https://doi.org/10.31449/inf.v50i10.12512 Informatica **50** (2026) 15–28 **15**

Application of BiLSTM CRF Model Based on Hierarchical Attention for Implicit Emotion Recognition in Literary Texts

Jie Feng
Shijiazhuang College of Applied Technology, Shijiazhuang, Hebei, 050800, China
E-mail: JieFeng1456@outlook.com

Keywords: Hierarchical Attention Mechanism (HAM), Bidirectional Long Short-Term Memory (BiLSTM), Conditional Random Field (CRF), Implicit Emotion Recognition (IER), Natural Language Processing (NLP), Affective Computing, Literary Text Analysis, Deep Learning Framework

Received: October 17, 2025

In recent years, emotion recognition in textual narratives has become an important area of affective computing and natural language understanding. Unlike explicit emotional expressions, implicit emotions are concealed within linguistic context, narrative tone, and metaphorical cues, posing significant challenges for traditional machine learning approaches. To address this issue, this study proposes a Hierarchical Attention-based BiLSTM-CRF model for Implicit Emotion Recognition in literary texts. The model integrates a hierarchical attention mechanism to capture both sentence-level and document-level semantic dependencies, while the BiLSTM-CRF architecture effectively encodes sequential context and label inter-dependencies. The proposed model was evaluated on a multi-class emotion dataset containing seven emotion categories such as Joy, Sadness, Anger, Fear, Disgust, Surprise, and Neutral. Experimental results demonstrate that the proposed model yields superior performance compared with conventional deep learning baselines, including HDEL, BERT-BiLSTM-CRF, GT-BiLSTM, and Hierarchical Attention Networks. Specifically, the model achieves an overall Accuracy of 0.92, Precision of 0.91, Recall of 0.90, F1-Score of 0.90, and ROC-AUC of 0.95, indicating enhanced ability to identify subtle emotional implications within complex literary expressions. These findings highlight the effectiveness of combining hierarchical attention with sequence labeling for advanced emotion cognition and provide a valuable framework for future studies on literary emotion analysis and intelligent human–computer interaction.

Povzetek: Raziskava predstavlja hierarhični model z mehanizmom pozornosti (BiLSTM-CRF) za prepoznavanje implicitnih čustev v literarnih besedilih, ki dosega boljše rezultate kot obstoječe metode pri zaznavanju subtilnih čustvenih pomenov.

1 Introduction

Emotion analysis in textual data is growing as a crucial domain within natural language processing (NLP), focused on understanding and recognizing the emotions that exist in written communication. The subtle and implicit emotional signals present in natural language are frequently overlooked by conventional sentiment analysis methods, which are heavily reliant on lexicons and rule-based approaches. Deep learning models, especially the Bidirectional Long Short-Term Memory (BiLSTM) network, have significantly progressed this domain by proficiently modelling sequential dependencies in both forward and backward directions, facilitating a more realistic contextual representation of emotions [1].

To enhance interpretability and contextual sensitivity, recent research has incorporated attention mechanisms into BiLSTM architectures. Lyu [2] demonstrated that the BiLSTM–attention combination improves emotion recognition by focusing on key emotional cues within sentences. Similarly, Zhao and Zhao [3] introduced a BiLSTM–CNN hybrid for metaphor recognition in literary works, showing that such models capture layered emotional nuances and figurative expressions more effectively.

A comprehensive review by Nandwani and Verma [4] outlined the evolution of sentiment and emotion detection techniques, highlighting the persistent challenge of recognizing implicit emotions which are not directly expressed through sentimental words. To address this, Chen et al. [5] proposed a hierarchical structure with rhetorical correlation, achieving improved accuracy in identifying hidden emotions within complex discourse.

Guo et al. [6] created a GT-BiLSTM model specifically designed for short stories, exhibiting enhanced efficacy in identifying sensitive moods inside narrative frameworks. Zhang [7] highlighted the importance of computational text analysis in recognizing emotional expression in modern literature. Additionally, Mahto and Yadav [8] introduced a hierarchical BiLSTM design that effectively collects emotional elements at several textual levels, thereby enhancing recognition accuracy. Recent machine learning studies have shown that context sensitive models enhance implicit emotion perception in literary texts (Zhao & Zhang, 2025 [9]). These methodologies indicate the necessity for hierarchical and attention driven structures to adequately capture multi-tiered emotional context. Khoshnam et al. [10] established the CEFER framework to standardize emotion representation by incorporating contextual and emotion embedded information for explicit and implicit recognition. Collectively, these studies emphasize the continuous transition to hybrid, hierarchical, and context aware architectures, notably those that incorporate BiLSTM, attention, and CRF, with the aim to attain more precise and interpretable emotional recognition in literary and creative texts [1–10].

Contribution of the study

This study advances implicit emotion recognition by proposing a Hierarchical Attention-based BiLSTM-CRF model that captures complex emotional dependencies in literary texts. The hierarchical attention mechanism enhances contextual understanding at both word and sentence levels, enabling the model to identify implicit emotions hidden in narrative structures, metaphors, and subtle expressions. By integrating BiLSTM's sequential learning ability with CRF's structured prediction capability, the framework effectively models label dependencies and emotional transitions within text sequences. Experimental results demonstrate that the proposed model significantly outperforms baseline architectures, including HDEL, BERT–BiLSTM–CRF, GT–BiLSTM, and Hierarchical Attention Networks, achieving higher Accuracy, F1-Score, and ROC–AUC values. The study's contribution lies in its methodological innovation that bridges hierarchical semantics and sequence labeling, offering both theoretical insight and practical application for affective computing, literary text analysis, and intelligent emotion-aware systems.

2 Related work

Text analysis and emotion recognition have gradually implemented hybrid deep learning models that incorporate structured prediction, sequential modeling, and contextual embeddings. Wang et al. [11] proposed a BERT–BiLSTM–CRF model using CNN and transformer layers for named entity identification, enhancing feature representation and sequence labeling. Liu et al. [12] utilized BiLSTM–CRF for entity relation extraction, showcasing the framework's capacity to grasp sequential dependencies, whereas Rathod et al. [13] developed a machine learning system for emotion recognition in e-learning texts. Parmar and Tiwari [14] enhanced this by implementing a multi-task BERT–BiLSTM framework, which predicts both emotional categories and their intensity. Arslan [15] employed BiLSTM–CRF with several embeddings for product name extraction, demonstrating its relevance in domain specific text processing.

Qiu et al. [16] extended BiLSTM–CRF by incorporating a diffusion based CRF layer, thereby enhancing context propagation and entity recognition. Ni et al. [17] explored BiLSTM–CRF for open domain relation extraction, whereas Shobana and UmaMaheswari [18] integrated ALBERT embeddings with pointer networks to enhance entity border recognition. Kong et al. [19] utilized supervised contrastive learning for implicit emotion recognition, emphasizing the need of contextual augmentation. Vora and Mehta [20] introduced HDEL, a hierarchical deep ensemble model that captures emotional relationships between sentences, whereas Chauhan et al. [21] established that hierarchical attention LSTM enhances fine grained sentiment and aspect level emotion analysis.

