UTBNLP at Semeval-2025 Task 11: Predicting Emotion Intensity with BERT and VAD-Informed Attention.

Melissa Moreno Novoa, Edwin Puertas, Juan Carlos Martinez-Santos
Universidad Tecnológica de Bolívar
Cartagena, Colombia
{memoreno, epuerta,jcmartinezs}@utb.edu.co

Abstract

Emotion intensity prediction is crucial in affective computing, allowing for a more precise understanding of how emotions are conveyed in text. This study proposes a system that estimates the levels of intensity of emotions by integrating contextual language representations with numerical emotion-based characteristics derived from Valence, Arousal, and Dominance (VAD). The methodology combines BERT embeddings, predefined VAD values per emotion, and machine learning techniques to enhance emotion detection without relying on external lexicons. The system was evaluated on the SemEval-2025 Task 11 Track B dataset, predicting five emotions (anger, fear, joy, sadness, and surprise) on an ordinal scale.

The results highlight the effectiveness of integrating contextual representations with predefined VAD values, allowing a more nuanced representation of emotional intensity. However, challenges arose in distinguishing intermediate intensity levels, which affected the classification accuracy for specific emotions. Despite these limitations, the study provides insight into the strengths and weaknesses of combining deep learning with numerical emotion modeling. It contributes to developing more robust emotion prediction systems. Future research will explore advanced architectures and additional linguistic features to enhance model generalization across various textual domains.

1 Introduction

Emotions are essential in language and vary in intensity depending on the context. Understanding which emotions we express in a conversation and why they occur is crucial for improving the quality of human interactions (Wang et al., 2023). Given a text and a perceived emotion, the system must estimate its level within a predefined scale, considering the language and its variations, thereby enabling a precise analysis of affective language (Cuadrado et al., 2023). This article presents a system to predict the intensity of a set of emotions in a text fragment. This analysis builds upon previous models that have contributed to the identification of emotions present in textual comments. Some emotions are more fundamental in physiological and cognitive terms, allowing them to manifest at different intensity levels (Hsu et al., 2025). Their significance lies in enhancing the automatic understanding of emotions in artificial intelligence and affective language analysis.

To address this task, we proposed a system that relies on contextual language representations and emotional features that capture language representations and emotion-specific VAD values, quantifying valence, arousal, and dominance (Ghosh et al., 2023).

Participation in this task allowed for the evaluation of the system's performance in detecting the intensity of perceived emotions compared to other approaches. Although the system's performance was below the average, it successfully captured relevant patterns in emotional intensity variation. We observed that integrating contextual representations with VAD-based emotional features enabled an approximation to the objective despite limitations in classifying certain emotions. The primary challenges arose in differentiating intermediate intensity levels, which affected accuracy in some categories. Despite these challenges, the experience provided valuable insights for improving the modeling of emotional intensity in future studies. The code developed in this study is available for researchers and developers to access, analyze the implemented approaches, and replicate the experiments. By sharing the code, this work promotes the reproducibility of results. It fosters continuous improvement in the detection of perceived emotion intensity. [1]

[1] https://github.com/Novoa0599/SemEval2025

Proceedings of the The 19th International Workshop on Semantic Evaluation (SemEval-2025), pages 1217–1222
July 31 - August 1, 2025

text	Joy	Fear	Anger	Sadness	Surprise
None of us has mentioned the incident since.	0	1	0	2	1
was seven and woke up early, so I went to the ba...	1	0	0	0	0
“ My God ” “ What’s wrong with my arm? ”,	0	2	0	0	1

Table 1: Examples showing the data structure and some features.

2 Background

Emotion recognition in text is an area of growing interest due to the rise of digital communication and the inherent complexity of language in these environments. Previous research has approached this challenge from various perspectives, achieving significant advances in the semantic and syntactic understanding of texts and improving language representation across multiple languages (Suresh et al., 2024; Mohammad and Kiritchenko, 2018). Additionally, authors highlighted the importance of context, sentiment consistency, and common-sense knowledge to accurately detect emotions in such texts (Tu et al., 2022). These approaches emphasize the need for sophisticated models to capture the emotional richness of diverse digital interactions.

Emotion recognition models have evolved considerably, shifting from rule-based approaches to more adaptive methods, allowing a more precise and flexible interpretation of emotional language (Bhati et al., 2024). These advancements have enhanced the ability of models to capture linguistic nuances, recognize context, and handle variations in emotional expression, which is crucial for practical applications in various domains.

