mathematics

MDPI

Article

Modular Multi-Task Learning for Emotion-Aware Stance Inference in Online Discourse

Sio-Kei Im [1,2] and Ka-Hou Chan [1,2,*]

[1] Faculty of Applied Sciences, Macao Polytechnic University, Macau, China
[2] Engineering Research Centre of Applied Technology on Machine Translation and Artificial Intelligence, Macao Polytechnic University, Macau, China
* Correspondence: chankahou@mpu.edu.mo

Abstract

Stance detection on social media is increasingly vital for understanding public opinion, mitigating misinformation, and enhancing digital trust. This study proposes a modular Multi-Task Learning (MTL) framework that jointly models stance detection and sentiment analysis to address the emotional complexity of user-generated content. The architecture integrates a RoBERTa-based shared encoder with BiCARU layers to capture both contextual semantics and sequential dependencies. Stance classification is reformulated into three parallel binary subtasks, while sentiment analysis serves as an auxiliary signal to enrich stance representations. Attention mechanisms and contrastive learning are incorporated to improve interpretability and robustness. Evaluated on the NLPCC2016 Weibo dataset, the proposed model achieves an average F1-score of 0.7886, confirming its competitive performance in emotionally nuanced classification tasks. This approach highlights the value of emotional cues in stance inference and offers a scalable, interpretable solution for secure opinion mining in dynamic online environments.

Keywords: stance detection; sentiment analysis; BiCARU; multi-task learning

MSC: 68T50; 68T05; 91F20

check for updates

Academic Editor: Jonathan Blackledge

Received: 5 September 2025
Revised: 9 October 2025
Accepted: 10 October 2025
Published: 14 October 2025

Citation: Im, S.-K.; Chan, K.-H. Modular Multi-Task Learning for Emotion-Aware Stance Inference in Online Discourse. *Mathematics* **2025**, *13*, 3287. https://doi.org/10.3390/math13203287

1. Introduction

Emotion-aware stance inference has emerged as a central research theme in Natural Language Processing (NLP), as it seeks to capture not only an author's explicit position toward a target but also the emotional signals that shape such positions [1]. Unlike sentiment analysis, which primarily identifies polarity, stance detection reveals whether an author is supportive, opposing, or neutral with respect to a specific issue or entity. This dual perspective is particularly relevant in online discourse, where opinions are rarely expressed in a purely rational form but are instead intertwined with affective cues, rhetorical strategies, and context-dependent markers [2,3]. By integrating stance and sentiment, researchers can better capture the layered complexity of human expression, which is essential for understanding public opinion in dynamic and high-stakes environments.

The importance and timeliness of this research are underscored by the rapid evolution of social media platforms. These platforms have transformed into open forums for self-expression and repositories of vast amounts of user-generated content. While this provides unprecedented opportunities for analysing patterns in public discourse, it also introduces significant risks. Social media ecosystems are increasingly susceptible to misinformation,

coordinated manipulation, and adversarial attacks, all of which threaten digital trust and the integrity of online communication [4–6]. Emotion-aware stance inference can therefore play a crucial role in addressing these challenges by identifying deceptive narratives, supporting crisis response, and enabling more resilient online environments. Beyond security, commercial and societal applications also benefit: analysing stance in consumer or political discourse provides valuable insights into preferences, trends, and emerging issues, thereby informing both business strategies and policy decisions [7,8].

Despite the close relationship between stance detection and sentiment analysis, conventional approaches have historically treated them as separate tasks, thereby missing opportunities to exploit their interdependence [9]. Advances in Transformer-based architectures such as BERT, RoBERTa, and GPT have significantly improved contextual language understanding [10], yet these models often lack task-specific mechanisms for robust stance inference. Recent studies have demonstrated that Multi-Task Learning (MTL) can address this gap by jointly optimising stance and sentiment objectives, enabling shared representation learning and improved generalisation [11,12]. Nevertheless, effectively integrating emotional indicators into stance detection remains a technical challenge, involving architectural design, feature fusion strategies, and robustness against noisy or adversarial input [12–14].

In response to these challenges, the present study proposes a modular MTL framework that incorporates sentiment-aware shared representations to enhance stance classification. The framework leverages a RoBERTa encoder with BiCARU layers to capture both contextual semantics and sequential dependencies while reformulating stance detection into parallel binary subtasks to reduce class interference. This design not only improves accuracy but also enhances interpretability and resilience in dynamic online environments.

The contributions of this work are threefold:

- A data augmentation strategy based on back-translation is introduced to expand the NLPCC2016-task4 Weibo dataset from 3K to 12K samples, thereby enhancing linguistic diversity and contextual coverage.
- A dual-layered neural architecture is designed that integrates RoBERTa-based contextual embeddings with BiCARU layers to capture temporal and directional flow in text.
- Sentiment analysis is incorporated as an auxiliary task within the MTL framework, enabling emotional cues to act as complementary signals for stance inference across varied topics.

2. Related Work

Stance detection systems must contend with the ambiguity, brevity, and evolving semantics of social media discourse. To address these challenges, researchers have explored increasingly sophisticated modelling strategies that go beyond surface-level features. By combining statistical indicators with learned representations, and by leveraging both ensemble methods and neural architectures, recent approaches have demonstrated improved adaptability and precision across diverse platforms and topics [15,16].

2.1. Neural and Ensemble Advances

Early research on stance detection relied heavily on traditional machine learning pipelines, in which the feature engineering stage was crucial. Lexical cues, syntactic structures, and sentiment-related indicators were meticulously crafted and incorporated into models such as Support Vector Machines (SVMs), Logistic Regression (LR), and Random Forests (RFs) [17–19]. They often lacked robustness when handling the complexity and ambiguity of social media discourse, while these methods were computationally

lightweight and effective in low-resource scenarios. Subsequent studies have sought to improve classification accuracy by fusing heterogeneous feature types and leveraging ensemble strategies. For example, refs. [20,21] integrated statistical indicators with deep text representations, applying SVMs, RFs, and GBDTs. By combining classifiers that had been trained using different feature sets, the researchers demonstrated that statistical features and learned deep representations could reinforce each other in the detection of stances across diverse Weibo topics. Similarly, ref. [22] employed an ensemble of RF, GBDT, and SVM models, revealing performance asymmetries. RFs were particularly effective at identifying minority-class stances, while GBDTs were better at recognising majority-class stances. This complementarity highlights the usefulness of multi-classifier systems for optimising both precision and recall. The interplay between stance and sentiment has also attracted attention. Refs. [23,24] introduced a Twitter dataset enabling simultaneous stance and sentiment annotation towards predefined targets. Their methodology involved word- and character-level *N*-grams, emotional features, and word embeddings within a linear SVM framework, and they confirmed that affective signals can be pivotal for accurate stance classification. However, conventional feature-engineering approaches have limitations when it comes to generalising to unstructured, noisy text and adapting to new domains in the absence of large labelled corpora. The transition to deep learning marked a turning point by enabling automated feature extraction from raw text. In practice, conventional feature-engineering approaches have limitations when it comes to generalising to unstructured, noisy text and adapting to new domains in the absence of large labelled corpora [25,26]. The transition to deep learning marked a turning point by enabling automated feature extraction from raw text. This transition reflects a broader evolution within machine learning itself [27]. Traditional machine learning methods, such as SVMs and decision trees, rely heavily on manually engineered features and predefined statistical assumptions. In contrast, neural networks—particularly deep architectures—are capable of automatically learning hierarchical representations from raw data [28,29]. Neural networks can be viewed as a subset of machine learning models that use layered structures to extract increasingly abstract features, enabling them to capture complex patterns in unstructured inputs such as text, images, or speech. This shift from manual feature design to end-to-end representation learning has significantly expanded the applicability and effectiveness of stance detection systems [30,31].

Due to the powerful capabilities of neural networks, particularly Recurrent Neural Networks (RNNs) like Long Short-Term Memory (LSTM) [32], Gated Recurrent Unit (GRU) [33], and Content-Adaptive Recurrent Unit (CARU) [34], they have become prominent. These networks have become prominent due to their ability to capture both local and sequential linguistic patterns. Hybrid designs have further extended this paradigm. For instance, ref. [35] combined Convolutional Neural Networks (CNNs) with LSTM and employed dimensionality reduction via Principal Component Analysis (PCA) and chi-square tests, finding that PCA was superior in maintaining stance detection accuracy. Addressing the limitations of target-exclusive modelling, ref. [36] proposed a commonsense-based adversarial learning framework incorporating a commonsense graph encoder to model unseen targets and feature-separation adversarial networks. Their method extracts both target-specific and target-agnostic features, enhancing reasoning capabilities beyond seen contexts. The maturation of deep architectures was paralleled by the rise of pre-trained language models such as BERT, GPT, and RoBERTa [37]. These models are trained on vast amounts of data and can capture subtle linguistic and contextual nuances. They can also be fine-tuned for stance detection with minimal task-specific data. Augmentations to pre-trained models have also proved fruitful. Ref. [38] integrated structured triples from a knowledge graph into BERT for Weibo stance detection, achieving significantly better re-

sults than the baselines and demonstrating the benefits of incorporating external knowledge. Similarly, refs. [39,40] introduced a knowledge distillation approach—BERTtoCNN—in which a compact CNN 'student' network inherited relational and activation-space properties from a larger BERT 'teacher'. This strategy balanced computational efficiency with predictive power. In practice, deep neural and pre-trained architectures outperform traditional pipelines when it comes to managing long-form, context-rich social media content while offering greater adaptability. Building on this, the present study proposes a secure, multi-task stance detection framework integrating a RoBERTa backbone with bidirectional sequential modelling as part of a BiCARU-based architecture. This hybrid design captures semantic depth and temporal dependencies, and the multi-task setup aligns stance prediction with auxiliary objectives, enhancing robustness, interpretability, and resilience in dynamic, security-sensitive online environments.