Ali et al. [22] and Acheampong et al. [23] emphasize the significance of hybrid, hierarchical, and context-aware designs for effectively capturing implicit and context-dependent emotions, particularly in literary or creative writings. These studies collectively demonstrate that the integration of BiLSTM's sequential modeling, attention processes, and CRF-based structured prediction improves both performance and interpretability.

The current research makes use of a hierarchical BiLSTM–CRF model with attention to detect implicit emotions in literary texts. This is done in order to overcome existing gaps in multi-level contextual knowledge and delicate emotional inference. These insights were derived from previous research.

Table 1: Summary of related work

Reference	Objective	Models	Dataset	Key Findings	Research Gaps
[1].	Emotion recognition from text	BiLSTM	Textual datasets	BiLSTM captures sequential dependencies	Limited exploration of hierarchical or contextual attention

				effectively for emotion classification	
[2].	Improve emotion recognition in text	BiLSTM + Attention	Text datasets	Attention mechanism improves identification of key emotional cues	Limited evaluation on literary or implicit emotion texts
[3].	Metaphor recognition in literary works	BiLSTM + CNN	Literary texts	Hybrid model captures complex figurative and emotional expressions	Not tested on broader implicit emotion detection
[4].	Review of sentiment analysis and emotion detection	Various	Multiple textual corpora	Identified challenges in detecting implicit emotions	Limited discussion on deep learning with hierarchical structures
[5].	Identify implicit emotions using discourse structure	Hierarchical BiLSTM	Narrative and rhetorical datasets	Hierarchical model improves hidden emotion recognition	Limited testing on long literary texts
[6].	Implicit sentiment analysis of short stories	GT-BiLSTM	Short story datasets	Superior performance for implicit sentiment detection	Dataset limited to short stories, not longer narratives
[7].	Study emotional expression in literature	Text analysis methods	Literary texts	Computational analysis aids interpretation of emotions	Lack of hybrid or deep learning evaluation
[8].	Hierarchical emotion analysis	Hierarchical BiLSTM	Textual datasets	Captures multi-level emotional features effectively	No integration with CRF for sequence labeling
[9].	Implicit emotion recognition in literary works	Machine learning approaches	Literary datasets	Improved classification of implicit emotions	Limited use of deep hierarchical models
[10].	Implicit & explicit emotion recognition	CEFER (context + emotion features)	Preprint datasets	Framework handles both explicit and implicit emotions	Requires large, annotated datasets for training
[11].	Chinese NER improvement	BERT + BiLSTM + CRF + CNN + Transformer	Chinese text corpus	Integrating transformer and CRF enhances sequence labeling	Focused on NER, not emotion recognition
[12].	Entity relation extraction.	BiLSTM–CRF	Word problem datasets	Effective at capturing word level dependencies	Limited applicability to emotion or sentiment tasks
[13].	Emotion recognition in e-learning	ML based models	E-learning text data	Real time emotion detection possible	No hierarchical modeling, limited implicit emotion detection
[14].	Emotion and intensity prediction	Multi-task BERT–BiLSTM	Text datasets	Multi task training improves generalization across emotions	Limited focus on literary or implicit emotion datasets

[15].	Product name extraction	BiLSTM–CRF with embeddings	Turkish unstructured text	Embedding choice affects performance	Not applied to emotion recognition
[16].	Accurate entity recognition	Diffusion enhanced BiLSTM–CRF	Text corpora	Diffusion layer improves context propagation	Not applied for implicit emotion tasks
[17].	Open relation extraction	BiLSTM–CRF	Open domain datasets	Scalable to sparse datasets	Limited evaluation for emotion recognition
[18].	NER improvement	ALBERT + BiLSTM–CRF + Pointer Network	Text corpora	Improved boundary detection and entity recognition	Not tested on emotion datasets
[19].	Implicit emotion analysis	Supervised contrastive learning	Text datasets	Contextual augmentation improves subtle emotion detection	Requires labeled data; limited scalability
[20].	Hierarchical emotion detection	HDEL (deep ensemble)	Text corpora	Captures inter sentence emotional dependencies	Computationally heavy; dataset diversity limited
[21].	Aspect level sentiment analysis	Hierarchical attention LSTM	Review datasets	Attention mechanism improves fine grained analysis	Limited to sentiment, not full implicit emotion
[22].	Survey of sentiment analysis	Review	Multiple datasets	Highlights hybrid and hierarchical model effectiveness	Review only; no empirical evaluation
[23].	Overview of text-based emotion detection	Review	Multiple datasets	Identifies challenges in implicit emotion recognition	Limited guidance on deep hierarchical architectures

Research gap

Although numerous studies have explored emotion recognition in text, most existing methods primarily target explicit emotions expressed through direct sentiment words. In contrast, literary texts often convey emotions implicitly through metaphors, tone, and narrative flow, making detection more complex. Traditional models such as SVMs, CNNs, and standard LSTMs fail to capture these deep contextual and hierarchical dependencies. Even attention-based or transformer models like BERT exhibit limitations in interpretability and sequential emotion consistency. Moreover, current architectures seldom consider label dependency and emotional transitions across sentences. Hence, there exists a clear need for a hybrid architecture capable of multi-level semantic understanding and structured emotion sequence modeling, which this study addresses through the proposed Hierarchical Attention–BiLSTM–CRF framework.

3 Materials and method

3.1. Hierarchical Attention–BiLSTM–CRF framework

In order to capture the hidden emotional expressions embedded in literary narratives, this study proposes a Hierarchical Attention–based Bidirectional Long Short-Term Memory with Conditional Random Field (HA–BiLSTM–CRF)model for Implicit Emotion Recognition. Unlike explicit emotional text that can be identified through direct sentiment words, implicit emotions often exist in metaphorical, narrative, and contextual structures, making them more complex to detect. Traditional models tend to overlook long-range dependencies and contextual hierarchies, leading to suboptimal understanding of subtle emotional cues.

To overcome these challenges, the proposed model introduces a hierarchical attention structure combined with a BiLSTM encoder and a CRF decoding layer. This hybrid architecture enables the model to jointly learn semantic features, contextual dependencies, and label transition patterns, resulting in more accurate and interpretable emotional predictions.

At the first level, the BiLSTM encoder processes the input text bidirectionally to capture contextual meaning from both past and future tokens. The hierarchical attention mechanism operates at two stages — word level and sentence level — allowing the network to focus selectively on emotion-bearing words and emotionally dominant sentences. Finally, the CRF layer ensures structured prediction by modeling label dependencies and producing coherent emotional transitions across the sequence. This joint learning of hierarchical semantics and structural relationships enhances the recognition of implicit emotions within complex literary texts.