In this work, we participated in Track 2 of Task 11, which focuses on detecting and predicting emotion intensity in English text. The task consisted of identifying the perceived emotion in a text fragment and assigning it an intensity level based on a predefined ordinal scale. This approach allows a single comment to contain multiple emotions, each with an intensity level, enriching data analysis and interpretation.

The dataset used for this task consisted of 117 developing samples, 2768 training samples, and test validation samples, in which we analyzed five specific emotions: joy, sadness, fear, anger, and surprise. The ordinal scale assigns the following values: 0 for no emotion, 1 for low intensity, 2 for moderate intensity, and 3 for high intensity. For example, in the sentence "None of us has mentioned the incident since", the system assigned the following intensities: joy (0), fear (1), anger (0), sadness (2), and surprise (1), illustrating the data structure and the model’s response generation, which has been explored in previous studies Table 1 (Muhammad et al., 2025)

We reviewed prior sentiment and emotion analysis research to support this approach, providing a solid theoretical and methodological foundation (Garcia et al., 2024). Integrating advanced natural language processing techniques with emotion prediction models has opened new possibilities for a deeper understanding of written communication. These approaches enhance accuracy in emotion detection and facilitate their application in customer service, mental health, and trend analysis on social media. As these models evolve, we expect future research to expand and refine emotion analysis methods, offering increasingly robust and efficient solutions for automatic language processing.

3 System Overview

Analyzing and predicting emotions from text follows a structured approach based on multiple stages. The methodology begins with data loading and preprocessing, followed by text representation using embeddings, fine-tuning a BERT-based model, predicting valence, arousal, and dominance (VAD) values, and finally converting these values into discrete emotion intensities. We evaluated the model using metrics that compare predictions with reference values to measure its accuracy and reliability, as shown in Figure 1.

In the preprocessing stage, textual data undergoes thorough cleaning and normalization to ensure a standardized representation. First, the dataset is loaded, ensuring its structure is consistent and suitable for the model. Then, text normalization techniques are applied, including lowercase conversion, removal of special characters, punctuation, and non-alphabetic elements, and expansion of contractions to enhance semantic coherence. We also removed stopwords to optimize text representation. Additionally, we computed VAD values to represent the emotion labels.

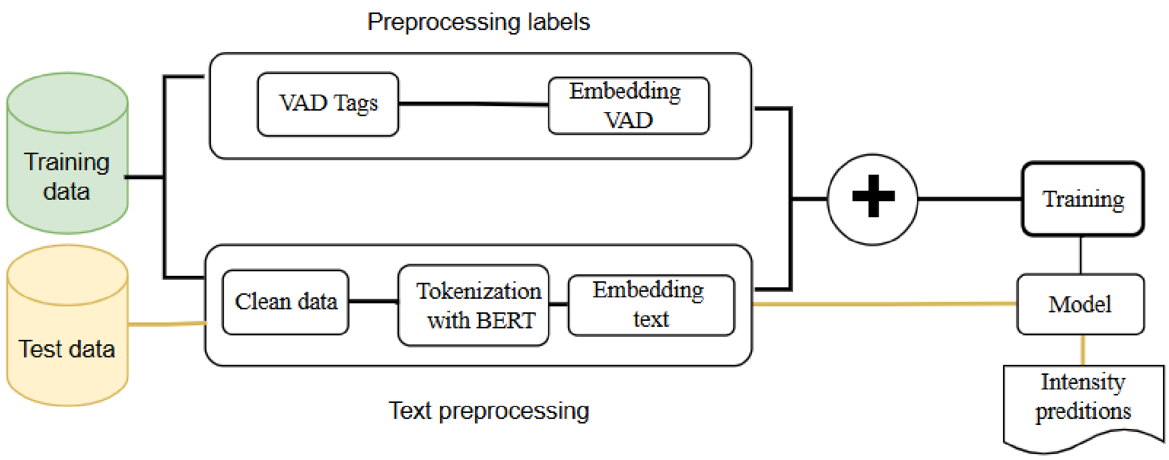

Figure 1: General pipeline system.

Once the text is processed, it is tokenized and represented using a pre-trained language model. In this case, BERT (Devlin et al., 2018) is employed to convert the text into token sequences and attention masks. These numerical representations capture semantic and contextual relationships within the text, enabling a deeper analysis of expressed emotions.

