machine learning & knowledge extraction

IMPACT FACTOR 8.4

CITESCORE 12.7

Article

RolEmo: A Role-Aware Commonsense-Augmented Contrastive Learning Framework for Emotion Classification

Muhammad Abulaish and Anjali Bhardwaj

MDPI

https://doi.org/10.3390/make8030079

machine learning & knowledge extraction

MDPI

Article

RolEmo: A Role-Aware Commonsense-Augmented Contrastive Learning Framework for Emotion Classification

Muhammad Abulaish * and Anjali Bhardwaj

Department of Computer Science and Engineering, South Asian University, New Delhi 110068, India; anjali_bhardwaj@students.sau.ac.in

* Correspondence: abulaish@sau.ac.in

Abstract

Emotion classification is a fundamental task in affective computing, with applications in human–computer interaction, mental health monitoring, and social media analysis. Although most existing methods formulate it as a flat classification problem, emotional expressions are inherently structured and grounded in semantic roles such as the emotion *cue*, *stimulus*, *experiencer*, and *target*. However, the relative contribution of these roles to emotion inference has not been systematically examined. Unlike prior models, we propose RolEmo, a role-aware framework for emotion classification that explicitly incorporates semantic role information. The framework employs a controlled role-masking strategy to analyze the contribution of individual roles, augments textual representations with external commonsense knowledge to capture implicit affective context, and applies supervised contrastive learning to structure the embedding space by bringing emotionally similar instances closer while separating opposing ones. We evaluate RolEmo on three benchmark datasets annotated with semantic roles. Experimental results demonstrate that RolEmo outperforms the strongest baseline across three datasets by up to 16.4%, 25.8%, and 23.2% in the *Full Text*, *Only Role*, and *Without Role* settings, respectively. The analysis further indicates that the *cue* and *stimulus* roles provide the most reliable signals for emotion classification, with their removal causing performance drops of up to 6.2% in macro f1-score, while *experiencer* and *target* roles exhibit more variable effects. These findings highlight the importance of structured semantic modeling and commonsense reasoning for robust and interpretable emotion understanding.

Keywords: emotion classification; commonsense reasoning; contrastive learning; machine learning; affective computing; natural language processing

Check for updates

Received: 27 January 2026
Revised: 9 March 2026
Accepted: 13 March 2026
Published: 19 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The exponential growth of user-generated content on the web has created unprecedented opportunities for large-scale emotion analysis, with applications in empathetic conversational agents, mental health monitoring, rumor detection, and social media analysis [1–3]. Emotions play a central role in human communication and narratives, aiding the comprehension of texts and enhancing the ability to empathize with others' experiences [4]. Consequently, both literary and web-based texts have been widely explored as valuable resources for emotion analysis [5]. However, accurately classifying fine-grained emotions remains challenging due to the implicit, subjective, and context-dependent nature of emotional expression, in which affect is often conveyed indirectly through events, participants, and situational cues rather than explicit emotion words [6].

Traditional approaches to emotion classification are typically modeled as simple classification tasks [7], where models directly map entire text to predefined emotion categories such as *anger*, *joy*, or *sadness*. Although this paradigm has proven effective for coarse-grained categorization, it fails to capture the structured semantic nature of emotional experience. Unlike sentiment analysis, which primarily focuses on polarity, fine-grained emotion understanding requires identifying *who* experiences the emotion, *what* triggers it, and *toward whom* it is directed [8].

Earlier studies have relied on lexicon-based methods and syntactic or hand-crafted features to approximate emotional content [9–11]. Although such approaches offer interpretability, they often struggle with coverage and generalization, especially when emotions are implicit or expressed figuratively. More recent transformer-based models have significantly improved performance by leveraging contextualized representations; however, substantial challenges remain in capturing implicit, context-sensitive, and role-dependent emotional expressions [12].

Semantic role labeling offers a structured way to address these challenges by decomposing emotional expressions into functional components: the *cue* (emotion word or phrase), *stimulus* (triggering event), *experiencer* (entity feeling the emotion), and *target* (entity toward whom the emotion is directed) [13]. Figure 1 illustrates this structure using an example from the GoodNewsEveryone (GNE) dataset [14], where the headline "Sudan protests: Outrage as troops open fire on protesters" contains multiple semantic roles that give rise to the expressed emotion *anger*. Although annotated resources for such roles exist across domains such as tweets and novels [15,16], it remains unclear which roles most reliably signal emotions and how their contributions can be systematically quantified.

Figure 1. An example from the GNE [14] dataset illustrating semantic roles. The *cue* (orange) expression is outrage, linked to the emotion *anger*. The roles involved are *experiencer* (blue, "Sudan protests"), *target* (red, "troops"), and *stimulus* (green, "troops open fire on protesters").

To address these limitations, we propose `RolEmo`, a semantic role-aware emotion classification framework that investigates how semantic roles contribute to emotion prediction. First, we encode semantic role spans using BIO (Beginning-Inside-Outside) tags as structured token-level features for sentence-level emotion classification, enabling the model to capture role-specific emotional semantics. Second, we analyze the importance of individual roles through a controlled role-masking strategy, quantifying how information from each role contributes to and constrains emotion inference. Third, we incorporate external commonsense knowledge to model implicit emotional contexts that may not be explicitly expressed in the text. Finally, we apply supervised contrastive learning to structure the embedding space by pulling semantically similar emotion instances closer while pushing apart opposing ones.

In this work, we specifically investigate whether certain semantic roles are inherently more informative, for example, whether repeated *experiencers* introduce dataset-specific biases, whether *stimuli* or *targets* alone are sufficient for emotion inference, or whether the presence of a *target* constrains the set of plausible emotions. Moreover, in datasets that provide explicit emotion cues, we analyze their relative importance given their direct lexical association with emotion categories.

In summary, the main contributions of `RolEmo` can be briefly described as follows:

- We propose a semantic role-aware emotion classification framework that encodes semantic role spans via BIO tagging as structured token-level features for sentence-level emotion classification.
- We leverage annotated semantic role structures to investigate the contribution of individual roles by analyzing which semantic roles are most informative for emotion prediction and how the information encoded in each role contributes to and constrains emotion inference through a controlled role-masking strategy.
- We integrate external commonsense inferences with supervised contrastive learning based on emotional opposition to enrich textual representations and structure the embedding space, enabling the modeling of implicit and context-dependent emotions while improving discriminability between emotionally similar and opposing instances.
- We conducted extensive experiments on three benchmark datasets, and we demonstrate that *cue* and *stimulus* roles are the most informative for emotion prediction and that combining the use of semantic role structure, commonsense knowledge, and contrastive learning achieves significant performance gains over baselines.

The rest of the paper is organized as follows. Section 2 provides a concise review of the related work. Section 3 presents a comprehensive description of the proposed role-aware, commonsense-augmented contrastive learning framework for emotion classification. Section 4 introduces the datasets, data pre-processing, evaluation metrics, and implementation details. A comparative analysis, baseline comparison, case study, discussion, and limitations are presented in Section 5. Finally, Section 6 summarizes our work and outlines future research directions.

2. Related Work

Early research in emotion analysis primarily formulated the task as a classification problem, assigning each document or sentence to one of several predefined emotion categories [17]. While effective for coarse-grained emotion detection, such approaches offer limited explanatory depth and provide little insight into how emotions are linguistically expressed. Linguistic and psychological studies have long argued that emotions exhibit systematic and interpretable linguistic patterns rather than appearing as isolated labels [18]. This perspective motivated the development of emotion-annotated corpora at progressively finer levels of granularity, including document, sentence, and token-level annotations [19].

To move beyond flat classification, subsequent work adopted more linguistically grounded representations, modeling emotions as structured events composed of interacting semantic components. From this viewpoint, emotional meaning arises through roles such as the *cue* that expresses affect, the *stimulus* (or cause) that triggers the emotion, the *experiencer* who feels it, and the *target* toward which it is directed. This semantic role-based interpretation aligns with theories in psychology and discourse analysis, which emphasize that emotions are grounded in interactions among entities, events, and contextual triggers rather than in isolated lexical items.

Several English-language resources implement this structured representation. Ghazi et al. [20] introduced the EmotionStimulus dataset by leveraging FrameNet emotion predicates and annotating the stimulus as a core argument. Mohammad et al. [15] introduced the ElectoralTweets (ET) dataset, while Gao et al. [16] proposed the NTCIR-13 Emotion Cause Analysis (ECA) corpus; both annotate tweets with multiple roles, including *cue*, *stimulus*, *experiencer*, and *target*. More recent corpora include Relational EMotion ANnotation (REMAN) [21], which provides fine-grained role annotations for literary text, and GNE [14], which annotates emotion roles in news headlines. In parallel, emotion cause extraction (ECE) emerged as a related research direction, focusing primarily on identifying stimulus

clauses that evoke emotions [22,23]. While ECE captures causal components, it typically does not model the full set of emotional participants.

Building on these annotated resources, modeling approaches have progressed from pipeline architectures to joint learning frameworks. Several studies demonstrate that jointly predicting emotions and their associated roles is mutually beneficial. For example, Xia et al. [23] and Wei et al. [24] show that integrating stimulus detection with emotion classification improves performance on both tasks. *EST* [25] further investigate role importance through masking experiments. More recently, Campagnano et al. proposed SRL4E [13], a unified sequence-tagging framework that aligns multiple emotion datasets under a common role schema.

Despite these advances, most recent emotion classification models, particularly transformer-based approaches, rely heavily on contextual embeddings and implicitly encoded semantics without explicitly analyzing the contribution of individual semantic roles. For instance, SRL4E employs a frozen BERT encoder with a BiLSTM decoder but does not explore role-wise masking [13]. SemBERT incorporates external semantic role annotations into BERT representations [26], but it was designed for general language understanding rather than emotion-specific analysis. Similarly, hybrid models such as TE-LSTM+SC integrate predicate-argument structures with BERT-based representations but report limitations in handling complex or noisy contexts [27]. Overall, most prior work treats semantic roles as auxiliary features rather than explicitly analyzing their individual contributions to emotion inference.

