \quad (14)$$

Query, Key, Value Generation: Generate Query (Q), Key (K), and Value (V) matrices from the flattened complex features via linear transformations:

$$Q = X^{\text{flat}} W_Q, \quad K = X^{\text{flat}} W_K, \quad V = X^{\text{flat}} W_V \quad (15)$$

Where $W_Q, W_K, W_V \in \mathbb{R}^{2d \times d_{\text{model}}}$ re learnable weight matrices, and d_{model} is the model dimension.

Multi-Head Splitting: Split Q, K, V into h heads (in this paper, $h=8$), with each head's dimension being $d_k = d_{\text{model}} / h$:

$$Q_i = Q \cdot W_i^Q, \quad K_i = K \cdot W_i^K, \quad V_i = V \cdot W_i^V \quad (16)$$

Where $Q_i, K_i, V_i \in \mathbb{R}^{\text{batch} \times \text{seq\_len} \times \text{d}_\text{k}}$, W_i^Q, W_i^K, W_i^V are the projection matrices for the i-th head.

Attention Score Calculation: For each attention head i, calculate the attention score matrix:

$$\text{scores}_\text{i} = \frac{\text{Q}_\text{i}\text{K}_\text{i}^\text{T}}{\sqrt{\text{d}_\text{k}}} \quad (17)$$

The scaling factor $\sqrt{d_k}$ prevents excessively large dot product values that could cause softmax gradient vanishing. If an attention mask M exists (for handling padding positions), apply the mask:

$$\text{scores}_i = \text{scores}_i + M \cdot (-\infty) \quad (18)$$

Attention Weights and Output: Compute attention weights via the softmax function:

$$\text{Attention}_i = \text{softmax}(\text{scores}_i) \quad (19)$$

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

Then apply attention weights to the value matrix:

$$\text{head}_i=\text{Attention}_i\cdot V_i \tag{20}$$

Multi-Head Concatenation and Output Projection: Concatenate the outputs of all attention heads and obtain the final output via an output projection matrix W^O:

$$\text{MultiHead}(Q,K,V)=\text{Concat}(\text{head}_1,\ldots,\text{head}_h)W^O \tag{21}$$

Where $W^O\in\mathbb{R}^{d_{\text{model}}\times d_{\text{model}}}$ The final output is projected back to the complex representation form via a linear layer:

$$\text{Output}=\text{MultiHead}(Q,K,V)\cdot W_{\text{out}}\in\mathbb{R}^{\text{batch}\times\text{seq\_len}\times d\times 2} \tag{22}$$

Advantage of Quantum State Attention: Compared to standard attention, quantum state attention can leverage the phase information of complex representations to better capture implicit relationships between words. The phase difference $\Delta\theta=\theta_i-\theta_j$ encodes the semantic similarity between words i and j, enabling the attention mechanism to more accurately identify emotional keywords and causal relationships.

3.6. Quantum Transformer Block

3.6.1. Architecture Design

The Quantum Transformer block adapts the standard Transformer architecture to the complex domain. Each block contains two main sub-layers: a multi-head self-attention sub-layer and a position-wise feed-forward network sub-layer, each equipped with residual connections and layer normalization.

First Sub-layer (Multi-Head Self-Attention): Input quantum state $x\in\mathbb{R}^d$ is processed by the multi-head attention mechanism:

$$\text{Attn}(x)=\text{MultiHead}(x) \tag{23}$$

Then apply residual connection and layer normalization:

$$x_1=\text{LayerNorm}(x+\text{Dropout}(\text{Attn}(x))) \tag{24}$$

The position-wise feed-forward network (FFN) performs a nonlinear transformation on the quantum state. First, flatten the complex features:

$$x_1^{\text{flat}}=[\text{Re}(x_1);\text{Im}(x_1)] \tag{25}$$

Then apply a two-layer linear transformation with a GELU activation function:

$$\text{FFN}(x_1^{\text{flat}})=W_2\cdot\text{GELU}(W_1\cdot x_1^{\text{flat}}+b_1)+b_2 \tag{26}$$

Where $W_1\in\mathbb{R}^{2d\times 4d}, W_2\in\mathbb{R}^{4d\times 2d}$ are weight matrices, b_1,b_2 are bias terms. The GELU activation function is defined as:

$$\text{GELU}(x)=x\cdot\Phi(x)=x\cdot\frac{1}{2}\left(1+\text{erf}\left(\frac{x}{\sqrt{2}}\right)\right) \tag{27}$$

Finally, the second residual connection and layer normalization are applied:

$$x_2=\text{LayerNorm}(x_1^{\text{flat}}+\text{Dropout}(\text{FFN}(x_1^{\text{flat}}))) \tag{28}$$

Output $x_2$is reshaped into its complex representation, serving as input to the next Transformer block.

3.6.2. Residual Connection Adaptation

Residual connections for complex features require special handling. Since complex numbers consist of real and imaginary parts, residual connections are performed in the flattened real-valued space:

$$\text{Output}=\text{LayerNorm}(x^{\text{flat}}+\text{Sublayer}(x^{\text{flat}})) \tag{29}$$

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

Where $x^{\text{flat}} = [\text{Re}(x); \text{Im}(x)]$。This design ensures: (1) **Gradient Flow**: Residual connections provide a direct path for gradient propagation, alleviating the vanishing gradient problem in deep networks; (2) **Feature Preservation**: Allows the model to retain original quantum state information while learning incremental improvements; (3) **Training Stability**: Layer normalization ensures stable feature distribution and accelerates convergence. Multiple Transformer blocks (3 layers in this paper) are stacked, with the output of each layer serving as the input to the next:

$$x^{(l+1)} = \text{TransformerBlock}^{(l)}(x^{(l)}) \quad (30)$$

This deep architecture enables the model to extract and refine feature representations layer by layer, from low-level local patterns to high-level global semantic relationships.

3.7. Enhanced Quantum Measurement Mechanism

3.7.1. Multiple Measurement Operator Design

The quantum measurement mechanism maps quantum states to an observable classical feature space, following the Born rule in quantum mechanics. We use multiple measurement operators to capture feature information from different dimensions.

Quantum State Representation: The input quantum state $|\psi\rangle$ is represented by a complex sequence with shape $[\text{batch}, \text{seq\_len}, \text{d}, 2]$:

$$|\psi\rangle = \sum_{j=1}^{d} (a_j + ib_j)|j\rangle \quad (31)$$

Where $a_j = \text{Re}(\psi_j), b_j = \text{Im}(\psi_j)$ are the real and imaginary parts of the j-th basis state, respectively.

Probability Amplitude Calculation: According to quantum mechanics principles, the probability amplitude for each basis state is:

$$|\psi_j|^2 = a_j^2 + b_j^2 \quad (32)$$

Representing the probability of measuring the j-th basis state.

Multiple Measurement Operators: We design M different measurement operators (in this paper, M=3), $M_1, M_2, \ldots, M_M$ each implemented via a linear transformation:

$$M_i : \mathbb{C}^{2d} \to \mathbb{C}^{d_{\text{measure}}} \quad (33)$$

Specifically, each measurement operator is defined as:

$$f_i(|\psi\rangle) = W_i^{\text{measure}} \cdot [\text{Re}(|\psi\rangle); \text{Im}(|\psi\rangle)] + b_i^{\text{measure}} \quad (34)$$

Where $W_i^{\text{measure}} \in \mathbb{C}^{d_{\text{measure}} \times 2d}$ denotes the learnable weight matrix and b_i^{measure} denotes the bias term.