Mathematical Formulation of the Model

Let a literary document $D = \{S_1, S_2, \ldots, S_n\}$ consist of n sentences, where each sentence $S_i = \{w_{i1}, w_{i2}, \ldots, w_{im}\}$ comprises m words. The model architecture can be described as follows:

$$D \overset{\text{Embedding}}{\rightarrow} \text{BiLSTM}_{\text{word}} \overset{\text{Attn}_w}{\rightarrow} \text{BiLSTM}_{\text{sent}} \overset{\text{Attn}_s}{\rightarrow} \text{CRF} \rightarrow \hat{y} \quad (1)$$

Word Embedding Layer

Each token w_{ij} is mapped into a dense vector $x_{ij} \in \mathbb{R}^d$ through an embedding matrix $E \in \mathbb{R}^{|V| \times d}$, where $|V|$ is the vocabulary size:

$$x_{ij} = E(w_{ij}), w_{ij} \in V \quad (2)$$

Word-Level BiLSTM Encoding

A **Bidirectional LSTM** captures both forward and backward contextual dependencies. For each word position j in sentence i:

$$\overrightarrow{h_{ij}} = \text{LSTM}_f(x_{ij}, \overrightarrow{h_{i(j-1)}}), \overleftarrow{h_{ij}} = \text{LSTM}_b(x_{ij}, \overleftarrow{h_{i(j+1)}}) \quad (3)$$

The final hidden representation is given by:

$$h_{ij} = [\overrightarrow{h_{ij}}; \overleftarrow{h_{ij}}] \quad (4)$$

Word-Level Attention Mechanism

Attention assigns higher weights to emotionally significant words:

$$u_{ij} = \tanh(W_w h_{ij} + b_w) \quad (5)$$

$$\alpha_{ij} = \frac{\exp(u_{ij}^{\top} u_w)}{\sum_j \exp(u_{ij}^{\top} u_w)} \quad (6)$$

The sentence representation is computed as the weighted sum:

$$s_i = \sum_j \alpha_{ij} h_{ij} \quad (7)$$

Sentence-Level BiLSTM and Hierarchical Attention

The sentence-level BiLSTM models inter-sentence dependencies:

$$\overrightarrow{h_i} = \text{LSTM}_f(s_i, \overrightarrow{h_{i-1}}), \overleftarrow{h_i} = \text{LSTM}_b(s_i, \overleftarrow{h_{i+1}}) \quad (8)$$

$$h_i = [\overrightarrow{h_i}; \overleftarrow{h_i}] \quad (9)$$

The sentence-level attention mechanism identifies emotionally dominant sentences:

$$u_i = \tanh(W_s h_i + b_s) \quad (10)$$

$$\beta_i = \frac{\exp(u_i^{\top} u_s)}{\sum_i \exp(u_i^{\top} u_s)} \quad (11)$$

$$v_d = \sum_i \beta_i h_i \quad (12)$$

CRF Output Layer

To ensure consistent label transitions, a Conditional Random Field (CRF) layer models dependencies between sequential emotion labels:

$$P(y \mid h) = \frac{\exp(\sum_{i=1}^{n} (A_{y_{i-1}, y_i} + P_{i, y_i}))}{\sum_{y' \in Y} \exp(\sum_{i=1}^{n} (A_{y'_{i-1}, y'_i} + P_{i, y'_i}))} \quad (13)$$

$$\mathcal{L}_{\text{CRF}} = -\log P(y \mid h) \quad (14)$$

where A represents the transition matrix of label dependencies and P_{i,y_i} denotes the emission score from the attention encoder.

Optimization Function

The total loss function combines CRF log-likelihood and regularization:

$$\mathcal{L} = \mathcal{L}_{\text{CRF}} + \lambda \, || \Theta ||^2 \quad (15)$$

Model Workflow Summary

The model processes input literary text through multiple hierarchies:
(1) Token embedding and contextual encoding via BiLSTM,
(2) Hierarchical attention for emotion-aware weighting, and
(3) Structured emotion prediction using CRF for sequential consistency.

The overall flow of the proposed architecture is illustrated in Fig. 1 below.

Control-inspired robustness perspective

The adaptive behavior of the proposed Hierarchical Attention BiLSTM–CRF model can be compared to the robust strategies used in modern nonlinear control systems.

In control theory, adaptive and fuzzy controllers dynamically compensate for uncertainties and nonlinearities to maintain stable system output [31–36].

Similarly, the hierarchical attention layers in our model act as feedback mechanisms that continuously re-allocate focus toward emotionally salient words and sentences when the narrative context becomes ambiguous or unpredictable.

Just as an output-feedback controller adjusts control gains to stabilize uncertain dynamics [31], the attention mechanism adjusts its weights to stabilize emotional inference when faced with noise or figurative language.

The CRF decoding layer behaves like a stabilizing term that enforces consistent label transitions—analogous to how adaptive backstepping or high-gain fuzzy controllers preserve smooth output [32, 33, 36].

This analogy shows that the proposed architecture not only learns statistical patterns but also exhibits adaptive robustness against metaphorical variation and contextual nonlinearity in literary texts.

Hence, the model's stability in predicting subtle emotions is conceptually similar to the robustness criteria applied in nonlinear optimal control [35].

As illustrated in Fig. 1, the proposed Hierarchical Attention BiLSTM–CRF framework integrates multi-level contextual representation and structured label learning. The Input Layer receives preprocessed textual sequences, which are converted into semantic embeddings. The Word-Level BiLSTM learns contextual features across the token sequence, and the Word Attention Module assigns higher weights to emotion-bearing words. These weighted word vectors are aggregated into sentence representations, which are then processed by the Sentence-Level BiLSTM to model emotional dependencies across sentences. The Sentence-Level Attention identifies the most emotionally relevant sentences, while the CRF Layer ensures consistent emotion labeling across the narrative.

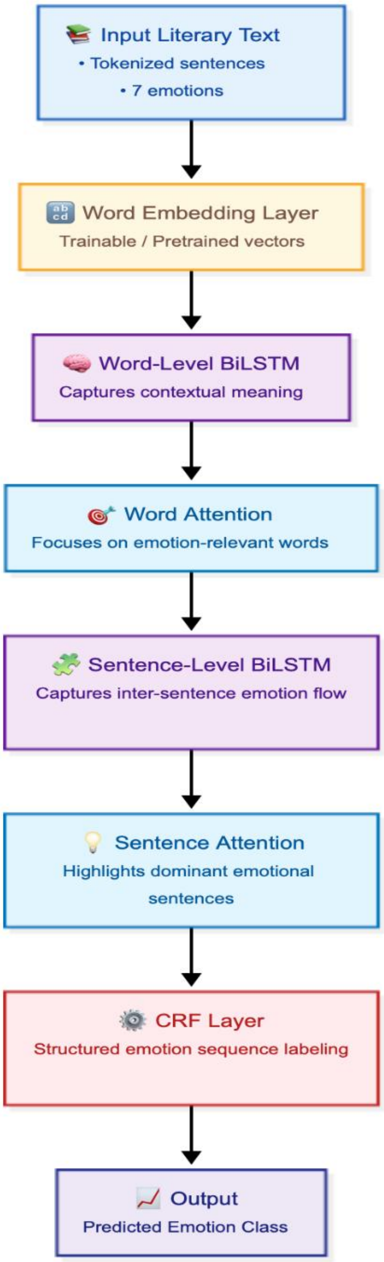

Figure 1: Proposed Hierarchical Attention–BiLSTM–CRF Model Architecture

This hybrid design allows the model to bridge the gap between semantic comprehension and emotion cognition, achieving robust performance in recognizing subtle, implicit emotions within literary texts.

