SemEval-2026 Task 2: Predicting Variation in Emotional Valence and Arousal over Time from Ecological Essays

Andrea Lolli, Amirhossein Lotf Ranaei, Xi Fan,
Bilal Hajji, Chiara Lunazzi
Politecnico di Torino
{andrea.lolli, s323336, s328208, s335370, s333949}@studenti.polito.it

Abstract

Longitudinal modeling of affect from text requires capturing both linguistic content and temporal emotional dynamics. SemEval 2026 Task 2 introduces a dataset of ecological essays and feeling words annotated with self-reported valence and arousal scores, enabling the study of affect as a time-evolving signal. In this work, we propose a neural architecture that combines pretrained Transformer encoders with temporal sequence modeling to predict continuous valence and arousal over user-specific timelines. Individual texts are encoded using a Transformer-based language model and aggregated through attention-based pooling before being processed by recurrent layers to capture longitudinal dependencies. To adapt pretrained representations under limited data conditions, we explore parameter-efficient fine-tuning strategies.

1 Introduction

The affective circumplex model proposes that emotions can be described in a two-dimensional space defined by *valence* and *arousal*(Russell, 1980). While most affective computing research focuses on static snapshots or external perception of emotion, real emotional experience is inherently *longitudinal*—it evolves over time, shaped by daily routines, life events, and individual patterns. SemEval-2026 Task 2 addresses this gap by introducing a longitudinal dataset of ecological essays and feeling words collected from U.S. service-industry workers over multiple years (2021–2024). This work focuses on **Subtask 1: Longitudinal Affect Assessment**.

Subtask 1: Longitudinal Affect Assessment

Given a chronological sequence of m texts $\{e_1, \ldots, e_m\}$ from a single user, the task is to predict valence and arousal scores for each text (Figure 1):

$$(v_1, a_1), \ldots, (v_m, a_m)$$

The test set contains both *seen users* (appearing in training at earlier timesteps) and *unseen users* (no prior data).

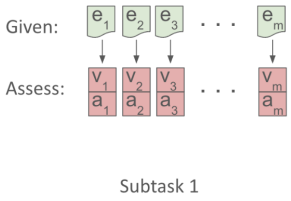


Figure 1: Task 1 Overview

Subtask 2a: Forecasting (future) Variation in Affect

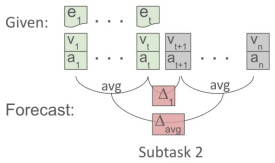


Figure 2: Task 1 Overview

While **Subtask 1** addresses the problem of assessing valence and arousal scores for each text concurrently, **Subtask 2a** is formulated as a forecasting problem (Figure 2). Given a sequence of posts $\{e_1, \ldots, e_t\}$ for a specific user and their corresponding historical affective states $\{y_1, \ldots, y_t\}$, the goal is to predict the future variation Δy_{t+1}. This objective extends the previous task by requiring the model to explicitly learn the trajectory of the user's emotional dynamics.

Data Characteristics

Users contribute varying numbers of texts, ranging from 2 to over 200 documents, with a median of 14 and a right-skewed distribution (Figure 3). Labels are discrete: valence takes 5 values in $\{-2, -1, 0, 1, 2\}$ while arousal takes 3 values in $\{0, 1, 2\}$. Valence is roughly balanced across

the range, whereas arousal concentrates at low values (Figure 4). This asymmetry motivates our loss function design.

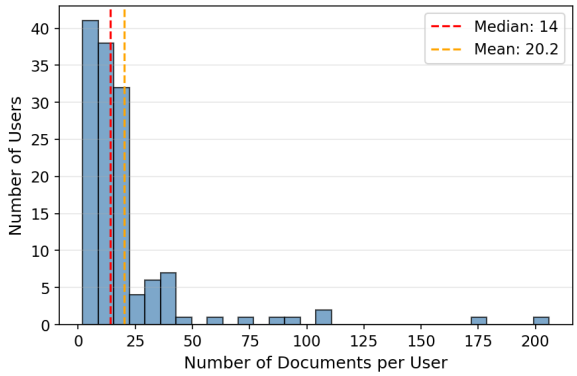


Figure 3: Number of documents per user in train data.