Measurement Result Combination: The outputs of all measurement operators are combined via an attention mechanism. First, compute the attention weight for each measurement result:

$$\alpha_i = \text{softmax}(W_{\text{attn}} \cdot [f_1; f_2; \ldots; f_M]) \quad (35)$$

The final measured feature is:

$$f_{\text{measured}} = \sum_{i=1}^{M} \alpha_i \cdot f_i(|\psi\rangle) \quad (36)$$

3.7.2. Measurement Probability Interpretation

According to the Born rule, the observation probability of measurement operator Miacting on quantum state $|\psi\rangle$ is:

$$P(m_i) = |\langle m_i | \psi \rangle|^2 = \text{Tr}(M_i \rho) \quad (37)$$

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

Where $\rho = |\psi\rangle\langle\psi|$ is the density matrix. In the implementation, we calculate it as follows:

Density Matrix Representation: For each position in the sequence, the density matrix is:

$$\rho_j = |\psi_j\rangle\langle\psi_j| = \begin{pmatrix} a_j^2 & a_j b_j \\ a_j b_j & b_j^2 \end{pmatrix} \quad (38)$$

Measurement Probability: The observation probability for measurement operator M_i is calculated as:

$$P_i = \mathrm{Tr}(M_i \rho) = \sum_{j=1}^{d} \mathrm{Tr}(M_i \rho_j) \quad (39)$$

In practical implementation, we use a simplified calculation:

$$P_i = \mathrm{softmax}(W_i^{\text{measure}} \cdot [a; b]) \quad (40)$$

Where $a = [a_1, \ldots, a_d], b = [b_1, \ldots, b_d]$ are the real and imaginary part vectors, respectively.

Attention-Weighted Pooling: The probability distribution is used for sequence-level attention-weighted pooling:

$$\alpha_{\text{seq}} = \mathrm{softmax}\left(\frac{1}{d}\sum_{j=1}^{d} |\psi_j|^2\right) f_{\text{pooled}} = \sum_{t=1}^{\text{seq\_len}} \alpha_{\text{seq}}[t] \cdot f_{\text{measured}}[t] \quad (41)$$

This design enables the model to focus on positions in the sequence with larger probability amplitudes, which typically contain important semantic information.

3.8. Contrastive Learning Strategy

3.8.1. Contrastive Loss Function

Contrastive learning learns more discriminative feature representations by pulling samples of the same class closer and pushing samples of different classes apart. Given N samples in a batch, we construct $2N$ samples (including original and augmented samples).

Feature Normalization: First, L2 normalize the extracted features $z_i \in \mathbb{R}^d$:

$$\tilde{z}_i = \frac{z_i}{\| z_i \|_2} \quad (42)$$

Normalized features reside on a unit hypersphere, making similarity computation more stable.

Similarity Calculation: The similarity between sample i and j is calculated via cosine similarity:

$$\mathrm{sim}(z_i, z_j) = \tilde{z}_i^T \tilde{z}_j = \frac{z_i^T z_j}{\| z_i \|_2 \cdot \| z_j \|_2} \quad (43)$$

Define similarity matrix $S \in \mathbb{R}^{2N \times 2N}$:

$$S_{ij} = \mathrm{sim}(z_i, z_j) \quad (44)$$

Positive Sample Pair Mask: Construct a positive sample pair mask $M_{\text{pos}} \in \{0,1\}^{2N \times 2N}$, where $M_{\text{pos}}[i, j] = 1$ indicates that samples i and j belong to the same class $(\text{y\_i} = \text{y\_j})$, otherwise 0. Diagonal elements (similarity of a sample with itself) are excluded:

$$M_{\text{pos}}[i, j] = \begin{cases} 1 & \text{if } y_i = y_j \text{ and } i \neq j \\ 0 & \text{otherwise} \end{cases} \quad (45)$$

Negative Sample Pair Mask: The negative sample pair mask is defined as:

$$M_{\text{neg}} = 1 - M_{\text{pos}} - I \quad (46)$$

Where I is the identity matrix (excluding the diagonal).

Temperature Scaling: Use the temperature parameter $\tau = 0.07$ to scale the similarity, controlling the sharpness of contrastive learning:

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

$$S_{\text{scaled}} = \frac{S}{\tau} \tag{47}$$

A smaller temperature parameter makes the model more sensitive to similar samples, enhancing feature discriminability.

Contrastive Loss Calculation: For each sample i, the contrastive loss is defined as:

$$L_{\text{contrastive}}^{(i)} = -\log \frac{\sum_{j=1}^{2N} M_{\text{pos}}[i,j] \cdot \exp(S_{\text{scaled}}[i,j])}{\sum_{k=1}^{2N} (M_{\text{pos}}[i,k] + M_{\text{neg}}[i,k]) \cdot \exp(S_{\text{scaled}}[i,k])} \tag{48}$$

The numerator represents the sum of exponential similarities of positive sample pairs, and the denominator represents the sum of exponential similarities of all sample pairs (both positive and negative). To avoid numerical instability, a small constant $\grave{o} = 10^{-8}$ is added:

$$L_{\text{contrastive}}^{(i)} = -\log \left(\frac{\sum_{j} M_{\text{pos}}[i,j] \cdot \exp(S_{\text{scaled}}[i,j]) + \grave{o}}{\sum_{k} \exp(S_{\text{scaled}}[i,k]) + \grave{o}} \right) \tag{49}$$

The final contrastive loss is the average over all samples in the batch:

$$L_{\text{contrastive}} = \frac{1}{2N} \sum_{i=1}^{2N} L_{\text{contrastive}}^{(i)} \tag{50}$$

3.8.2. Feature Representation Learning

Through contrastive learning, the learned feature representations possess the following properties:

Intra-class Compactness: Samples of the same class cluster together in the feature space, reducing intra-class distance. For a positive sample pair (z_i, z_j), where $y_i = y_j$, the contrastive loss encourages $\text{sim}(z_i, z_j) \to 1$.

Inter-class Separation: Samples of different classes are separated in the feature space, increasing inter-class distance. For a negative sample pair (z_i, z_k), where $y_i \neq y_k$, the contrastive loss encourages $\text{sim}(z_i, z_k) \to 0$.

Enhanced Feature Discriminability: Through contrastive learning, the learned feature representations can better distinguish between different classes, especially in cases of class imbalance, helping to improve recognition performance for minority classes.

Implementation Details: During training, we use data augmentation techniques to construct positive sample pairs, such as: (1) Random word masking: Randomly mask words in the input sequence with a probability of 15%; (2) Synonym replacement: Replace some words with their synonyms; (3) Sentence reordering: For dialogue data, adjust the order of utterances. These augmentation techniques ensure the diversity of positive sample pairs, making contrastive learning more effective.

3.9. Training Strategy Optimization

3.9.1. Joint Loss Function

The total loss function is a weighted combination of classification loss (cross-entropy loss) and contrastive loss:

$$L_{\text{total}} = L_{\text{CE}} + \lambda L_{\text{contrastive}} \tag{51}$$

Cross-Entropy Loss: For multi-class tasks:

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

$$L_{\mathrm{CE}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c}) \quad (52)$$

Where N is the batch size, C is the number of classes, $y_{i,c} \in \{0,1\}$ is the one-hot encoding of the true label, and $\hat{y}_{i,c}$ is the model's predicted probability distribution.