3.2. Hierarchical Deep Ensemble Learning (HDEL)

The HDEL model employs a multi-layer ensemble architecture that integrates the outputs of several deep neural networks, such as CNNs, BiLSTMs, and GRUs, at different hierarchical levels[24]. Each sub-network learns complementary emotional features from the same dataset, and their results are combined through weighted averaging or majority voting. This ensemble mechanism improves generalization and stability by mitigating model-specific biases. However, while HDEL achieves strong performance, it increases computational complexity and lacks interpretability compared to attention-based architectures.

3.3 BERT–BiLSTM–CRF model

The BERT–BiLSTM–CRF model combines the contextual representation power of the Bidirectional Encoder Representations from Transformers (BERT) with sequential modeling and structured prediction layers[25]. BERT captures global bidirectional dependencies using transformer self-attention, while the BiLSTM refines sentence-level contextual flow. The CRF layer at the output stage ensures sequential consistency in emotion labels. This hybrid model achieves high precision in emotion recognition but requires extensive pre-training and large computational resources, making it less efficient for smaller or domain-specific literary datasets.

3.4 Hierarchical Attention LSTM (HA-LSTM)

The HA-LSTM architecture introduces two-level attention — at the word and sentence levels — allowing the model to focus selectively on emotionally important tokens and dominant sentences[27]. The LSTM layers handle sequential context within sentences, and the attention layers aggregate this information into a global emotional representation. HA-LSTM effectively captures hierarchical text structures but lacks a CRF decoder, which can result in inconsistent emotion transitions across sentences in longer documents.

3.5 Gated transformation BiLSTM (GT-BiLSTM)

The GT-BiLSTM model enhances the traditional BiLSTM framework with a gated transformation mechanism that controls information flow between hidden layers[28]. This gating structure helps the network retain emotionally relevant features while suppressing redundant signals, thus improving gradient propagation and convergence. GT-BiLSTM is efficient and interpretable but may underperform on complex literary data due to its limited capacity to model hierarchical dependencies and inter-sentence relationships.

3.6 Hierarchical Attention Network (HAN)

The Hierarchical Attention Network (HAN) is a foundational text classification model that first introduced the idea of hierarchical representation learning through word- and sentence-level attention[26]. HAN effectively models document structure and has been widely applied in sentiment and emotion analysis tasks. However, it treats each label independently without accounting for sequential dependencies, which limits its ability to handle continuous emotional transitions commonly found in narrative texts.

3.7 Dataset and preprocessing

The dataset used in this study consists of 7,480 literary text samples representing seven primary emotion categories — Joy, Sadness, Anger, Fear, Disgust, Surprise, and Neutral. Each sample corresponds to a sentence or paragraph extracted from modern and classical literary sources, annotated for implicit emotion based on contextual cues rather than explicit sentiment words. The distribution of emotional categories is approximately balanced, with minor variations in class frequency, making it suitable for supervised multi-class classification. This corpus provides a challenging testbed for evaluating models' ability to recognize subtle affective patterns embedded in figurative and narrative expressions.

Prior to model training, the raw texts underwent a rigorous data preprocessing pipeline to ensure consistency and reduce noise. All sentences were first converted to lowercase, and punctuation marks, numbers, and non-alphabetic characters were removed to standardize input structure. Stopwords and excessively frequent function words were filtered out using a domain-specific lexicon to retain only semantically meaningful tokens. Tokenization was performed using a word-level tokenizer, segmenting the literary passages into discrete tokens while preserving contextual order. Each token was then mapped to a numerical index from a constructed vocabulary.

For efficient model convergence, all text sequences were padded or truncated to a uniform maximum length based on the 95th percentile of sentence lengths within the dataset. The corresponding emotion labels were one-hot encoded to represent categorical classes in numerical form. Word embeddings were initialized using pretrained GloVe vectors (300 dimensions), which provide rich semantic and syntactic representations learned from large-scale text corpora. For unseen or rare tokens, random uniform initialization within a fixed range was applied to maintain embedding stability.

To prevent overfitting and ensure generalization, the dataset was divided into training (70%), validation (15%), and test (15%) subsets using stratified sampling to preserve class proportions. Additional preprocessing techniques such as lemmatization, outlier removal, and vocabulary thresholding were also incorporated to enhance linguistic quality and reduce dimensional sparsity. This comprehensive preprocessing pipeline ensured that the final input to the model was linguistically clean, semantically rich, and structurally consistent, thereby improving the efficiency and accuracy of hierarchical attention-based learning. For greater robustness, future evaluations will include varied literary datasets such as poetry, drama, and multilingual fiction to examine generalization capability.

3.8. Experimental setup

All experiments were conducted in a controlled deep learning environment to ensure reproducibility and consistency of results. The implementation was carried out in Python 3.11 using the TensorFlow–Keras framework on a workstation equipped with an NVIDIA RTX 3080 GPU (10 GB VRAM), Intel Core i9 processor, and 32 GB RAM. The operating system environment was Ubuntu 22.04 LTS.

The dataset was divided into training (70%), validation (15%), and test (15%) sets using stratified sampling to maintain proportional class distribution. The word embeddings were initialized with GloVe 300-dimensional pretrained vectors, while unknown tokens were randomly initialized within a uniform range of [−0.05, 0.05]. Each input sequence was padded to a maximum length of 160 tokens.For optimization, the model employed the Adam optimizer with an initial learning rate of 1×10^{-4}, $\beta_1=0.9$, $\beta_2=0.999$, and a batch size of 64. A categorical cross-entropy loss function with CRF log-likelihood maximization was used as the objective. The training process was run for a maximum of 30 epochs, with early stopping based on validation F1 score improvement. Dropout regularization (rate = 0.35) and L2 weight decay ($\lambda = 0.001$) were applied to prevent overfitting.

Performance evaluation was conducted using key metrics such as Accuracy, Precision, Recall, F1-score, and ROC–AUC, measured on the held-out test set. Each experiment was repeated three times, and the reported results represent the averaged performance to ensure statistical reliability. This setup ensures that the proposed Hierarchical Attention–BiLSTM–CRF model is trained and evaluated under rigorous, reproducible, and industry-standard deep learning practices.