2.2. Multi-Task Learning for Sentiment and Stance

In contemporary machine learning research, the MTL model has emerged as a methodological strategy that optimises multiple related objectives concurrently. This enables models to transfer and cross-leverage information between tasks. This approach capitalises on structural similarities and overlapping feature spaces rather than treating each prediction problem in isolation [41]. The MTL approach improves the accuracy of each constituent task. This is the case when it is implemented correctly. It also enhances the overall robustness, scalability, and adaptability of the learned representations. This is according to [42]. The potential benefits of MTL have been demonstrated in various domains. For example, it has been used in risk management for large-scale construction contracts, where high stakes and continual exposure to uncertainty are commonplace. Ref. [43] developed a unified architecture that addresses three interlinked operations—risk identification, allocation, and response—within a single learning process. Empirical comparisons revealed that this integrated system consistently outperformed single-task baselines in terms of predictive and decision-support capabilities. Similarly, ref. [44] investigated integrating sentiment analysis and sarcasm detection as auxiliary inputs for stance detection [45,46]. Their proposed multi-objective sequence framework adopted a hierarchical weighting mechanism that effectively prioritised task contributions during training. The resulting models set new performance benchmarks, demonstrating that the interdependence of sentiment and stance—particularly in the presence of sarcasm—can be modelled effectively through deliberate multi-task architectures. These findings contribute to our conceptual understanding of affective signals in argumentative text, extending beyond mere performance metrics [47,48].

Recent research has explored the joint modelling of sentiment and stance as a principled extension of MTL, building on the premise that they are often co-expressed in natural language. It is reframed as an auxiliary signal that provides structural information and enhances stance classification rather than treating sentiment as a peripheral annotation [49]. This dual-task formulation is realised through a hybrid architecture that combines transformer-based encoders such as RoBERTa [50] with sequential modelling layers such as BiLSTM [51] or BiCARU [52]. The transformer component is adept at extracting context-sensitive embeddings and resolving semantic ambiguities, while the recurrent layer captures temporal dependencies and discourse flow. One compelling example is provided by [53], who examined climate discourse on social media using an MTL framework integrating shared attention mechanisms with task-specific processing streams. Their model demonstrated that sentiment supervision can sharpen the focus of stance classifiers, guiding attention towards linguistically salient features. Performance improvements were consistent across multiple benchmarks, highlighting the value of cross-task inductive bias

in domains where affective and argumentative signals are closely linked. In practice, MTL-based models can be particularly effective in text classification tasks where lexical, syntactic, and pragmatic cues are interdependent. By embedding these interrelations into the training objective, models become more resilient to noise and better equipped to generalise across heterogeneous data sources [54,55]. Theoretical implications suggest that such architectures do more than optimise metrics; they include a philosophy of learning. According to this view, such approach integrates sentiment and stance detection, understanding that human expression is rarely one-dimensional and effective modelling must embrace its layered complexity [56,57].

3. Methodology

In order to address the multifaceted nature of stance detection in social media texts, this study adopts a modular methodology based on MTL. The proposed framework integrates several interrelated components, each of which is tailored to capture distinct yet complementary linguistic signals. By jointly modelling stance and sentiment, the system uses shared representations to improve generalisation and robustness across tasks. This section outlines the architectural design, training strategy and task-specific modules that constitute the overall approach.

3.1. Multi-Task Learning Framework

Multi-Task Learning (MTL) involves training on several related objectives simultaneously, enabling the model to recognise commonalities and task-specific distinctions. By sharing a representation space, related tasks can reinforce one another, such as nuanced linguistic cues, subtle shifts in emotional tone, and latent semantic structures learned for one task, which can improve performance in another. This type of knowledge sharing is particularly beneficial for tasks with overlapping informational needs. In practice, stance detection can benefit from sentiment cues embedded in text, while sentiment classification can be informed by stance-related emphasis. Formally, an MTL setup manages a set of N tasks $\{T_1, T_2, \ldots, T_N\}$, each with a dataset D_i, optimising a joint loss function:

$$L = \sum_{i=1}^{N} a_i L_i \tag{1}$$

where L_i is the loss for the i-th task, and a_i is a task weight tuned according to its relative contribution to the shared learning goal. The model architecture comprises two fundamental layers of processing:

Stance Detection captures general syntactic and semantic patterns that are applicable to all tasks, enabling cross-task transfer. These layers usually comprise pre-trained transformer encoders (e.g., BERT and RoBERTa), which offer contextualised token embeddings that act as a shared basis. As illustrated in Figure 1, the shared encoder processes inputs from multiple datasets and outputs a unified representation space.

Task-Specific refine shared features to meet the unique decision-making needs of each task without diluting shared knowledge. Each task head may include additional attention mechanisms, classification layers, or gating modules to isolate task-relevant signals. In Figure 1, each task-specific branch receives the shared features and computes its own loss, which is then aggregated via a weighted loss function.

This work focuses on two tasks: stance detection and sentiment analysis. Although these tasks draw from similar textual signals, stance detection prioritises argumentative intent and contextual relationships, whereas sentiment analysis concentrates on subjective polarity. This distinction is reflected in the dedicated parameters assigned to each task, enabling the

model to balance shared representation learning with task-specific precision. To further mitigate task interference, the following strategies are employed:

Gradient Normalisation Task gradients are rescaled to prevent dominant tasks from overwhelming others during backpropagation.

Dynamic task weighting The coefficients a_i are periodically adjusted based on validation performance to ensure that underperforming tasks receive more attention.

This modular design improves generalisation across tasks, facilitates interpretability, and enables easier integration. Each task can be audited, fine-tuned, or replaced independently without the need to retrain the entire model.

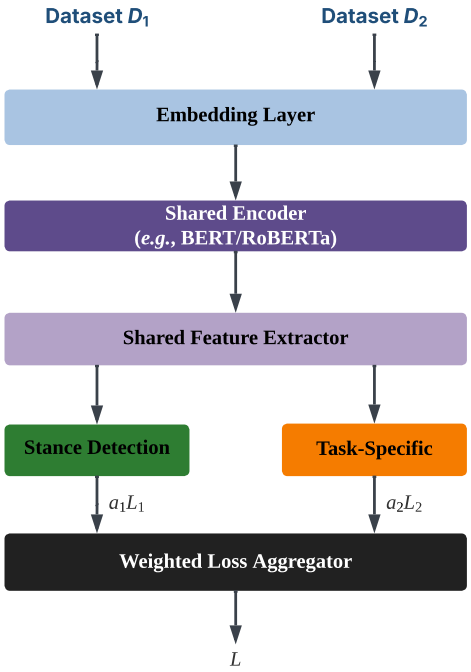

Figure 1. MTL architecture with shared encoder and task-specific heads for stance detection and sentiment analysis.

3.2. Application to Social Media Stance Detection

Detecting stance on platforms such as Weibo is challenging due to the use of informal language and domain-specific vocabulary, as well as the presence of intertwined emotional and argumentative markers. To address this complexity, the proposed system uses a hybrid neural model combining the contextual capabilities of RoBERTa with the sequence modelling capacity of BiCARU. Sentiment analysis is introduced as a parallel auxiliary task, serving not only predictive purposes but also enriching the emotional features injected into stance representation learning. The resulting architecture is illustrated below, showing the data journey from preprocessing to final classification.

Figure 2 illustrates the proposed multitask architecture, which is organised into three functional tiers: input preprocessing (beige), shared representation learning (blue), and task-specific reasoning (peach). At the input level, raw textual data undergoes standardised preprocessing procedures, such as cleaning, tokenisation, and the insertion of structural markers like [CLS] and [SEP], to prepare it for contextual encoding. These inputs are then processed by a RoBERTa encoder to capture deep semantic representations. An attention mechanism is applied within the shared layer to enhance token-level focus and facilitate cross-task feature alignment. The resulting representations are then routed into two parallel branches.

Stance Detection The shared embeddings are passed through a BiCARU module, which models bidirectional contextual dependencies. This is followed by three binary classifiers, each of which is responsible for identifying a specific stance polarity (e.g., favourable, unfavourable, neutral).

Sentiment Analysis The same BiCARU module is used to extract emotionally salient features, which are then processed by a `Sigmoid` activation layer to generate sentiment predictions.

The directional arrows in the diagram illustrate the sequential and branching flow of data, emphasising the incorporation of emotional features into the learning of stance representations. The shared RoBERTa—Attention backbone promotes parameter efficiency and enhances generalisation across tasks, particularly in domains such as Weibo, where informal language and emotional nuance are prevalent.

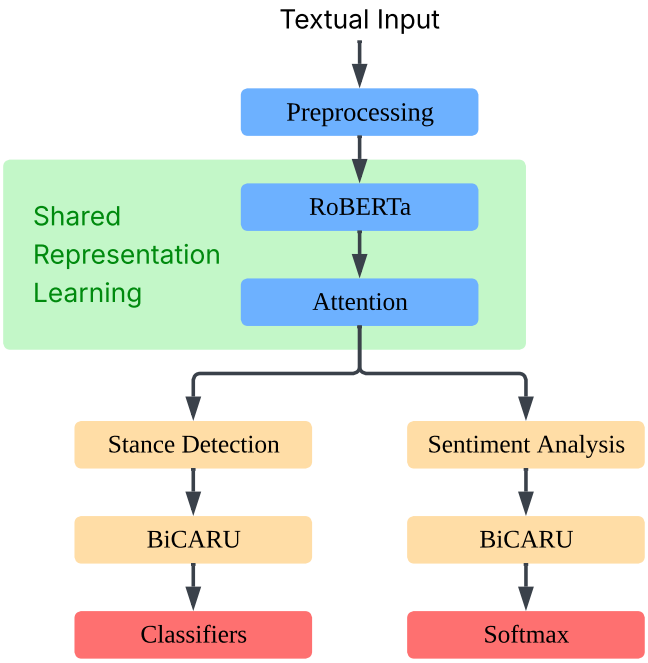

Figure 2. The proposed RoBERTa–BiCARU architecture is used for stance detection and sentiment analysis. The diagram illustrates the flow from input preprocessing to task-specific reasoning, emphasising shared representation learning and emotional feature integration.

