Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2024) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

9th International Conference on Computer Science and Computational Intelligence 2025 (ICCSCI 2025)

Emotion Intensity Prediction using Transformer-Based Models

Damien Herlnata[a, *], Nabila Danisha Putri[a], Renji Earl Kurniawan[a], Bakti Amirul Jabar[a], Jurike V. Moniaga[a]

[a]*Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480*

Abstract

Emotion detection has been increasingly applied in various fields, including chatbots, virtual assistants and affective computing systems. However, most existing emotion detection models primarily focus on one task, which is to identify types of emotion expressed using emotion classification. While this is essential, a complete understanding of emotional expression also requires understanding the intensity of emotion expressed by the author. In this paper, we propose a Bidirectional Encoder Representations of Transformer (BERT) based model that can understand the type and strength of an emotion expressed through a single text. We approached this problem with two learning methods: Single-Task Learning, where each task is trained independently, and Multi-Task Learning, where both tasks are learned simultaneously through a shared encoder. Experiment results show that the multi-task learning model performs better than single-task learning.

Keywords: *Emotion classification, emotion intensity prediction, Single-Task Learning, Multi-Task Learning, BERT Transformer*

*Corresponding author.
E-mail address: damien.herlnata@binus.ac.id

1. Introduction

Emotion is a basic component in human communication; it influences how people communicate, make decisions and perceive the world [1]. While traditional emotion classification focuses on identifying text into different emotion categories—like the typical emotion categories proposed by Ekman and Plutchik [2] such as *joy, anger, sadness, fear, disgust, surprise*—it often overlooks the varying degrees to which these emotions are expressed. In various industries, such as mental health monitoring, virtual assistants, empathetic chatbot [3] and human-computer interaction, being able to understand just the type of emotion expressed is not enough. It's also crucial to understand how strongly those emotions are expressed. Emotion intensity prediction aims to address this gap by measuring the magnitude or strength of an emotion expressed, providing more insight into emotional states beyond just emotion categories. This task is particularly challenging due to the subjective and context-dependent state of emotional expression [4], but recent advances in natural language processing and machine learning models have enabled promising approaches to tackle this problem.

Emotional expressions can be seen through various forms, including facial expressions [5]. Although the universality of emotional expression has been mentioned in several studies, language plays an important role in emotional design and how emotions are expressed [5]. Text not only communicates informative content, but also conveys information about the attitude and emotion expressed by the author [6]. Word choice can reflect emotional intensity—for example, both *"angry"* and *"furious"* expresses anger, but *"furious"* conveys a stronger emotion. Take the following two sentences:

(1) The boss is angry with the new intern

(2) The boss is very furious with the new intern

Both of these examples would be labelled with the emotion *anger*. However, it is clear that the second sentence expresses more magnitude/intensity of *anger*. This suggests that emotions are embedded in language, and by analyzing word choice, we can infer both the type and intensity of a person's feelings.

The purpose of emotional intensity prediction is to identify not only the type of emotion, but also the intensity level of the emotion expressed. The ability to accurately predict not just the emotion category but also the severity of those emotions can significantly enhance the emotional intelligence of AI systems [7] such as virtual assistant, mental health applications, sentiment analysis, et cetera. Through this research, we hope to contribute to the development of models that go beyond basic classification in order to bridge the gap between raw textual expression and nuanced emotional understanding, resulting in more natural and effective human-computer interaction.

2. Literature Review

Understanding and interpreting emotional expressions in text has become an essential task in the development of emotionally intelligent systems across domains such as affective computing [8,9], health care [10], customer service, and social media monitoring. While traditional emotion classification aims to categorize text into basic emotional types [5], recent research has expanded this scope to include the prediction of emotion intensity, a more nuanced and complex task that captures the degree to which an emotion is expressed in language. This advancement is crucial, as the strength of an emotional expression can often carry more significance than the emotion category itself [11,12].

Research efforts in this field have explored various methodologies, ranging from lexicon-based techniques and classical machine learning models to deep neural networks and transformer-based approaches. This chapter outlines key contributions and advancements from previous studies in emotion intensity prediction, particularly those relevant to textual data.

A major milestone in this field was marked by this research [13], which introduced the task of estimating the intensity of emotions in social media text, particularly Twitter posts. In this task, Saif M. Mohammad and Felipe Bravo-Marquez presented an annotation strategy called Best-Worst Scaling (BWS), which enabled the construction of a fine-grained dataset for four key emotions: anger, fear, joy and sadness [13].

Participants were asked to assign real-valued intensity scores (ranging from 0 to 1) to tweets based on how strongly an emotion was expressed. The results demonstrated that combining dense word embeddings and affective lexical resources was effective in modelling emotion intensity. The top-performing systems achieved high correlation scores when compared to human annotations.