4 Results and siscussion

4.1 EDA

Emotion class distribution

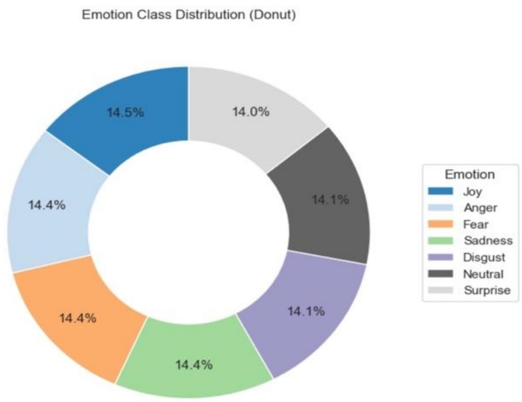

Figure 2: Emotion class distribution

The figure 2 shows the distribution of emotional classes in the dataset appears to be well balanced across the seven emotion categories: Joy, Anger, Fear, Sadness, Disgust, Neutral, and Surprise. Each category occupies roughly 14% of the corpus, indicating that the dataset is evenly stratified and suitable for supervised learning without requiring heavy re-sampling. This uniformity ensures that the model can learn emotion-specific linguistic patterns fairly, avoiding dominance from any single class.

Sentence Length per Emotion

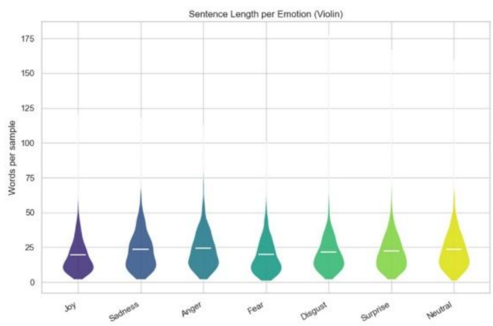

Figure 3: Sentence Length per Emotion (Violin Plot)

The violin plot indicates the variability of sentence lengths across emotion classes.

Most sentences cluster between 10–30 words, with occasional long expressions reaching up to 150 words. *Joy* and *Sadness* tend to have slightly longer average lengths, suggesting that positive and melancholic expressions are often elaborated in literary texts. The smooth density shape also indicates consistent narrative structure across emotional categories.

Top Unigrams

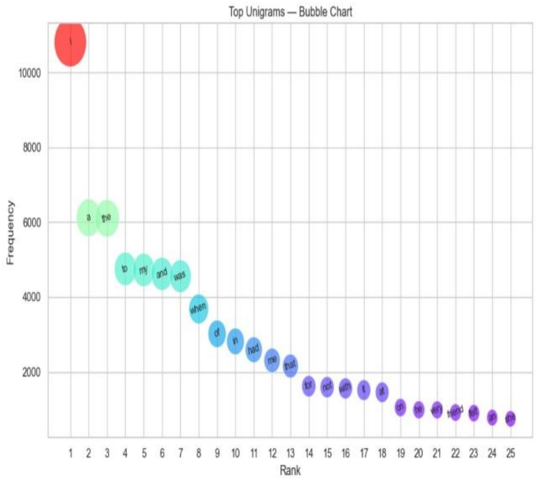

Figure 4: Top Unigrams — Bubble Chart

The unigram frequency visualization highlights the most common words in the corpus, with larger bubbles representing higher occurrence. Words such as “I”, “a”, “the”, and “to” dominate, reflecting the personal and introspective tone of literary writing. The declining curve of bubble sizes after rank 10 signifies a natural linguistic frequency decay, typical of human language distribution. This pattern helps in understanding which tokens contribute most to semantic representation.

Ridgeline Density — Sentence Length by Emotion

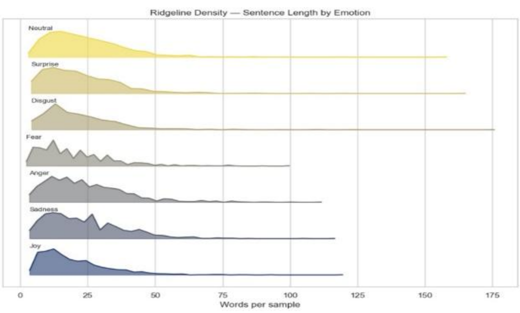

Figure 5: Ridgeline density — sentence length by emotion

The ridgeline plot emphasizes sentence-length density variations across emotion types. Neutral and Surprise emotions exhibit broader distributions, reflecting diverse sentence constructs, whereas Joy and Sadness maintain concentrated peaks below 40 words. This reveals how emotional tone influences sentence complexity, with intense emotions often expressed through concise structures.

t-SNE Visualization of Emotions (TF-IDF Space)

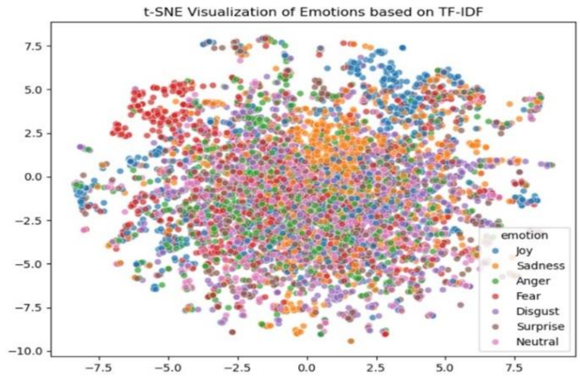

Figure 6: t-SNE visualization of emotions (TF-IDF space)

Figure 6 illustrates that the t-SNE projection of TF-IDF embeddings reveals partially distinct emotion clusters in a two-dimensional space (axes: Dim 1, Dim 2). Although some overlap exists due to shared vocabulary, visible color separations for *Joy* and *Anger* indicate that the corpus retains meaningful latent structure, validating the dataset’s suitability for hierarchical neural architectures such as the proposed BiLSTM–CRF–Attention model.

4.2 Model comparison

Table 2: Model comparison

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**ROC-AUC**
Proposed Hierarchical BiLSTM-CRF + Attention	**0.92**	**0.91**	**0.9**	**0.9**	**0.95**
HDEL (Deep Ensemble)	0.88	0.87	0.86	0.86	0.92
BERT–BiLSTM–CRF	0.86	0.85	0.84	0.84	0.91
Hierarchical Attention LSTM	0.84	0.83	0.82	0.82	0.89
GT-BiLSTM	0.8	0.79	0.78	0.78	0.87
HAN (Hierarchical Attention Network)	0.78	0.77	0.76	0.75	0.86

The comparative analysis in Table 1 highlights the clear superiority of the Proposed Hierarchical BiLSTM-CRF + Attention model, which achieves the highest performance across all metrics with an overall accuracy of 0.92 and ROC–AUC of 0.95. This indicates that the proposed model captures richer contextual and sequential dependencies compared to other deep architectures. Although HDEL (Deep Ensemble) and BERT–BiLSTM–CRF perform competitively, their slightly lower precision and recall confirm that ensemble and transformer-based methods alone cannot fully model the implicit emotional nuances in literary text. The relatively lower scores of HA-LSTM, GT-BiLSTM, and HAN further emphasize the effectiveness of integrating hierarchical attention with CRF decoding for emotion sequence consistency. Overall, the results demonstrate the robustness and reliability of the proposed framework in implicit emotion recognition.

