Momentum at SemEval-2026 Task 2: LongVA-RoBERTa Transformer-Based Longitudinal Valence and Arousal Modeling

Supriya Nadiger, Sunil Saumya, Rahul Pujari
Veeresh S. Hiremath, Kiran A. Chikaraddi, Anoop U. Kadkol
IIIT Dharwad & KLE Technological University
supriya.nadiger@iiitdwd.ac.in, sunil.saumya@iiitdwd.ac.in
rahulpujari8775@gmail.com, vshhiremath2004@gmail.com
chikaraddikiran03@gmail.com, kadakolanoopuk@gmail.com

Abstract

Emotion modeling is an inherently dynamic phenomenon that changes over time and is influenced by users' affective tendencies. While traditional methods emphasize categorical emotion classification, real-world emotions extend beyond categorical labels, thus can be represented in a continuous valence and arousal space. We propose a method for SemEval 2026 task 2 to address this challenge. The proposed model **LongVA-RoBERTa** is a transformer-based architecture for Subtask 1 and Subtask 2A. Subtask 1 is a supervised regression problem, in which we predict continuous values for valence and arousal in a two-dimensional space employing partially fine-tuned RoBERTa with attention-pooling mechanism and Multi-Layer Perceptron (MLP). Subtask 2A extends the architecture with a BiLSTM and MLP while keeping the RoBERTa layer unchanged to forecast the short-term longitudinal changes in valence and arousal values. The evaluation metrics used are Pearson correlation coefficient and Mean Absolute Error (MAE). The results for Subtask 1 improved over baseline with competitive performance for Subtask 2A.

1 Introduction

Emotions play a crucial role in human interactions and decision-making processes. Personal experiences, cultural influences, social interactions, and psychological factors all play a role in shaping a person's ways of perceiving and responding to emotions. Emotion recognition has become a key area of research in natural language processing, as it serves as a foundation for interactive intelligent systems, personalized computing, and empathetic systems (LaGrandeur, 2015).

Conventional methods frequently utilise categorical emotion classification (Tripathi et al., 2017; Xie et al., 2019), which are discrete labels. However, the emotions expressed in real-world are often subtle, nuanced and hard to articulate. Moreover, emotional states are inherently dynamic, they evolve across time and are influenced by prior affective experiences. In contrast to categorical emotion classification approaches, the widely used affective circumplex model (Russell, 1980) posits that all emotions can be characterized as points in a two-dimensional continuous space defined by valence and arousal as shown in Figure 1. Valence relates to the degree of pleasantness, and arousal refers to its intensity of the emotion. This representation allows for a more flexible modeling of subtle, ambiguous, and implicitly expressed emotions that are difficult to capture using discrete categorical methods.

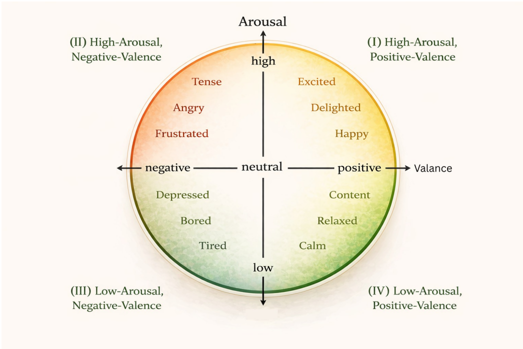

Figure 1: Affective Circumplex Model illustrating emotions in a two-dimensional valence–arousal space by James Russell.

Beyond their representation in a continuous affective space, emotions also exhibit strong temporal dependencies. Research in affect dynamics and longitudinal psychology emphasizes concepts such as emotional inertia and variability, suggesting that emotions unfold and evolve over time (Kuppens et al., 2010; Houben et al., 2015). In real-world longitudinal data, such as essays and personal diaries, emotional expressions often reflect long-term patterns rather than isolated snapshots. Incorporating temporal modeling therefore facilitates a deeper understanding of how emotions fluctuate, persist, and develop across time.