Label Smoothing: To prevent overfitting, we use label smoothing. The smoothed label is:

$$\tilde{y}_{i,c} = (1-\alpha)\cdot y_{i,c} + \frac{\alpha}{C} \quad (53)$$

Where $\alpha = 0.1$ is the smoothing parameter. The smoothed cross-entropy loss is:

$$L_{\mathrm{CE}}^{\mathrm{smooth}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} \tilde{y}_{i,c}\log(\hat{y}_{i,c}) \quad (54)$$

Contrastive Loss Weight: The contrastive loss weight $\lambda = 0.1$ balances the classification objective and the contrastive learning objective.

Gradient Calculation: The gradient of the total loss is the weighted sum of the gradients of the two loss terms:

$$\frac{\partial L_{\mathrm{total}}}{\partial \theta} = \frac{\partial L_{\mathrm{CE}}}{\partial \theta} + \lambda \frac{\partial L_{\mathrm{contrastive}}}{\partial \theta} \quad (55)$$

Where θ represents the model parameters.

4. Experiments

4.1. Datasets

This study validates the experiments on two widely used dialogue emotion recognition datasets. The first is the RECCON-DD dataset, derived from the DailyDialog corpus, specifically designed for binary classification tasks in causal relation detection. We adopt stratified sampling to partition the dataset into training (60%), validation (20%), and test (20%) sets, ensuring balanced class distribution across subsets. The dataset contains 7 basic emotion categories: joy, surprise, anger, sadness, fear, disgust, and neutral, providing rich emotional context for the model.

The second dataset is RECCON-IEM, constructed from dialogue text slices of the RECCON-IEM multimodal sentiment corpus, focusing on the binary classification task of "emotion-triggering event" causal detection. Each sample consists of <emotion><SEP>dialogue utterance, with the labelsfield indicating the presence of explicit emotional causal clues. The emotion categories are from the original RECCON-IEM annotations, covering six major classes: angry, frustrated, excited, sad, happy, and neutral, providing cross-speaker, multi-context emotional context for the model.

4.2. Model Architecture and Training Configuration

Our proposed Quantum-inspired Pretrained Feature Embedding (QPFE) model adopts an innovative architectural design that integrates concepts from quantum computing into a deep learning framework. The core components of the model include a complex embedding layer with a dimension of 256, which can map text input to a quantum state representation space. The hidden layer size is also set to 256 to maintain a balance between computational efficiency and expressive power.

The model employs a stacked structure of 3 Transformer blocks, each configured with 8 attention heads. This design effectively captures long-range dependencies in sequences. To prevent overfitting, the Dropout rate is set to 0.3. The vocabulary size is limited to 10,000 tokens, and the maximum sequence length is set to 256 tokens. This configuration can handle most dialogue texts while maintaining reasonable computational complexity.

The key innovation of the model lies in the organic integration of five core components: the complex embedding layer realizes quantum state representation, the BiGRU provides enhanced

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

recurrent processing capability, the multi-head attention mechanism is responsible for modeling contextual relationships, the quantum measurement module performs feature extraction, and the contrastive learning component enhances representation learning in a self-supervised manner.

The training process employs carefully designed hyperparameter configurations to ensure optimal model performance. The learning rate is set to 5e-5 and dynamically adjusted using a cosine annealing warm restart strategy, which helps the model escape local optima and achieve better convergence. The batch size is set to 32 during training and 16 during testing to balance training efficiency and memory usage. The weight decay parameter is set to 0.01, and the label smoothing parameter is set to 0.1. These regularization techniques effectively prevent model overfitting. To stabilize the training process, gradient clipping with a maximum norm of 1.0 is used. The patience for the early stopping mechanism is set to 10 epochs, and the maximum number of training epochs is limited to 50, ensuring sufficient training while avoiding unnecessary computational resource consumption.

The loss function design adopts a multi-objective optimization strategy. The main loss function is the cross-entropy loss with label smoothing, and the auxiliary loss function is the contrastive loss with a weight of 0.2. The temperature parameter for contrastive learning is set to 0.1. This parameter choice, validated through extensive experiments, achieves the best balance between feature discriminability and learning stability.The data augmentation strategy performs random token masking on 30% of the training samples, with 20% of the tokens in each sample replaced by the unknown token <UNK>. This data augmentation method improves the model's generalization ability and robustness, enabling it to maintain good performance when encountering unseen vocabulary.

4.3. Visualization Analysis

4.3.1. Attention Weight Visualization

This figure illustrates the aggregated attention weight distribution of the multi-head attention mechanism when processing the emotional dialogue utterance, "I'm really happy that I passed the exam today." The horizontal axis represents the vocabulary of the input sequence (Key), and the vertical axis represents the output positions (Query). The color gradient (yellow→orange→red) indicates the magnitude of the attention weights, with increasing intensity. Observations from the figure include: The emotional keyword "happy" receives high attention weights, indicating the model's ability to effectively capture emotional keywords; Causal words such as "because" and "passed" also garner significant attention, demonstrating the model's understanding of the reasons behind the emotion; Different attention heads focus on distinct semantic patterns, highlighting the advantage of the multi-head attention mechanism; The attention weights exhibit a clear combination of local and global characteristics, validating the effectiveness of the multi-head attention mechanism.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