The next step involves training the model. We fine-tuned a BERT-based model for predicting continuous VAD values, and we adapted a pre-trained model specialized in language comprehension through a fine-tuning process. During this phase, the model learns to associate textual patterns with corresponding VAD values, optimizing its weights via a gradient descent algorithm and using a regression loss function.

We trained the model using standard hyperparameters to fine-tune BERT in classification tasks. We used a learning rate 2e-5 and three epochs to ensure convergence without overfitting. We set a batch size of 8 to optimize memory usage when processing long sequences. We employed AdamW as the optimizer. A learning rate scheduler with an initial warm-up phase and linear decay was also applied to improve training stability and convergence. We selected these values based on best practices for emotional intensity prediction.

After training, the model generates VAD predictions for new data. These predictions provide a continuous estimation of the emotion present in each sentence, offering insights into the emotional intensity of the text. However, we converted these VAD values into discrete emotional intensities to facilitate result interpretation. We achieved this through a mapping mechanism based on Euclidean distances. In this process, we compared the predicted VAD values with predefined reference values for each emotion, and we assigned the closest intensity on a scale from 0 to 3.

Finally, we evaluated the model using performance metrics. We calculated the Pearson correlation to measure the relationship between predicted and reference VAD values, assessing the model's capability to correctly identify emotional intensities. The results of these metrics help identify strengths and limitations, guiding future improvements in the emotion prediction system.

The proposed methodology integrates advanced natural language processing techniques with deep learning models to predict emotions in text. By leveraging BERT for language representation and a conversion system for mapping VAD values to discrete intensities, a robust and efficient model is developed for analyzing emotions in textual conversations.

4 Experimental Setup

We got three datasets: training, development, and testing, ensuring that the model is trained and evaluated on entirely independent data. We applied techniques such as normalization, contraction expansion, pattern removal, and stopword elimination during preprocessing. We used tools like spaCy and the contractions library to clean and standardize the text before using BertTokenizer. Hyperparameter tuning, including adjustments to the learning rate and number of epochs, is performed on the training set. In contrast, we used the development set for iterative validation and model selection. Finally, the Pearson correlation coefficient is calculated individually for each dimension of the VAD space, which is then transformed into emotional intensities to assess the quality of the predictions.

5 Result

In our experiments, we evaluated the model over three training epochs, where we monitored the Pearson correlation coefficient for each dimension of the VAD space (Valence, Arousal, and Dominance). During the first epoch, a global accuracy of 0.6552 was obtained, with Pearson's r values of 0.5133, 0.5467, and 0.6429 for V, A, and D, respectively. In the second epoch, the metrics improved significantly, reaching a precision of 0.7726 and Pearson coefficients of 0.5765, 0.5898, and 0.6910 for V, A, and D, demonstrating an increased ability of the model to capture emotional intensity at this stage. In the third epoch, the accuracy stabilized at 0.7455, and the Pearson coefficients were 0.5609, 0.5874, and 0.6821, respectively. The final average metrics in all epochs showed a precision of 0.7244 and average Pearson values of 0.5502 (V), 0.5746 (A) and 0.6720 (D), as shown in Table 2. These results suggest that the model more accurately captures perceived intensity regarding control and authority, whereas positivity (Valence) and emotional activation (Arousal) exhibit more significant variability in predictions.

The performance of the model varied significantly between emotions. As shown in Table 3, the model achieved better predictions for emotions such as Anger (0.6418) and Sadness (0.55), suggesting that these emotions exhibit more distinct patterns in the VAD space. In contrast, Fear obtained the lowest score (0.0514), indicating difficulties in accurately identifying this emotion. Joy and Surprise showed intermediate values, with scores of 0.5408 and 0.1782, respectively. In particular, the low score for Surprise suggests that this emotion presents an additional challenge, possibly due to the subjectivity and implicit contextual cues required for proper interpretation.

Furthermore, Table 5 presents the comparison of validation metrics between the baseline model, without VAD information integration, and the improved model that incorporates these affective features. Although the precision of the micrometrics, the recall, F1 and the overall accuracy (0.7437) remained constant, the macrometrics showed substantial improvements: the macro precision increased from 0.5959 to 0.7494, the macro recall from 0.5535 to 0.7528, and macro F1 from 0.5715 to 0.7434. These results demonstrate a more balanced performance between classes, indicating that the incorporation of VAD information significantly improved the quality and fairness of the model.