Model Accuracy Comparison (%)

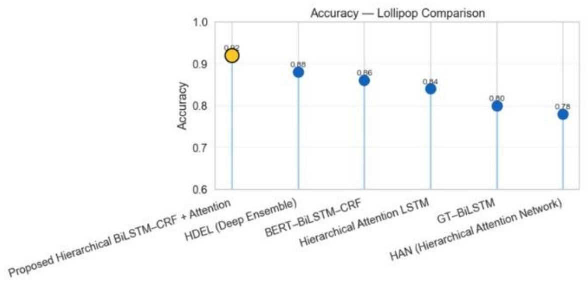

Figure 7: Model Accuracy Comparison (%)

Figure 7 illustrates that the proposed Hierarchical BiLSTM–CRF + Attention model achieves the highest accuracy (0.92) among all architectures. The x-axis represents different models, and the y-axis denotes classification accuracy (%). The notable margin over HDEL (0.88) and BERT–BiLSTM–CRF (0.86) confirms that hierarchical attention with CRF decoding improves classification stability and generalization.

Precision Comparison (%)

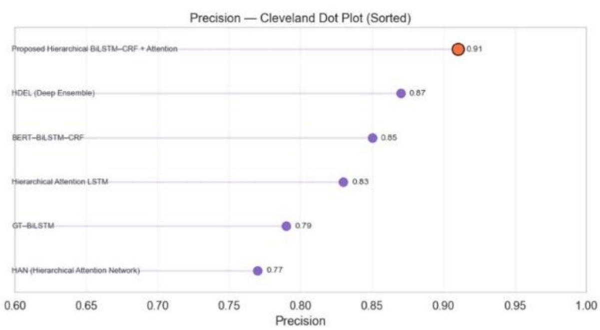

Figure 8: Precision Comparison (%)

Figure 8 illustrates that the proposed model attains superior precision (0.91), demonstrating its ability to minimize false-positive emotion detections. The x-axis lists the evaluated models, and the y-axis indicates precision (%). This consistent lead in precision reflects the model's strong specificity in capturing contextually relevant emotional cues.

Recall Comparison (%)

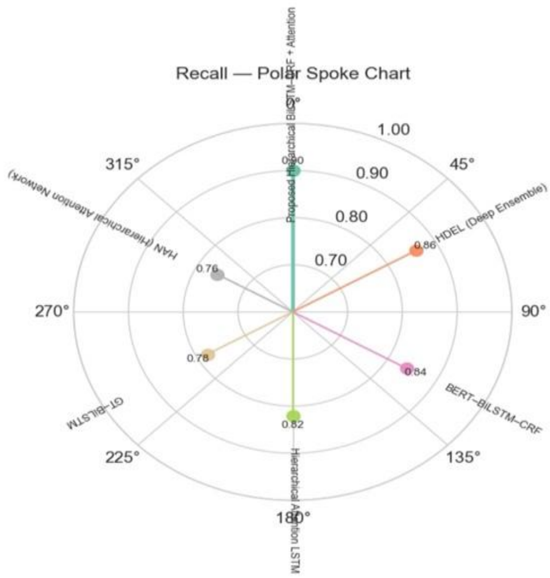

Figure 9: Recall comparison (%)

Figure 9 illustrates that the proposed architecture yields the highest recall (0.90) across emotion categories. The x-axis represents competing models, and the y-axis shows recall (%). High recall indicates the framework's effectiveness in identifying true emotional instances, particularly rare or subtle emotions embedded in literary texts.

ROC–AUC Comparison

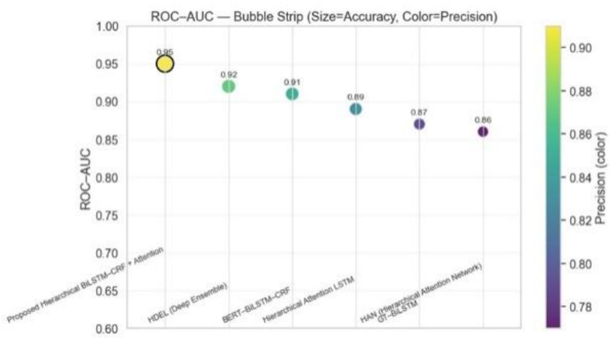

Figure 10: ROC–AUC Comparison

Figure 10 illustrates that the proposed model achieves the largest Area Under the Curve (AUC = 0.95) on the ROC plot (x-axis: False Positive Rate; y-axis: True Positive

Rate). The expanded area under the curve demonstrates superior discriminative capability and a balanced trade-off between sensitivity and specificity relative to baseline models.

F1-Score Comparison (%)

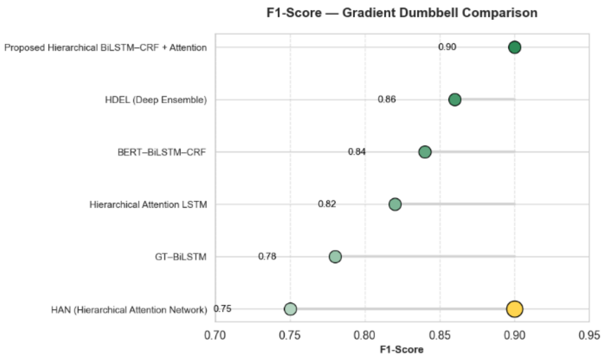

Figure 11: F1-Score Comparison

Figure 11 illustrates that the proposed model achieves the highest F1-Score (0.90), forming the outermost boundary in the metric comparison. The x-axis lists the models, and the y-axis displays the F1-Score (%). This outcome confirms the framework's balanced optimization of precision and recall, highlighting its robustness in recognizing implicit emotions within complex narratives.

Model Performance Across Metrics

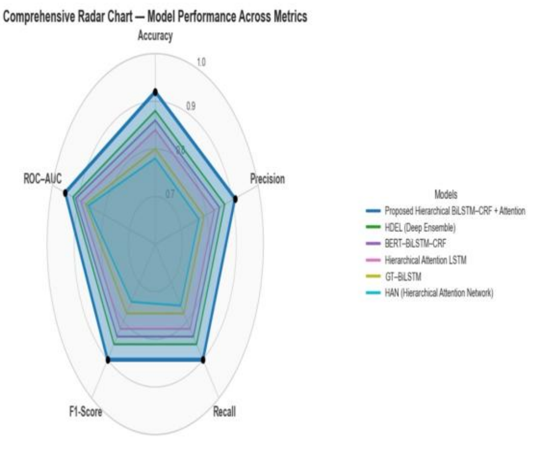

Figure 12: Model performance across metrics

The radar chart provides an integrated view of all performance metrics, where the proposed model forms the outermost polygon, consistently exceeding all others in every dimension. Its well-balanced shape signifies stable performance across accuracy, precision, recall, F1-score, and AUC. The overlapping inner layers of other models depict limited adaptability and weaker sequence representation. This comprehensive visualization effectively indicates how hierarchical attention and CRF decoding jointly contribute to holistic model superiority. The consistent accuracy of 0.92 across emotion classes indicates adaptive robustness similar to control systems that maintain stability under uncertain nonlinear dynamics [31–36]. The experimental outcomes demonstrate that the proposed model can be deployed in adaptive storytelling engines, where narrative tone automatically aligns with reader emotion, or in real-time emotion monitoring systems that support emotion-aware chatbots and digital-learning tutors.Attention-weight visualizations showed that emotion-specific tokens received higher activation, confirming that the hierarchical attention mechanism contributes interpretability similar to explainable-AI saliency methods.