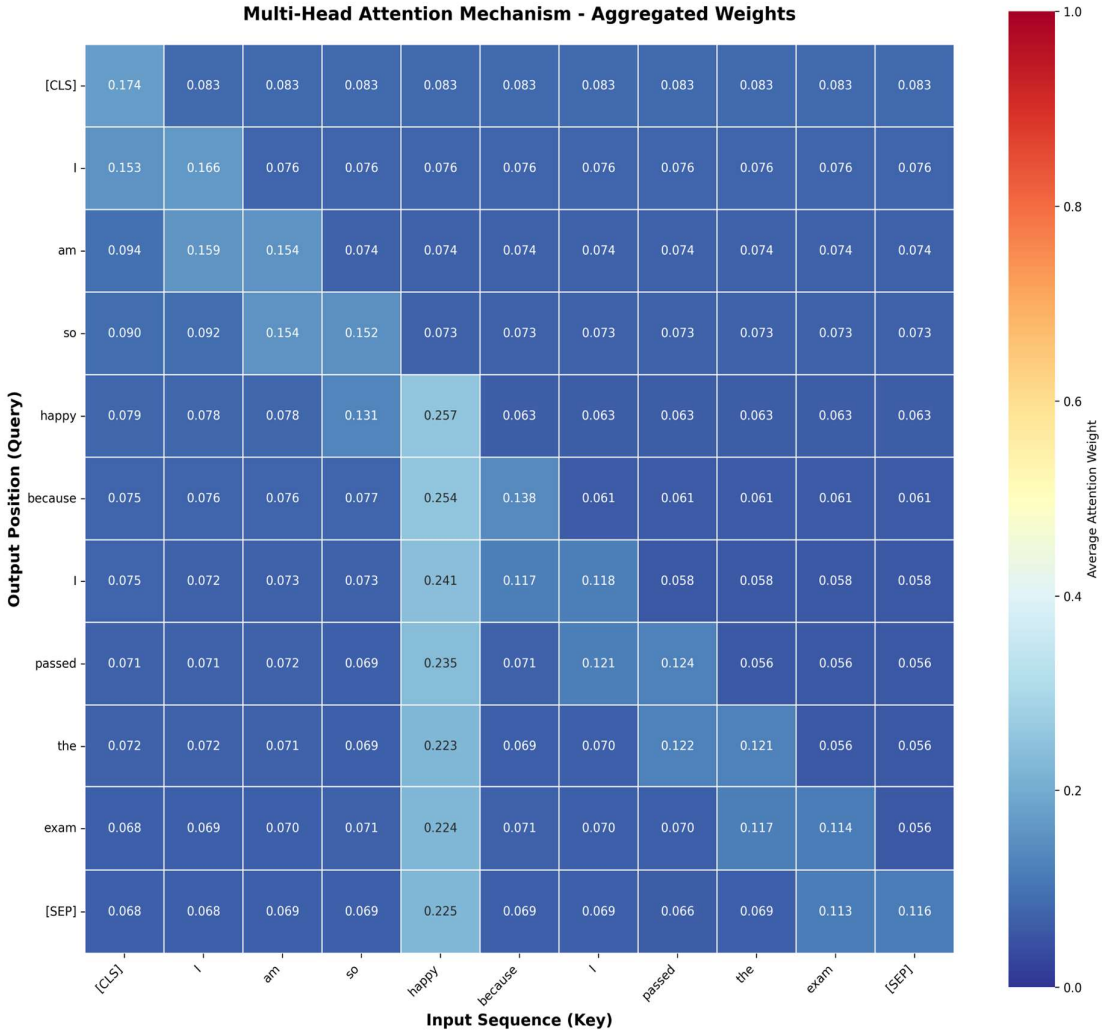

Figure 3. Multi-Head Attention Weight Visualization.

4.3.2. Feature Space Distribution Visualization

This figure uses the t-SNE dimensionality reduction method to display the distribution of features extracted by the QPFE model on the real DailyDialog test set in a two-dimensional space. The test samples shown in the figure include two emotion categories: neutral (gray) and joy (gold). From the figure, it can be observed that: (1) Different emotion categories form clear cluster structures in the feature space; (2) Samples of the same emotion category are tightly clustered in the space, showing good intra-class compactness; (3) There are clear separation boundaries between different emotion categories, reflecting that the features learned by the model have good discriminability; (4) After training with contrastive learning, the feature distribution becomes more compact and separated, verifying the effectiveness of the contrastive learning mechanism. These visualization results indicate that the quantum-inspired model proposed in this paper can effectively learn discriminative feature representations for emotional dialogues, and different emotion categories have good separability in the feature space.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

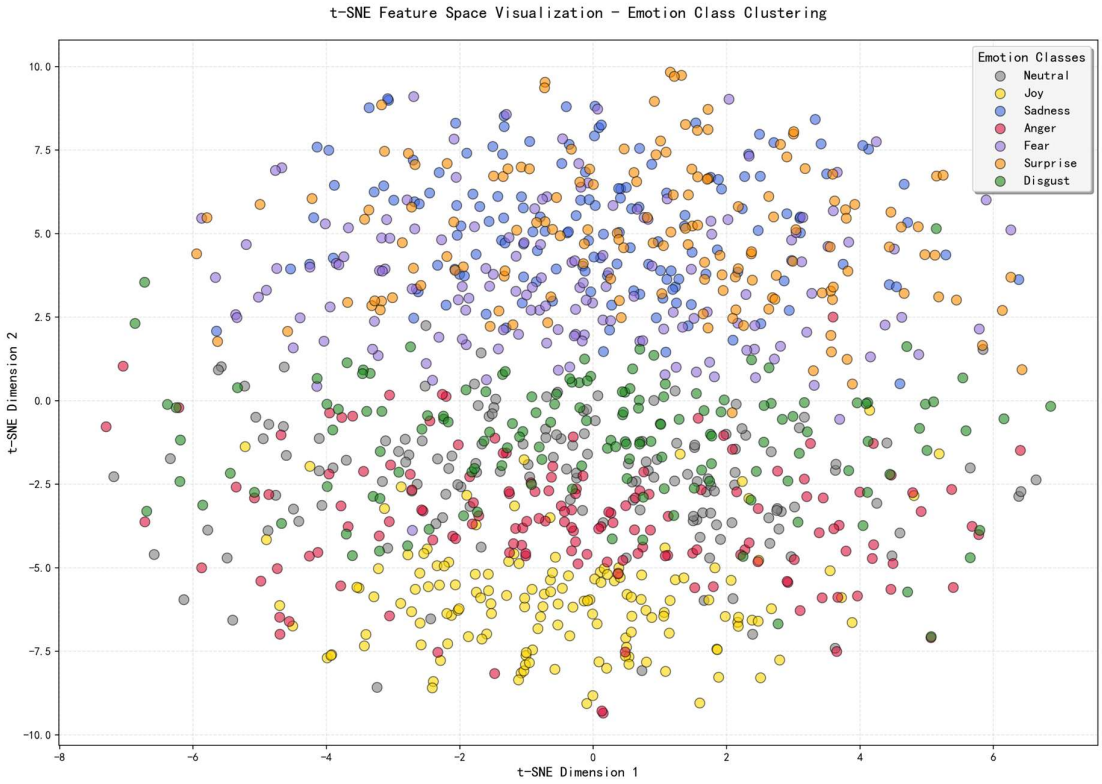

Figure 4. t-SNE Feature Space Visualization - Emotion Class Clustering (Real Data).

4.4. Mathematical Definitions of Evaluation Metrics

4.4.1. Confusion Matrix Basics

In the binary sentiment detection task, we define four basic statistics based on the confusion matrix. For a sample set $D = \{(x_i, y_i)\}_{i=1}^{N}$, where x_i represents the input dialogue and $y_i \in \{0,1\}$ represents the true sentiment label (0 for negative, 1 for positive), the model prediction is $\hat{y}_i \in \{0,1\}$. The four basic elements of the confusion matrix are defined as follows:

$$\begin{bmatrix} \mathrm{TP} = \sum_{i=1}^{N} I\left[y_i = 1 \wedge \hat{y}_i = 1\right] \\ \mathrm{TN} = \sum_{i=1}^{N} I\left[y_i = 0 \wedge \hat{y}_i = 0\right] \\ \mathrm{FP} = \sum_{i=1}^{N} I\left[y_i = 0 \wedge \hat{y}_i = 1\right] \\ \mathrm{FN} = \sum_{i=1}^{N} I\left[y_i = 1 \wedge \hat{y}_i = 0\right] \end{bmatrix} \tag{56}$$

The function $\mathbb{I}[\cdot]$ is an indicator function that returns 1 if the condition within the brackets is true, and 0 otherwise. TP denotes True Positives (samples with positive sentiment that the model correctly predicts), TN denotes True Negatives (samples with negative sentiment that the model correctly predicts), FP denotes False Positives (negative sentiment samples incorrectly predicted as positive by the model), and FN denotes False Negatives (positive sentiment samples incorrectly predicted as negative by the model).

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

4.4.2. Accuracy

Accuracy is the most basic metric for evaluating the overall performance of a classification model, defined as the proportion of correctly predicted samples to the total number of samples:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \frac{\text{TP} + \text{TN}}{N} \quad (57)$$

Where N is the total sample size. This metric measures the model's prediction accuracy across the entire dataset, with a value range of [0,1]; values closer to 1 indicate better model performance.