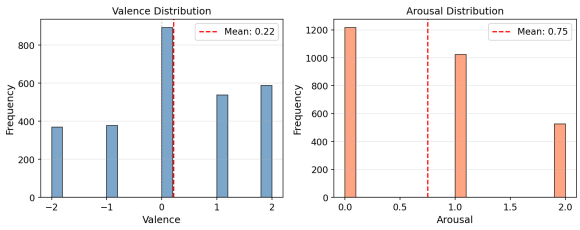


Figure 4: Valence and Arousal distribution in train data.

Modeling Hypothesis

We hypothesize that affective states exhibit a temporal duality. Valence reflects rapid, reactive variations triggered by immediate events and explicit emotional language. Arousal, in contrast, follows a slower, inertial dynamic influenced by routines, persistent stress, and general well-being. This distinction guides our architectural choices.

Contributions

We propose a hierarchical architecture combining BERT, Set Transformer attention (ISAB/PMA), and bidirectional LSTM for longitudinal affect modeling. The attention mechanism extracts salient affective cues from documents, avoiding dilution from mean pooling. We design a hybrid loss combining MSE and CCC with asymmetric weighting to address the valence-arousal difficulty gap. We also provide empirical analysis of parameter-efficient fine-tuning (BitFit, LoRA) for small longitudinal datasets.

2 Background

Mendes and Martins (2023) empirically investigate continuous affect prediction in multilingual settings, assessing the capacity of pre-trained Transformer architectures to quantify valence and arousal across diverse domains. Their comparative analysis demonstrated that model size is a decisive factor in prediction quality, with larger architectures such as XLM-RoBERTa significantly outperforming smaller, distilled variants like DistilBERT. This finding supports the necessity of employing full-scale encoder backbones rather than lightweight alternatives to ensure robust representational power. The study further outlined a standard methodology for this regression task, utilizing a pre-trained Transformer where the pooled representation of the encoder's final layer activations is projected through a linear layer. Regarding optimization, the authors benchmarked several objective functions, including Mean Squared Error (MSE), Concordance Correlation Coefficient Loss (CCCL), and complex alternatives, establishing these metrics as standard objectives for training models on equally weighted valence and arousal targets.

Christ et al. (2022) propose a hierarchical architecture for longitudinal affect modeling, utilizing a pretrained ELECTRA model to encode sentences into fixed-size representations via the standard `[CLS]` token. To capture temporal dependencies, they freeze the encoder and feed sequences of these `[CLS]` embeddings into a bidirectional LSTM, effectively treating the task as a two-stage process. As an alternative to recurrence, they also experimented with Transformer layers using restricted attention masks to simulate a sliding window over the narrative. In both configurations, the system is optimized using a summed Mean Squared Error (MSE) loss on valence and arousal values.

Lee et al. (2019) address the challenge of processing set-structured data, where the model output must be invariant to the permutation of input elements. Unlike Recurrent Neural Networks (RNNs), which are sensitive to input order, or standard feed-forward networks that require fixed-size inputs, the authors propose the **Set Transformer**, an attention-based architecture designed to model interactions among elements in an input set of arbitrary size. A key contribution of their framework is the reduction of computational complexity associated with standard self-attention. By introducing an attention scheme inspired by inducing point methods from sparse Gaussian processes, they reduce the operation cost from quadratic $O(n^2)$ to linear $O(nm)$ (where m is a fixed number of inducing points).

This efficiency allows the architecture to scale to large sets while naturally encoding pairwise and higher-order interactions between elements, making it superior to simple aggregation baselines for complex set-input problems.

3 Methodology

3.1 Data Preprocessing

Efficiently processing longitudinal data requires addressing the variability in both the number of documents per user and the length of individual texts. To enable parallelized batch processing, we implement a hierarchical padding strategy that standardizes input dimensions into a (B, S, L) tensor, where B is the batch size, S is the maximum number of documents per user, and L is the maximum token length.