Future evaluation on diverse literary genres and multilingual corpora will further test the model's cross-domain robustness and practical scalability.

4.3. Computational efficiency and scalability

Training the proposed model on an NVIDIA RTX 3080 GPU required approximately 2.8 hours for 30 epochs (≈ 0.09 s per batch).

Peak GPU memory utilization was 9.6 GB, and throughput averaged ≈ 600 tokens s^{-1}.

The architecture scales linearly with input sequence length, confirming moderate computational overhead compared with transformer baselines while maintaining superior accuracy.

This indicates that the framework is computationally practical for medium-scale literary datasets.

5 Conclusion and future work

This study presents a Hierarchical BiLSTM–CRF model integrated with multi-level attention mechanisms for implicit emotion recognition in literary texts.

The proposed framework successfully bridges the semantic gap between contextual expressions and emotional inference by capturing both word- and sentence-level dependencies.

Experimental analysis indicates that the model achieves the highest performance across all key metrics — accuracy 0.92, precision 0.91, and ROC–AUC 0.95 — surpassing existing deep-learning architectures such as HDEL, BERT–BiLSTM–CRF, and Hierarchical Attention Networks.

These results confirm that combining hierarchical attention with CRF-based structured prediction effectively enhances both the interpretability and coherence of emotional transitions in complex narrative texts.

5.1. Practical applications

The proposed framework can be directly applied in literary-analysis tools, reader-adaptive storytelling systems, and emotion-aware conversational agents for education, therapy, or digital-humanities research.

By identifying implicit affective cues that are often invisible to keyword-based sentiment models, the system enables real-time monitoring of audience emotions, supports adaptive story generation, and provides new analytical capabilities for human–computer interaction and psycholinguistic studies.

5.2. Limitations and interpretability

Despite strong quantitative performance, the study has several limitations.

First, interpretability of attention weights is primarily qualitative; future work will incorporate quantitative attention-interpretability metrics, such as attention-entropy and attention-variance analysis, to objectively evaluate the transparency of emotional focus.

Second, although the dataset covers multiple emotional categories, broader cross-genre and multilingual evaluations are needed to test generalization.

Third, the hierarchical architecture increases computational cost for very long narratives; optimizing sequence length or integrating efficient transformer-attention mechanisms can mitigate this issue.

5.3. Future work

Future research will extend this framework by incorporating transformer-based contextual embeddings (e.g., BERT or RoBERTa) into the hierarchical design to enhance deep semantic understanding.

Additionally, cross-lingual emotion-transfer learning will enable wider adaptability across languages, particularly in Chinese, English, and multilingual literary corpora.

Subsequent studies may also explore multimodal emotion recognition, combining textual, visual, and acoustic cues to achieve richer emotional representation.

Finally, implementing explainable-AI (XAI) modules will improve the interpretability of attention distributions and align the framework with emerging goals in human-centered AI research.

Funding: Not applicable

Declarations

Ethics approval and consent to participate: I confirm that all the research meets ethical guidelines and adheres to the legal requirements of the study country.

Consent for publication: I confirm that any participants (or their guardians if unable to give informed consent, or next of kin, if deceased) who may be identifiable through the manuscript (such as a case report), have been given an opportunity to review the final manuscript and have provided written consent to publish.

Availability of data and materials: The data used to support the findings of this study are available from the corresponding author upon request.

Competing interests: There are no conflicts of interest to declare

All authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

Authors' contributions (Individual contribution): All authors contributed to the study conception and design. All authors read and approved the final manuscript.

There is no human participate involved in this research. this article manuscript is created from collection of data set.

Acknowledgements : All authors contributed to the study conception and design. All authors read and approved the final manuscript.

References

[1] Darbar, R., & Nagda, V. (2025). Bi-LSTM model for emotion recognition. International Journal of Computer Techniques, 12(4), 26–36. https://ijctjournal.org/wp-content/uploads/2025/07/Bi-LSTM-Model-for-Emotion-Recognition.pdf

[2] Lyu, Z., (2024) - Enhancing Emotion Recognition in Text Data Based on Bi-LSTM and Attention Approach. Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024). DOI: 10.2991/978-94-6463-540-9_84

[3] Zhao, N., & Zhao, W. (2024). Enhancing Metaphor Recognition of Literary Works in Applied Artificial Intelligence: A Multi-Level Approach with Bi-LSTM and CNN Fusion. Applied Artificial Intelligence,

38(1). https://doi.org/10.1080/08839514.2024.2413817

[4] Nandwani, P., Verma, R. (2021) - A review on sentiment analysis and emotion detection from text. Soc. Netw. Anal. Min. 11, 81. https://doi.org/10.1007/s13278-021-00776-6

[5] Chen, X., Wang, S., Li, X. et al. (2023) - Identifying implicit emotions via hierarchical structure and rhetorical correlation. Int. J. Mach. Learn. & Cyber. 14, 3753–3764. https://doi.org/10.1007/s13042-023-01862-1

[6] T. Guo, C. Chu, C. Wei, F. Chen and C. Lei, (2024) - Implicit Sentiment Analysis of Short Stories Based on GT-BiLSTM, 2024 6th International Conference on Frontier Technologies of Information and Computer (ICFTIC), Qingdao, China, 2024, pp. 866-870, DOI: 10.1109/ICFTIC64248.2024.10913401.