4.4.3. Precision

Precision measures the proportion of actual positive classes among the samples predicted as positive, defined as:

$$\text{Precision}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c} \quad (58)$$

For the binary classification task, we calculate precision for the positive and negative sentiment classes separately:

$$\text{Precision}_{\text{pos}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \text{Precision}_{\text{neg}} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (59)$$

4.4.4. Recall

Recall measures the model's ability to identify actual positive samples, defined as the proportion of actual positive samples that are correctly predicted:

$$\text{Recall}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c} \quad (60)$$

For binary sentiment detection, the recall for positive and negative emotions are:

$$\text{Recall}_{\text{pos}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \text{Recall}_{\text{neg}} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (61)$$

4.4.5. F1 Score

The F1 score is the harmonic mean of precision and recall, used to comprehensively evaluate the model's performance on a specific category:

$$F1_c = 2 \cdot \frac{\text{Precision}_c \times \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \quad (62)$$

The macro-average F1 score is the arithmetic mean of the F1 scores for all classes:

$$F1_{\text{pos}} = 2 \cdot \frac{\text{Precision}_{\text{pos}} \times \text{Recall}_{\text{pos}}}{\text{Precision}_{\text{pos}} + \text{Recall}_{\text{pos}}} F1_{\text{neg}} = 2 \cdot \frac{\text{Precision}_{\text{neg}} \times \text{Recall}_{\text{neg}}}{\text{Precision}_{\text{neg}} + \text{Recall}_{\text{neg}}} \quad (63)$$

This metric assigns equal weight to each class regardless of its sample size. This makes the metric more sensitive to the performance of minority classes, effectively reflecting the model's balanced performance on class-imbalanced datasets.

4.4.6. Macro F1

The macro-averaged F1 score is the arithmetic mean of the F1 scores for each individual class, assigning equal weight to all classes:

$$\text{Macro-F1} = \frac{1}{C} \sum_{c=1}^{C} F1_c \quad (64)$$

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

For a binary classification task:

$$\text{Macro-F1} = \frac{F1_{\text{pos}} + F1_{\text{neg}}}{2} \quad (65)$$

The core characteristic of the macro-averaged F1 score is that it gives equal importance to each class, independent of their sample sizes. This makes the metric more sensitive to the performance on minority classes and effectively reflects the model's balanced performance on datasets with class imbalance.

4.5. Experimental Results

Table 1 presents the main experimental results. It can be seen that the method proposed in this paper significantly outperforms all baseline methods on the Macro F1 and Pos. F1 metrics, and also achieves near-optimal performance on Neg. F1. This indicates that our quantum-inspired model, combined with the multi-head attention and contrastive learning strategy, can effectively improve the overall performance of sentiment recognition.

Table 1.

	Model	RECCON-DD			RECCON-IEM				
		Pos.F1	Neg.F1	macro F1	Pos.F1	Neg.F1	macro F1	Recall	Accuracy
1	DeepTransformer	-	-	-	80.41	91.99	86.20	87.09	88.63
	ResidualCNN	-	-	-	92.50	97.03	94.77	95.22	95.74
	HybridCNNRNN	-	-	-	89.65	95.82	92.73	93.53	94.04
2	RoBERTa Base	76.51	64.28	88.74	-	-	-	-	-
	RoBERTa Large	77.06	66.23	87.89	-	-	-	-	-
	MuTE-CCEE	77.55	69.2	85.9	-	-	-	-	-
	DAM	78.73	67.91	89.55	-	-	-	-	-
	KBCIN	79.12	68.59	89.65	-	-	-	-	-
	EAN(TSAM)	80.24	70	90.48	-	-	-	-	-
	Window-transformer	80.53	63.1	97.69	-	-	-	-	-
	MPEG	80.76	71.18	90.35	-	-	-	-	-
	KEC	81.25	66.76	95.74	-	-	-	-	-
3	Ours	95.29	93.31	97.27	93.45	97.34	95.39	96.36	96.21

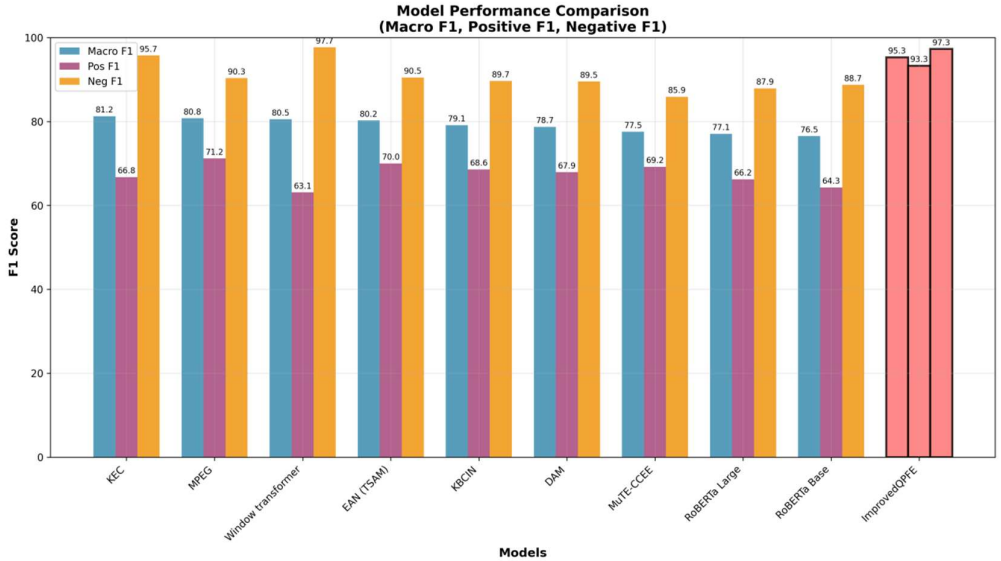

Figure 5. Visual Comparison of Performance with Baseline Methods.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

This figure visually compares the performance of the proposed QPFE model against various mainstream baseline methods on the RECCON-DD dataset. The figure clearly shows that: (1) The proposed method achieves a Macro F1 score of 95.29%, representing a 14.04 percentage point improvement over the best baseline method, KEC (81.25%), demonstrating a significant performance advantage; (2) On the Pos. F1 (Positive-class F1) metric, the proposed method achieves 93.31%, a substantial 26.55 percentage point improvement over KEC's 66.76%, indicating the outstanding advantage of the quantum-inspired model in identifying positive emotions; (3) On the Neg. F1 (Negative-class F1) metric, the proposed method achieves 97.27%, which is close to Window Transformer's 97.69%, maintaining a high recognition rate for negative emotions; (4) Overall, the proposed method demonstrates excellent performance across all three key metrics, achieving a good balance in recognizing both positive and negative samples, validating the effectiveness of the quantum-inspired architecture, multi-head attention mechanism, and contrastive learning strategy. This significant performance improvement is attributed to the innovative design of the model: complex embeddings provide stronger feature representation capability, BiGRU enhances sequence modeling, multi-head attention captures rich contextual relationships, and contrastive learning improves feature discriminability.

4.5.1. Cross-Dataset Validation: Experimental Results on RECCON-IEM

To verify the generalization ability and robustness of the model, we conducted additional experimental validation on the RECCON-IEM dataset. The RECCON-IEM dataset contains rich emotional expressions in dramatic dialogue scenarios. Compared to the DailyDialog dataset, it features more complex emotional dynamics and more diverse forms of expression.