Hierarchical Masking We generate a binary sequence mask $M_{seq} \in \{0, 1\}^{B \times S}$ to distinguish between genuine user texts and padding. This mask is propagated throughout the network to ensure that padded positions do not contribute to attention computations or the loss function.

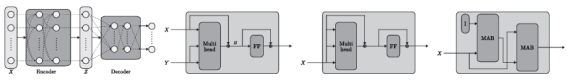
Token Truncation At the document level, we utilize the `bert-base-uncased` tokenizer. Due to the quadratic complexity of attention mechanisms and GPU memory constraints, we limit the maximum token length to $L = 128$. This threshold balances semantic coverage with the memory efficiency required for processing multiple documents per user.

3.2 Model Architecture

Our architecture is a hierarchical neural pipeline that transforms longitudinal text into continuous valence and arousal predictions. It comprises four modules: (1) a Transformer encoder, (2) attention-based aggregation, (3) temporal modeling, and (4) a prediction head.

3.2.1 Encoder Backbone

We employ `bert-base-uncased` as the semantic encoder. Given an input of L tokens, the encoder outputs contextualized embeddings $H_{enc} \in R^{L \times d}$, where $d = 768$. To reduce computational cost and prevent overfitting, we employ Parameter-Efficient Fine-Tuning (PEFT) strategies described in Section 3.4.



(a) Our model (b) MAB (c) SAB (d) ISAB

Figure 1. Diagrams of our attention-based set operations.

Figure 5: Overview of the Set Transformer attention mechanisms (Lee et al., 2019). We utilize the Multihead Attention Block (b) for feature extraction and the Induced Set Attention Block (d) to reduce complexity using latent inducing points.

3.2.2 Attention-based Aggregation (ISAB & PMA)

To aggregate the variable-length sequence H_{enc} into a fixed-size document vector, we reject simple mean pooling in favor of the Set Transformer framework (Lee et al., 2019). This allows the model to learn which tokens carry affect ("signal") versus neutral context ("noise").

Multihead Attention Block (MAB) The fundamental building block, illustrated in Figure 5(b), is a variant of the Transformer block without positional encodings. Given a Query set Q and Key-Value set K, we define:

$$\text{MAB}(Q, K) = \text{LayerNorm}(H + \text{FFN}(H)) \quad (1)$$

where the intermediate attention output H is:

$$H = \text{LayerNorm}(Q + \text{Multihead}(Q, K, K)) \quad (2)$$

Here, FFN denotes a standard row-wise feed-forward network with GELU activations. We fix the number of attention heads to $h = 8$ across all modules.

Induced Set Attention Block (ISAB) Standard self-attention scales quadratically $O(L^2)$. To efficiently filter the input tokens, we use ISAB (Figure 5(d)), which approximates the attention mechanism using m learnable inducing points $I \in R^{m \times d}$:

$$\text{ISAB}_m(X) = \text{MAB}(X, H) \in R^{L \times d} \quad (3)$$

where the latent prototypes H are computed as:

$$H = \text{MAB}(I, X) \in R^{m \times d} \quad (4)$$

This reduces complexity to $O(Lm)$. The inducing points I first compress the input X into a lower-dimensional summary H, which is then used to contextualize the original input tokens.

Pooling by Multihead Attention (PMA) To compress the filtered sequence into a fixed-size representation, we apply PMA using a set of k learnable seed vectors $S \in R^{k \times d}$:

$$V_{doc} = \text{PMA}_k(X) = \text{MAB}(S, X) \in R^{k \times d} \quad (5)$$

In our configuration, we set $k = 8$. The resulting output is flattened into a vector $v_{doc} \in R^{k \cdot d}$ to serve as the input for the temporal model.

3.2.3 Multimodal Fusion (Subtask 2a)

To address the forecasting objective, we adapted the temporal modeling architecture to integrate historical user data. We introduced two key modifications:

Multimodal Input Fusion: We employ a feature-level fusion strategy to combine textual and affective signals. At each timestep t, the aggregated document embedding is concatenated with the corresponding historical valence and arousal values. This enables the LSTM to jointly model the linguistic content and the user's affective trajectory.