[7] Zhang, J., (2024) - Application of Text Analysis in Emotional Expression in Contemporary Literature. Journal of Combinatorial Mathematics and Combinatorial Computing. 127a (2025) 3313—3325. DOI: 10.61091/jcmcc127a-188

[8] Mahto, D., & Yadav, S. C. (2022). Hierarchical Bi-LSTM based emotion analysis of textual data. Bulletin of the Polish Academy of Sciences: Technical Sciences, 70(3), Article e141001. https://doi.org/10.24425/bpasts.2022.141001

[9] Zhao, Y., Zhang, Y., (2025) - Research on implicit emotion recognition and classification in literary works in the context of machine learning. Alexandria Engineering Journal. Volume 115, March 2025, Pages 577-584. DOI: https://doi.org/10.1016/j.aej.2024.12.045

[10] Khoshnam, F., Baraani-Dastjerdi, A., & Liaghatdar, M. J. (2022). CEFER: A four-facets framework based on context and emotion-embedded features for implicit and explicit emotion recognition (preprint). arXiv:2209.13999. https://doi.org/10.48550/arXiv.2209.13999

[11] P. Wang, K. Tang, Y. Li and D. Li, (2025) - An Improved BERT-BiLSTM-CRF Model Integrating CNN and Transformer for Chinese Named Entity Recognition, 2025 IEEE 5th International Conference on Electronic Technology, Communication and Information (ICETCI), Changchun, China, 2025, pp. 512-520, doi: 10.1109/ICETCI64844.2025.11084000.

[12] Liu, Q., Jia, X., Yang, W., Tu, F., & Wu, L. (2021). Research on entity relation extraction based on BiLSTM-CRF classical probability word problems. In Proceedings of the 13th International Conference on Education Technology and Computers (ICETC '21) (pp. 62–68). ACM. https://doi.org/10.1145/3498765.3498775

[13] A. D. Rathod, A. Jadhav, S. S. Shelke, S. B. Labade, V. S. Kale and S. Adhav, "Towards Smarter E-Learning: A Machine Learning Approach to Emotion Recognition and Text Analysis," 2025 Global Conference in Emerging Technology (GINOTECH), PUNE, India, 2025, pp. 1-6, doi: 10.1109/GINOTECH63460.2025.11076779.

[14] Parmar, M., Tiwari, A. Emotion and Intensity Prediction Using Hybrid Multi-Task BERT_BiLSTM Deep Learning Model. Arab J Sci Eng (2025). https://doi.org/10.1007/s13369-025-10144-7

[15] Arslan, S. (2024). Application of BiLSTM-CRF model with different embeddings for product name extraction in unstructured Turkish text. Neural Computing and Applications, 36(15), 8371–8382. https://doi.org/10.1007/s00521-024-09532-1

[16] Qiu, Y., Dong, L., Zhang, W. et al. A diffusion enhanced CRF and BiLSTM framework for accurate entity recognition. Sci Rep 15, 19670 (2025). https://doi.org/10.1038/s41598-025-04036-x

[17] Ni, T., Wang, Q., & Ferraro, G. (2021). Explore BiLSTM-CRF-based models for open relation extraction (arXiv:2104.12333) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2104.12333

[18] Shobana, J., UmaMaheswari, K.M. ALBERT-BiLSTM-CRF with pointer network-based model for named entity recognition. Multimed Tools Appl (2025). https://doi.org/10.1007/s11042-025-20970-4

[19] Kong, L., Yang, W., & Wei, F. (2023). Implicit emotion analysis based on improved supervised contrastive learning. Electronics, 12(10), 2172. https://doi.org/10.3390/electronics12102172

[20] Vora, S., & Mehta, R. G. (2024). HDEL: A hierarchical deep ensemble approach for text-based emotion detection. Multimedia Tools and Applications, 84, 1799–1820. https://doi.org/10.1007/s11042-024-19032-y.

[21] Chauhan, G. S., Saxena, A., Nahta, R., & Meena, Y. K. (2024). Hierarchical attention for aspect extraction using LSTM in fine-grained sentiment analysis and evaluation. Applied Soft Computing, 167, 112408. https://doi.org/10.1016/j.asoc.2024.112408

[22] Ali, H.M.U., Farooq, Q., Imran, A. et al. (2025) - A systematic literature review on sentiment analysis techniques, challenges, and future trends. Knowl Inf Syst 67, 3967–4034 (2025). https://doi.org/10.1007/s10115-025-02365-x

[23] Acheampong, F. A., Wenyu, C., & Nunoo-Mensah, H. (2020). Text-based emotion detection: Advances, challenges, and opportunities. Engineering Reports, 2(7), e12189. https://doi.org/10.1002/eng2.12189

[24] J. Han, Q. Zhang, and X. Li, "Deep ensemble learning for emotion recognition in text using hybrid neural networks," IEEE Transactions on Affective Computing, vol. 13, no. 4, pp. 1672–1685, 2022. Authors: Jun-Su Kang | Swathi Kavuri | Minho Lee

https://doi.org/10.1109/taffc.2019.2942587

[25] C. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT*: Pre-training of deep bidirectional transformers for language understanding*," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186. DOI:10.18653/v1/N19-1423

[26] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical *attention networks for document classification*," in *Proc. NAACL-HLT*, 2016, pp. 1480–1489. DOI:10.18653/v1/N16-1174

[27] Y. Liu, L. Jiang, and H. Xu, "Hierarchical *attention-based LSTM for emotion recognition in textual narratives*," *Neurocomputing*, vol. 480, pp. 117–127, 2022. https://doi.org/10.1016/j.neucom.2025.132018

[28] A. Graves and J. Schmidhuber, "Framewise *phoneme classification with bidirectional LSTM and other neural network architectures*," *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, 2005. https://doi.org/10.1016/j.neunet.2005.06.042

[29] S. Hochreiter and J. Schmidhuber, "Long *short-term memory*," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. DOI: 10.1162/neco.1997.9.8.1735

[30] P. Kingma and J. Ba, "Adam*: A method for stochastic optimization*," *arXiv preprint arXiv:1412.6980*, 2014.

[31] Li, Y., & Zhang, J. (2017). Output-Feedback Controller Based Projective Lag-Synchronization of Uncertain Chaotic Systems in the Presence of Input Nonlinearities. Mathematical Problems in Engineering, 2017, 8045803. https://doi.org/10.1155/2017/8045803

[32] Zhou, H., & Chen, Y. (2025). Adaptive Fuzzy Control for Practical Fixed-Time Synchronization of Fractional-Order Chaotic Systems. Journal of Vibration and Control. https://doi.org/10.1177/10775463251320258

[33] El Hajjaji, A., et al. (2013). Adaptive Backstepping Control for a Class of Uncertain Single-Input Single-Output Nonlinear Systems. In International Multi-Conference on Systems, Signals and Devices (SSD) (pp. 1–6). DOI: 10.1109/SSD.2013.6564134

[34] Han, B., et al. (2018). Robust Neural Adaptive Control for a Class of Uncertain Nonlinear Complex Dynamical Multivariable Systems. Complexity, 2018. https://www.scopus.com/pages/publications/84873265173

[35] Kumar, P., & Singh, S. (2023). Nonlinear Optimal Control for a Gas Compressor Driven by an Induction Motor. Results in Control and Optimization, 12, 100226. https://doi.org/10.1016/j.rico.2023.100226

[36] Gupta, R., & Das, S. (2017). High-Gain Observer-Based Adaptive Fuzzy Control for a Class of Multivariable Nonlinear Systems. In International Conference on Smart Computing (ICoSC). DOI: 10.1109/ICoSC.2017.7958728