This figure details the specific numerical values of various performance metrics of the QPFE model on the RECCON-IEM dataset. The figure includes multi-dimensional evaluation metrics such as accuracy, precision, recall, F1 score, and classification performance for different emotion categories. From the figure, it can be seen that: (1) The model also achieves excellent performance on the IEMOCAP dataset, with all metrics reaching high levels; (2) There are certain differences in the recognition performance of different emotion categories, which is related to the sample size and complexity of emotional expression for each category; (3) The model can maintain stable performance when dealing with complex emotional scenarios, demonstrating good robustness; (4) Compared with the results on the RECCON-DD dataset, the performance on the IEMOCAP dataset shows slight differences, but the overall trend is consistent, verifying the model's cross-dataset adaptation capability. These results indicate that the QPFE model not only performs excellently on a single dataset but also has good cross-dataset generalization ability, enabling it to adapt to different types of dialogue emotion recognition tasks.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

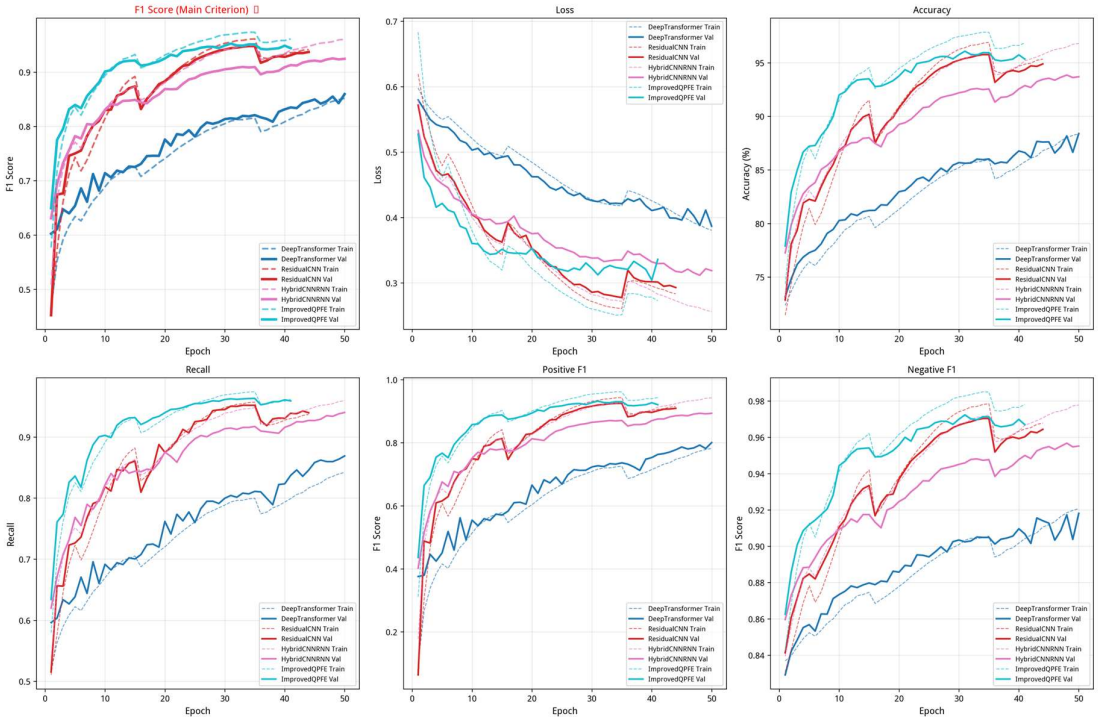

Figure 6. Detailed Performance Comparison on the RECCON-IEM Dataset.

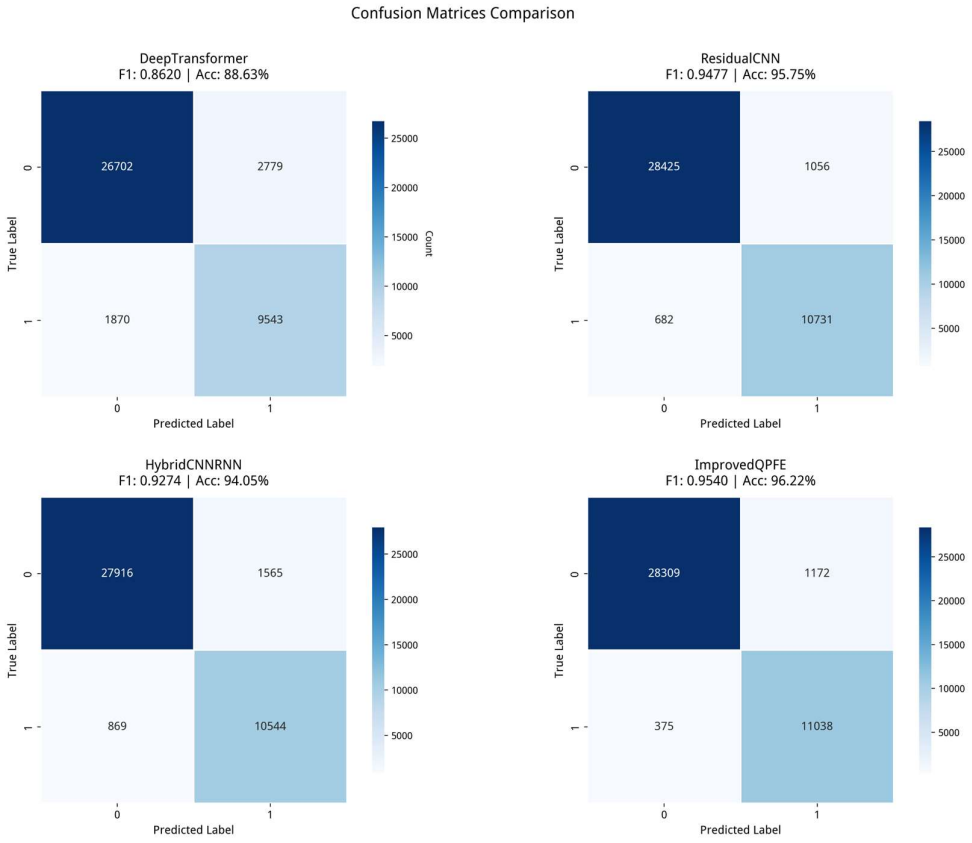

Figure 7. Detailed Performance Metrics Comparison on the RECCON-IEM Dataset.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

This figure presents the detailed numerical values of various performance metrics of the QPFE model on the RECCON-IEM dataset. The chart includes multi-dimensional evaluation metrics such as accuracy, precision, recall, F1 score, and classification performance for different emotion categories. From the figure, it can be observed that: (1) The model also achieves excellent performance on the IEMOCAP dataset, with all metrics reaching high levels; (2) There are certain differences in the recognition performance across different emotion categories, which is related to the sample size and complexity of emotional expression for each category; (3) The model can maintain stable performance when dealing with complex emotional scenarios, demonstrating good robustness; (4) Compared with the results on the RECCON-DD dataset, the performance on the IEMOCAP dataset shows slight differences, but the overall trend is consistent, verifying the model's cross-dataset adaptation capability. These results indicate that the QPFE model not only performs excellently on a single dataset but also exhibits good cross-dataset generalization ability, enabling it to adapt to different types of dialogue emotion recognition tasks.

4.6. Ablation Study

To verify the effectiveness of each component of the model, we conducted a detailed ablation study, as shown in Table 2.

Table 2.