State-Aware Prediction Head: The prediction mechanism is conditioned on the most recent observable state. We concatenate the final LSTM hidden state with the last known affective scores before the output projection. This design explicitly informs the model of the prior state, effectively simplifying the task to predicting the *variation* (or delta) required to reach the future state, rather than regressing the absolute values from scratch.

3.2.4 Temporal Modeling and Inference

To capture the longitudinal dynamics of affect, the sequence of document embeddings is processed by an LSTM with a hidden dimension of 256. We fix the depth to 1 layer to mitigate overfitting. We treat directionality as a hyperparameter, evaluating both Unidirectional and Bidirectional configurations. The resulting temporal states are projected via two separate linear heads to predict valence and arousal respectively.

3.3 Optimization Objective

Predicting continuous affect presents a tension between two objectives: minimizing absolute error (magnitude accuracy) and preserving structural trends (correlation). A model trained solely on Mean Squared Error (MSE) tends to regress toward the mean, while pure correlation objectives can be numerically unstable.

The Zero-Variance Problem The gradient of correlation-based losses is inversely proportional to the prediction standard deviation $\sigma_{\hat{y}}$. In early training, if the model predicts near-constant values ($\sigma_{\hat{y}} \to 0$), gradients can explode, causing training instability.

Concordance Correlation Coefficient (CCC) To address this, we employ Lin's Concordance Correlation Coefficient (CCC), which measures both correlation and agreement in mean/variance:

$$CCC = \frac{2\sigma_{\hat{y}}\sigma_y\rho}{\sigma_{\hat{y}}^2 + \sigma_y^2 + (\mu_{\hat{y}} - \mu_y)^2} \quad (6)$$

By including the ground truth variance σ_y^2 in the denominator, CCC remains stable even when $\sigma_{\hat{y}} \approx 0$.

Combined Loss Our loss function combines MSE for magnitude accuracy and CCC for structural alignment:

$$\mathcal{L}_{dim}(y, \hat{y}) = \lambda \cdot \mathcal{L}_{MSE}(y, \hat{y}) + (1-\lambda) \cdot (1 - CCC(y, \hat{y})) \quad (7)$$

The total loss is computed separately for valence and arousal:

$$\mathcal{L}_{total} = \omega_v \mathcal{L}_{dim}(v, \hat{v}) + (1 - \omega_v)\mathcal{L}_{dim}(a, \hat{a}) \quad (8)$$

We set $\lambda = 0.15$ (favoring CCC) and $\omega_v = 0.2$ (emphasizing arousal, which we found more challenging to predict).

3.4 Parameter-Efficient Fine-Tuning (PEFT)

Full fine-tuning of large language models on small longitudinal datasets often leads to overfitting. To address this, we freeze the majority of the encoder parameters and employ two efficient adaptation strategies.

Bias-term Fine-tuning (BitFit) Following Ben-Zaken et al. (2022), we freeze all attention and feed-forward weight matrices, training *only* the bias vectors. The trainable parameter set is defined as:

$$\Theta_{trainable} = \{b^{(l)} \mid l \in Layers\} \cup \Theta_{head} \quad (9)$$

where Θ_{head} denotes the task-specific parameters (ISAB, LSTM, Prediction Head). This approach drastically reduces the number of trainable parameters while allowing the model to shift activation distributions to align with the affect prediction task.

Weight-Decomposed Low-Rank Adaptation (DoRA) To capture more complex dependencies than bias updates allow, we employ DoRA (Liu et al., 2024). While standard LoRA approximates weight updates using low-rank matrices ($\Delta W = BA$), DoRA further decomposes the pre-trained weights into magnitude and direction components. This decomposition improves learning stability and capacity compared to standard LoRA. We apply DoRA to all linear layers with a rank of $r = 8$ and a scaling factor of $\alpha = 16$.