Proceedings of the The 20th International Workshop on Semantic Evaluation (2026), pages 546–552
July 3-4, 2026 ©2026 Association for Computational Linguistics

SemEval 2026 Task 2 "Predicting Variations in Emotional Valence and Arousal over Time from Ecological Essays" (Soni et al., 2026) represents modeling emotions in temporal space incorporating affective circumplex model. It has two subtasks:

- *Subtask 1- Longitudinal Affect Assessment:* The goal is to automatically predict continuous valence and arousal scores from text. Valence represents emotional polarity (negative to positive), while arousal represents emotional intensity (calm to excited). The model predicts signed real-valued scores in the normalized range [-1, +1].

- *Subtask 2- Forecasting(future) Variations in Affect:* The goal is to forecast the changes in valence and Arousal values. This subtask is divided further into *Subtask 2A (State Change):* predicting the variation in affect between consecutive time steps. *Subtask 2B Disposition Change:* predicting the long-term shifts in emotional disposition of a person.

This paper addresses *Subtask 1* and *Subtask 2A*. We propose *LongVA-RoBERTa* (Longitudinal Valence and Arousal for Emotion), a transformer-based framework designed to model affect in a continuous and longitudinal setting. The model leverages RoBERTa as the backbone encoder to obtain contextualised text representations, while task-specific components are incorporated to address the objectives of each subtask.

For Subtask 1 (Longitudinal Affect Assessment), we apply an attention-based pooling mechanism over the RoBERTa embeddings, followed by a two-layer regression head to predict continuous valence and arousal scores. In contrast, *Subtask 2A (State Change Forecasting)* introduces explicit temporal modelling through a BiLSTM layer to capture dispositional affective changes over time.

2 Related Work

Emotion is fundamental to human expression and interaction, exhibiting both temporal variation and dispositional change. Early research in emotion prediction predominantly relied on categorical emotion theories, where emotions are represented as discrete classes such as anger, fear, sadness, joy, disgust, and surprise (Ekman, 1992). This framework has served as the foundation for many classification-based emotion recognition systems. However, given the contextual dependence of emotions and their sensitivity to individual dispositions, categorical models often struggle to capture nuanced and evolving affective states. To address these limitations, Russell's affective circumplex model (Russell, 1980) was introduced, representing emotions within a continuous two-dimensional space defined by valence and arousal. Although proposed earlier, this dimensional perspective has gained increasing prominence in recent years due to its flexibility in modelling subtle and mixed emotional expressions.

Building upon the dimensional representation of affect, several shared tasks and datasets have advanced continuous emotion modelling. The SemEval-2018 Task 1 on Affect in Tweets explored emotion detection in a continuous space by predicting emotion intensity scores alongside discrete labels (Mohammad et al., 2018). Similarly, the EmoBank dataset provides large-scale annotations of text in terms of valence, arousal, and dominance (VAD), facilitating dimensional emotion analysis (Buechel and Hahn, 2017). Mohammad (Mohammad, 2016) further proposed methods for detecting emotion and sentiment intensity in text. Subsequent works, including those in the EmoInt-2017 shared task (Mohammad and Bravo-Marquez, 2017), explored various modelling approaches. For instance, Madisetty (Madisetty and Desarkar, 2017) proposed an ensemble model based on Support Vector Regression (SVR), while Goel and Koper (Goel et al., 2017; Koper et al., 2017) employed CNN–LSTM architectures for intensity prediction.

Despite these advances in dimensional emotion modelling, most existing approaches treat affective states as independent instances, overlooking their temporal dependencies. However, research in affect dynamics highlights concepts such as emotional inertia and dispositional variability (Houben et al., 2015), emphasising that emotions evolve and persist over time. Consequently, modelling affect in longitudinal textual data requires capturing short-term fluctuations as well as longer-term trends. Nevertheless, only a few recent studies have systematically addressed the problem of forecasting affective variation in longitudinal textual settings.

Recent work in dimensional affect modeling uses transformer-based architectures for predicting continuous valence and arousal representations (Mendes and Martins, 2023; Troiano et al., 2023). While these approaches achieve better performance,

	Total	**Train**	**Test**
Samples	5,285	2,764	1,737
Users	182	137	91

Table 1: Overall Dataset Statistics

	Total	**Train**	**Test**
Avg. essays	53.1	40.3	53.3
Median essays	18.0	14.0	50.0
Avg. feeling-words	48.3	42.0	46.6
Median feeling-words	18.0	16.0	40.0

Table 2: Per-User Text Statistics

individual texts are treated independently, limiting their ability to capture temporal dependencies in longitudinal settings.