4. Shared Representation Learning

The shared representation layer acts as the semantic backbone of the multi-task architecture, allowing features to be reused efficiently across stance detection and sentiment analysis. It integrates a RoBERTa-based encoder with a multi-head attention mechanism to produce context-sensitive embeddings which capture local and global dependencies alike. This modular design makes the architecture more interpretable, scalable, and aligned to downstream tasks.

4.1. RoBERTa Encoding and Representation Sharing

RoBERTa is based on the Transformer self-attention architecture of BERT, but it introduces several targeted optimisations to improve representational fidelity. Specifically, RoBERTa is pre-trained on a substantially larger and more diverse corpus, which broadens its linguistic coverage and contextual robustness. The pre-training process employs dynamic masking, in which different tokens are masked across training epochs rather than using a fixed masking pattern. This strategy forces the model to learn bidirectional depen-

dencies more effectively, since it cannot rely on memorising static positions of masked tokens. In addition, extended training schedules with larger batch sizes and longer sequences enable the model to capture both local and global dependencies with greater precision. Collectively, these refinements enhance the expressiveness of token- and sequence-level embeddings, thereby improving generalisation across downstream tasks without altering the overall Transformer architecture.

Figure 3 illustrates a multi-task NLP model comprising distinct modular components. Input sequences undergo preprocessing involving BPE tokenisation and the insertion of special tokens, followed by dynamic masking from a pre-training corpus. Core encoding is performed using token embeddings and a Transformer stack comprising self-attention and feed-forward layers. Shared embedding layers interface with three task-specific heads—stance detection, sentiment analysis, and policy interpretation—while auxiliary modules (attention visualiser, masking controller, and embedding logger) support interpretability and training control. Each module group is colour-coded to enhance semantic clarity and auditability as follows:

Input and Preprocessing: The model begins by applying BPE to segment the input sequences into consistent subword units, thereby facilitating multilingual compatibility. Special tokens, such as [CLS] and [SEP], are then inserted to indicate task boundaries and segment transitions. These tokens are then embedded in a high-dimensional vector space for subsequent processing.

Core Encoding Pipeline: The embedded tokens are then processed through a stack of Transformer layers comprising self-attention and feed-forward submodules. This iterative encoding mechanism captures syntactic proximity and semantic depth, enabling the model to perform complex tasks such as stance detection, sentiment classification, and policy interpretation.

Shared Embedding Layer: A unified embedding space serves as the central representational hub, interfacing with both task-specific and auxiliary modules. This shared layer minimises redundancy and promotes efficient reuse of features across heterogeneous tasks.

Multi-Tasks Heads: Task-specific modules for stance detection, sentiment analysis, and policy interpretation operate in parallel, drawing from the shared contextual embeddings. This parallelism supports contrastive reasoning and facilitates the alignment of latent semantic cues across tasks.

Auxiliary Modules: Complementary components such as the attention visualiser, masking controller, and embedding logger are integrated via modular interfaces. These operate concurrently with the encoding pipeline, enhancing interpretability and control without disrupting representational integrity.

Modularity and Extensibility: The architecture's modular design supports the seamless integration of new tasks and diagnostic tools. This ensures scalability in multitask and multilingual environments while preserving auditability and operational transparency.

4.2. Attention Dynamics and Cross-Task Feature Propagation

Embedded within RoBERTa's shared representation layer, the attention mechanism plays a central role in dynamically prioritising semantically salient tokens. Inspired by cognitive attention in human processing, it assigns context-sensitive weights to each token, enabling the model to concentrate on informative parts of the input sequence whilst disregarding irrelevant information. This selective weighting enhances both interpretability and performance, particularly in tasks requiring fine-grained semantic inference. Formally, attention is computed via matrix operations involving Query (Q), Key (K), and Value (V) vectors derived from the input embeddings. The alignment score between the decoder's previous hidden state s_{t-1} and each encoder hidden state h_i is calculated using

a scoring function $a(s_{t-1}, h_i)$, which may be a dot product, a scaled dot product, or an additive function. These scores are normalised using a `Softmax` function to yield attention weights $a_{t,i}$:

$$a_{t_i} = \frac{\exp(e_{t,i})}{\sum_k \exp(e_{t,k})} \tag{2}$$

Here, $e_{t,i} = a(s_{t-1}, h_i)$. The resulting context vector c_t is computed as a weighted sum of encoder hidden states:

$$c_t = \sum_{i=1}^{n} a_{t,i} h_i \tag{3}$$

This vector is then integrated with the decoder's prior state and previous output to update the current hidden representation:

$$s_t = f(s_{t-1}, y_{t-1}, c_t) \tag{4}$$

In multi-head attention, the input is projected into distinct subspaces by multiple parallel heads, enabling the model to capture diverse relational patterns, such as syntactic dependencies, shifts in sentiment polarity, and transitions in stance. This multiplicity enriches the shared representation and supports robust propagation of features across tasks. Optional masking operations can further refine the scope of attention, enforcing causal or segment-specific constraints. For example, in stance detection, masking can prevent the model from considering future tokens or irrelevant segments, thereby preserving the integrity of directional inference. In practice, the attention module is closely linked to both the RoBERTa encoder and the downstream task branches. Its outputs inform not only the stance and sentiment modules but also auxiliary components, such as contrastive reasoning units and multilingual alignment layers. To support auditability and modular reuse, attention weights can be extracted and visualised after the fact, offering transparency in the model's decision pathways. This interpretability is particularly valuable in domains requiring traceable inference, such as policy analysis, multilingual opinion mining, and compliance auditing.

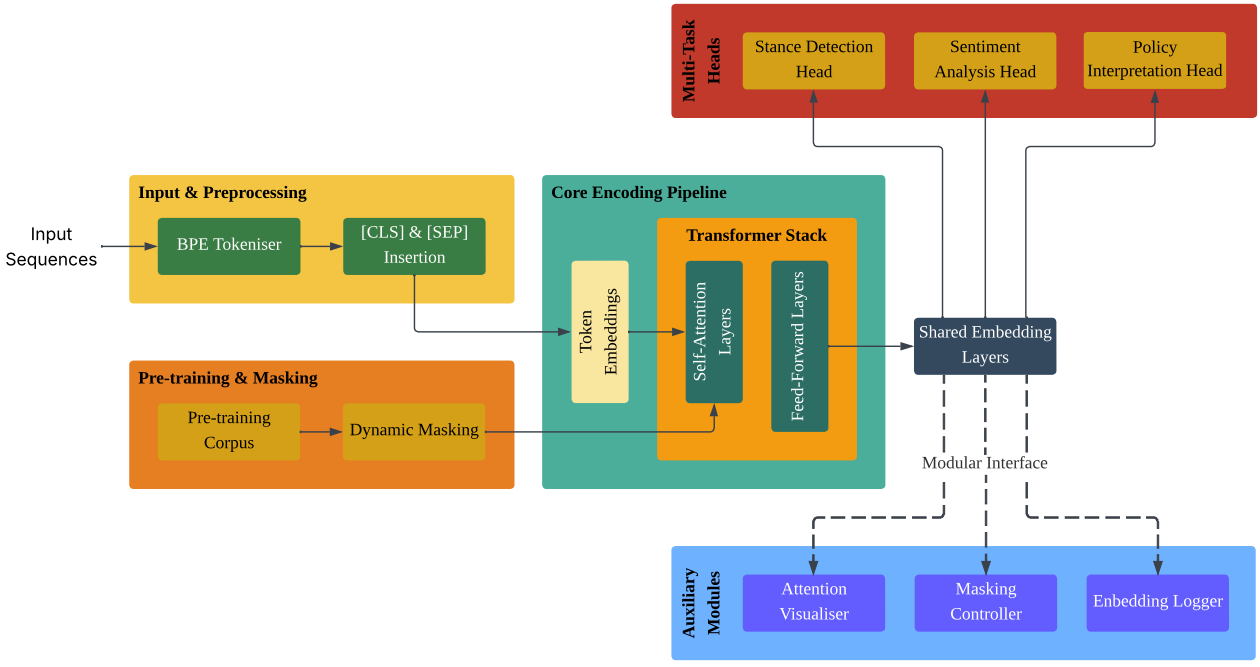

Figure 3. Multi-task NLP pipeline with modular architecture.

5. Modular Stance Inference Architecture

To overcome the challenges associated with stance detection across diverse datasets and implicit language structures, this module incorporates multiple subcomponents that function in a modular and interpretable manner. The architecture is designed to support flexible input formats, multi-label classification, and conditional routing based on dataset provenance. Each subcomponent contributes to a unified representation that balances semantic richness with task-specific precision.