Neuro-symbolic approaches have recently gained traction in emotion understanding, combining neural networks with symbolic reasoning to enhance interpretability and logical consistency [28,29]. Representative works include Zhang et al. [30], who employed knowledge graphs with logical constraints for fine-grained emotion detection and Liu et al. [31], who integrated fuzzy logic for handling ambiguous emotional expressions. SenticNet 7 [32] is a neuro-symbolic framework that combines subsymbolic techniques with symbolic representations to support explainable sentiment analysis. More directly related to conversational emotion, researchers have employed probabilistic soft logic to encode explicit structural dependencies for emotion recognition in dialogue [33], defining weighted logical rules that capture label propagation, speaker consistency, and utterance similarity to guide inference toward emotionally coherent predictions. While effective, such approaches rely on explicitly defined logical rules that must be specified a priori, introducing potential limitations in handling nuanced expressions, covering complex linguistic-emotional relationships, and adapting to diverse domains.

`RolEmo` takes a fundamentally different approach by implicitly capturing semantic and causal structures through the integration of semantic role information and commonsense knowledge, rather than enforcing explicit logical constraints. The model learns to attend to role-based cues and leverage commonsense inferences to discover underlying emotional patterns directly from data, offering greater flexibility, data-driven discovery of subtle patterns, and natural capture of semantic relationships without requiring formalization as logical statements.

Across these approaches, the relative importance of individual semantic roles in emotion inference remains underexplored. In particular, it is unclear whether certain roles consistently dominate prediction, whether specific roles introduce dataset-specific biases, or how the presence or absence of particular roles constrains plausible emotions across domains. Our work addresses this gap through a controlled corpus-level analysis that systematically masks or highlights individual roles, enabling a role-wise performance analysis of their contribution to emotion classification and informing the design of more interpretable, role-aware emotion models.

3. Proposed Approach

This paper investigates the role of semantic annotations in emotion classification. We focus on semantic roles to analyze how they guide models in inferring emotions. To systematically assess their contribution, we employ a controlled masking framework over annotated corpora, which enables us to identify the roles that most reliably signal emotion. The inputs are further enriched with commonsense inferences, allowing the model to integrate both linguistic structure and external knowledge. Finally, we train the classifier with supervised contrastive learning to obtain a semantically structured embedding space. The overall framework is illustrated in Figure 2.

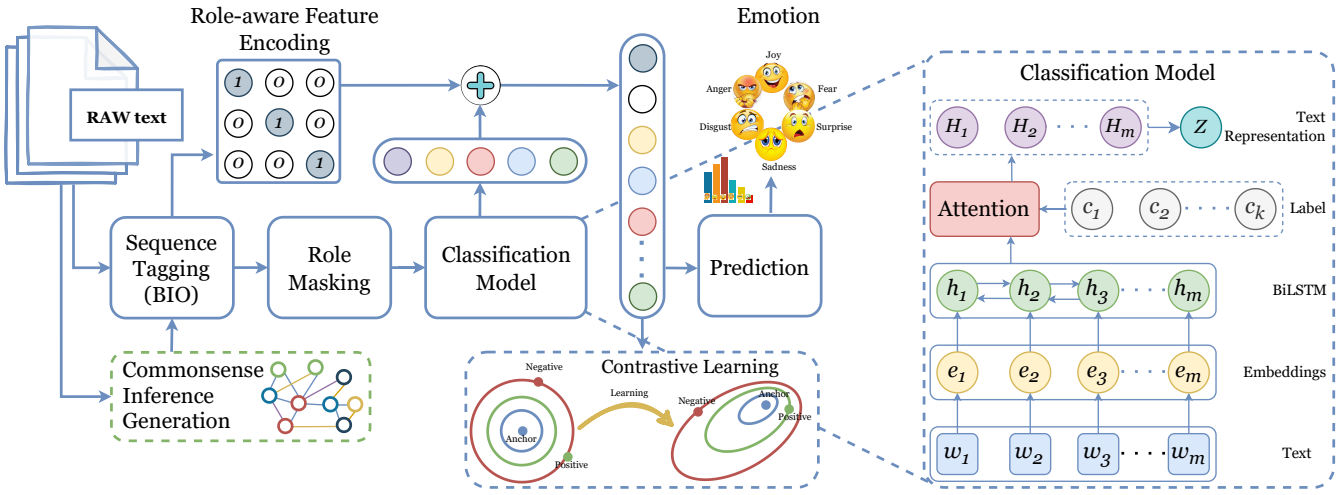

Figure 2. The workflow of the proposed framework `RolEmo`.

3.1. Problem Definition

Let $\mathcal{D} = \{d_1, d_2, \ldots, d_n\}$ denote a corpus of n sentences, and let $K = \{c_1, c_2, \ldots, c_k\}$ be the set of k predefined emotion categories. The task is to learn a mapping

$$f : \mathcal{D} \to K, \quad f(d_j) = c_k \tag{1}$$

where $c_k \in K$ is the predicted emotion for input d_j. Each sentence d_j is tokenized as $d_j = \{w_1, w_2, \ldots, w_m\}$ with m tokens. Each token w_i is augmented with semantic role labels and commonsense inferences to capture contextual and structural information. The enriched sequence is then encoded by a deep model and optimized with a joint classification–contrastive loss, which enhances both predictive accuracy and the structure of the embedding space.

3.2. Semantic Role Representation

3.2.1. Roles and Definitions

Semantic roles are labels assigned to sentence spans that characterize their functional relationship to the expressed emotion. In this work, a sentence is decomposed into four primary semantic roles: which word actually shows the emotion (*cue*), what triggers the emotion (*stimulus*), who feels the emotion (*experiencer*), and whom the emotion is directed toward (*target*). Following prior studies, we adopt these roles to investigate which are most informative for predicting emotions in text.

An example illustrating the annotated semantic roles is shown in Figure 1, where each span corresponds to one of the defined roles. Table 1 summarizes the role definitions and the emotion categories considered. For each instance, role-specific spans are extracted from the dataset. In cases where inconsistencies occur between annotated spans and the original

text, we manually review the dataset to verify and correct the role markings so that the extracted spans remain consistent with the sentence boundaries.

Table 1. Semantic role definitions and emotion categories.

Role	Definition
Cue	A word or phrase (explicit or implicit) that signals the presence of an emotion.
Stimulus	The event, entity, or action that evokes the emotion in the experiencer.
Experiencer	The entity or individual who feels the emotion.
Target	The entity or object toward which the emotion is directed.
Emotions	anger, disgust, fear, joy, sadness, and surprise

3.2.2. Sequence Tagging with BIO

Semantic role spans are identified via token-level sequence labeling using the BIO scheme. Each token w_i in sentence d_j is assigned a label: tokens beginning a role span are marked with "B," those continuing the span with "I," and tokens outside any span with "O." For a given role R, let the annotated span be $S_R = [s_R, t_R] \subseteq \{1, \ldots, m\}$, where $\{1, \ldots, m\}$ denotes the token indices and s_R, t_R are the start and end indices, respectively. The corresponding BIO tag sequence is defined as

$$\text{TagSeq}_R(w) = (y_i)_{i=1}^{m}, \qquad y_i = \begin{cases} B, & i = s_R, \\ I, & s_R < i \leq t_R, \\ O, & otherwise. \end{cases} \tag{2}$$

If $S_R = \varnothing$ (i.e., the role is absent), then $y_i = O$, $\forall\, i$. The tagging is applied independently for each role, extracting multiple roles from the same sentence and providing role-specific labeling so the model can learn to recognize emotion.

3.2.3. Role Masking

To identify which semantic roles most reliably signal emotions, we employ a controlled masking framework on annotated corpora. This approach manipulates sentences by selectively masking or marking role spans while keeping the model architecture fixed, using a special mask token `X` and boundary markers `<R>` and `</R>` to highlight role-specific phrases. Following [25], we define four input settings for each role R, as summarized in Table 2:

Table 2. An overview of the experimental configurations. The symbols X, ⌊, and ⌉ represent boundary markers inserted into the input as each setting dictates.

Input Setting	Model Input								
Full Text	Sudan	protests	outrage	as	troops	open	fire	on	protesters
Only cue	X	X	outrage	X	X	X	X	X	X
Only stimulus	X	X	X	X	X	open	fire	on	protesters
Only experiencer	Sudan	protests	X	X	X	X	X	X	X
Only target	X	X	X	X	troops	X	X	X	X
Without cue	Sudan	protests	X	as	troops	open	fire	on	protesters
Without stimulus	Sudan	protests	outrage	as	troops	X	X	X	X
Without experiencer	X	X	outrage	as	troops	open	fire	on	protesters
Without target	Sudan	protests	outrage	as	X	open	fire	on	protesters
Cue indicator	Sudan	protests	⌊outrage⌉	as	troops	open	fire	on	protesters
Stimulus indicator	Sudan	protests	outrage	a	troops	⌊open	fire	on	protesters⌉
Experiencer indicator	⌊Sudan	protests⌉	outrage	as	troops	open	fire	on	protesters
Target indicator	Sudan	protests	outrage	as	⌊troops⌉	open	fire	on	protesters

- *Full Text*: the complete sentence without masking (Equation (3)).
- *Only Role*: preserves tokens within the role span S_R and masks all others with X (Equation (4)).
- *Without Role*: masks the tokens within S_R but retains all others (Equation (5)).
- *Role Indicator*: retains the full sentence and inserts boundary markers around S_R (Equation (6)).

$$\text{Full}(d_j) = (w_1, w_2, \ldots, w_m) \tag{3}$$

$$\text{Only}_R(d_j) = (y_i)_{i=1}^{m}, \qquad y_i = \begin{cases} w_i, & i \in S_R, \\ \text{X}, & i \notin S_R \end{cases} \tag{4}$$

$$\text{Without}_R(d_j) = (y_i)_{i=1}^{m}, \qquad y_i = \begin{cases} \text{X}, & i \in S_R, \\ w_i, & i \notin S_R \end{cases} \tag{5}$$

$$\text{Ind}_R(d_j) = (y_i)_{i=1}^{m+2}, \qquad y_i = \begin{cases} w_i, & i < s_R, \\ \texttt{<R>}, & i = s_R, \\ w_{i-1}, & s_R < i \leq t_R + 1, \\ \texttt{</R>}, & i = t_R + 2, \\ w_{i-2}, & i > t_R + 2 \end{cases} \tag{6}$$

The $\text{Ind}_R(d_j)$ setting effectively wraps the role span $\{w_{s_R}, \ldots, w_{t_R}\}$ with explicit boundary markers without removing any contextual information. Each setting allows us to analyze the contribution of role R by comparing performance across these controlled manipulations.