To address this, recent studies have explored sequence-based modeling methods that capture evolving affective behavior over time (Ganesan et al., 2026; Tseriotou et al., 2023). These approaches highlight challenges in affective persistence and temporal variability, indicating that modeling emotional trajectories requires both contextual and sequential understanding.

In addition, personalization has emerged as an important factor in affect modeling, with approaches incorporating user representations to capture persistent affective tendencies (Mireshghallah et al., 2022). On the other hand, simpler approaches based on frozen transformer embeddings combined with regression models remain competitive baselines, demonstrating that strong contextual representations alone can capture significant affective signals, albeit without explicit temporal or user-level modeling.

3 Dataset Description

The *SemEval 2026 Task 2* dataset comprises longitudinal ecological essays designed for affect modelling over extended periods of time. The textual data are paired with self-reported Ecological Momentary Assessment (EMA) ratings, providing continuous affect annotations grounded in participants' real-time emotional experiences.

The training and evaluation split consist of 5,285 longitudinal text entries authored by 182 participants between 2021 and 2024. The dataset includes both structured feeling-word responses and free-form real-time essays. Since the task requires to predict the valence and arousal values for seen and unseen data, seen users are the ones that appear in train and test dataset and unseen users are those which were not part of model training. The distribution of data across these two formats, along with participant-level statistics, is summarized in Table 1 and Table 2.

4 Methodology

This section describes the modelling techniques, architectural design, and data preprocessing strategies employed for longitudinal affect computation in *SemEval 2026 Task 2*. Our approach addresses both *Subtask 1* (Longitudinal Affect Assessment) and *Subtask 2A* (State Change Forecasting).

4.1 Data Preprocessing

The preprocessing pipeline involved removing null entries, correcting malformed timestamps, and resolving formatting inconsistencies within the dataset. For *Subtask 1*, a strategic data-splitting scheme was adopted under a seen-user setting, where the same users may appear in both training and validation sets, but with different textual instances. This setup enables the model to learn user-specific affect patterns while preventing data leakage at the instance level. For *Subtask 2A*, preserving temporal consistency was essential, as affect forecasting is inherently sequential. Accordingly, all textual entries were chronologically ordered based on their timestamps for each user prior to modelling. To ensure numerical stability during optimisation, the valence, arousal, and state-transition labels were normalised to a fixed range. Additionally, a set of surface-level handcrafted features, such as text length and the frequency of exclamation and question marks, were extracted and incorporated into the modelling process. To distinguish between feeling words and essays in the dataset another boolean feature is_word is incorporated to adapt to variations in expressive granularity.

4.2 Text Representation using RoBERTa

For both subtasks, the textual inputs are first tokenised using the RoBERTa tokenizer and subsequently passed through a pretrained RoBERTa encoder (Liu et al., 2019) to generate contextual embeddings. Let the input token sequence be defined as

$$x = \{t_1, t_2, \ldots, t_T\}.$$

The transformer encoder produces contextual embeddings:

$$H = \text{RoBERTa}(x), \tag{1}$$

where $H \in \mathbb{R}^{T \times d}$, T denotes the sequence length, and d denotes the hidden dimension.

To balance generalisation and task-specific adaptation, we employ a partial fine-tuning strategy. Specifically, the lower layers of the transformer are frozen to preserve the general linguistic knowledge acquired during large-scale pretraining, while the upper layers are fine-tuned to adapt to the affect prediction and forecasting objectives.

4.3 Subtask 1: Longitudinal Affect Assessment

Subtask 1 is formulated as a supervised regression problem aimed at predicting continuous valence and arousal scores for each individual text instance. Let the dataset be defined as

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N},$$

where x_i denotes a text instance and $y_i = (v_i, a_i)$ represents the corresponding normalised valence and arousal values within the range $[-1, 1]$.

4.3.1 Attention Pooling Mechanism

To derive a fixed-length sentence representation from token-level embeddings, we employ a single-head attention pooling mechanism. Given the contextual embeddings $H = \{h_1, h_2, \dots, h_T\}$, attention weights are computed as:

$$\alpha_i = \frac{\exp(W^\top h_i)}{\sum_{j=1}^{T} \exp(W^\top h_j)}, \tag{2}$$

where h_i denotes the contextual embedding of the i-th token produced by RoBERTa, and $W \in \mathbb{R}^d$ is a learnable parameter vector that assigns importance to each token.