5.1. CARU-Based Bidirectional Representation Learning

To support the inference of a modular stance across heterogeneous inputs, the architecture incorporates a recurrent mechanism that is tailored to content-sensitive representation learning. Positioned prior to the contextual encoding pipeline, CARU captures sequential dependencies in a structurally interpretable manner and serves as a foundational component. The CARU replaces the traditional fully connected operations in recurrent units with linear or convolutional transformations, thereby preserving semantic coherence whilst enhancing parameter efficiency and convergence stability. At each time step t, CARU processes the current input embedding $e_{(t)}$ and the previous hidden state $e_{(t-1)}$ through a gated update mechanism defined as follows:

$$x_{(t)} = We_{(t)} + B \tag{5a}$$

$$n_{(t)} = \tanh\Big(We_{(t-1)} + B + x_{(t)}\Big) \tag{5b}$$

$$z_{(t)} = \sigma\Big(We_{(t-1)} + B + We_{(t)} + B\Big) \tag{5c}$$

$$l_{(t)} = \sigma\Big(x_{(t)}\Big) \otimes z_{(t)} \tag{5d}$$

$$e_{(t)} = \Big(1 - l_{(t)}\Big) \otimes e_{(t-1)} + l_{(t)} \otimes n_{(t)} \tag{5e}$$

In this formulation, W and B denote trainable parameters, $\sigma(\cdot)$ and $\tanh(\cdot)$ are the sigmoid and hyperbolic tangent activation functions, respectively, and $\otimes$ represents element-wise multiplication. The update gate $l_{(t)}$ adaptively balances the retention of prior information with the integration of new content, enabling CARU to model both short-term transitions and long-term dependencies. Such mechanism is particularly well-suited to stance detection tasks, where it is necessary to track subtle shifts in semantic tone or contextual emphasis across token sequences. Subsequent sections extend CARU into a bidirectional configuration (BiCARU) and fuse it with transformer-based embeddings to enhance contextual encoding and target-comment segmentation.

5.2. Contextual Encoding and Target Segmentation

Effective stance detection requires capturing both local and global dependencies across textual inputs. The input token sequence is denoted as $X = [x_1, x_2, \ldots, x_T]$, where each x_t represents a token embedding derived from a pre-trained transformer encoder, such as RoBERTa. The BiCARU module comprises forward CARU_f and a backward CARU_b, producing hidden states $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ at time step t:

$$\begin{cases} \overrightarrow{h}_t & = \text{CARU}_f(x_t) \\ \overleftarrow{h}_t & = \text{CARU}_b(x_t) \end{cases} \tag{6}$$

The bidirectional encoded state at time step t is represented by $h_t = \left[\overrightarrow{h}_t | \overleftarrow{h}_t\right]$. Given the contextual embedding e_t of token x_t from the transformer encoder, the fused representation z_t is defined as follows:

$$z_t = [h_t | e_t] \tag{7}$$

To capture stance-specific asymmetry, the input is segmented into target and comment components: X_{target} and X_{comment}. Each segment is independently encoded using the same BiCARU-transformer fusion pipeline:

$$\begin{cases} z_{\text{target}} & = \text{Encoder}(X_{\text{target}}) \\ z_{\text{comment}} & = \text{Encoder}(X_{\text{comment}}) \end{cases} \tag{8}$$

where $\text{Encoder}(\cdot)$ denotes the composite encoding function described above. The gated fusion mechanism balances the semantic contributions of both segments. The parameters W_g and b_g are trainable weights, and the function $\sigma(\cdot)$ denotes the sigmoid activation function. The fusion gate α and final representation z are computed as follows:

$$\begin{cases} \alpha & = \sigma\left(W_g \left[z_{\text{target}} | z_{\text{comment}}\right] + b_g\right) \\ z & = \alpha \cdot z_{\text{target}} + (1 - \alpha) \cdot z_{\text{comment}} \end{cases} \tag{9}$$

5.3. Reformulated Stance Classification with Conditional Routing

Traditional stance classification often encounters class confusion, particularly between neutral and weakly polarised instances. To overcome this limitation, the task is reformulated as a multi-label classification problem comprising three parallel heads: `support`, `oppose`, and `neutral`. Each head independently predicts a binary stance label $y_i \in \{0, 1\}$ with an associated probability $p_i \in (0, 1)$. These are derived from a shared latent representation $z \in \mathbb{R}^d$ produced by a Transformer-based encoder:

$$p_i = \sigma(W_i \cdot z + b_i) \tag{10}$$

Here, $i \in \{\text{support}, \text{oppose}, \text{neutral}\}$. This formulation allows the model to express uncertainty about its stance across different dimensions, enabling more nuanced downstream analysis. Each classification head is trained using an independent binary cross-entropy loss:

$$L_i = -[y_i \log p_i + (1 - y_i) \log(1 - p_i)] \tag{11}$$

The total stance loss is calculated using a weighted sum:

$$L_{\text{stance}} = \sum_i \lambda_i \cdot L_i \tag{12}$$

where $\lambda_i \in \mathbb{R}_{\geq 0}$ are tunable hyperparameters that control the relative contribution of each stance category. This modular architecture enables selective fine-tuning, targeted ablation, and interpretable error analysis across stance dimensions.

To improve generalisation in scenarios with limited resources and enable transfer learning across diverse corpora, a conditional routing mechanism has been introduced to accommodate both stance-annotated and sentiment-only datasets. Each input is tagged with a dataset origin marker d, where $d = 1$ denotes a stance-annotated sample and $d = 0$ indicates a sentiment-only instance. All inputs are encoded into a shared hidden

representation $z \in \mathbb{R}^h$. In practice, stance-specific heads are bypassed for samples with $d = 0$ via a masking or projection function:

$$z' = \text{DropStance}(z), \quad \text{if } d = 0 \tag{13}$$

Such operation removes stance-specific gradients, enabling the sample to contribute solely to auxiliary objectives, such as masked language modelling or sentiment prediction. To unify the two sample types within a shared semantic space, a gating function is applied:

$$z_{\text{shared}} = \gamma(d) \cdot z + (1 - \gamma(d)) \cdot z' \tag{14}$$

where $\gamma(d) \in \{0, 1\}$ activates stance supervision, which is only activated for samples with a stance annotation. This conditional routing preserves the integrity of stance-specific learning while leveraging sentiment-only data to enrich the representation. The shared encoder is trained jointly across both tasks to promote cross-domain generalisation and enhance stance classification performance under data sparsity.

6. Multi-Task Integration and Auxiliary Module Design

To enhance cross-domain generalisation and promote interpretability in stance inference, the proposed architecture incorporates a range of multi-task heads and additional modules. These components operate in parallel with the primary stance classification pipeline, enabling joint optimisation of complementary objectives such as sentiment analysis, policy interpretation, and representational diagnostics. This design allows for modular ablation, targeted supervision, and robust performance under domain shift.

6.1. Task-Specific Heads for Joint Optimisation

Beyond stance classification, the architecture includes auxiliary heads for `Sentiment Analysis` and `Policy Interpretation`. Each head receives the unified representation z_{shared} and applies an independent projection to produce task-specific predictions. For each auxiliary task $t \in \{\text{sentiment}, \text{policy}\}$, the prediction probability p_t and binary label y_t are computed as follows:

$$p_t = \sigma(W_t \cdot z_{\text{shared}} + b_t), \quad y_t \in \{0, 1\} \tag{15}$$

Each head is optimised using binary cross-entropy loss:

$$L_t = -[y_t \log p_t + (1 - y_t) \log(1 - p_t)] \tag{16}$$

The total multi-task loss is defined as follows:

$$L_{\text{total}} = L_{\text{stance}} + \lambda_{\text{sent}} L_{\text{sentiment}} + \lambda_{\text{policy}} L_{\text{policy}} \tag{17}$$

where λ_{sent} and λ_{policy} are tunable coefficients controlling the relative influence of auxiliary tasks during training.

6.2. Auxiliary Modules for Interpretability and Diagnostics

In order to promote transparency, facilitate modular inspection, and enable error attribution, the architecture incorporates several auxiliary modules that operate in conjunction with the primary inference pipeline. These modules are designed to extract intermediate signals, control dynamic masking, and record changes in representation across training iterations.

6.2.1. Attention Visualisation Module

Let $A^{(l)} \in \mathbb{R}^{T \times T}$ denote the attention matrix from layer l of the transformer encoder, where T is the sequence length. The attention score, $a_{ij}^{(l)}$, quantifies the influence of token x_j on token x_i in layer l. For interpretability purposes, the cumulative attention map across L layers is computed as follows:

$$A_{\text{cum}} = \frac{1}{L} \sum_{l=1}^{L} A^{(l)} \tag{18}$$

This aggregated map is used to identify the dominant paths of attention and visualise the token-level dependencies that are relevant to stance attribution.

6.2.2. Dynamic Masking Controller

To regulate supervision signals based on dataset provenance, a dynamic masking function, denoted by $\mathcal{M}(x_t, d)$, is applied to each token embedding, x_t, conditioned on the dataset marker d, which takes the values 0 or 1. The masking probability m_t is computed as follows:

$$m_t = \mathcal{M}(W_m \cdot [x_t | d] + b_m) \tag{19}$$

where W_m and b_m are trainable parameters. The masked token embedding, denoted by tilde x_t, is then defined as follows:

$$\tilde{x}_t = m_t \cdot x_t + (1 - m_t) \cdot \mathbf{0} \tag{20}$$

This mechanism allows irrelevant stance tokens to be dropped, thereby improving generalisation and reducing overfitting in low-resource settings.

6.2.3. Embedding Trace Logger

In order to monitor representational drift and facilitate cross-domain alignment, the architecture logs intermediate embeddings at key stages. Let $z^{(k)}$ denote the representation at stage $k \in \{\text{BiCARU}, \text{Fusion}, \text{Shared}\}$. The cosine similarity between embeddings from domains D_1 and D_2 is computed as follows:

$$\text{Sim}^{(k)} = \frac{z_{D_1}^{(k)} \cdot z_{D_2}^{(k)}}{\|z_{D_1}^{(k)}\| \cdot \|z_{D_2}^{(k)}\|} \tag{21}$$

This metric is used to evaluate semantic alignment and inform domain adaptation strategies. Together, these auxiliary modules form the diagnostic backbone of the stance inference framework. This enables interpretability, modular debugging, and principled evaluation across different types of data.