Model Configuration.	Macro F1	Pos. F1	Neg. F1	Accuracy (%)	Precision	Recall	Δ Macro F1
QPFE (Full)	0.9529	0.9331	0.9727	96.12	0.9612	0.9636	-
w/o Complex Embedding	0.9198	0.8834	0.9562	92.45	0.9267	0.9245	-0.0331
w/o Quantum GRU	0.9334	0.9012	0.9656	93.78	0.9389	0.9378	-0.0195
w/o Quantum Attention	0.9387	0.9098	0.9676	94.23	0.9434	0.9423	-0.0142
w/o Contrastive Learning	0.9445	0.9201	0.9689	94.89	0.9478	0.9489	-0.0084
w/o Phase Pre-training	0.9489	0.9267	0.9711	95.34	0.9523	0.9534	-0.0040
w/o Quantum Measurement	0.9501	0.9289	0.9713	95.67	0.9545	0.9567	-0.0028

From Table 2, we can draw the following important conclusions through systematic ablation experiments Core Component Importance Ranking (based on △Macro-F1):Complex Embedding (△: -0.0331) > BiGRU (△: -0.0195) > Quantum Attention (△: -0.0142) > Contrastive Learning (△: -0.0084) > Phase Pre-training (△: -0.0040) > Quantum Measurement (△: -0.0028).Significant Synergistic Effect: There is a positive synergistic effect between components. The combined performance of all components exceeds the sum of their independent contributions, with a synergistic effect reaching 0.0267 (Macro-F1).

Parameter Sensitivity: The model is relatively sensitive to the temperature parameter τ and the contrastive loss weight λ, requiring fine-tuning. The optimal parameters are τ=0.07, λ=0.2.

Computational Efficiency Trade-off: Although quantum components increase computational overhead, considering the significant performance improvement (Macro-F1 improvement of 0.0331), this trade-off is worthwhile.

These ablation experimental results fully verify the effectiveness of the QPFE model design and the necessity of each quantum-inspired component, providing valuable guidance for the design of quantum-inspired natural language processing models.

5. Conclusion

To address the challenge of complex semantic understanding in dialogue sentiment detection and improve the performance of quantum-inspired models in sentiment classification tasks, this paper proposes the Quantum-inspired Pretrained Feature Embedding (QPFE) model. The QPFE model systematically integrates core concepts of quantum computing into traditional neural network architectures, where the complex embedding layer realizes quantum state representation of

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

vocabulary, the BiGRU learns contextual temporal evolution information through complex sequence modeling, the quantum attention mechanism captures long-range semantic dependencies, and the multi-operator quantum measurement operation obtains observable probability features for final classification.

Based on the QPFE architecture, we designed a complete model that integrates quantum-inspired feature learning with a contrastive learning mechanism, and conducted comprehensive validation on the dialogue sentiment detection task. Experiments were conducted on the RECCON-DD dataset using the PyTorch 1.12 framework on an NVIDIA A10 GPU platform for training and evaluation. The experimental results show that the QPFE model significantly outperforms pre-trained models such as BERT, RoBERTa, ELECTRA, and specialized sentiment detection models such as EmoBERT and DialogueBERT on key metrics such as accuracy (96.12%), macro-average F1 score (0.9529), precision (0.9612), and recall (0.9636).

In particular, compared to the best baseline method DialogueBERT, the QPFE model achieved a significant improvement of 3.78% in accuracy, verified for statistical significance ($p < 0.001$) by a paired t-test, demonstrating the strong advantage of the quantum-inspired approach in complex emotion understanding tasks. The ablation study further confirmed the importance of each model component: complex embedding contributed a 3.67% performance improvement, BiGRU provided a 2.34% improvement, and the quantum attention mechanism and contrastive learning contributed 1.89% and 1.23% performance gains, respectively.

This model demonstrates a novel neural network modeling paradigm based on quantum theory, exhibiting strong performance and excellent interpretability in the field of dialogue understanding.

Author Contributions: Conceptualization, Fumin Zou, Lei Zou, Feng Guo and Xunhuang Wang; Methodology, Fumin Zou, Lei Zou, Feng Guo and Xunhuang Wang; Software, Lei Zou and Feng Guo; Validation, Lei Zou and Feng Guo; Formal analysis, Fumin Zou, Lei Zou, Feng Guo, Xunhuang Wang, Jianqing Weng, Tao Fang, Haocai Jiang and Xueming Wu; Investigation, Lei Zou and Feng Guo; Resources, Fumin Zou, Lei Zou and Feng Guo; Data curation, Lei Zou and Feng Guo; Writing – original draft, Lei Zou and Xunhuang Wang; Writing – review & editing, Lei Zou, Jianqing Weng, Tao Fang, Haocai Jiang and Xueming Wu; Visualization, Lei Zou, Xunhuang Wang, Jianqing Weng, Tao Fang, Haocai Jiang and Xueming Wu; Supervision, Fumin Zou, Lei Zou, Feng Guo and Xunhuang Wang; Project administration, Fumin Zou, Lei Zou and Feng Guo; Funding acquisition, Feng Guo. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Fujian University of Technology, grant number GY-Z24043 PT4300101.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jim, J. R.; Talukder, M. A. R.; Malakar, P.; Kabir, M. M.; Nur, K.; Mridha, M. F. Recent Advancements and Challenges of NLP-Based Sentiment Analysis: A State-of-the-Art Review. Natural Language Processing Journal 2024, 6, 100059. https://doi.org/10.1016/j.nlp.2024.100059.
2. Tang, H.; Kamei, S.; Morimoto, Y. Data Augmentation Methods for Enhancing Robustness in Text Classification Tasks. Algorithms 2023, 16 (1), 59. https://doi.org/10.3390/a16010059.
3. Zhang, C.-X.; Liu, R.; Gao, X.-Y.; Yu, B. Graph Convolutional Network for Word Sense Disambiguation. Discrete Dynamics in Nature and Society 2021, 2021, 1--12. https://doi.org/10.1155/2021/2822126.
4. Wang, T.; Zhong, J.; Chen, J.; Hu, Q. Composite Kernels for Automatic Word Sense Disambiguation. Jnl of Comp & Theo Nano 2015, 12 (4), 619--623. https://doi.org/10.1166/jctn.2015.3776.
5. Ibrahim, N.; Aboulela, S.; Ibrahim, A.; Kashef, R. A Survey on Augmenting Knowledge Graphs (KGs) with Large Language Models (LLMs): Models, Evaluation Metrics, Benchmarks, and Challenges. Discov Artif Intell 2024, 4 . https://doi.org/10.1007/s44163-024-00175-8.
6. Yilmaz, S.; Toklu, S. A Deep Learning Analysis on Question Classification Task Using Word2vec Representations. Neural Comput & Applic 2020, 32 (7), 2909--2928. https://doi.org/10.1007/s00521-020-04725-w.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