When analyzing the development set, the Pearson coefficients for each emotion reflect a more significant discrepancy between predicted and actual intensities. Anger (-0.0617), Fear (0.0141), Joy (-0.0023), Sadness (-0.0252), and Surprise (-0.1307), with an average of -0.0412 (Table 4). This significant difference in correlation at the emotional level suggests that, while the model captures some coherence in the VAD space, the conversion to emotional intensities still presents inconsistencies. Furthermore, the negative correlation in some emotions indicates that the model may be predicting intensities in opposite directions to those expected, highlighting the need for refinement in transforming VAD into discrete emotional categories.

A key finding in our analysis is that integrating VAD values improved the detection of emotions with well-defined characteristics in valence, arousal, and dominance. However, for emotions with less distinct distributions in this space, such as Fear and Surprise, the model tended to predict values closer to neutrality, which may explain the lower scores for these categories. The difference in performance between emotions could be attributed to an imbalance in the distribution of the class within the training set, limiting the model's ability to learn appropriate representations for less frequent or more ambiguous emotions Table 6.

While the model demonstrates promising performance in predicting certain emotions, the results suggest that additional adjustments are needed to improve its generalization capability. Techniques such as class balancing, loss function adjustments, or incorporating additional linguistic features—such as deep semantic analysis or multimodal models integrating supplementary signals—could be explored to address these limitations. Additionally, evaluating the model on more diverse datasets would allow for assessing its robustness across domains and contexts.

6 Conclusions

This study presented a system for predicting the intensity of perceived emotions in a text by inte-

	Pearson r Score
Valence	0.5502
Arousal	0.5746
Dominance	0.6720

Table 2: Pearson r VAD Assessment Score.

Emotion	Pearson r
Anger	-0.0617
Fear	0.0141
Joy	-0.0022
Sadness	-0.0252
Surprise	-0.1307
Average	-0.0412

Table 3: Development model evaluation.

Emotion	Pearson r
Anger	0.6418
Fear	0.0514
Joy	0.5408
Sadness	0.55
Surprise	0.1782
Average	0.40

Table 4: Test model evaluation.

Metric	Without VAD	With VAD
Accuracy	0.7437	0.7437
Precision (Micro)	0.7437	0.7437
Precision (Macro)	0.5959	0.7494
Recall (Micro)	0.7437	0.7437
Recall (Macro)	0.5535	0.7528
F1 (Micro)	0.7437	0.7437
F1 (Macro)	0.5715	0.7434

Table 5: Validation results comparison without and with VAD.

Text	Gold Label					Prediction Label				
	Joy	Fear	Anger	Sadness	Surprise	Joy	Fear	Anger	Sadness	Surprise
So... for reasons unknown...	0	1	0	0	2	1	1	0	0	1
None of us has mentioned the incident since.	0	1	0	2	1	0	1	0	1	0
I stopped a couple times to stretch out my calves and quads.	0	0	0	0	0	1	1	0	0	1

Table 6: Comparison of gold and predicted emotion labels.

grating contextual linguistic representations with a numerical modeling approach based on predefined VAD values. By combining BERT embeddings and machine learning techniques, the system successfully captured relevant patterns in emotional intensity variation, achieving acceptable performance in predicting certain emotions such as anger and sadness.

The results highlight the impact of integrating VAD values into emotional representation, allowing for a more nuanced capture of emotion intensity compared to discrete classification approaches. However, the system encountered difficulties distinguishing intermediate intensity levels, particularly in emotions such as fear and surprise, suggesting further refinement to improve its accuracy in these categories.

Despite these limitations, the applied methodology offers a promising approach to text-based emotion analysis, emphasizing the importance of continuous representations in emotion modeling. This study provides a solid foundation for future research, underscoring the need to explore advanced techniques further to enhance the prediction of emotional intensity.

7 Future Work

Future research should focus on optimizing the integration of VAD values within BERT's attention mechanism, dynamically adjusting them based on semantic context to better capture subtle emotional nuances. Calibration techniques like ordinal regression and isotonic calibration could also improve the accuracy of intermediate intensity classifications. Additionally, enhancing text preprocessing through efficient tokenization and embedding strategies that merge semantic and affective information could further strengthen predictive performance. A detailed error analysis is recommended to refine the model based on misclassification patterns. Finally, applying interpretability methods such as SHAP and LIME would provide greater transparency and insights into the model's decision-making in affective computing.