This research [14] developed a supervised regression model that combined three key sources of information: handcrafted affective lexicons, automatically expanded lexical features, and a deep learning model based on CNN-

LSTM architecture. Each feature set independently performed well, but their integration led to a highly accurate system for predicting emotion intensity in Twitter data. They also showed that tailoring lexical resources to the language of social media improved predictive performance.

This research [15] proposed a neural network design that used convolutional and fully connected layers arranged in parallel, a novel setup at the time for NLP tasks. Their approach leveraged transfer learning and affective features to enhance emotion intensity prediction. The results indicated a significant boost in prediction performance compared to previous state-of-the-art systems.

Similarly to this research [16], introduced a multi-task ensemble framework that addressed multiple objectives simultaneously, including emotion classification and intensity prediction. By combining CNN, LSTM, and GRU architectures with handcrafted features into a multilayer perceptron (MLP), they built a robust system capable of generalizing across various emotion analysis tasks.

With the advent of transformer-based models, emotion recognition has taken a leap forward in contextual understanding. This research [1] explored the use of BERT for contextual emotion detection, demonstrating that such models could effectively capture complex emotional subtleties, especially when emotions are implied rather than explicitly stated.

In a similar vein, this research [17] proposed DeepEmotex, a transfer learning approach leveraging BERT fine-tuned on emotional-labeled Twitter data. This method proved to be highly effective across multiple domains and achieved high accuracy even with limited labelled data.

This research [18] also developed a model named TA-MERT, which applied text augmentation strategies along with transformer encoders to improve emotion recognition in conversations. Their model, tested on the MELD dataset, showed improved performance particularly in handling data imbalance.

To address the lack of annotated emotional data in languages other than English, This research [19] examined cross-lingual emotion intensity prediction. They compared approaches such as machine translation, cross-lingual embeddings, and unsupervised methods to transfer emotion models from English to Spanish and Catalan. Interestingly, their findings revealed that models with minimal dependency on parallel data often outperformed those relying on large bilingual corpora.

This research [20] proposed a selective feature approach using a Bi-LSTM model that focused solely on the most emotionally impactful words in a sentence. Their method integrated CNN, LSTM and Bi-LSTM to generate vector representations of dominant emotional words, resulting in better performance when classifying core emotions like joy, fear, anger, and sadness.

This research [21] emphasized the importance of incorporating contextual information into lexical features. Their approach used a recurrent neural network (RNN) to dynamically adjust word sentiment scores based on syntactic cues such as negation and intensification. Unlike traditional bag-of-words models, this method was better at capturing compositional and improved sentiment analysis accuracy.

3. Methodologies

In order to build an emotion classification and intensity prediction model, we use a Transformer-based approach to model the type and strength of emotion expressed in a text, leveraging the capabilities of Bidirectional Encoder Representations from Transformers (BERT). BERT is a pre-trained language model that has shown a state-of-art performance on several NLP benchmarks showing its robustness across a wide range of tasks [22]. Due to its ability to capture rich contextual relationships between words through bidirectional context, BERT offers a solid and robust foundation for modeling both our task of emotion classification and intensity prediction.

Using BERT, we propose two learning approaches, namely, Single-Task Learning (STL) and Multi-Task Learning (MTL). In the STL setup, separate models are trained independently for the tasks of emotion classification and intensity prediction. Meanwhile, in the MTL setup, a single model is trained to solve both tasks simultaneously. Through these two approaches, we compare both results and aim to explore the potential benefits and trade-offs of MTL in improving generalization, using STL as a baseline for evaluation.