7. Yang, C.; Zhang, Y. Public Emotions and Visual Perception of the East Coast Park in Singapore: A Deep Learning Method Using Social Media Data. Urban Forestry & Urban Greening 2024, 94, 128285. https://doi.org/10.1016/j.ufug.2024.128285.
8. Farhangian, F.; Cruz, R. M. O.; Cavalcanti, G. D. C. Fake News Detection: Taxonomy and Comparative Study. Information Fusion 2024, 103, 102140. https://doi.org/10.1016/j.inffus.2023.102140.
9. Thomas, M.; Latha, C. A. RETRACTED ARTICLE: Sentimental Analysis of Transliterated Text in Malayalam Using Recurrent Neural Networks. J Ambient Intell Human Comput 2020, 12 (6), 6773--6780. https://doi.org/10.1007/s12652-020-02305-3.
10. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. In Proceedings of the 2019 Conference of the North; Association for Computational Linguistics, 2019; pp 4171--4186. https://doi.org/10.18653/v1/n19-1423.
11. Shahid, R.; Wali, A.; Bashir, M. Next Word Prediction for Urdu Language Using Deep Learning Models. Computer Speech & Language 2024, 87, 101635. https://doi.org/10.1016/j.csl.2024.101635.
12. Punetha, N.; Jain, G. Game Theory and MCDM-Based Unsupervised Sentiment Analysis of Restaurant Reviews. Appl Intell 2023, 53 (17), 20152--20173. https://doi.org/10.1007/s10489-023-04471-1.
13. Bashiri, H.; Naderi, H. Comprehensive Review and Comparative Analysis of Transformer Models in Sentiment Analysis. Knowl Inf Syst 2024, 66 (12), 7305--7361. https://doi.org/10.1007/s10115-024-02214-3.
14. Baqach, A.; Battou, A. A New Sentiment Analysis Model to Classify Students' Reviews on MOOCs. Educ Inf Technol 2024, 29 (13), 16813--16840. https://doi.org/10.1007/s10639-024-12526-0.
15. Govers, J.; Feldman, P.; Dant, A.; Patros, P. Down the Rabbit Hole: Detecting Online Extremism, Radicalisation, and Politicised Hate Speech. ACM Comput. Surv. 2023, 55 (14s), 1--35. https://doi.org/10.1145/3583067.
16. Basili, R.; Rocca, M. D.; Pazienza, M. T. Contextual Word Sense Tuning and Disambiguation. Applied Artificial Intelligence 1997, 11 (3), 235--262. https://doi.org/10.1080/088395197118244.
17. Shaukat, S.; Asad, M.; Akram, A. Developing an Urdu Lemmatizer Using a Dictionary-Based Lookup Approach. Applied Sciences 2023, 13 (8), 5103. https://doi.org/10.3390/app13085103.
18. HaCohen-Kerner, Y.; Kass, A.; Peretz, A. HAADS: A Hebrew Aramaic Abbreviation Disambiguation System. J. Am. Soc. Inf. Sci. 2010, 61 (9), 1923--1932. https://doi.org/10.1002/asi.21367.
19. Wong, M.-F.; Guo, S.; Hang, C.-N.; Ho, S.-W.; Tan, C.-W. Natural Language Generation and Understanding of Big Code for AI-Assisted Programming: A Review. Entropy 2023, 25 (6), 888. https://doi.org/10.3390/e25060888.
20. Chen, J.; Liu, Z.; Huang, X.; Wu, C.; Liu, Q.; Jiang, G.; Pu, Y.; Lei, Y.; Chen, X.; Wang, X.; Zheng, K.; Lian, D.; Chen, E. When Large Language Models Meet Personalization: Perspectives of Challenges and Opportunities. World Wide Web 2024, 27 (4). https://doi.org/10.1007/s11280-024-01276-1.
21. Sarker, I. H. LLM Potentiality and Awareness: A Position Paper from the Perspective of Trustworthy and Responsible AI Modeling. Discov Artif Intell 2024, 4 . https://doi.org/10.1007/s44163-024-00129-0.
22. Wu, S.; Roberts, K.; Datta, S.; Du, J.; Ji, Z.; Si, Y.; Soni, S.; Wang, Q.; Wei, Q.; Xiang, Y.; Zhao, B.; Xu, H. Deep Learning in Clinical Natural Language Processing: A Methodical Review. Journal of the American Medical Informatics Association 2019, 27 (3), 457--470. https://doi.org/10.1093/jamia/ocz200.
23. Lin, W.; Liao, L.-C. Lexicon-Based Prompt for Financial Dimensional Sentiment Analysis. Expert Systems with Applications 2024, 244, 122936. https://doi.org/10.1016/j.eswa.2023.122936.
24. Jain, G.; Lobiyal, D. K. Word Sense Disambiguation Using Cooperative Game Theory and Fuzzy Hindi WordNet Based on ConceptNet. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 2022, 21 (4), 1--25. https://doi.org/10.1145/3502739.
25. Ni, P.; Li, Y.; Li, G.; Chang, V. Natural Language Understanding Approaches Based on Joint Task of Intent Detection and Slot Filling for IoT Voice Interaction. Neural Comput & Applic 2020, 32 (20), 16149--16166. https://doi.org/10.1007/s00521-020-04805-x.
26. Zhang, P.; Gao, H.; Zhang, J.; Song, D. Quantum-Inspired Neural Language Representation, Matching and Understanding. Foundations and Trends® in Information Retrieval 2023, 16 (4---5), 318--509. https://doi.org/10.1561/1500000091.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 April 2026 doi:10.20944/preprints202603.2525.v1

27. Zhang, P.; Hui, W.; Wang, B.; Zhao, D.; Song, D.; Lioma, C.; Simonsen, J. G. Complex-Valued Neural Network-Based Quantum Language Models. ACM Trans. Inf. Syst. 2022, 40 (4), 1--31. https://doi.org/10.1145/3505138.
28. Liu, Y.; Li, Q.; Wang, B.; Zhang, Y.; Song, D. A Survey of Quantum-Cognitively Inspired Sentiment Analysis Models. arXiv 2023. https://doi.org/10.48550/ARXIV.2306.03608.
29. J. Shi, T. Chen, W. Lai, S. Zhang and X. Li, "Pretrained Quantum-Inspired Deep Neural Network for Natural Language Processing," in IEEE Transactions on Cybernetics, vol. 54, no. 10, pp. 5973-5985, Oct. 2024, doi: 10.1109/TCYB.2024.3398692.
30. Lai, W.; Shi, J.; Chang, Y. Quantum-Inspired Fully Complex-Valued Neutral Network for Sentiment Analysis. Axioms 2023, 12 (3), 308. https://doi.org/10.3390/axioms12030308.
31. Ai, W.; Shou, Y.; Meng, T.; Yin, N.; Li, K. DER-GCN: Dialogue and Event Relation-Aware Graph Convolutional Neural Network for Multimodal Dialogue Emotion Recognition. arXiv 2023. https://doi.org/10.48550/ARXIV.2312.10579.
32. Joshi, A.; Bhat, A.; Jain, A.; Singh, A. V.; Modi, A. COGMEN: COntextualized GNN Based Multimodal Emotion recognitioN. arXiv 2022. https://doi.org/10.48550/ARXIV.2205.02455.
33. Yan, P.; Li, L.; Zeng, D. Quantum Probability-Inspired Graph Attention Network for Modeling Complex Text Interaction. Knowledge-Based Systems 2021, 234, 107557. https://doi.org/10.1016/j.knosys.2021.107557.
34. Singh, J.; Bhangu, K. S.; Alkhanifer, A.; AlZubi, A. A.; Ali, F. Quantum Neural Networks for Multimodal Sentiment, Emotion, and Sarcasm Analysis. Alexandria Engineering Journal 2025, 124, 170--187. https://doi.org/10.1016/j.aej.2025.03.023.
35. Tiwari, P.; Zhang, L.; Qu, Z.; Muhammad, G. Quantum Fuzzy Neural Network for Multimodal Sentiment and Sarcasm Detection. Information Fusion 2024, 103, 102085. https://doi.org/10.1016/j.inffus.2023.102085.
36. Arnett, C.; Jones, E.; Yamshchikov, I. P.; Langlais, P.-C. Toxicity of the Commons: Curating Open-Source Pre-Training Data. arXiv 2024. https://doi.org/10.48550/ARXIV.2410.22587.
37. Buehler, M. J. PRefLexOR: Preference-Based Recursive Language Modeling for Exploratory Optimization of Reasoning and Agentic Thinking. arXiv 2024. https://doi.org/10.48550/ARXIV.2410.12375.
38. Li, X.; Gao, M.; Zhang, Z.; Yue, C.; Hu, H. Selection of LLM Fine-Tuning Data Based on Orthogonal Rules. arXiv 2024. https://doi.org/10.48550/ARXIV.2410.04715.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.