3.3. Commonsense Inference Generation

Emotions are often conveyed through implicit context requiring commonsense reasoning. To capture these latent inferences, we incorporate COMET, a neural language model trained on ATOMIC-2020 [34], a large-scale knowledge graph of *if–then* inferences about everyday social situations. ATOMIC-2020 organizes knowledge into three broad categories: social-interaction relations, event–entity relations, and physical-centered relations.

We focus on four social and event-centered relations from ATOMIC-2020 that are most relevant for emotion analysis: *xReact* describes how the agent (PersonX) emotionally reacts to an event, *xEffect* captures the direct consequences experienced by the agent, *xAttr* infers traits or characteristics of the agent from their actions, and *Causes* identifies causal events that trigger the situation. For example, given the headline from the GNE [14] dataset, "Sudan protests: Outrage as troops open fire on protesters," COMET-ATOMIC-2020 generates inferences such as PersonX feels angry (*xReact*), gets arrested (*xEffect*), appears violent (*xAttr*), and that others feel angry (*Causes*). These inferences reveal emotional and causal dimensions that are not explicitly stated in the text.

Formally, for each input sequence d_j, ATOMIC-2020 generates inferences for each selected relation $r \in \{xReact, xEffect, xAttr, Causes\}$. Let $r(d_j)$ denote the inference text under relation r. The resulting commonsense knowledge set is defined as

$$\text{CSK}(d_j) = \{\, r(d_j) \mid r \in \{xReact, xEffect, xAttr, Causes\}\}\tag{7}$$

where $\text{CSK}(d_j)$ denotes the set of inferences generated for d_j. Finally, these inferences are concatenated with the original input so that the model conditions on both the explicit text and external commonsense cues:

$$\tilde{d}_j = d_j \oplus \text{CSK}(d_j) \tag{8}$$

3.4. Classification Model

We adopt pre-trained `GloVe` [35] embeddings to encode each token into a dense semantic vector. To enhance these representations, we further incorporate semantic role annotations and commonsense features, thereby capturing both lexical semantics and structural cues essential for emotion classification.

3.4.1. Input Representation

Given an enriched input sentence with commonsense knowledge (Equation (8)), we denote it as $\tilde{d}_j = \{\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_m\}$. Each token $\tilde{w}_i$ is mapped to its embedding vector as

$$g_i = G(\tilde{w}_i) \in \mathbb{R}^l \tag{9}$$

where G denotes the `GloVe` embedding function that produces l-dimensional vector. To inject structural knowledge, each token is further augmented with role-based features. Specifically, BIO tags or semantic role indicators are assigned to each token, denoted as $t_i \in \{\text{B}, \text{I}, \text{O}\}$, and represented in one-hot form:

$$b_i^{tag} = \begin{cases} [1,0,0], & t_i = \text{B}, \\ [0,1,0], & t_i = \text{I}, \\ [0,0,1], & t_i = \text{O} \end{cases} \tag{10}$$

where $b_i^{tag} \in \mathbb{R}^u$ and u is the dimensionality of the BIO feature space. The role-aware embedding of token $\tilde{w}_i$ is then formed by concatenation:

$$e_i = g_i \oplus b_i^{tag} \in \mathbb{R}^{l+u} \tag{11}$$

3.4.2. Contextual Encoding

The sequence $\{e_1, e_2, \ldots, e_m\}$ is processed by a bidirectional LSTM (BiLSTM) sequence encoder implemented using AllenNLP [36]. The forward and backward hidden states are computed as

$$\overrightarrow{h_i} = LSTM_{fwd}(e_i, \overrightarrow{h_{i-1}}), \qquad \overleftarrow{h_i} = LSTM_{bwd}(e_i, \overleftarrow{h_{i+1}}) \tag{12}$$

Furthermore, they are concatenated to yield the contextualized embedding

$$h_i = \overrightarrow{h_i} \oplus \overleftarrow{h_i} \in \mathbb{R}^{2d'} \tag{13}$$

Thus, the BiLSTM outputs $H = \{h_1, h_2, \ldots, h_m\}$, which encode lexical meaning, structural role information, and bidirectional context.

3.4.3. Attention Mechanism

An additive attention mechanism is then applied to highlight the most informative tokens. The attention score for each hidden state h_i is computed as

$$\alpha_i = \frac{\exp(v^\top \tanh(W_a h_i))}{\Sigma_{j=1}^m \exp(v^\top \tanh(W_a h_j))} \tag{14}$$

where $W_a \in \mathbb{R}^{d_a \times 2d'}$ projects hidden states into the attention space and $v \in \mathbb{R}^{d_a}$ is a trainable scorer vector. The aggregated context vector is

$$\hat{h} = \Sigma_{i=1}^m \alpha_i h_i \in \mathbb{R}^{2d'} \tag{15}$$

https://doi.org/10.3390/make8030079

3.4.4. Prediction Layer

Finally, $\hat{h}$ is passed through a fully connected layer with softmax activation to obtain the probability distribution over k emotion categories:

$$\hat{y} = softmax(W\hat{h} + b), \quad \hat{y} \in \mathbb{R}^k \tag{16}$$

where $W \in \mathbb{R}^{k \times 2d'}$ and $b \in \mathbb{R}^k$. The predicted label corresponds to the class with the maximum probability. Separate models are trained for different input settings to assess the impact of semantic structure on emotion inference.

3.5. Contrastive Learning

We employ supervised contrastive learning [37] to refine the embedding space for emotion classification. The objective is to pull semantically and emotionally similar sentences closer while pushing dissimilar ones farther apart. To identify semantically related texts, we use Sentence-BERT (SBERT) https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2 (accessed on 11 December 2025) [38], which encodes sentences into contextual embeddings suitable for similarity estimation.

For each anchor sentence $\tilde{d}_j$, the positive sample p^+ is selected from the same emotion class as $\tilde{d}_j$ with the highest cosine similarity. The negative sample n^- is selected from an opposing emotion class with the highest similarity, forming a hard negative. Following prior work [39], we identify emotional oppositions grounded in psychological theory using Plutchik's emotion wheel [40]. We use the GNE dataset (in its original Plutchik label space), to identify oppositional instances for hard negative samples. These samples are used only for the contrastive objective, whereas the final classification and evaluation are performed under the unified Ekman taxonomy [41]. This design encourages representations that capture both semantic similarity and emotional opposition.

Formally, the positive and hard negative samples are selected as

$$p^+ = \underset{y:label(y)=label(\tilde{d}_j)}{\arg\max} \; sim(E_\theta(\tilde{d}_j), E_\theta(y)) \tag{17}$$

$$n^- = \underset{y:label(y)\in\mathcal{O}(label(\tilde{d}_j))}{\arg\max} \; sim(E_\theta(\tilde{d}_j), E_\theta(y)) \tag{18}$$

where $E_\theta(\cdot)$ denotes the sentence encoder, $sim(\cdot,\cdot)$ is cosine similarity, and $\mathcal{O}(label(\tilde{d}_j))$ is the set of emotions opposing the label of $\tilde{d}_j$. Each triplet $(\tilde{d}_j, p^+, n^-)$ is then used for training.

During training, the model encodes the anchor, positive, and negative sentences into $z_{\tilde{d}_j}$, z_{p^+}, and z_{n^-}, respectively. To implement discriminative structure in the embedding space, we adopt the *triplet margin loss* with margin $\mu > 0$:

$$L_{ctr} = \max(0, d(z_{\tilde{d}_j}, z_{p^+}) - d(z_{\tilde{d}_j}, z_{n^-}) + \mu) \tag{19}$$

where $d(\cdot,\cdot)$ denotes the Euclidean distance (ℓ_2 norm). This objective ensures that the anchor is closer to its positive counterpart than to its negative counterpart by at least a margin μ, thereby structuring the embedding space according to emotional similarity and opposition.

3.6. Training Objective

During training, the model learns to extract emotion by minimizing errors using a loss function. We optimize the model with a joint loss that combines a standard classification objective with a contrastive learning signal. The classification component facilitates the

accurate prediction of emotion labels, while the contrastive component ensures a structured embedding space that separates semantically dissimilar inputs.

For emotion classification, we adopt the cross-entropy loss, which penalizes the negative log-likelihood of the predicted probability distribution $\hat{y}$ over k classes with respect to the true one-hot label vector y:

$$L_{cls} = -\Sigma_{c=1}^{k} y_c \log \hat{y}_c \tag{20}$$

where $y_k \in \{0, 1\}$ and $\Sigma_{c=1}^{k} y_k = 1$. The overall training objective combines classification and contrastive loss:

$$L = L_{cls} + \alpha L_{ctr} \tag{21}$$

where $\alpha \geq 0$ is a weighting hyperparameter that balances predictive accuracy with embedding space regularization. This joint formulation encourages the model to assign correct emotion labels through L_{cls} while simultaneously learning a discriminative and semantically structured representation space through L_{ctr}.

4. Experimental Setup

This section describes the experimental setup used to identify which semantic roles within the dataset contribute most significantly to emotion classification performance.

4.1. Datasets

We evaluate the proposed approach on three publicly available English benchmark datasets: ECA, ET, and GNE. All three datasets include *stimulus* and *cue* annotations, while *experiencer* and *target* spans are available only in the ET and GNE datasets. Table 3 summarizes the key statistics of each dataset, including the total number of instances and the average token length of each role. Due to the heterogeneity in emotion taxonomies across datasets, we standardized annotations according to Ekman's [41] six basic emotions, a widely adopted psychological framework. This mapping ensures a unified emotion taxonomy, enabling a deeper understanding and more consistent and comprehensive evaluations of emotion across datasets. The resulting emotion-wise instance distributions are reported in Table 4.