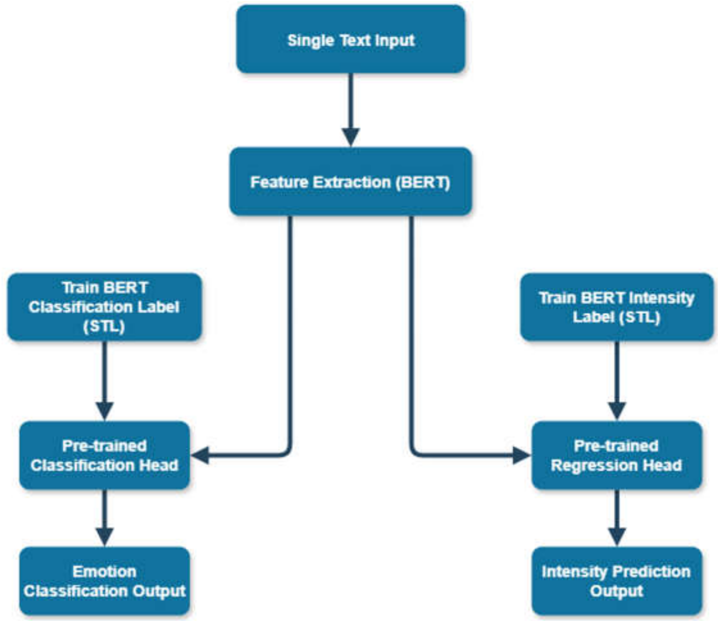

Fig. 1. Proposed Single-Task BERT framework

3.1. *Single-Task Learning using BERT Transformer*

For the single-task learning approach, we employ the architecture illustrated in Figure 1. First, we develop two independent models: one for emotion classification and another for intensity prediction. Both models use bert-base-uncased as the base encoder, each fine-tuned using its respective set of labels without any parameter sharing between tasks. Each model also incorporates additional lexicon-based features concatenated with the BERT pooled output before passing through a dropout layer and final output head. The architecture is the same for both model pipelines, with only the final output head differing to suit their nature of the task.

For emotion classification, we use a linear classification head that outputs raw logits. These logits are then passed to a cross-entropy loss function, which internally applies softmax activation to produce a probability distribution between discrete labels over emotion categories. For intensity regression, we use a regression head that produces a single scalar output, scaled through tanh activation function to map the output to a normalized intensity score between 0 and 1.

The dataset is preprocessed to isolate the relevant task. For emotion classification, we only use text samples and emotion labels. For intensity prediction, we only use text samples and intensity scores. Each task is trained independently using the standard train-test-split. Cross-entropy loss is used to calculate the loss function for the classification task, and a differentiable Pearson correlation-based loss is used for the regression task.

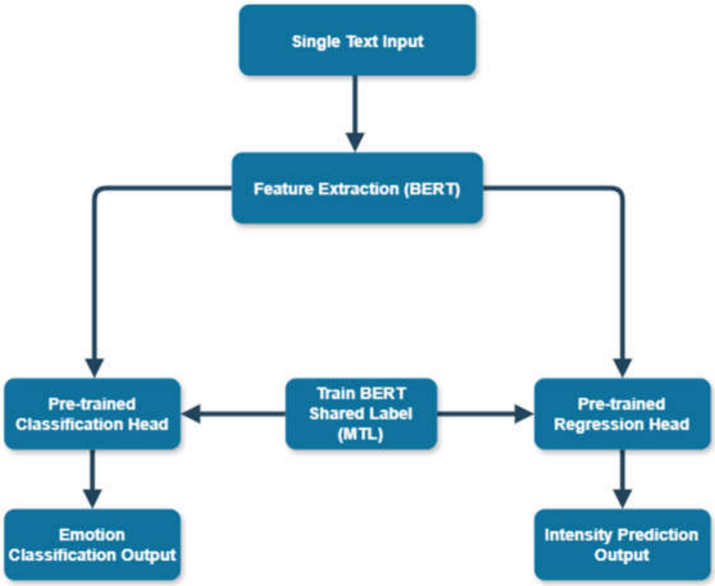

Fig. 2. Proposed Multi-Task BERT framework

3.2. *Multi-Task Learning using BERT Transformer*

For the multi-task learning approach, we employ the architecture illustrated in Figure 2. First, we design a model that performs both emotion classification and intensity prediction simultaneously. The model utilizes a shared BERT encoder fine-tuned jointly on both tasks, allowing it to improve generalization from the input text. On top of the shared BERT encoder, we add two task-specific output heads: one for classification and one for regression.

For emotion classification, we use the classification head that consists of a fully connected layer followed by sigmoid activation to produce independent probabilities for each emotion category, enabling multi-label classification. For intensity prediction, we use the regressor head that consists of a fully connected layer followed by a scaled tanh activation to constrain and produce a normalized intensity score between 0 and 1. This architecture allows the model to optimize both tasks in parallel instead of having each task done separately, potentially benefiting from shared representations.

The model is trained on samples that include both an emotion label and an intensity score—unlike the single-task learning approach where each model is trained independently. A joint loss function is used, combining binary cross-entropy loss for multi-label emotion classification and a differentiable Pearson correlation-based loss for intensity prediction. The joint loss function is defined as:

$$Ltotal = \lambda \cdot\ Lclassification\ +\ (1\ - \lambda) \cdot\ Lregression \tag{1}$$

where λ ∈ [0,1] controls the weight between cross-entropy loss and Pearson's correlation coefficient. Model training is performed using the standard AdamW optimizer with gradient updates applied to the shared encoder. This shared training setup allows the model to learn shared representation optimized for both classification and regression tasks, potentially improving overall generalization.

3.3. *Dataset*

This research utilizes the SemEval-2018 dataset [23], which is a benchmark dataset for affective computing. The dataset consists of English tweets annotated for both emotion classification and intensity prediction, making it suitable for both single-task and multi-task learning approaches.

The dataset consists of multiple subtasks in which we chose one relevant subtask, namely, Emotion Intensity Prediction (El-reg), where tweets are annotated with a real-valued score representing intensity of the emotion expressed, ranging between 0 and 1.