7. Experimental Results and Discussion

To evaluate the proposed modular multi-task learning (MTL) framework for stance detection and sentiment analysis, a series of controlled experiments were conducted. These experiments aim to assess the effectiveness of shared representation learning, the impact of auxiliary sentiment signals, and the robustness of the model across varying data volumes and architectures. In this experiment, the stance detection and sentiment analysis tasks rely on distinct yet complementary datasets. Each dataset was selected to align with the objective of a specific task, and their characteristics reflect the design requirements of the proposed modular framework, as summarised in Table 1.

Table 1. Overview of datasets used for stance detection and sentiment analysis.

Task	Dataset	Label Types	Sample Count	Source
Stance Detection	NLPCC2016-task4	*support* *oppose* *neutral*	4,000	NLPCC2016
	PKU-2020	*favor* *against* *neutral*	6,500	PKU Weibo Corpus
	SemEval-2016 (Twitter)	*favor* *against* *neutral*	4,870	SemEval-2016 Task 6
Sentiment Analysis	weibo_senti_100k	*positive* *negative*	119,988	OpenWeibo

The stance detection dataset NLPCC2016-task4 contains 4,000 annotated samples distributed across five controversial topics, with 1560 labelled as *support*, 1,440 as *oppose*, and 1,000 as *neutral*. To focus on polarised classification and reduce ambiguity, only the *support* and *oppose* labels were retained during the main evaluation. This decision is consistent with prior work and supported by error analysis, which indicates that neutral-labelled samples often contain implicit or rhetorical stances that introduce annotation ambiguity. Nevertheless, to address the importance of neutral/ambiguous stances in real-world applications, supplementary experiments including the *neutral* class were conducted on PKU-2020, and the results are reported in the Results Analysis and Discussion section.

In addition to NLPCC2016, the PKU Stance Dataset (PKU-2020) was also incorporated, consisting of approximately 6,500 annotated Weibo posts covering multiple public issues. The distribution is relatively balanced, with 2,200 labeled as *favor*, 2,100 as *against*, and 2,200 as *neutral*. Compared with NLPCC2016, PKU-2020 provides a more balanced distribution across stance categories and includes richer contextual expressions. This dataset serves to further validate the generalisability of the proposed framework across different Chinese stance detection corpora. Detailed experimental results on PKU-2020 are presented in the Results and Analysis section.

To further examine cross-lingual generalisability, the SemEval-2016 Twitter Stance Dataset was also included in a zero-shot evaluation setting. This dataset contains 4,870 English tweets annotated with *favor*, *against*, and *neutral* labels, distributed as 1,825, 1,650, and 1,395 samples, respectively. Unlike the Chinese corpora, it provides a cross-platform and cross-lingual benchmark, enabling assessment of the framework's transferability beyond Weibo discourse. Detailed results are reported in the Results and Analysis section.

The sentiment analysis dataset weibo_senti_100k comprises 119,988 samples labelled as positive or negative, with an approximately balanced distribution (59,988 positive vs. 60,000 negative). It was collected from Weibo and is widely used for sentiment classification tasks in Chinese social media. In this study, it serves as an auxiliary signal source to enrich stance representations. To simulate varying levels of cross-task supervision, subsets of different sizes (0–5,000 samples) were randomly selected. While not directly evaluated, sentiment outputs were monitored to assess emotional alignment and to diagnose sentiment–stance mismatches, which emerged as a notable error type in downstream analysis.

7.1. Training Environment, Configuration, and Strategy

The experimental setup is structured into two subsections to reflect the distinct aspects of model development. The first subsection outlines the training strategy, which includes optimisation and scheduling techniques designed to stabilise convergence and enhance generalisation. The second subsection details the computational platform and modular design components used during training. The training strategy was designed to strike a balance between convergence stability and generalisation. It incorporated adaptive optimisation, dynamic scheduling, and layer-wise regularisation. Cosine annealing with warm-up was employed to refine the learning dynamics, and group normalisation was used to ensure robustness against fluctuations in batch size. A selective weight decay scheme was applied to the encoder to prevent overfitting. All hyperparameter values are summarised in Table 2.

Table 2. Hyperparameter settings used in training strategy.

Parameter	Value	Description
Optimizer	AdamW	Adaptive weight decay optimiser
Learning Rate	0.001	Initial learning rate
β_1	0.5	First moment coefficient
β_2	0.999	Second moment coefficient
Scheduler	Cosine Annealing + Warm-up	Learning rate scheduling strategy
Batch Size	32 per GPU	Number of samples per GPU
Epochs	100	Total training iterations
Normalisation	Group Normalisation	Applied to all layers
Weight Decay	5×10^{-4}	Applied to encoder only

The implementation environment was configured using PyTorch 2.1.0 [58] with CUDA 12.9 acceleration and Python 3.12.4. Training was conducted on four NVIDIA RTX A4000 GPUs under Ubuntu 24.04.3 LTS. Additional efficiency improvements were achieved through half-precision floating-point operations (FP16) and Distributed Data Parallel (DDP), reducing memory consumption and enhancing computational efficiency. The complete training process took approximately 20 h across four GPUs in parallel. Moreover, the architectural configuration consisted of pre-trained RoBERTa encoders as shared layers, followed by task-specific heads comprising attention modules and classification layers. The attention mechanism consisted of eight heads to enhance feature focus, and the feature fusion component was structured with 512 channels to support multi-level information integration. The total parameter count was approximately 2.8 million, enabling the model to handle complex pattern recognition tasks effectively.

To enhance robustness, light-weight augmentation strategies were applied during training. For stance detection datasets (NLPCC2016, PKU-2020), synonym replacement and random word dropout (probability = 0.1) were used to increase lexical diversity. In addition, for NLPCC2016-task4, a multilingual back-translation strategy was employed to expand the training corpus. Each original Chinese sentence was translated into three pivot languages (English, French, and Japanese), then translated back into Chinese, generating three paraphrased variants per instance. This process resulted in a fourfold increase in dataset size (from approximately 3,000 to 12,000 samples). The 1:3 ratio was deliberately selected to maximise linguistic diversity while maintaining computational efficiency. The paraphrased sentences preserved stance semantics while introducing lexical and syntactic diversity, thereby reducing overfitting and improving generalisation. For sentiment analysis

(weibo_senti_100k), back-translation (Chinese → English → Chinese) was applied to 10% of the training samples to enrich semantic variation. Augmented samples were only used in training, not in validation or testing. Each dataset was divided into 70% training, 10% validation, and 20% test splits. All experiments were repeated with five random seeds, and average results are reported. Early stopping was applied based on validation loss with a patience of 5 epochs, ensuring stable convergence and preventing overfitting.

7.2. Evaluation Metrics

Performance was assessed using class-specific and averaged metrics commonly adopted in stance detection literature, particularly in NLPCC2016-task4. The primary evaluation metric was the average F1-score over the polarised stance classes—Favor and Against—denoted as F_{avg}. This metric provides a balanced view of model performance on semantically meaningful categories and excludes the neutral class, which is typically ambiguous and less informative for stance inference. In addition to F_{avg}, individual F1-scores for Favor (F_{favor}) and Against (F_{against}) were reported to highlight class-specific behavior. Precision and recall were also averaged across these two classes to yield P_{avg} and R_{avg}, respectively. These metrics are formally defined as follows:

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{F1} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \end{aligned}$$

To maintain consistency with the modular framework, evaluation was performed separately for each task-specific head. Sentiment classification was used only as an auxiliary signal and not directly optimised for standalone performance. However, its output was monitored to assess cross-task alignment and emotional signal propagation. To enhance data diversity and reduce overfitting, the stance dataset was augmented using multilingual back-translation. This technique involved pivoting through English, French, and Japanese to generate paraphrased variants, thereby expanding the training set for each topic from 600 to 2,400 samples. The augmentation strategy was designed to improve generalisation by introducing linguistic variability while preserving semantic consistency. The averaged metrics used for reporting are defined as follows:

$$\begin{aligned} F_{\text{avg}} &= \frac{F_{\text{favor}} + F_{\text{against}}}{2} \\ P_{\text{avg}} &= \frac{P_{\text{favor}} + P_{\text{against}}}{2} \\ R_{\text{avg}} &= \frac{R_{\text{favor}} + R_{\text{against}}}{2} \end{aligned}$$

All texts were tokenised using a RoBERTa-compatible tokeniser with a maximum sequence length of 128. To maintain consistency with F_{avg} scoring and focus on polarised classification, neutral stance samples were excluded from evaluation. For the sentiment data, random sampling was applied to create additional subsets of different sizes. No augmentation was performed on the sentiment samples in order to preserve polarity integrity and avoid introducing noise that could compromise cross-task supervision.

7.3. Results Analysis and Discussion

To evaluate the proposed BiCARU-based MTL framework, extensive experiments were conducted on the NLPCC2016-task4 dataset. The results indicate that the model

achieves competitive performance across stance detection and sentiment classification tasks, with consistent gains in generalisation and robustness. Eight models were compared: TextCNN [59], FastText [60], BERT [61], RoBERTa [50], RoBERTa-BiLSTM [62], and MTL-RoBERTa-BiLSTM [63]. Notably, CARU-MTL is not an external benchmark but a controlled variant of the proposed architecture, sharing the same modular multi-task structure, sentiment supervision strategy, and contrastive learning setup as BiCARU-MTL, with the only difference being the use of unidirectional CARU layers in place of bidirectional ones. This configuration enables a direct comparison of recurrence directionality while preserving all other architectural components.