- The *ECA* https://research.nii.ac.jp/ntcir/ntcir-13/index.html (accessed on 11 November 2025) dataset, introduced by Gao et al. [16], contains clause-level and token-level annotations for both English and Mandarin texts. In this work, we consider only the English subset, which consists of 300 English instances extracted from novels. The dataset is annotated with Ekman's emotion categories and provides explicit annotations for emotion *cues* as well as their corresponding *stimuli* (causes), making it suitable for analyzing the role of emotional triggers in text.
- The *ET* https://saifmohammad.com/WebDocs/ElectoralTweetsData.zip (accessed on 11 November 2025) dataset, proposed by Mohammad et al. [15], is inspired by Frame Semantics and comprises 4185 tweets related to electoral events. Among these, 2427 tweets include token-level annotations for emotional stimuli obtained via crowdsourcing. In addition to stimuli, ET provides annotations for multiple semantic roles, including cue, experiencer, and target, enabling fine-grained role-based analysis in informal social media text.
- The *GNE* https://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/goodnewseveryone/ (accessed on 11 November 2025) dataset, introduced by Bostan et al. [14], comprises 5000 news headlines collected from 82 news sources and annotated through crowdsourcing. The GNE dataset provides semantic role annotations as well as writer and reader emotion labels, following Plutchik's emotion model,

ensuring consistency with other datasets. Furthermore, GNE provides semantic role annotations for key components of emotion expression, including the *cue*, *stimulus*, *experiencer*, and *target*, making it a rich resource for studying structured emotional information in news text.

In our approach, each configuration was evaluated using a 5-fold cross-validation setup. In each fold, 20% of the data was used for testing, while the remaining 80% was further split into 70% for training and 10% for validation using stratified sampling. Performance was reported using macro-averaged metrics to account for class imbalance across emotion categories.

Table 3. Statistics of datasets with semantic role annotations. # denotes the number of instances containing the corresponding role, len denotes the average token length of the role spans, and '–' indicates that the role is not annotated in the dataset.

Dataset	Emotion Category	Instances		Cue		Stimulus		Experiencer		Target	
		#	len	#	len	#	len	#	len	#	len
ECA	Ekman	300	40.99	294	1.01	288	10.49	–	–	–	–
ET	19 categories	4185	17.77	4167	5.18	2723	5.74	4156	1.00	3623	1.57
GNE	extended Plutchik	5000	13.00	4806	1.43	4799	7.04	3467	1.79	4477	4.50

Table 4. Emotion-wise distribution of instances across datasets in our framework. # denotes the number of instances.

Dataset	# Instances	# Anger (%)	# Disgust (%)	# Fear (%)	# Joy (%)	# Sadness (%)	# Surprise (%)
ECA	300	44 (15)	14 (5)	25 (8)	105 (35)	58 (19)	54 (18)
ET	4185	635 (15)	1324 (32)	288 (7)	1407 (33)	279 (7)	252 (6)
GNE	5000	886 (18)	501 (10)	419 (8)	835 (17)	884 (18)	1475 (29)

4.2. Data Pre-Processing

Before using the corpus in our experiments, we conducted pre-processing to handle the informal and noisy nature of social media text. Each raw sentence first underwent HTML entity decoding, where common entities such as "&" were replaced with "and", and "<" and ">" are converted to their literal characters, "*lessthan*" and "*greaterthan*". Next, all newline characters (\n), extra spaces, tabs, and non-breaking spaces in the input text were replaced with a single space to normalize the text. User mentions (e.g., @username) were anonymized by replacing them with a generic token "name" to protect privacy while preserving the mention context. Emojis and emoticons, increasingly prevalent in online communication, provide valuable context for emotion classification. Emoticons (e.g., :-(as a sad face) are combinations of punctuation marks, numbers, and letters used to create pictorial representations that often directly convey emotions or sentiments. In contrast, emojis are pictographs representing objects, symbols (e.g., ☺, for a "smiling face"), and faces. Both serve to convey emotions and sentiments [42]. To retain this information, we encoded emoticons and emojis into textual representations using the `emot` (version 3.1) library https://github.com/NeelShah18/emot (accessed on 15 November 2025). The cleaned text was then tokenized into smaller pieces (called tokens) using either the Twitter-specific Twokenize http://www.ark.cs.cmu.edu/TweetNLP (accessed on 15 November 2025) for tweets or spaCy's tokenizer for other domains. To improve semantic clarity, we expanded common English contractions (e.g., "can't" becomes "cannot") using the `pycontractions` (version 2.0.1) library https://github.com/ian-beaver/pycontractions (accessed on 15 November 2025), which understands the sentence context to choose the right expansion. For example, it knows that "ain't" could mean "am not/are not/is not/has not/have not" and outperforms the standard

`Python` contractions library as it considers the text's grammar. Finally, we removed non-alphabetic words (including numbers and symbols) and URLs, handled UTF-8 characters, and eliminated duplicate instances to further clean the text.

4.3. Implementation Details

All experiments were conducted using Python 3.8 and the open-source PyTorch deep learning framework. Word representations were initialized using pre-trained 300-dimensional Stanford `GloVe` https://nlp.stanford.edu/projects/glove/ (accessed on 15 November 2025) embeddings [35]. Each token was augmented with BIO tag features represented as a 3-dimensional one-hot vector, resulting in a 303-dimensional input representation. These embeddings were provided as input to a two-layer BiLSTM [43], implemented using AllenNLP [36]. Each BiLSTM layer employed a hidden dimension of 128 with a dropout rate of 0.3. Model parameters were optimized using the AdamW optimizer with a learning rate of 1×10^{-4} and a weight decay of 0.01, along with an additional $L2$ regularization coefficient of $\alpha = 0.0012$. Training was performed with a batch size of 32 for a maximum of 40 epochs. Early stopping with a patience of 3 epochs, based on validation performance, was used to prevent overfitting. Figure 3 illustrates the training and validation loss curves for the ECA dataset under the *Role indicator* setting, confirming stable training dynamics across all 5 folds.

Hyperparameters were selected using grid search to obtain the optimal configuration. The search ranges included margin $\mu \in \{0.2, 0.5, 1.0, 2.0\}$, contrastive loss weight $\alpha \in \{0.2, 0.3, 0.4, 0.5\}$, and hidden size $\in \{64, 128, 256, 512\}$. To ensure fair and consistent evaluation, the same training setup and hyperparameter configuration were applied uniformly across all datasets. The final hyperparameters were selected based on validation performance on the GNE dataset and then fixed for all other datasets without further per-dataset tuning. This ensures that performance differences reflect dataset characteristics rather than dataset-specific hyperparameter optimization.

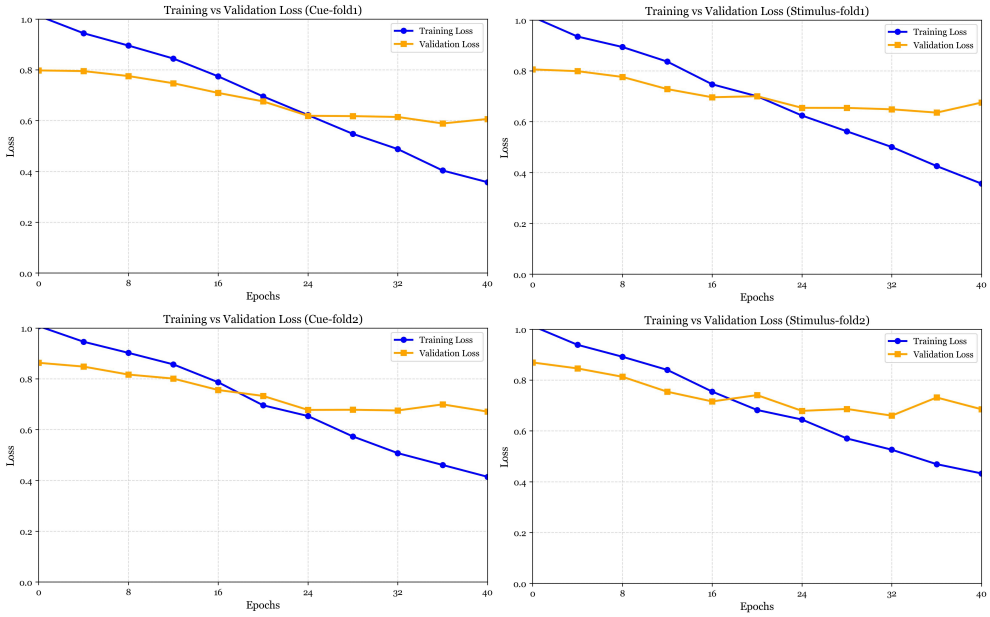

Figure 3. *Cont.*

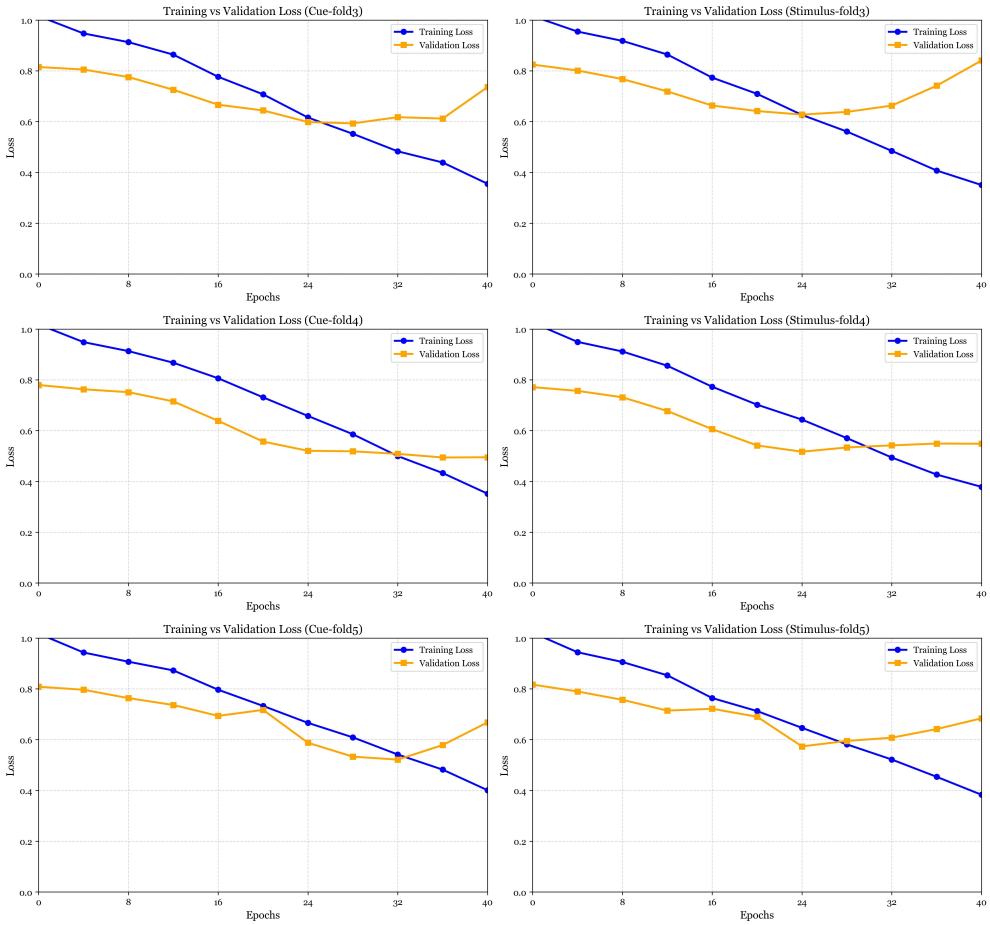

Figure 3. Training and validation loss curves for the ECA dataset under the *Role indicator* setting across 5 folds for the *cue* (**left**) and *stimulus* (**right**) roles.