3.5 Implementation Details

The system is trained end-to-end using mixed-precision computation (FP16) to improve efficiency. Due to the memory overhead of the Set Transformer and LSTM unrolling, we use a micro-batch size of 1 with gradient accumulation to simulate an effective batch size of 16. Gradient clipping (1.0) is applied to stabilize optimization.

Differential Learning Rates When utilizing PEFT strategies, we apply a differential optimization schedule. We use a lower learning rate (5×10^{-6}) for the pre-trained encoder parameters to preserve linguistic knowledge, and a higher rate (1×10^{-4}) for the randomly initialized task-specific modules (ISAB, PMA, LSTM, prediction head). This prevents catastrophic forgetting in the backbone while allowing the new attention mechanisms to converge efficiently.

All hyperparameters, including loss weighting coefficients (λ, ω_v) and pooling configurations, are managed via experiment configuration files to ensure reproducibility.

4 Results

While training utilized the hybrid loss function described in Section 3, model selection was performed using the official SemEval evaluation metrics on the validation set. For each dimension (Valence and Arousal), we compute the composite correlation by combining between-user and within-user Pearson correlations via Fisher's z-transformation. This composite correlation is the official ranking metric for the challenge. The overall score is the average of Valence and Arousal composite correlations. Results are summarized in Table **??**.

4.1 Overall Performance

4.1.1 Task 1

The best performing model achieved a composite score of $r = 0.6802$ (Sim 17). This configuration combines a BERT-base-uncased encoder adapted via LoRA, an Induced Set Attention Block (ISAB) with 32 inducing points, Pooling by Multihead Attention (PMA) with 8 seeds, and a unidirectional LSTM for temporal modeling. When ISAB and PMA are disabled, document representations are obtained via mean pooling over the Transformer's final layer token embeddings.

4.1.2 Impact of Attention Mechanisms

The introduction of ISAB and PMA provided significant performance gains over LSTM-only baselines. PMA alone (Sim 6, $r = 0.6211$) substantially outperformed the baseline without attention mechanisms (Sim 3, $r = 0.5776$). Adding ISAB further improved results: with 16 inducing points (Sim 10) the score increased to $r = 0.6555$, and with 32 inducing points (Sim 11) it reached $r = 0.6697$. Notably, PMA primarily improved Valence correlation, while ISAB corresponded to improvements in Arousal correlation. We hypothesize that ISAB filters noise in the user's textual history by attending to the most salient posts, which is particularly beneficial for the more challenging Arousal dimension.

4.1.3 Parameter-Efficient Fine-Tuning

We compared two PEFT strategies: LoRA and BitFit. LoRA (Sim 17, $r = 0.6802$) outperformed both BitFit (Sim 14, $r = 0.6640$) and the baseline without PEFT (Sim 11, $r = 0.6697$). The improvement from LoRA was primarily driven by Valence correlation ($r = 0.8049$), suggesting that adapting encoder weights is necessary to capture semantic nuances of emotional polarity. Arousal correlation also improved ($r = 0.5556$), though to a lesser extent. BitFit, while competitive, proved insufficient for optimal domain adaptation. Combining both strategies (Sim 18, $r = 0.6636$) caused a slight performance decline compared to LoRA alone, suggesting interference between the two adaptation methods.

4.1.4 Temporal Modeling

The effect of LSTM bidirectionality depended on architectural complexity. In simpler configurations without ISAB, bidirectional LSTMs provided improvements: Sim 5 (Bi-LSTM, $r = 0.6340$) out-