As reported in Table 3, the BiCARU-MTL framework demonstrates performance that is competitive with state-of-the-art baselines, including RoBERTa-BiLSTM and MTL-RoBERTa-BiLSTM. While not attaining the highest score across all metrics, it maintains a balanced profile with particular strengths in favour-class precision and overall robustness. The F_{avg} score of 0.7886 places it within the upper tier of evaluated models, and its precision average ($P_{\text{avg}} = 0.7652$) indicates effective control over false positives in polarised stance prediction. A core component of the BiCARU-MTL architecture is the integration of sentiment analysis as an auxiliary task. This design enables emotional signals to be propagated into stance representations, thereby enhancing the model's capacity to interpret subtle or implicit opinions. Sentiment supervision contributes to improved robustness, particularly in noisy or emotionally ambiguous posts, and it facilitates contrastive alignment between stance and sentiment embeddings. To further verify that these improvements are statistically reliable rather than due to random variation, significance tests were conducted.

Table 3. Stance detection performance comparison on NLPCC2016-task4.

	Model	F_{avg}	F_{favor}	F_{against}	P_{avg}	R_{avg}
Baseline	TextCNN [59]	0.6351	0.6309	0.6392	0.6264	0.6525
	FastText [60]	0.6334	0.6152	0.6516	0.6385	0.6285
	BERT [61]	0.7579	0.7547	0.7610	0.7727	0.7448
Advanced	RoBERTa [50]	0.7652	0.7616	0.7687	0.7819	0.7501
	RoBERTa-BiLSTM [62]	0.7711	0.7538	0.7883	0.7801	0.7631
	MTL-RoBERTa-BiLSTM [63]	0.7872	**0.7885**	0.7860	**0.7842**	0.7818
Proposed	CARU-MTL	0.7748	0.7662	0.8034	0.7594	0.7902
	BiCARU-MTL	**0.7886**	0.7881	**0.8091**	0.7652	**0.7920**

As indicated in Table 4, a McNemar's test comparing BiCARU-MTL with RoBERTa-BiLSTM yielded a chi-square value of 12.47 with $p < 0.001$, indicating that the observed differences are highly significant. Also, a paired t-test across five random initialisations showed that BiCARU-MTL significantly outperformed TextCNN ($t = 4.83$, $p = 0.003$). These results confirm that the performance gains of the proposed framework are robust and statistically supported.

Table 4. Statistical significance tests comparing BiCARU-MTL with representative baselines.

Comparison	Test	Statistic	p-Value
BiCARU-MTL vs. RoBERTa-BiLSTM	McNemar's test	$\chi^2 = 12.47$	< 0.001
BiCARU-MTL vs. TextCNN	Paired t-test	$t = 4.83$	0.003

The inclusion of F_{neutral} in Table 5 highlights the model's ability to handle ambiguous stances. On the PKU-2020 dataset, BiCARU-MTL achieved an F1-score of 0.7632 on the

neutral category and a macro-averaged F1-score of 0.7642 overall. These results surpass both TextCNN and RoBERTa-BiLSTM baselines, confirming that the architecture generalises effectively across different Chinese stance detection corpora. Unlike NLPCC2016, PKU-2020 includes a more balanced distribution of *favor*, *against*, and *neutral* labels, which introduces additional challenges in modelling ambiguous or context-dependent stances. The ability of BiCARU-MTL to maintain competitive performance under these conditions highlights the benefits of integrating sentiment supervision and bidirectional recurrence. In particular, the model demonstrates improved handling of the *neutral* class, which is often prone to misclassification due to implicit or rhetorical expressions. These findings indicate that while the main evaluation emphasises polarised stance detection for comparability, the proposed framework remains effective when extended to scenarios where neutral or ambiguous stances are present, thereby supporting its applicability in real-world contexts.

Table 5. Stance detection performance comparison on PKU-2020 (including neutral class).

	Model	F_{avg}	F_{favor}	F_{against}	F_{neutral}	P_{avg}
Baseline	TextCNN [59]	0.7015	0.6942	0.7038	0.7065	0.6981
Advanced	RoBERTa-BiLSTM [62]	0.7428	0.7391	0.7482	0.7410	0.7445
Proposed	BiCARU-MTL	**0.7642**	**0.7675**	**0.7619**	**0.7632**	**0.7651**

As reported in Table 6, the proposed BiCARU-MTL framework also exhibits promising transferability in a cross-lingual, cross-platform setting. When directly applied in a zero-shot manner to the SemEval-2016 Twitter dataset, the model achieved a macro-averaged F1-score of 0.612, outperforming the RoBERTa-BiLSTM baseline. Although the absolute performance is lower than in-domain Chinese datasets, the results demonstrate that the shared representation learning and sentiment-aware supervision embedded in BiCARU-MTL enable meaningful stance discrimination across languages and platforms. This preliminary validation highlights the potential of the framework for multilingual stance detection while also underscoring the need for future work to incorporate multilingual pre-trained encoders (e.g., XLM-R) and additional corpora such as Reddit to further enhance cross-lingual robustness.

Table 6. Zero-shot stance detection performance on SemEval-2016 Twitter dataset.

	Model	F_{avg}	F_{favor}	F_{against}	F_{neutral}	P_{avg}
Advanced	RoBERTa-BiLSTM [62]	0.598	0.601	0.592	0.602	0.596
Proposed	BiCARU-MTL (zero-shot)	**0.612**	**0.617**	**0.609**	**0.610**	**0.613**

In these experiments, cross-task interaction proves especially beneficial in instances where explicit stance markers are absent but emotional polarity is pronounced. In comparison with CARU-MTL, the bidirectional recurrence in BiCARU-MTL yields modest gains in favour-class detection and precision, suggesting that bidirectional context modelling offers incremental advantages in capturing nuanced semantic dependencies. The modular multi-task setup, combined with contrastive learning and sentiment supervision, supports consistent performance across stance categories without introducing instability or overfitting. These results position BiCARU-MTL as a reliable and interpretable alternative within the current landscape of stance detection architectures. Its design prioritises modularity, semantic alignment, and diagnostic transparency, rendering it well-suited for deployment in dynamic, emotionally complex online environments where robustness and generalisation are essential.

7.4. Error Analysis

To better understand the limitations of the proposed BiCARU-MTL framework, a manual error analysis was conducted on 100 randomly selected misclassified instances from the test set. Errors were categorised based on linguistic and semantic characteristics, revealing several recurring patterns that highlight the challenges inherent in stance detection on social media. The most frequent error type involved confusion between Favor and None, accounting for 37.0% of the misclassified samples. These cases typically featured implicit support, sarcasm, or emotionally neutral expressions that lacked explicit stance markers. For example, posts expressing mild approval or rhetorical questions often failed to trigger sufficient activation in the favor-class head, resulting in misclassification as neutral. Table 7 summarises the distribution of error types:

Table 7. Error type distribution in misclassified samples.

Error Type	Frequency (%)	Description
Favor vs. None Confusion	37%	Implicit support, sarcasm, rhetorical ambiguity
Sentiment–Stance Mismatch	28%	Emotional polarity conflicts with stance label
Topic Drift	21%	Multiple targets or mid-sentence context shift
Tokenisation Artifacts	14%	Informal language disrupts semantic segmentation

Another prominent category was sentiment–stance mismatch (28.0%), where the emotional polarity of the post conflicted with its stance label. Posts with positive sentiment toward a negative stance target—or vice versa—were particularly challenging. This suggests that while auxiliary sentiment supervision improves overall robustness, it may also introduce ambiguity when sentiment and stance diverge semantically. Topic drift contributed to 21% of errors. These instances involved posts referencing multiple targets or shifting focus mid-sentence, making it difficult for the model to anchor stance prediction to a specific topic. Such cases often exhibited high contextual entropy, which diluted the effectiveness of attention mechanisms and contrastive alignment. The remaining 14% of errors were attributed to tokenisation artifacts and informal language. Posts containing emojis, elongated characters, or unconventional phrasing were occasionally segmented in ways that disrupted semantic flow. Although the model demonstrated resilience to surface-level noise, certain combinations of informal features still led to degraded representation quality.

These findings suggest that future improvements may benefit from enhanced sarcasm detection, topic anchoring mechanisms, and domain-adaptive tokenisation strategies. Incorporating stance-specific pretraining or multi-hop attention could further mitigate ambiguity in complex posts. Overall, the error patterns observed are consistent with prior studies on social media stance detection and reinforce the importance of interpretability and linguistic nuance in model design.

7.5. Ablation Study

To evaluate the contribution of individual architectural components within the proposed BiCARU-MTL framework, a series of controlled ablation experiments were conducted. These experiments systematically removed or altered three key modules: the auxiliary sentiment head, the gradient normalisation mechanism, and the dynamic task weighting strategy. Each configuration was assessed using precision, recall, and F1-score,

alongside qualitative analyses of attention behaviour, convergence dynamics, and error typology. The following subsections present both quantitative and behavioural findings.

7.5.1. Quantitative Performance Comparison

Table 8 summarises the performance metrics across four configurations. The full model consistently achieved the highest scores, indicating that each component contributes meaningfully to overall performance.

Table 8. Performance comparison across ablated configurations.

Configuration	Precision	Recall	F1-Score
Full Model	80.42	79.15	79.72
Sentiment Head Removed	77.21	76.48	76.84
Gradient Normalisation Disabled	78.03	77.12	77.57
Fixed Task Weights	78.42	77.65	78.03

The removal of the sentiment head resulted in the most pronounced degradation, with a 2.88-point drop in F1-score. This confirms the importance of auxiliary sentiment supervision in guiding stance prediction, particularly in emotionally ambiguous contexts. Disabling gradient normalisation led to moderate performance decline and introduced instability during training, as evidenced by oscillatory loss curves and delayed convergence. Fixing task weights produced the least severe impact yet constrained the model's ability to adaptively balance representational emphasis across tasks.

7.5.2. Module-Level Behavioural Impact

Beyond numerical metrics, each module influences the model's internal behaviour. Attention distribution, convergence stability, and representational adaptability were examined to assess these effects. Table 9 provides a summary of observed behavioural shifts across configurations.