5. Performance Evaluation Results and Comparative Analysis

This section presents the experimental results of the proposed approach, followed by a comparative analysis against baseline methods, a case study, and a discussion that interprets the findings and validates the effectiveness of the proposed approach. Selecting appropriate baselines was challenging due to the inadequate number of directly comparable approaches, so previous studies were carefully reviewed to choose suitable baselines. For fair comparison, the EST baseline was reimplemented based on the description reported in the original paper and evaluated using the same datasets, data splits, and evaluation settings as `RolEmo`. All training configurations were kept consistent to minimize implementation bias. Experiments were conducted on three publicly available datasets: ECA, GNE, and ET. Tables 5–7 report the P_{mac}, R_{mac}, and $F1_{mac}$ obtained under four input settings: *Full Text*, *Only Role*, *Without Role*, and *Role indicator*. Figures 4–6 visualize emotion-wise performance under the *Full Text* and *Role indicator* settings using heatmaps. A paired t-test conducted on $F1_{mac}$ confirms that the improvements achieved by `RolEmo` are statistically significant at $p < 0.05$ across datasets.

Table 5. Comparative performance of `RolEmo` with baseline methods under four input settings on the ECA dataset. '–' indicates that results are not reported for methods that do not incorporate explicit emotion cues required by the corresponding setting. Bold values indicate improvements over the corresponding role-specific settings.

Approach	Role	Full Text			Only Role			Without Role			Role Indicator		
		P_{mac}	R_{mac}	$F1_{mac}$	P_{mac}	R_{mac}	$F1_{mac}$	P_{mac}	R_{mac}	$F1_{mac}$	P_{mac}	R_{mac}	$F1_{mac}$
RolEmo$_{NAM}$	Cue	32.5	41.9	36.3	**22.0**	19.0	12.3	26.0	38.0	30.4	45.9	44.3	41.8
	Stimulus				**41.9**	37.5	36.0	50.5	48.4	48.4	57.7	57.7	57.7
RolEmo$_{NCK}$	Cue	33.1	38.0	34.9	6.4	16.7	9.2	22.3	19.0	12.6	34.7	41.2	36.1
	Stimulus				11.7	22.0	14.6	23.7	29.0	22.4	47.3	47.3	47.3
RolEmo$_{NCL}$	Cue	33.1	40.2	35.9	6.4	16.7	9.2	32.1	35.7	28.2	41.0	44.9	41.2
	Stimulus				17.6	22.1	18.3	35.2	42.5	37.0	56.7	56.7	56.7
EST	Cue	41.0	39.0	39.0	–	–	–	–	–	–	–	–	–
	Stimulus				30.0	25.0	22.9	48.0	48.0	48.0	52.0	51.0	51.0
`RolEmo`	Cue	**43.3**	**44.9**	**44.1**	14.4	**19.0**	**12.6**	**44.0**	**44.8**	**43.5**	**57.0**	**51.1**	**50.0**
	Stimulus				33.5	**45.9**	**38.4**	**56.1**	**53.1**	**53.6**	**61.7**	**61.3**	**58.9**

Table 6. Comparative performance of `RolEmo` with baseline methods under four input settings on the GNE dataset. Bold values indicate improvements over the corresponding role-specific settings.

Approach	Role	Full Text			Only Role			Without Role			Role Indicator		
		P_{mac}	R_{mac}	$F1_{mac}$	P_{mac}	R_{mac}	$F1_{mac}$	P_{mac}	R_{mac}	$F1_{mac}$	P_{mac}	R_{mac}	$F1_{mac}$
RolEmo$_{NAM}$	Cue	27.5	28.5	27.3	23.0	29.2	25.3	27.7	29.2	27.6	49.3	27.8	28.1
	Stimulus				21.4	26.5	23.1	21.8	26.1	23.6	44.1	27.8	35.2
	Experiencer				25.4	27.4	23.2	30.6	32.1	30.4	41.1	28.0	31.4
	Target				21.8	25.2	22.5	29.9	32.1	30.7	36.3	21.8	27.9
RolEmo$_{NCK}$	Cue	23.2	27.8	24.7	10.0	16.8	8.4	8.6	16.8	8.6	44.8	21.3	17.3
	Stimulus				8.8	16.9	8.2	10.7	16.7	8.9	37.5	20.2	21.9
	Experiencer				5.2	16.5	7.9	21.1	24.1	22.0	37.4	23.4	26.8
	Target				10.4	16.6	9.0	11.7	16.9	8.8	48.0	26.2	22.6
RolEmo$_{NCL}$	Cue	26.3	28.5	26.3	9.3	16.7	8.3	13.7	20.2	15.2	50.1	**29.1**	26.0
	Stimulus				14.0	17.1	10.3	20.0	21.5	19.6	46.3	28.1	31.2
	Experiencer				7.7	16.6	8.2	27.8	27.2	26.3	50.9	27.0	30.4
	Target				18.0	17.1	9.1	21.0	26.7	23.5	51.1	29.0	25.7
EST	Cue	**34.0**	14.0	12.0	**93.0**	10.0	5.0	**62.0**	13.0	10.0	**64.0**	13.0	10.0
	Stimulus				**85.0**	11.0	7.0	**93.0**	10.0	5.0	**60.0**	13.0	9.0
	Experiencer				**93.0**	10.0	5.0	**55.0**	18.0	15.0	**63.0**	15.0	13.0
	Target				**93.0**	10.0	5.0	**86.0**	12.0	8.0	**62.0**	14.0	11.0
`RolEmo`	Cue	29.2	**28.9**	**28.4**	23.3	**29.3**	**25.4**	33.2	**30.0**	**27.9**	31.4	27.9	**41.6**
	Stimulus				24.4	**27.5**	**25.3**	26.5	**27.3**	**26.1**	34.6	**28.3**	**38.6**
	Experiencer				22.7	**27.8**	**24.6**	35.3	**35.7**	**34.7**	42.8	**29.1**	**32.5**
	Target				22.2	**28.2**	**24.6**	30.1	**33.4**	**31.2**	38.5	**34.7**	**34.8**

Table 7. Comparative performance of `RolEmo` with baseline methods under four input settings on the ET dataset. Bold values indicate improvements over the corresponding role-specific settings.

Approach	Role	Full Text			Only Role			Without Role			Role Indicator		
		P_{mac}	R_{mac}	$F1_{mac}$	P_{mac}	R_{mac}	$F1_{mac}$	P_{mac}	R_{mac}	$F1_{mac}$	P_{mac}	R_{mac}	$F1_{mac}$
RolEmo$_{NAM}$	Cue	23.7	29.8	25.6	26.2	27.1	21.9	26.2	27.1	21.9	19.2	25.9	20.1
	Stimulus				20.7	27.1	21.8	20.7	27.1	21.8	16.5	24.4	18.6
	Experiencer				18.3	27.0	21.7	20.4	28.2	22.8	17.3	25.9	19.9
	Target				34.8	**27.4**	22.4	40.0	28.6	23.7	19.2	26.0	20.2
RolEmo$_{NCK}$	Cue	18.7	28.5	22.4	12.6	19.1	15.1	12.6	19.1	15.1	17.3	26.1	20.0
	Stimulus				12.0	18.9	14.7	12.0	18.9	14.7	16.0	24.4	18.5
	Experiencer				12.2	16.7	8.8	19.4	28.7	23.0	17.0	26.1	20.0
	Target				13.6	21.2	16.5	41.1	28.2	23.3	17.2	26.4	20.2
RolEmo$_{NCL}$	Cue	19.1	28.4	22.7	17.7	26.9	21.3	17.7	26.9	21.3	18.0	26.2	20.2
	Stimulus				25.8	27.2	21.5	25.8	27.2	21.5	16.6	24.4	18.7
	Experiencer				18.1	27.0	21.6	19.5	**28.6**	23.0	17.2	26.0	19.9
	Target				34.7	27.2	22.2	35.6	28.3	23.6	17.3	26.5	20.2
EST	Cue	**51.0**	26.0	25.0	**79.0**	18.0	15.0	**63.0**	23.0	22.0	**62.0**	25.0	23.0
	Stimulus				**59.0**	15.0	11.0	**50.0**	23.0	21.0	**57.0**	27.0	27.0
	Experiencer				**80.0**	12.0	7.0	**53.0**	26.0	**24.0**	**48.0**	23.0	20.0
	Target				**64.0**	16.0	14.0	**56.0**	27.0	**26.0**	**65.0**	24.0	21.0
`RolEmo`	Cue	27.5	**30.6**	**28.0**	24.3	**27.8**	**23.4**	24.3	**27.8**	**23.4**	35.2	**32.0**	**28.1**
	Stimulus				26.6	**27.2**	**22.1**	26.6	**27.2**	**22.1**	29.2	**31.5**	**27.8**
	Experiencer				18.3	**27.3**	**21.8**	20.2	28.5	23.0	33.8	**31.6**	**28.3**
	Target				40.3	27.3	**22.5**	41.3	**29.4**	24.6	27.3	**30.7**	**26.0**

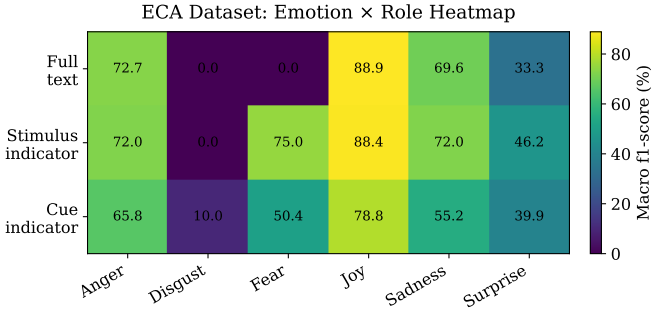

Figure 4. Emotion-wise *macro f1-score* ($F1_{mac}$) performance (%) under the *Full Text* and *Role indicator* settings for ECA dataset.