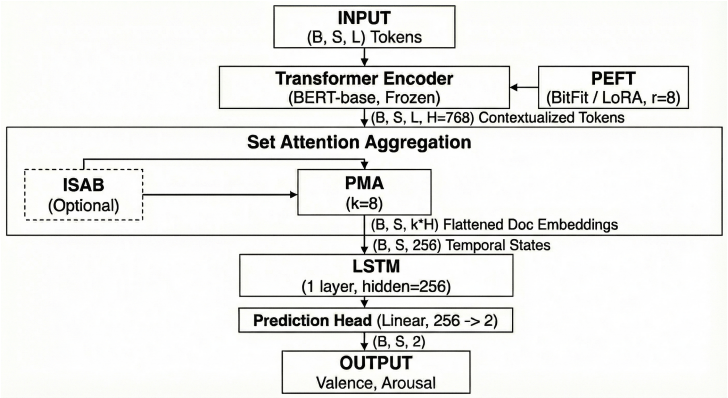


Figure 6: Overview of the proposed hierarchical architecture. The model processes longitudinal text sequences using a shared BERT encoder (optimized via DoRA), aggregates features via Induced Set Attention (ISAB) and Pooling by Multihead Attention (PMA), and models temporal dynamics with a Bidirectional LSTM.

Table 1: Simulation results with different configuration for Task 1. r represent the average composite score.

Sim.	BitFit	LoRA	Inducing (I)	PMA (k)	Bi-LSTM	Batch/Acc	Score (r)	Val. (r)	Aro. (r)
3	FALSE	FALSE			FALSE	1 / 16	0.5776	0.6896	0.4655
4	FALSE	FALSE			TRUE	1 / 16	0.5791	0.6849	0.4733
5	FALSE	FALSE		8	TRUE	1 / 16	0.634	0.7608	0.5071
6	FALSE	FALSE		8	FALSE	1 / 16	0.6211	0.7898	0.4525
9	FALSE	FALSE	16	8	TRUE	1 / 16	0.6438	0.7683	0.5193
10	FALSE	FALSE	16	8	FALSE	1 / 16	0.6555	0.7961	0.515
11	FALSE	FALSE	32	8	FALSE	1 / 16	0.6697	0.795	0.5444
12	FALSE	FALSE	32	8	TRUE	1 / 16	0.6689	0.7835	0.5543
13	FALSE	FALSE	32	8	TRUE	4 / 4	0.6405	0.7811	0.4999
14	TRUE	FALSE	32	8	FALSE	1 / 16	0.664	0.8068	0.5211
15	FALSE	FALSE	32	8	FALSE	4 / 4	0.6291	0.7639	0.4943
16	FALSE	FALSE		8	FALSE	4 / 4	0.6204	0.7717	0.4691
17	FALSE	TRUE	32	8	FALSE	1 / 16	0.6802	0.8049	0.5556
18	TRUE	TRUE	32	8	FALSE	1 / 16	0.6636	0.788	0.5392
19	FALSE	FALSE			FALSE	4 / 4	0.5532	0.6596	0.4468

performed Sim 6 (Uni-LSTM, $r = 0.6211$). However, once ISAB was introduced, unidirectional LSTMs performed better: Sim 10 (Uni-LSTM, $r = 0.6555$) outperformed Sim 9 (Bi-LSTM, $r = 0.6438$), and similarly Sim 11 ($r = 0.6697$) outperformed Sim 12 ($r = 0.6689$). We attribute this to functional redundancy. Both ISAB and bidirectional LSTMs model long-range dependencies; when ISAB handles global attention, the LSTM need only capture local sequential progression. Additionally, BERT already provides bidirectional contextualization at the token level, further reducing the benefit of a bidirectional recurrent layer. The unidirectional LSTM also has fewer parameters, mitigating overfitting on our small dataset.

4.1.5 Results on Subtask 2a

We evaluated different model configurations to identify the optimal strategy for forecasting. To this end, models were trained minimizing the **Mean Squared Error (MSE)** loss, while the **Pearson Correlation Coefficient** (r) was employed as the primary evaluation metric. Results are summarized in Table 2. As a baseline, we employed a **History-Aware Model** which utilizes the previous valence and arousal values concatenated with the BERT embeddings, but lacks the specialized attention mechanisms (ISAB, PMA) and adaptation layers (LoRA) of our full architecture. The best performing model (sim. 6) achieved an average Score of **0.6488**, utilizing the winning architecture from Subtask 1 augmented with a Bidirectional LSTM. Comparing our best model to the baseline, we observe that the introduction of historical data is not sufficient for balance performance since it fails to understand the Arousal trend. Regarding the best configuration, the introduction of **bidirectionality** led to a substantial improvement in **Valence** prediction ($+0.09$) compared to the unidirectional counterpart, effectively balancing the model's performance. Although the Arousal prediction experienced a slight decline, the overall composite score improved due to the better alignment of the valence

Table 2: Simulation results sorted by configuration for Task 2a. **Bold** indicates the best overall performance.