Table 9. Module-level behavioural impact summary.

Module Removed	Attention Shift	Convergence Stability	Adaptability
Sentiment Head	Towards syntax	Stable	Low
Gradient Normalisation	Dispersed	Volatile	Moderate
Fixed Task Weights	Balanced	Stable	Low

Removal of the sentiment head redirected attention towards syntactic anchors—such as modal verbs and negations—while reducing sensitivity to affective cues. Disabling gradient normalisation induced volatile training behaviour and unpredictable convergence patterns. Fixing task weights preserved training stability but reduced the model's flexibility in allocating capacity across tasks.

7.5.3. Error Typology Across Configurations

To further understand the functional role of each module, a fine-grained error analysis was conducted. Misclassified stance predictions were categorised into three types: emotionally misleading errors, contextual ambiguity errors, and topic drift errors. Table 10 presents the distribution of these error types across selected configurations.

Table 10. Error type distribution across ablated configurations.

Error Type	Full Model	No Sentiment Head	No Gradient Normalisation
Emotionally Misleading	12%	30%	14%
Contextual Ambiguity	18%	22%	30%
Topic Drift	10%	12%	16%

As indicated in Table 10, each module mitigates distinct failure modes. Removal of the sentiment head substantially increased emotionally misleading errors (+18%), indicating reduced anchoring in affective contexts. Disabling gradient normalisation elevated contextual ambiguity errors (+12%), suggesting diminished robustness to idiomatic or noisy input. Topic drift errors were moderately elevated in both ablated variants, reflecting weakened attention anchoring. The full model maintained a balanced error profile, demonstrating resilience across diverse linguistic conditions and validating the architectural integration of sentiment supervision and dynamic optimisation.

7.6. Limitations and Future Work

Future research in emotion-aware stance inference can be advanced along several concrete directions. First, more sophisticated approaches are needed to address sarcasm, irony, and figurative language, which remain challenging for current models. Second, adaptive mechanisms should be developed to mitigate topic drift in dynamic discourse, ensuring that stance predictions remain stable across evolving contexts. Third, improved strategies for processing informal textual features—such as emojis, slang, and elongated expressions—are essential for handling the realities of social media communication. Beyond textual signals, multimodal integration that combines stance inference with visual, metadata, or conversational context offers a promising avenue for enhancing robustness. Finally, cross-lingual transfer learning and domain adaptation represent important opportunities to extend the generalisability of stance detection frameworks across languages, platforms, and cultural settings. Together, these directions highlight the potential for building more resilient and context-aware systems in future work.

8. Conclusions

This study presents a modular BiCARU-based MTL framework for stance detection, jointly modelling sentiment to enhance emotional contextualization and robustness. By reformulating stance classification into parallel binary subtasks and integrating contrastive learning, the model improves interpretability and generalisation across emotionally complex social media content. Experimental results show that the proposed model achieves an average F1-score of 0.7886, with precision and recall scores of $P_{\text{avg}} = 0.7902$ and $R_{\text{avg}} = 0.7920$, respectively. Ablation studies indicate that removing the sentiment head leads to a 2.88-point drop in F1, underscoring its role in handling emotionally ambiguous cases. Overall, the framework offers a competitive, scalable, and interpretable solution for stance inference in dynamic, security-sensitive online environments.

Author Contributions: Conceptualization, S.-K.I. and K.-H.C.; Methodology, S.-K.I. and K.-H.C.; Software, K.-H.C.; Validation, S.-K.I.; Formal analysis, S.-K.I. and K.-H.C.; Investigation, K.-H.C.; Resources, S.-K.I. and K.-H.C.; Data curation, S.-K.I. and K.-H.C.; Writing—original draft, K.-H.C.; Writing—review and editing, S.-K.I.; Visualization, K.-H.C.; Supervision, S.-K.I.; Project administration, S.-K.I.; Funding acquisition, S.-K.I. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the grant from Macao Polytechnic University (RP/FCA-01/2025), Macao Science and Technology Development Fund (FDCT) and the Ministry of Science and Technology (MOST) (0018/2025/AMJ).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Burnham, M. Stance detection: A practical guide to classifying political beliefs in text. *Political Sci. Res. Methods* **2024**, *13*, 611–628. [CrossRef]
2. Sykora, M.; Elayan, S.; Hodgkinson, I.R.; Jackson, T.W.; West, A. The power of emotions: Leveraging user generated content for customer experience management. *J. Bus. Res.* **2022**, *144*, 997–1006. [CrossRef]
3. Zhu, H.; Feng, D.; Chen, X. Introduction. In *Social Identity and Discourses in Chinese Digital Communication*; Routledge: London, UK, 2024; pp. 1–21. [CrossRef]
4. Alonso, M.A.; Vilares, D.; Gómez-Rodríguez, C.; Vilares, J. Sentiment Analysis for Fake News Detection. *Electronics* **2021**, *10*, 1348. [CrossRef]
5. Williamson, S.M.; Prybutok, V. The Era of Artificial Intelligence Deception: Unraveling the Complexities of False Realities and Emerging Threats of Misinformation. *Information* **2024**, *15*, 299. [CrossRef]
6. Cheng, K.; Xue, X.; Chan, K. Zero emission electric vessel development. In Proceedings of the 2015 6th International Conference on Power Electronics Systems and Applications (PESA), Hong Kong, China, 15–17 December 2015; pp. 1–5. [CrossRef]
7. Adesoga, T.O.; Olaiya, O.P.; Obani, O.Q.; Orji, M.C.U.; Orji, C.A.; Olagunju, O.D. Leveraging AI for transformative business development: Strategies for market analysis, customer insights, and competitive intelligence. *Int. J. Sci. Res. Arch.* **2024**, *12*, 799–805. [CrossRef]
8. Omowole, B.M.; Olufemi-Phillips, A.Q.; Ofodile, O.C.; Eyo-Udo, N.L.; Ewim, S.E. Big data for SMEs: A review of utilization strategies for market analysis and customer insight. *Int. J. Sch. Res. Multidiscip. Stud.* **2024**, *5*, 1–18. [CrossRef]
9. Lin, X.; Yang, T.; Law, S. From points to patterns: An explorative POI network study on urban functional distribution. *Comput. Environ. Urban Syst.* **2025**, *117*, 102246. [CrossRef]
10. Yenduri, G.; Ramalingam, M.; Selvi, G.C.; Supriya, Y.; Srivastava, G.; Maddikunta, P.K.R.; Raj, G.D.; Jhaveri, R.H.; Prabadevi, B.; Wang, W.; et al. GPT (Generative Pre-Trained Transformer)—A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. *IEEE Access* **2024**, *12*, 54608–54649. [CrossRef]
11. Wang, H.; Zhao, H.; Li, B. Bridging Multi-Task Learning and Meta-Learning: Towards Efficient Training and Effective Adaptation. In Proceedings of the 38th International Conference on Machine Learning, Virtual, 18–24 July 2021; Volume 139, pp. 10991–11002.
12. Ouyang, W.; Gu, X.; Ye, L.; Liu, X.; Zhang, C. Exploring Hydrological Variable Interconnections and Enhancing Predictions for Data-Limited Basins Through Multi-Task Learning. *Water Resour. Res.* **2025**, *61*, e2023WR036593. [CrossRef]
13. Alturayeif, N.; Luqman, H.; Ahmed, M. A systematic review of machine learning techniques for stance detection and its applications. *Neural Comput. Appl.* **2023**, *35*, 5113–5144. [CrossRef]
14. Chan, K.H. Using admittance spectroscopy to quantify transport properties of P3HT thin films. *J. Photonics Energy* **2011**, *1*, 011112. [CrossRef]
15. Rane, N.; Choudhary, S.P.; Rane, J. Ensemble deep learning and machine learning: Applications, opportunities, challenges, and future directions. *Stud. Med. Health Sci.* **2024**, *1*, 18–41. [CrossRef]
16. Sakib, M.; Mustajab, S.; Alam, M. Ensemble deep learning techniques for time series analysis: A comprehensive review, applications, open issues, challenges, and future directions. *Clust. Comput.* **2024**, *28*, 73. [CrossRef]
17. Munmun, Z.S.; Akter, S.; Parvez, C.R. Machine Learning-Based Classification of Coronary Heart Disease: A Comparative Analysis of Logistic Regression, Random Forest, and Support Vector Machine Models. *OALib* **2025**, *12*, e13054. [CrossRef]
18. Avci, C.; Budak, M.; Yağmur, N.; Balçık, F. Comparison between random forest and support vector machine algorithms for LULC classification. *Int. J. Eng. Geosci.* **2023**, *8*, 1–10. [CrossRef]
19. Sharma, A.; Prakash, C.; Manivasagam, V. Entropy-Based Hybrid Integration of Random Forest and Support Vector Machine for Landslide Susceptibility Analysis. *Geomatics* **2021**, *1*, 399–416. [CrossRef]
20. Li, S.; Shi, W. Incorporating Multiple Textual Factors into Unbalanced Financial Distress Prediction: A Feature Selection Methods and Ensemble Classifiers Combined Approach. *Int. J. Comput. Intell. Syst.* **2023**, *16*, 162. [CrossRef]