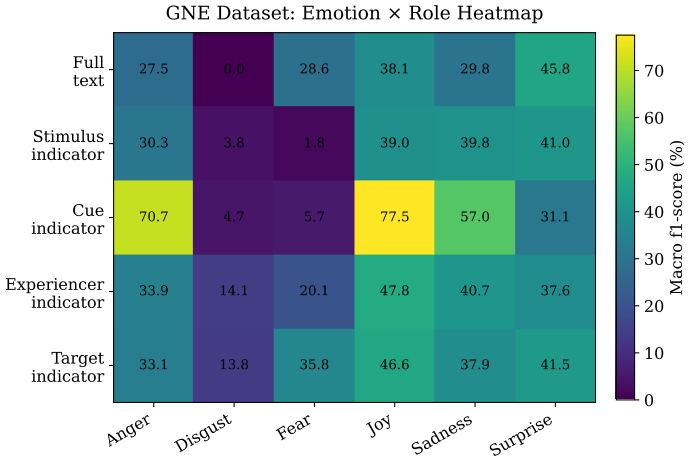

Figure 5. Emotion-wise *macro f1-score* ($F1_{mac}$) performance (%) under the *Full Text* and *Role indicator* settings for GNE dataset.

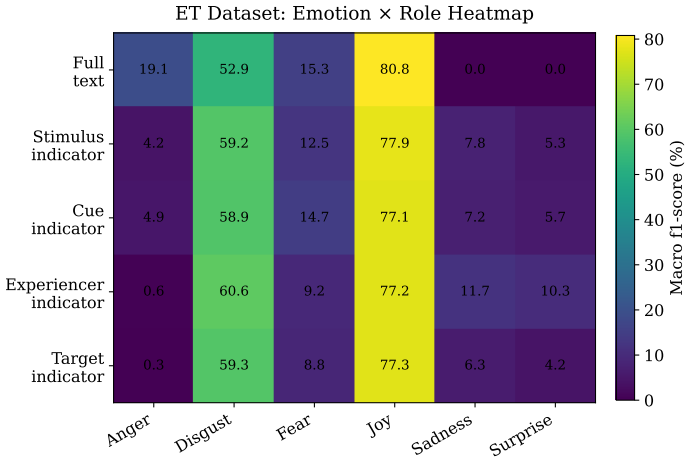

Figure 6. Emotion-wise *macro f1-score* ($F1_{mac}$) performance (%) under the *Full Text* and *Role indicator* settings for ET dataset.

5.1. Role-Based Input Configuration Analysis

As shown in Tables 5–7, the *Full Text* setting outperforms both the *Only Role* and *Without Role* settings across all datasets as it provides complete contextual information required for accurate emotion classification. Conversely, the *Role indicator* setting further improves performance over *Full Text* in most cases, demonstrating that explicitly marking semantic role boundaries enhances the model's ability to use structural information and capture interactions among roles. Nevertheless, performance under the *Full Text* setting varies considerably across datasets, highlighting the influence of corpus-specific characteristics such as genre, annotation quality, and role coverage.

In the *Only Role* setting, where the model is restricted to a single role span, performance consistently degrades across all datasets due to the absence of surrounding context and information from other roles. This degradation is particularly observed in the ECA dataset, where using only the *cue* span results in a substantial performance drop of 31.5%, indicating that *cue* expressions in narrative text are highly context-dependent. Similarly, relying solely on the *stimulus* role leads to performance decreases of 5.7%, 3.1%, and 5.9% for ECA, GNE, and ET, respectively, demonstrating that in the GNE and ET datasets, *cue*, *experiencer*, and *target* spans are often missing, which limits the model's access to these signals. Additional drops are observed when only the *cue*, *experiencer*, or *target* roles are used. In GNE, the corresponding declines are 3.0%, 3.8%, and 3.8%, while in ET they reach 4.6%, 6.2%, and 5.5%. These results indicate that isolated role spans are insufficient for accurate emotion prediction.

In the *Without Role* setting, removing a role span can sometimes help a model improve its performance. For instance, in GNE, masking the *experiencer* role achieves a performance increase of 6.3%, suggesting a potential bias toward particular *experiencers* and suggesting that this role may act as a confounder. A similar improvement of 2.8% is observed when removing the *target* role. In ECA, excluding the *stimulus* role results in a notable performance gain of 9.5%, highlighting the importance when other roles are present. This suggests that the *cue* captures most of the emotion information. In contrast, for all datasets, the *cue* role is less dominant, although excluding it still results in performance degradation. Except for the ECA dataset, masking the *stimulus* results in performance decreases. In GNE, the performance drops the most when removing the *stimulus* span and improves the most when only the *stimulus* is available. While in the ET corpus, removing the *stimulus* span causes a performance drop of 5.9%, while providing only the *stimulus* achieves a comparable improvement. As expected, the *stimulus* role plays an important role in emotion classification.

Finally, in the *Role indicator* setting, all information remains available, but explicit positional markers are added for the role under investigation. This configuration leads to performance gains in most cases, demonstrating that explicitly encoding role positions enables the model to utilize structural information more effectively. The strongest improvements are observed for the *stimulus* role in ECA and GNE, with gains of 14.8% and 10.2%, respectively, and for the *cue* role in ECA and GNE, with gains of 5.9% and 13.2%, respectively. In GNE, adding positional indicators for the *experiencer* role also results in a moderate improvement of 4.1%. In contrast, in the ET dataset, adding positional information for the *target* role results in a performance decrease of 2.0%; it leads to a marginal decline for the *stimulus* (0.2%) due to noisy or inconsistent span annotations, while performance improves by 0.3% in the case of *experiencer* role. These results suggest that the effectiveness of positional markers depends on annotation consistency and the quality of role span, which vary across datasets.

5.2. Baseline Comparison

We further compared `RolEmo` against baseline methods (different components of the model). $RolEmo_{NAM}$, which removes the attention mechanism, weakens the model's ability to identify informative tokens. $RolEmo_{NCK}$, which excludes commonsense knowledge, limits the model's ability to capture implicit and contextual emotional cues. $RolEmo_{NCL}$, without contrastive learning, reduces the model's capacity to separate opposing emotional states. *EST* [25], which investigated the impact of semantic roles for predicting emotions.

Tables 5–7 show that compared to $RolEmo_{NAM}$, `RolEmo` improves performance by 1.1–7.8% in the *Full Text* setting, by 0.1–2.4% in the *Only Role* settings, and 0.1–13.1% in the *Without Role* settings, showing that attention mechanisms are essential for identifying informative tokens. Removing commonsense knowledge ($RolEmo_{NCK}$) leads to larger performance drops of 3.7–9.2%, 3.4–23.8%, and 0.1–31.2% across the respective settings. This confirms that pre-trained commonsense inferences provide valuable implicit cues that improve emotion understanding. Similarly, excluding contrastive learning ($RolEmo_{NCL}$) degrades performance by 2.1–8.2% in the *Full Text* setting, 0.2–20.1% in the *Only Role* settings, and 0–16.6% in the *Without Role* settings, indicating its role in improving class separability.

Finally, `RolEmo` outperforms the method of *EST* [25] by 3.0–16.4% in the *Full Text* setting, 8.4–25.8% in the *Only Role* settings, and by 1.1–23.2% across *Without Role* settings, with minor exceptions for the *experiencer* and *target* roles in the ET dataset. Overall, these results demonstrate that `RolEmo` consistently achieves superior performance by jointly leveraging attention, commonsense augmentation, and contrastive learning. Among the semantic roles, *cue* and *stimulus* emerge as the most informative for emotion inference.

5.3. Emotion-Wise Performance Analysis

Figures 4–6 illustrate the impact of incorporating role indicator (positional) information on emotion classification performance. Overall, augmenting the *Full Text* representation with explicit semantic role indicators leads to consistent performance gains, although the improvements vary across datasets and emotion categories. In the ECA dataset, which consists of literary narratives, role-aware representations enable the model to better discriminate across all emotion classes. Notable improvements are observed for *fear*, *sadness*, and *surprise* when *stimulus indicators* are introduced, while *cue indicators* further enhance performance for *disgust*, *fear*, and *surprise*.

In the GNE dataset, improvements over the *Full Text* setting are consistently observed for *anger*, *disgust*, *joy*, and *sadness*, although the extent of these gains is role-dependent. In particular, *cue indicators* achieve the largest gains for *anger* and *joy*, whereas *experiencer* and *target indicators* provide smaller but consistent improvements across emotions. Simi-

larly, in the ET dataset, role indicators improve overall performance relative to the *Full Text* setting, particularly for *disgust*, *sadness*, and *surprise*.

These performance variations can be attributed to dataset-specific characteristics. Manual inspection revealed noisy samples and ambiguous labeling, particularly for overlapping emotions such as *disgust* (often confused with *anger*) and *fear* (frequently overlapping with *sadness*). This issue is further compounded by class imbalance as low-frequency emotions such as *disgust* and *fear* have limited training instances and often lack or are weakly annotated with roles. Moreover, the predictive value of semantic roles varies across emotions and datasets. Certain emotions depend heavily on specific roles; when those roles are absent or weakly expressed, the model lacks sufficient discriminative cues. Consequently, in some configurations, the model fails to correctly identify instances of these under-represented emotions, resulting in zero recall and corresponding zero *macro f1-scores*.

To further assess model performance under class imbalance conditions, Figure 7 presents the ROC curves and AUC scores for the ECA dataset under the *Role Indicator* setting for the *cue* and *stimulus* roles. The curves reveal substantial variation across emotions: *joy* achieve relatively high AUC scores (e.g., *joy* AUC = 0.845 in *cue*, *fear* AUC = 0.848 in *stimulus*), while *disgust* and *surprise* consistently fall near or below the random baseline (e.g., *disgust* AUC = 0.121 in *cue*, *surprise* AUC = 0.139 in *stimulus*), reflecting the difficulty of distinguishing these emotions due to class imbalance and overlapping linguistic patterns. The macro-average AUC is slightly higher for the *stimulus* role (0.518) than for the *cue* role (0.478).