Sim.	LoRA	BitFit	ISAB	Bi-LSTM	PMA	Best Epoch	Valence (r)	Arousal (r)	Avg Score (r)
6	Yes	No	32	Yes	8	1	0.6195	0.6782	**0.6488**
1	Yes	No	32	No	8	12	0.5257	0.7307	0.6282
2	No	No	-	No	-	17	0.4873	0.7579	0.6226
3	No	Yes	32	No	8	10	0.4576	0.7790	0.6183
7	No	No	-	Yes	8	1	0.4613	0.7606	0.6109
8	Yes	No	-	Yes	8	2	0.5444	0.6673	0.6058
5	No	No	-	No	8	13	0.4443	0.6598	0.5521
4	No	No	32	No	8	15	0.4994	0.5943	0.5468
9	No	Yes	32	Yes	8	10	0.4199	0.7109	0.5654

trajectory. The previous findings regarding the necessity of Parameter-Efficient Fine-Tuning (LoRA) for semantic encoding and the capacity of Attention Mechanisms (ISAB and PMA) are confirmed also in the forecasting setting.

5 Conclusion

We presented a hierarchical architecture for longitudinal affect assessment that combines BERT encoding, Set Transformer attention (ISAB/PMA), and recurrent temporal modeling. Our best configuration achieved a composite correlation of $r = 0.6802$ on the validation set. Three factors proved critical for performance: (1) attention-based pooling (PMA) over mean pooling for document aggregation, with ISAB providing additional gains by filtering noisy historical context; (2) LoRA adaptation of the encoder, which outperformed both frozen backbones and BitFit; (3) unidirectional LSTMs when combined with ISAB, avoiding functional redundancy with the attention mechanism. Our approach has several limitations. First, the model treats each document as a flat token sequence, ignoring internal structure such as paragraphs or discourse segments. Second, the encoder relies solely on contextual embeddings without explicit affective knowledge. Third, evaluation was conducted only on the validation split; generalization to the official test set remains to be verified.

5.1 Future Work

Two directions appear promising for extending this work. Incorporating external affective lexicons (e.g., NRC, LIWC) as auxiliary features or through lexicon-guided pretraining could provide explicit emotional grounding. This may improve robustness, particularly for arousal prediction where contextual cues alone proved insufficient. Splitting documents into paragraphs or sentences and applying a second level of attention could capture intra-document structure. This hierarchical approach would allow the model to first aggregate sentence-level affect signals, then combine them into document representations, potentially improving sensitivity to localized emotional expressions.

References

Elad Ben-Zaken, Shauli Ravfogel, and Yoav Goldberg. 2022. Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. *Preprint*, arXiv:2106.10199.

Lukas Christ, Shahin Amiriparian, Manuel Milling, Ilhan Aslan, and Björn W. Schuller. 2022. Automatic emotion modelling in written stories. *Preprint*, arXiv:2212.11382.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. 2019. Set transformer: A framework for attention-based permutation-invariant neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, pages 3744–3753.

Shih-Yang Liu, Chien-Yi Wang, Hongxu Yin, Pavlo Molchanov, Yu-Chiang Frank Wang, Kwang-Ting Cheng, and Min-Hung Chen. 2024. Dora: Weight-decomposed low-rank adaptation. *Preprint*, arXiv:2402.09353.

Gonçalo Azevedo Mendes and Bruno Martins. 2023. Quantifying valence and arousal in text with multilingual pre-trained transformers. *Preprint*, arXiv:2302.14021.

James Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178.