The aggregated sentence representation is then computed as:

$$s = \sum_{i=1}^{T} \alpha_i h_i. \tag{3}$$

This attention mechanism enables the model to emphasise emotionally salient tokens rather than relying solely on the `[CLS]` representation.

4.3.2 Multilayer Perceptron (MLP) Regression Head

The attention-pooled representation s is passed through a two-layer Multilayer Perceptron (MLP) with Layer Normalisation to predict the valence and arousal scores. The regression function is defined as:

$$\hat{y} = W_2\, \sigma(W_1 s + b_1) + b_2, \tag{4}$$

where W_1 and W_2 are weight matrices, b_1 and b_2 are bias terms, and $\sigma(\cdot)$ denotes the ReLU activation function.

The final output is given by

$$\hat{y} = (\hat{v}, \hat{a}),$$

where $\hat{v}$ and $\hat{a}$ denote the predicted valence and arousal scores, respectively. The MLP thus maps high-level contextual representations to continuous affective dimensions.

4.4 Subtask 2A: Forecasting Future Variation in Affect

Subtask 2A is formulated as a regression problem aimed at predicting short-term changes in emotional valence and arousal from longitudinal ecological text data. The objective is to model how affective states evolve over time for each individual user.

Let $V_u(t)$ and $A_u(t)$ denote the valence and arousal ratings of user u at time step t. The prediction targets correspond to the immediate next-step affective variation, defined as:

$$\begin{aligned} \Delta V_u(t) &= V_u(t+1) - V_u(t), \\ \Delta A_u(t) &= A_u(t+1) - A_u(t). \end{aligned} \tag{5}$$

4.4.1 Sequential Refinement using BiLSTM

Although transformer encoders capture rich contextual representations, we introduce an additional sequential modelling layer to better account for temporal dependencies in emotional trajectories. The contextual embeddings obtained from RoBERTa, denoted by H, are passed through a bidirectional Long Short-Term Memory (BiLSTM) network:

$$Z = \text{BiLSTM}(H), \tag{6}$$

where Z represents the sequence of hidden states. The BiLSTM captures both forward and backward dependencies, enabling improved modelling of emotional flow across time. Since the input to BiLSTM at each time step t, is a sliding window

Model	Valence (V)						Arousal (A)					
	r_{comp}	$r_{between}$	r_{within}	mae_{comp}	$mae_{between}$	mae_{within}	r_{comp}	$r_{between}$	r_{within}	mae_{comp}	$mae_{between}$	mae_{within}
Baseline (linear BERT)	.557	.659	.435	.743	.472	.886	.299	.343	.253	.459	.311	.585
Proposed Model	**.6381**	**.7147**	**.5463**	**.6455**	**.4159**	**.7977**	**.4552**	**.5075**	**.3996**	**.3998**	**.2561**	**.5262**

Table 3: Subtask 1 Results on the Main Test Set for Valence (V) and Arousal (A)

$Z_t = (z_{t-k+1}, \ldots, z_t)$, of k past utterances, the model strictly adheres to this historical context with no access to future time steps $t+1, \ldots$ during forward or backward processing. The final hidden representation is extracted as the pooled affective representation:

$$z = Z_{\text{last}}. \tag{7}$$

This representation integrates semantic and temporal emotional information.

4.4.2 Feature Fusion and Prediction

To incorporate multiple sources of information, the final feature vector is constructed via concatenation:

$$x = z \oplus e_u \oplus V_u(t) \oplus A_u(t) \oplus s_t, \tag{8}$$

where e_u denotes the learned user embedding, $V_u(t)$ and $A_u(t)$ provide the current affective state, and s_t represents surface linguistic features.

The fused representation is passed through a shared multilayer perceptron with two separate regression heads to predict changes in valence and arousal:

$$\begin{aligned} \Delta\hat{V}_u(t) &= W_V x + b_V, \\ \Delta\hat{A}_u(t) &= W_A x + b_A, \end{aligned} \tag{9}$$

where W_V and W_A are weight matrices, and b_V and b_A are bias terms. The EMA data is recorded sporadically. However, the present model treats all the time steps uniformly, which does not capture the elapsed time gap during emotion modelling.