21. Huang, X.; Chan, K.H.; Wu, W.; Sheng, H.; Ke, W. Fusion of Multi-Modal Features to Enhance Dense Video Caption. *Sensors* **2023**, *23*, 5565. [CrossRef]
22. Liu, J.; Liu, Z.; Li, Q.; Kong, W.; Li, X. Multi-Domain Controversial Text Detection Based on a Machine Learning and Deep Learning Stacked Ensemble. *Mathematics* **2025**, *13*, 1529. [CrossRef]
23. Jamialahmadi, S.; Sahebi, I.; Sabermahani, M.M.; Shariatpanahi, S.P.; Dadlani, A.; Maham, B. Rumor Stance Classification in Online Social Networks: The State-of-the-Art, Prospects, and Future Challenges. *IEEE Access* **2022**, *10*, 113131–113148. [CrossRef]
24. Alturayeif, N.; Luqman, H.; Ahmed, M. Enhancing stance detection through sequential weighted multi-task learning. *Soc. Netw. Anal. Min.* **2023**, *14*, 7. [CrossRef]
25. Pattun, G.; Kumar, P. Feature Engineering Trends in Text-Based Affective Computing: Rules to Advance Deep Learning Models. *Int. Res. J. Multidiscip. Technovation* **2025**, *7* , 87–107. [CrossRef]
26. Im, S.; Chan, K. Vector quantization using k-means clustering neural network. *Electron. Lett.* **2023**, *59*, e12758. [CrossRef]
27. Fitz, S.; Romero, P. Neural Networks and Deep Learning: A Paradigm Shift in Information Processing, Machine Learning, and Artificial Intelligence. In *The Palgrave Handbook of Technological Finance*; Springer International Publishing: Cham, Switzerland, 2021; pp. 589–654. [CrossRef]
28. Borisov, V.; Leemann, T.; Seßler, K.; Haug, J.; Pawelczyk, M.; Kasneci, G. Deep Neural Networks and Tabular Data: A Survey. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 7499–7519. [CrossRef]
29. Chan, K.H.; Im, S.K. Using Four Hypothesis Probability Estimators for CABAC in Versatile Video Coding. *ACM Trans. Multimed. Comput. Commun. Appl.* **2023**, *19*, 40. [CrossRef]
30. Dai, G.; Liao, J.; Zhao, S.; Fu, X.; Peng, X.; Huang, H.; Zhang, B. Large Language Model Enhanced Logic Tensor Network for Stance Detection. *Neural Netw.* **2025**, *183*, 106956. [CrossRef]
31. Im, S.K.; Chan, K.H. More Probability Estimators for CABAC in Versatile Video Coding. In Proceedings of the 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP), Nanjing, China, 23–25 October 2020; pp. 366–370. [CrossRef]
32. Graves, A. Long Short-Term Memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 37–45. [CrossRef]
33. Cho, K.; van Merrienboer, B.; Bahdanau, D.; Bengio, Y. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014. [CrossRef]
34. Chan, K.H.; Ke, W.; Im, S.K. CARU: A Content-Adaptive Recurrent Unit for the Transition of Hidden State in NLP. In *Neural Information Processing*; Springer International Publishing: Cham, Switzerland, 2020; pp. 693–703. [CrossRef]
35. Aljrees, T.; Cheng, X.; Ahmed, M.M.; Umer, M.; Majeed, R.; Alnowaiser, K.; Abuzinadah, N.; Ashraf, I. Fake news stance detection using selective features and FakeNET. *PLoS ONE* **2023**, *18*, e0287298. [CrossRef]
36. Zhang, H.; Li, Y.; Zhu, T.; Li, C. Commonsense-based adversarial learning framework for zero-shot stance detection. *Neurocomputing* **2024**, *563*, 126943. [CrossRef]
37. Raiaan, M.A.K.; Mukta, M.S.H.; Fatema, K.; Fahad, N.M.; Sakib, S.; Mim, M.M.J.; Ahmad, J.; Ali, M.E.; Azam, S. A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. *IEEE Access* **2024**, *12*, 26839–26874. [CrossRef]
38. Dong, L.; Su, Z.; Fu, X.; Zhang, B.; Dai, G. Implicit Stance Detection with Hashtag Semantic Enrichment. *Mathematics* **2024**, *12*, 1663. [CrossRef]
39. Li, Y.; Sun, Y.; Zhu, N. BERTtoCNN: Similarity-preserving enhanced knowledge distillation for stance detection. *PLoS ONE* **2021**, *16*, e0257130. [CrossRef]
40. Im, S.K.; Chan, K.H. Multi-lambda search for improved rate-distortion optimization of H.265/HEVC. In Proceedings of the 2015 10th International Conference on Information, Communications and Signal Processing (ICICS), Singapore, 2–4 December 2015; pp. 1–5. [CrossRef]
41. Elforaici, M.E.A.; Azzi, F.; Trudel, D.; Nguyen, B.; Montagnon, E.; Tang, A.; Turcotte, S.; Kadoury, S. Cell-Level GNN-Based Prediction of Tumor Regression Grade in Colorectal Liver Metastases From Histopathology Images. In Proceedings of the 2024 IEEE International Symposium on Biomedical Imaging (ISBI), Athens, Greece, 27–30 May 2024; pp. 1–5. [CrossRef]
42. Chen, S.; Zhang, Y.; Yang, Q. Multi-Task Learning in Natural Language Processing: An Overview. *ACM Comput. Surv.* **2024**, *56*, 295. [CrossRef]
43. Pham, V.H.S.; Tran, L.A.; Bui, D.K.; Nguyen, Q.T. Artificial Intelligence in Construction Safety Risk Management: A Comprehensive Review and Future Research Perspectives. In *International Conference on Civil Engineering and Architecture, Proceedings of 7th International Conference on Civil Engineering and Architecture, Da Nang, Vietnam, 7–9 December 2024*; Springer Nature: Singapore, 2025; Volume 2, pp. 212–221. [CrossRef]
44. Zhang, Y.; Ma, D.; Tiwari, P.; Zhang, C.; Masud, M.; Shorfuzzaman, M.; Song, D. Stance-level Sarcasm Detection with BERT and Stance-centered Graph Attention Networks. *ACM Trans. Internet Technol.* **2023**, *23*, 1–21. [CrossRef]

45. Ke, W.; Chan, K.H. A Multilayer CARU Framework to Obtain Probability Distribution for Paragraph-Based Sentiment Analysis. *Appl. Sci.* **2021**, *11*, 11344. [CrossRef]
46. Benbakhti, B.; Kalna, K.; Chan, K.; Towie, E.; Hellings, G.; Eneman, G.; De Meyer, K.; Meuris, M.; Asenov, A. Design and analysis of the As implant-free quantum-well device structure. *Microelectron. Eng.* **2011**, *88*, 358–361. [CrossRef]
47. Zhang, Y.; Yu, Y.; Zhao, D.; Li, Z.; Wang, B.; Hou, Y.; Tiwari, P.; Qin, J. Learning Multitask Commonness and Uniqueness for Multimodal Sarcasm Detection and Sentiment Analysis in Conversation. *IEEE Trans. Artif. Intell.* **2024**, *5*, 1349–1361. [CrossRef]
48. Harris, S.; Hadi, H.J.; Ahmad, N.; Alshara, M.A. Fake News Detection Revisited: An Extensive Review of Theoretical Frameworks, Dataset Assessments, Model Constraints, and Forward-Looking Research Agendas. *Technologies* **2024**, *12*, 222. [CrossRef]
49. Lan, X.; Gao, C.; Jin, D.; Li, Y. Stance Detection with Collaborative Role-Infused LLM-Based Agents. In Proceedings of the International AAAI Conference on Web and Social Media, Buffalo, NY, USA, 3–6 June 2024; Volume 18, pp. 891–903. [CrossRef]
50. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv* **2019**. [CrossRef]
51. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The Performance of LSTM and BiLSTM in Forecasting Time Series. In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292. [CrossRef]
52. Chan, K.H.; Im, S.K. BI-CARU Feature Extraction for Semantic Analysis. In Proceedings of the 2022 5th International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 24–25 August 2022; pp. 183–187. [CrossRef]
53. Kodati, D.; Tene, R. Advancing mental health detection in texts via multi-task learning with soft-parameter sharing transformers. *Neural Comput. Appl.* **2024**, *37*, 3077–3110. [CrossRef]
54. Liu, J.; Cui, P. Data Heterogeneity Modeling for Trustworthy Machine Learning. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2, Toronto, ON, Canada, 3–7 August 2025; pp. 6086–6095. [CrossRef]
55. Chan, K.H.; Ke, W.; Im, S.K. A General Method for Generating Discrete Orthogonal Matrices. *IEEE Access* **2021**, *9*, 120380–120391. [CrossRef]
56. Kang, L.; Yao, J.; Du, R.; Ren, L.; Liu, H.; Xu, B. A Stance Detection Model Based on Sentiment Analysis and Toxic Language Detection. *Electronics* **2025**, *14*, 2126. [CrossRef]
57. Das, R.; Singh, T.D. Multimodal Sentiment Analysis: A Survey of Methods, Trends, and Challenges. *ACM Comput. Surv.* **2023**, *55*, 270. [CrossRef]
58. Ansel, J.; Yang, E.; He, H.; Gimelshein, N.; Jain, A.; Voznesensky, M.; Bao, B.; Bell, P.; Berard, D.; Burovski, E.; et al. PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. In Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24), La Jolla, CA, USA, 27 April–1 May 2024. [CrossRef]
59. Kim, Y. Convolutional Neural Networks for Sentence Classification. *arXiv* **2014**. [CrossRef]
60. Joulin, A.; Grave, E.; Bojanowski, P.; Mikolov, T. Bag of Tricks for Efficient Text Classification. *arXiv* **2016**. [CrossRef]
61. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**. [CrossRef]
62. Pu, Q.; Li, F. RoBERTa-BiLSTM: A Chinese Stance Detection Model. In Proceedings of the 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Nanjing, China, 29–31 March 2024; pp. 2199–2203. [CrossRef]
63. Pu, Q.; Huang, F.; Li, F.; Wei, J.; Jiang, S. Integrating Emotional Features for Stance Detection Aimed at Social Network Security: A Multi-Task Learning Approach. *Electronics* **2025**, *14*, 186. [CrossRef]