Acknowledgments

The authors express their gratitude to the Call 933 "Training in National Doctorates with a Territorial, Ethnic and Gender Focus in the Framework of the Mission Policy — 2023" of the Ministry of Science, Technology and Innovation (Minciencia). In addition, we thank the team of the Artificial Intelligence Laboratory VerbaNex [2], affiliated with the UTB, for their contributions to this project.

References

Mansi Bhati, Kanika Sharma, Dhyanendra Jain, Anjani Gupta, and Shruti Agarwal. 2024. Integrating nlp with random forest for emotion detection. In *2024 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, pages 489–494. IEEE.

Juan Cuadrado, Elizabeth Martinez, Juan Carlos Martinez-Santos, and Edwin Puertas. 2023. Team utb-nlp at finances 2023: Financial targeted sentiment analysis using a phonestheme semantic approach.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Santiago Garcia, Elizabeth Martinez, Juan Cuadrado, Juan Martinez-santos, and Edwin Puertas. 2024. VerbaNexAI lab at SemEval-2024 task 10: Emotion recognition and reasoning in mixed-coded conversations based on an NRC VAD approach. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 1332–1338, Mexico City, Mexico. Association for Computational Linguistics.

Soumitra Ghosh, Asif Ekbal, and Pushpak Bhattacharyya. 2023. Vad-assisted multitask transformer framework for emotion recognition and intensity prediction on suicide notes. *Information Processing Management*, 60:103234.

Chia-Feng Hsu, Sriyani Padmalatha Konara Mudiyanselage, Rismia Agustina, and Mei-Feng Lin. 2025. Basic emotion detection accuracy using artificial intelligence approaches in facial emotions recognition system: A systematic review. *Applied Soft Computing*, 172:112867.

Saif Mohammad and Svetlana Kiritchenko. 2018. Understanding emotions: A dataset of tweets to study interactions between affect categories. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Shamsuddeen Hassan Muhammad, Nedjma Ousidhoum, Idris Abdulmumin, Jan Philip Wahle, Terry Ruas, Meriem Beloucif, Christine de Kock, Nirmal Surange, Daniela Teodorescu, Ibrahim Said Ahmad, David Ifeoluwa Adelani, Alham Fikri Aji, Felermino D. M. A. Ali, Ilseyar Alimova, Vladimir Araujo, Nikolay Babakov, Naomi Baes, Ana-Maria Bucur, Andiswa Bukula, Guanqun Cao, Rodrigo Tufino Cardenas, Rendi Chevi, Chiamaka Ijeoma Chukwuneke, Alexandra Ciobotaru, Daryna Dementieva, Murja Sani Gadanya, Robert Geislinger, Bela Gipp, Oumaima Hourrane, Oana Ignat, Falalu Ibrahim Lawan, Rooweither Mabuya, Rahmad Mahendra, Vukosi Marivate, Andrew Piper, Alexander Panchenko, Charles Henrique Porto Ferreira, Vitaly Protasov, Samuel Rutunda, Manish Shrivastava, Aura Cristina Udrea, Lilian Diana Awuor Wanzare, Sophie Wu, Florian Valentin Wunderlich, Hanif Muhammad Zhafran, Tianhui Zhang, Yi Zhou, and Saif M. Mohammad. 2025. Brighter: Bridging the gap in human-annotated textual emotion recognition datasets for 28 languages. *Preprint*, arXiv:2502.11926.

Tharun Suresh, Ayan Sengupta, Md Shad Akhtar, and Tanmoy Chakraborty. 2024. A comprehensive understanding of code-mixed language semantics using hierarchical transformer. *IEEE Transactions on Computational Social Systems*, 11:4139–4148.

Geng Tu, Jintao Wen, Cheng Liu, Dazhi Jiang, and Erik Cambria. 2022. Context- and sentiment-aware networks for emotion recognition in conversation. *IEEE Transactions on Artificial Intelligence*, 3:699–708.

Fanfan Wang, Zixiang Ding, Rui Xia, Zhaoyu Li, and Jianfei Yu. 2023. Multimodal emotion-cause pair extraction in conversations. *IEEE Transactions on Affective Computing*, 14:1832–1844.

[2]https://github.com/VerbaNexAI