5 Experimental Setup and Results

This section describes the experimental configuration for Subtask 1 and Subtask 2A, along with the corresponding evaluation results. Model performance is assessed using the Pearson correlation coefficient (r) (Benesty et al., 2009) and Mean Absolute Error (MAE) (Willmott and Matsuura, 2005). We employ the RoBERTa-base as the fundamental encoder for all the experiments. To ensure consistent input representation of feelings words and essays, all inputs are tokenized for a fixed maximum length of 128 tokens with standard padding and truncation. To mitigate overfitting, the lower eight transformer layers are frozen, while the top four layers are fine-tuned during training.

Model	Valence (V)		Arousal (A)	
	r	mae	r	mae
Baseline (linear(prev))	.615	1.168	.670	0.638
Proposed Model	.5529	1.1389	.5887	0.6979

Table 4: Subtask 2A Results

5.1 Subtask 1: Longitudinal Affect Assessment

Subtask 1 involves predicting continuous valence and arousal scores for each text instance using `user_id`, text, and timestamp information, with affect annotations available only in the training set. The model is trained using L1 loss with the AdamW optimizer (learning rate 2×10^{-5}, weight decay 0.01), a dropout rate of 0.2, and a batch size of 16 for up to 10 epochs with early stopping. Results are compared against the official baseline in Table 3.

The table reports performance under composite ($comp$), between-user ($between$), and within-user ($within$) settings, corresponding to overall, inter-user, and intra-user evaluation, respectively. The proposed model consistently outperforms the baseline for valence, achieving higher correlation and lower MAE across all settings. For arousal, similar gains are observed, particularly in composite and between-user correlation, demonstrating improved modelling of both global affect patterns and user-specific temporal dynamics.

5.2 Subtask 2A: State Change Forecasting

Subtask 2A focuses on predicting short-term changes in valence and arousal. Since multiple temporal entries exist for each user, preventing data leakage is critical. Therefore, we adopt a GroupKFold cross-validation strategy, ensuring that each user appears exclusively in either the training or validation split.

The model is trained using the Mean Squared Error (MSE) loss function with the AdamW optimizer, a learning rate of 1×10^{-4}, and a weight decay of 0.01. The batch size is set to 16, and early stopping is applied to prevent overfitting.

Table 4 presents the performance comparison for

Variant	Valence (r)	Arousal (r)
R	0.366	0.312
R + U	0.361	0.314
R + B	0.406	0.329
R + B + U	0.420	0.337
Full Model	**0.553**	**0.589**

Table 5: Ablation study for Subtask 2A on the validation setting. R: RoBERTa, B: BiLSTM, and U: User Embeddings.

short-term affect forecasting along the official baseline. For valence, the proposed model achieves a lower MAE (1.1389 vs. 1.168), indicating reduced prediction error, although the Pearson correlation is slightly lower than the baseline. For arousal, while the baseline attains higher correlation, the proposed model remains competitive in error magnitude. Further analysis suggests that arousal prediction is challenging compared to valence due to its reliance on emotional intensity. Additionally, BiLSTM represents a contextual information across observations during forward and backward pass, which can over-smoothen the localized high energy fluctuations, leading to competitive error compared to baseline. These results suggest that forecasting short-term affective variation is more challenging than static prediction, but the proposed framework demonstrates improved robustness in terms of absolute error reduction.

5.3 Ablation Study

To understand the contribution of all the features such as temporal and user-level modules to the RoBERTa in the proposed framework we conducted an ablation study. Table 5 reports the Pearson correlation (r) for valence and arousal for all the settings.

The ablation results show a consistent improvement in performance with the addition of model components. The RoBERTa (R) baseline provides moderate performance, while the inclusion of user embeddings (U) results in only marginal changes. In contrast, the addition of the BiLSTM (B) layer leads to a clear improvement in both valence and arousal, highlighting the importance of temporal modeling. Combining BiLSTM with user embeddings further improves the performance, indicating that temporal and user-level features provide complementary information. The full model achieves the best results, demonstrating the benefit of integrating contextual, temporal, and user-specific features for longitudinal affect prediction.