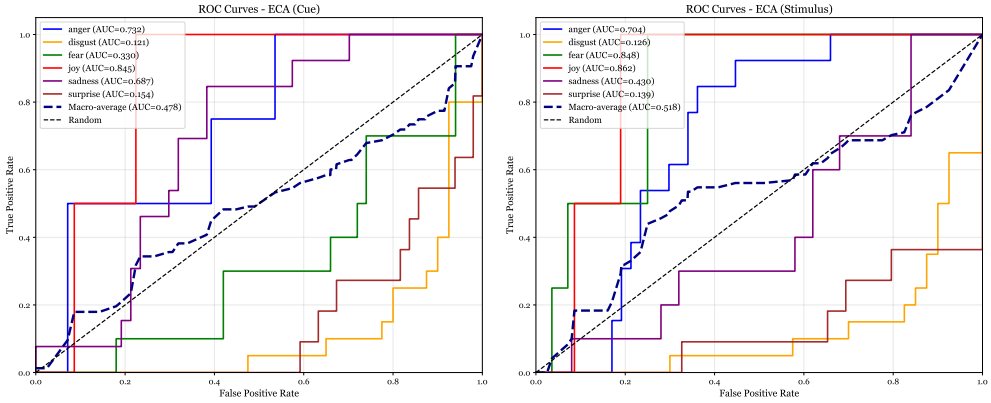

Figure 7. ROC curves for the ECA dataset under the *Role indicator* setting for the *cue* (**left**) and *stimulus* (**right**) roles.

Overall, the heatmap and ROC analyses confirm that role-aware modeling improves emotion classification performance. However, its effectiveness depends on dataset characteristics, annotation quality, and the distribution of emotions.

5.4. Case Study

We present a case study to illustrate the practical implications of the proposed `RolEmo` approach and to demonstrate how individual semantic roles contribute to emotion inference, as shown in Figure 8. We randomly choose a news headline from the GNE dataset [14], annotated with the emotion *anger*: "Sudan protests: Outrage as troops open fire on protesters." In this example, "outrage" serves as the *cue*, "open fire on protesters" as the *stimulus*, "Sudan protests" as the *experiencer*, and "troops" as the *target*.

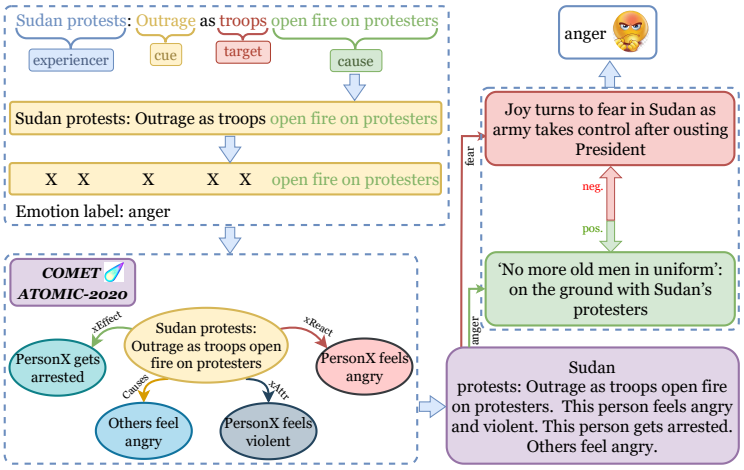

Figure 8. Case study example demonstrating role-aware emotion inference using `RolEmo` on a news headline from the GNE [14] dataset.

To evaluate the informational contribution of the *stimulus* role, we apply the *Only Stimulus* setting, in which only *stimulus* tokens remain visible to the model while all other tokens are masked. This setting directly tests whether the *stimulus* alone provides sufficient information for accurate emotion classification. The phrase "open fire on protesters" alone conveys a clear and explicit emotional signal, enabling correct prediction of *anger* even in isolation.

To further capture implicit contextual cues, commonsense knowledge is incorporated using COMET-ATOMIC-2020 [34]. Given the full headline as input, COMET generates inferences such as *xReact* (PersonX feels angry), *xEffect* (PersonX gets arrested), *xAttr* (PersonX appears violent), and *Causes* (others feel angry). These inferences formalize the intuition that violent suppression of protests conventionally evokes *anger* among both participants and observers, confirming that the *stimulus* alone encapsulates a strong and explicit emotional signal. Additionally, contrastive learning is utilized to refine the emotion embedding space by selecting positive samples from the same emotion class and negative samples from opposing emotion classes, based on cosine similarity. This contrastive structure enables the model to distinguish between semantic similarity and emotional opposition more effectively, thereby enhancing its discriminative capability.

Overall, this case study supports the broader experimental findings by demonstrating that (i) the *stimulus* role can independently encode strong emotional signals, (ii) commonsense knowledge enhances the model's ability to interpret causal emotion understanding, and (iii) role-aware modeling enables interpretable and robust emotion inference in real-world text, where emotional cues are often implicit and context-dependent.

5.5. Discussion

This study demonstrates that emotion classification is inherently structured and role-dependent and that explicitly modeling semantic roles leads to consistent improvements across datasets. The BiLSTM model with `GloVe` non-contextualized embeddings is suitable for our experiments, where systematic manipulation of the input token sequence is required. As shown in Tables 5–7, the substantial performance degradation in the *Only Role* setting confirms that isolated role spans are insufficient for reliable emotion inference. Instead, effective emotion understanding emerges from interactions among multiple semantic roles and their surrounding context. In contrast, the *Role indicator* configuration consistently improves performance by preserving complete contextual information while explicitly encoding role boundaries, allowing the model to use structural cues more effectively.

The comparative analysis further highlights the contributions of attention mechanisms, commonsense knowledge, and contrastive learning. The performance drop observed when removing attention mechanisms underscores their importance in prioritizing emotionally informative tokens. Similarly, excluding commonsense knowledge leads to substantial degradation, underscoring the role of implicit world knowledge in resolving contextual and underspecified emotional cues. Contrastive learning further improves class separability, particularly for emotions that are semantically similar or frequently confused. Together, these components strengthen the representational capacity of `RolEmo` and contribute to consistent gains over baseline methods. Dataset-specific trends also indicate that role-aware modeling is especially beneficial in structurally rich and context-dependent domains (e.g., narratives and news), whereas improvements are relatively moderate in noisier settings where role annotations may be inconsistent or ambiguous, reinforcing the importance of annotation quality for structured emotion modeling.

While transformer-based models (e.g., BERT) provide powerful contextual representations, their multi-layer self-attention mixes information across tokens and spans, which can blur fine-grained role signals with broader sentence-level context. In a role-focused analysis, such contextual blending may reduce the transparency of controlled role ablations, since traces of masked role content can still influence neighboring representations and complicate interpretation. To more cleanly isolate and quantify the contribution of explicit semantic roles, we therefore employ pre-trained `GloVe` embeddings. Their static and interpretable nature preserves clearer token-level boundaries, facilitating role separation when combined with role indicators and masking strategies. Moreover, this lightweight design reduces computational overhead, making it better suited to low-resource settings and practical deployment scenarios.

To assess generalizability, we conducted additional experiments by fine-tuning DistilBERT-base-uncased [44] using the identical role-augmented input format as `RolEmo`, evaluated on the ECA dataset under the *Role Indicator* setting for the *cue* and *stimulus* roles. The results are reported as average macro-F1 across 5 folds. For the *stimulus* role, `RolEmo` (58.9%) outperforms DistilBERT (46.4%), suggesting that role indicators combined with commonsense augmentation provide stronger signal when causal context is present. For the *cue* role, DistilBERT (59.6%) outperforms `RolEmo` (50.0%), possibly reflecting the benefit of contextual representations for surface-level affective cues.

Overall, the results confirm that emotion classification extends beyond surface-level lexical cues to structured semantic inference by explicitly modeling *who* experiences an emotion, *what* triggers it, and *toward whom* it is directed. Additionally, the analysis reveals that (i) the *cue* and *stimulus* roles are the most informative for emotion inference, underscoring the importance of affective expressions and emotional triggers. In contrast, the *experiencer* and *target* roles exhibit more variable contributions, depending on dataset characteristics and annotation quality; (ii) explicit role encoding via the *Role indicator* setting consistently strengthens learning; and (iii) commonsense augmentation and contrastive learning improve discriminability by enhancing implicit reasoning and representation separation. Together, these findings demonstrate that `RolEmo` is an effective framework for interpretable and role-aware emotion understanding.

A limitation of our framework is its dependency on accurate semantic role annotations, which can be resource-intensive to obtain and may not be available for all datasets. The model also exhibits sensitivity to noisy commonsense knowledge, where missing or incorrect inferences can introduce misleading cues and degrade performance. Furthermore, `RolEmo` faces challenges in distinguishing closely related emotion pairs such as *anger-disgust* and *fear-sadness*, which share overlapping linguistic patterns and are often ambiguously labeled.

6. Conclusions and Future Work

In this paper, we proposed `RolEmo`, a role-aware emotion classification framework that integrates semantic role modeling, external commonsense knowledge, and supervised contrastive learning. Through controlled role-masking experiments on multiple benchmark datasets, we systematically analyzed the contribution of individual semantic roles to emotion inference. The results show that role contributions are both role and dataset-dependent. The *cue* and *stimulus* roles are the most informative, while the *experiencer* role exhibits more variable behavior and may act as a confounding factor in certain contexts. Furthermore, explicitly encoding role boundaries through positional indicators improves performance, highlighting the importance of preserving both contextual information and structural cues in emotion modeling. Overall, these findings demonstrate that effective emotion classification extends beyond surface lexical features and benefits from structured semantic representations and commonsense augmentation. To address current limitations in future work. First, exploring semi-supervised or weakly supervised approaches may reduce the dependency on manually annotated semantic roles and improve applicability in low-resource settings. Second, incorporating confidence mechanisms to filter noisy commonsense inferences could enhance robustness and reliability. Finally, developing finer-grained emotion representations may help better distinguish overlapping emotional categories, and integrating contextual embeddings (e.g., BERT) could further improve classification performance and scalability in ambiguous settings.