6 Conclusion

This study demonstrates the effectiveness of transformers for modelling continuous and longitudinal affect in ecological text data. The incorporation of an attention pooling mechanism enhances the model's ability to focus on emotionally salient cues within text, improving affect prediction performance. Furthermore, the integration of a BiLSTM layer in Subtask 2A enables better modelling of temporal dynamics, supporting short-term affect forecasting.

Overall, the proposed framework highlights the importance of combining contextual representation learning with temporal modelling for longitudinal emotion analysis. In Subtask 2 future direction of work will focus on normalizing time gap in consecutive observations based on the emotional intensity at time t to better capture the temporal irregularities in logitudinal data. The findings provide a foundation for future research in affective computing, with potential applications in mental health monitoring, therapeutic support systems, and emotion-aware technologies.

References

Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer.

Sven Buechel and Udo Hahn. 2017. Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis. In *Proceedings of the 15th Conference of the European Chapter of the ACL.*

Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion*, 6(3-4):169–200.

Adithya V Ganesan and 1 others. 2026. From word sequences to behavioral sequences: Adapting modeling and evaluation paradigms for longitudinal nlp. *arXiv preprint arXiv:2601.07988.*

Pranav Goel, Devang Kulshreshtha, Prayas Jain, and Kaushal Kumar Shukla. 2017. Prayas at emoint 2017: An ensemble of deep neural architectures for emotion intensity prediction in tweets. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 58–65.

Marieke Houben, Peter Kuppens, and Francis Tuerlinckx. 2015. Emotion variability and psychological well-being: A meta-analysis. *Psychological Bulletin.*

Maximilian Koper, Evgeny Kim, and Roman Klinger. 2017. Ims at emoint-2017: Emotion intensity prediction with affective norms, automatically extended resources and deep learning. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 50–57.

Peter Kuppens, Francis Tuerlinckx, and Marc A. B. van Mechelen. 2010. Emotion dynamics and emotional inertia. *Psychological Science*.

Kevin LaGrandeur. 2015. Emotion, artificial intelligence, and ethics. In *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide*, pages 97–109. Springer.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Sreekanth Madisetty and Maunendra Sankar Desarkar. 2017. Nsemo at emoint-2017: An ensemble to predict emotion intensity in tweets. In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 219–224.

Goncalo Mendes and Bruno Martins. 2023. Quantifying valence and arousal in text with multilingual pretrained transformers. In *ECIR*.

Fatemehsadat Mireshghallah and 1 others. 2022. Useridentifier: Implicit user representations for personalized sentiment analysis. In *NAACL*.

Saif M. Mohammad. 2016. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion Measurement*, pages 201–237. Elsevier.

Saif M. Mohammad and Felipe Bravo-Marquez. 2017. Wassa-2017 shared task on emotion intensity. *arXiv preprint arXiv:1708.03700*.

Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 task 1: Affect in tweets. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 1–17.

James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178.

Nikita Soni, H. Andrew Schwartz, Ryan L. Boyd, Phi Long Bui, Syeda Mahwish, August Håkan Nilsson, Adithya V Ganesan, Lyle Ungar, Niranjan Balasubramanian, and Saif M. Mohammad. 2026. SemEval-2026 task 2: Predicting variation in emotional valence and arousal over time from ecological essays. In *Proceedings of the 20th International Workshop on Semantic Evaluation (SemEval-2026)*. Association for Computational Linguistics.

Samarth Tripathi, Shrinivas Acharya, Ranti Sharma, Sudhanshi Mittal, and Samit Bhattacharya. 2017. Using deep and convolutional neural networks for accurate emotion classification on deap data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4746–4752.

Laura Troiano and 1 others. 2023. Quantifying valence and arousal in text with multilingual pre-trained transformers. *arXiv preprint arXiv:2302.14021*.

Talia Tseriotou, Adam Tsakalidis, and 1 others. 2023. Sequential path signature networks for personalised longitudinal language modeling. In *Findings of ACL*.

Cort J Willmott and Kenji Matsuura. 2005. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82.

Yue Xie, Zseiyu Liang, Zhenlin Liang, Chengwei Huang, Cairong Zou, and Björn Schuller. 2019. Speech emotion classification using attention-based lstm. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(11):1675–1685.