Author Contributions: Conceptualization, M.A. and A.B.; methodology, A.B.; validation, M.A. and A.B.; formal analysis, A.B.; data curation, A.B.; writing—original draft preparation, A.B.; writing—review and editing, M.A. and A.B.; visualization, A.B.; supervision, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The study did not involve the creation of new datasets. All datasets used in this research are publicly available and have been referenced in the manuscript.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT-5.2 for grammar correction and language refinement. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

References

1. Plaza-del Arco, F.M.; Cercas Curry, A.A.; Cercas Curry, A.; Hovy, D. Emotion Analysis in NLP: Trends, Gaps and Roadmap for Future Directions. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italy, 20–25 May 2024; pp. 5696–5710.
2. Chakma, K.; Datta, S.; Jamatia, A.; Rudrapal, D. Semantic Role Labelling: A Systematic Review of Approaches, Challenges, and Trends for English and Indian Languages. *Expert Syst.* **2025**, *42*, e13838. [CrossRef]
3. Haque, A.; Abulaish, M. CARES: A commonsense knowledge-enriched and graph-based contextual learning approach for rumor detection on social media. *Expert Syst. Appl.* **2025**, *266*, 125965. [CrossRef]
4. Bhardwaj, A.; Abulaish, M. MINDS: A Multi-label Emotion and Sentiment Classification Dataset Related to COVID-19. In Proceedings of the WNLPe-Health 2022: The First Workshop on Context-aware NLP in eHealth, IIIT Delhi, New Delhi, India, 15–18 December 2022; pp. 3–13.
5. Kim, E.; Klinger, R. A Survey on Sentiment and Emotion Analysis for Computational Literary Studies. *Z. Digit. Geisteswiss.* **2019**, 7. [CrossRef]
6. Zentner, S.; Barradas Chacon, A.; Wriessnegger, S.C. The Impact of Light Conditions on Neural Affect Classification: A Deep Learning Approach. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 199–214. [CrossRef]
7. Ghazi, D.; Inkpen, D.; Szpakowicz, S. Hierarchical versus Flat Classification of Emotions in Text. In Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Los Angeles, CA, USA, 5 June 2010; pp. 140–146.

8. Mohammad, S.; Zhu, X.; Martin, J. Semantic Role Labeling of Emotions in Tweets. In Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Baltimore, MD, USA, 27 June 2014; pp. 32–41. [CrossRef]
9. Rebora, S. Sentiment Analysis in Literary Studies. A Critical Survey. *Digit. Humanit. Q.* **2023**, *17*, 1–17.
10. Bhardwaj, A.; Wasi, N.A.; Abulaish, M. Unlocking Emotions in Text: A Fusion of Word Embedding and Lexical Knowledge for Emotion Classification. In Proceedings of the 20th International Conference on Natural Language Processing (ICON), Goa University, Goa, India, 14–17 December 2023; pp. 766–772.
11. Bhardwaj, A.; Abulaish, M. EmoFusion: An integrated machine learning model leveraging embeddings and lexicons to improve textual emotion classification. *Mach. Learn. Appl.* **2025**, *21*, 100693. [CrossRef]
12. Rezapour, M. Emotion Detection with Transformers: A Comparative Study. *arXiv* **2024**, arXiv:2403.15454. [CrossRef]
13. Campagnano, C.; Conia, S.; Navigli, R. SRL4E—Semantic Role Labeling for Emotions: A Unified Evaluation Framework. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Dublin, Ireland, 22–27 May 2022; pp. 4586–4601. [CrossRef]
14. Bostan, L.A.M.; Kim, E.; Klinger, R. GoodNewsEveryone: A Corpus of News Headlines Annotated with Emotions, Semantic Roles, and Reader Perception. In Proceedings of the Twelfth Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; pp. 1554–1566.
15. Mohammad, S.M.; Zhu, X.; Kiritchenko, S.; Martin, J. Sentiment, Emotion, Purpose, and Style in Electoral Tweets. *Inf. Process. Manag.* **2015**, *51*, 480–499. [CrossRef]
16. Gao, Q.; Hu, J.; Xu, R.; Gui, L.; He, Y.; Wong, K.F.; Lu, Q. Overview of NTCIR-13 ECA Task. In Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies, Tokyo, Japan, 5–8 December 2017; pp. 361–366.
17. Alm, C.O.; Roth, D.; Sproat, R. Emotions from Text: Machine Learning for Text-based Emotion Prediction. In Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Vancouver, BC, Canada, 6–8 October 2005; pp. 579–586.
18. Johnson-Laird, P.N.; Oatley, K. The language of emotions: An analysis of a semantic field. *Cogn. Emot.* **1989**, *3*, 81–123. [CrossRef]
19. Thennakoon Mudiyanselage, A.U.G.; Zhang, J.; Li, Y. Context-Aware Emotion Gating and Modulation for Fine-Grained Sentiment Classification. *Mach. Learn. Knowl. Extr.* **2026**, *8*, 9. [CrossRef]
20. Ghazi, D.; Inkpen, D.; Szpakowicz, S. Detecting Emotion Stimuli in Emotion-Bearing Sentences. In Proceedings of the Computational Linguistics and Intelligent Text Processing, Cairo, Egypt, 14–20 April 2015; pp. 152–165. [CrossRef]
21. Kim, E.; Klinger, R. Who Feels What and Why? Annotation of a Literature Corpus with Semantic Roles of Emotions. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018; pp. 1345–1359.
22. Gui, L.; Xu, R.; Wu, D.; Lu, Q.; Zhou, Y. Event-Driven Emotion Cause Extraction with Corpus Construction. In *Social Media Content Analysis: Natural Language Processing and Beyond*; World Scientific: Singapore, 2018; pp. 145–160. [CrossRef]
23. Xia, R.; Ding, Z. Emotion-Cause Pair Extraction: A New Task to Emotion Analysis in Texts. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 1003–1012. [CrossRef]
24. Wei, P.; Zhao, J.; Mao, W. Effective Inter-Clause Modeling for End-to-End Emotion-Cause Pair Extraction. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 3171–3181. [CrossRef]
25. Oberländer, L.A.M.; Reich, K.; Klinger, R. Experiencers, Stimuli, or Targets: Which Semantic Roles Enable Machine Learning to Infer the Emotions? In Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media, Online, 13 December 2020; pp. 119–128.
26. Zhang, Z.; Wu, Y.; Zhao, H.; Li, Z.; Zhang, S.; Zhou, X.; Zhou, X. Semantics-aware BERT for Language Understanding. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 9628–9635.
27. Nag, P.K.; Priya, R.V. Text-based emotion recognition using contextual phrase embedding model. *Multimed. Tools Appl.* **2023**, *82*, 35329–35355.
28. Wang, W.; Yang, Y.; Wu, F. Towards Data-and Knowledge-Driven AI: A Survey on Neuro-Symbolic Computing. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *47*, 878–899. [CrossRef] [PubMed]
29. Nawaz, U.; Anees-ur Rahaman, M.; Saeed, Z. A review of neuro-symbolic AI integrating reasoning and learning for advanced cognitive systems. *Intell. Syst. Appl.* **2025**, *26*, 200541. [CrossRef]
30. Zhang, X.F.; Beauchamp, N.; Wang, L. Generative Entity-to-Entity Stance Detection with Knowledge Graph Augmentation. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 9950–9969. [CrossRef]
31. Mar, J.; Liu, J. From Cognitive to Computational Modeling: Text-based Risky Decision-Making Guided by Fuzzy Trace Theory. In Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2022, Seattle, WA, USA, 10–15 July 2022; pp. 391–409. [CrossRef]

32. Cambria, E.; Liu, Q.; Decherchi, S.; Xing, F.; Kwok, K. SenticNet 7: A Commonsense-based Neurosymbolic AI Framework for Explainable Sentiment Analysis. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, Marseille, France, 20–25 June 2022; pp. 3829–3839.
33. Augustine, E.; Pryor, C.; Jandaghi, P.; Albalak, A.; Dickens, C.; Wang, W.; Getoor, L. Emotion Recognition in Conversation using Probabilistic Soft Logic. *arXiv* **2022**, arXiv:2207.07238. [CrossRef]
34. Hwang, J.D.; Bhagavatula, C.; Le Bras, R.; Da, J.; Sakaguchi, K.; Bosselut, A.; Choi, Y. (COMET-)ATOMIC$^{20}_{20}$: On Symbolic and Neural Commonsense Knowledge Graphs. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 6384–6392. [CrossRef]
35. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, Doha, Qatar, 25–29 October 2014; pp. 1532–1543. [CrossRef]
36. Gardner, M.; Grus, J.; Neumann, M.; Tafjord, O.; Dasigi, P.; Liu, N.F.; Peters, M.; Schmitz, M.; Zettlemoyer, L. AllenNLP: A Deep Semantic Natural Language Processing Platform. In Proceedings of the Workshop for NLP Open Source Software (NLP-OSS), Melbourne, Australia, 20 July 2018; pp. 1–6. [CrossRef]
37. Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised contrastive learning. In Proceedings of the 34th International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 6–12 December 2020; pp. 18661–18673. Available online: https://dl.acm.org/doi/abs/10.5555/3495724.3497291 (accessed on 11 December 2025).
38. Wang, W.; Wei, F.; Dong, L.; Bao, H.; Yang, N.; Zhou, M. MINILM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers. In Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–12 December 2020; pp. 5776–5788. Available online: https://dl.acm.org/doi/abs/10.5555/3495724.3496209 (accessed on 11 December 2025).
39. Lee, S.; Lee, D.B.; Hwang, S.J. Contrastive Learning with Adversarial Perturbations for Conditional Text Generation. *arXiv* **2021**, arXiv:2012.07280. [CrossRef]
40. Plutchik, R. A general psychoevolutionary theory of emotion. In *Theories of Emotion*; Elsevier: Amsterdam, The Netherlands, 1980; pp. 3–33.
41. Ekman, P. Basic Emotions. *Handb. Cogn. Emot.* **1999**, *98*, 45–60.
42. Dresner, E.; Herring, S.C. Functions of the nonverbal in CMC: Emoticons and illocutionary force. *Commun. Theory* **2010**, *20*, 249–268. [CrossRef]
43. Schuster, M.; Paliwal, K.K. Bidirectional Recurrent Neural Networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]
44. Sanh, V.; Debut, L.; Chaumond, J.; Wolf, T. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *arXiv* **2019**, arXiv:1910.01108.

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.