4. Result & Discussions

4.1. *Implementation Details*

Word Embeddings: We use pre-trained word embeddings available from BERT tokenizer and utilize them to preprocess the input text by tokenizing it into subwords and then converting it into token IDs. The token IDs are then passed into the proposed BERT model to obtain contextualized word embeddings. We find it more suitable over other pre-trained models because our approach is based on BERT architecture, thus ensuring compatibility between the tokenization process and the model's input requirements.

Preprocessing: Before extracting the word embeddings from the input text, we used some preprocessing procedures to clean the data. These include converting all text into lowercase letters, replacing emojis with expression words, expanding contractions, removing URLs present in the text, removing user mentions, removing non-alphabetic characters (except for ! and ?), keeping hashtags and removing excess whitespace.

Synonym Replacement & Noise Injection: A safe synonym replacement and noise injection are applied to create more training sets to balance the number of datasets representing each emotion. Synonym replacement involves selecting random text from the dataset and replacing certain words with their synonyms using the WordNet English Vocabulary [24]. The replacement probability of the word being set as 0.3. We also add noise injection into those new synonym sets to mimic real-life typo. Types of noise injected include skipping random characters and swapping characters. Character skip and swap probability are both set as 0.05. These settings are applied across all our models to ensure the augmented samples are meaningfully distinct from the original dataset.

Lexicons: We incorporate emotion-related features using three types of lexicons, including NRC Emotion Lexicon (EmoLex) [5], NRC Valence-Arousal-Dominance (VAD) Lexicon [25], and NRC Hashtag Emotion Lexicon (HashEmo) [26].

The EmoLex emotion-word association lexicon is used to extract affective features from the text into ten expression vectors: anger, anticipation, disgust, fear, joy, sadness, surprise, trust, positive, and negative. Each word in the input text is matched against the EmoLex dictionary and the corresponding emotion categories are incremented accordingly.

The NRC VAD Lexicon is used to extract affective features from the text that are grouped into three psychological dimensions: *Valence, Arousal,* and *Dominance*. For each word in the text found in the VAD lexicon, its respective scores are accumulated. The final vector representation for the text is computed by averaging the scores across all matched words for each of the three dimensions.

The HashEmo Lexicon is used to extract affective features from the text into 8 basic emotions consisting of: *anger, anticipation, disgust, fear, joy, sadness, surprise,* and *trust*. Each word in the input text is matched against the HashEmo dictionary and the corresponding emotion categories are incremented accordingly. The final vector representation for the text is computed by averaging the scores across all matched words for each of the 8 categories.

All the lexicon features are then scaled using Standard Scaler and then combined together into one vector which consists of 21 dimensions per instance.

Training: We train each model with AdamW optimizer [cite] with learning rate of 2e-5 and weight decay of 0.01. Training proceeds for a maximum of 10 epochs using cosine annealing scheduler with warm restarts. Early stopping is applied based on validation performance with a patience of 3 epochs.

For single-task learning, we developed two models: emotion classification model and intensity regression model. The emotion classification model was trained using Cross-Entropy Loss with label smoothing of 0.1 to mitigate confidence, avoid overfitting and improve generalization.

The intensity regression model was trained using a composite loss function that combines the Huber Loss with delta 0.3 and Pearson correlation coefficient to balance between absolute prediction error with correlation trend to the ground truth. The final loss function is defined as:

$$Ltotal = \alpha \cdot LHuber + (1 - \alpha) \cdot LPearson \tag{2}$$

$$LPearson = 1 - \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2} + \varepsilon} \tag{3}$$

where $\alpha = 0.3$, x and y denote predicted and ground truth intensity values, respectively, and $\varepsilon = 1e\text{-}8$ which is a small constant used for numerical stability.

For multi-task learning, we developed 1 multi-task model that could do both emotion classification and intensity regression through shared parameter training. The model was trained using a composite loss function by combining the classification and regression losses.

For the classification loss, we used Binary Cross-Entropy loss. For the regression loss, we used a combination of Huber Loss and Pearson Correlation Loss, identical to the setup of the single-task regression model. The final loss function is defined as:

$$Ltotal = \alpha \cdot LClassification + (1 - \alpha) \cdot LRegression, where\ \alpha = 0.3 \quad (4)$$

4.2. *Evaluation*

Each classification model is evaluated using precision, recall, f1-score per emotion class and overall accuracy of the model. On the other hand, each regression model is evaluated using Pearson's correlation score. The SemEval2018 El-reg dataset exhibits an imbalanced emotion distribution, with fear being the most represented emotion, followed by anger, joy, and sadness.

4.2.1. *Classification Model Evaluation*

Table 1. Comparison classification scores.

	Single-Task Learning	Multi-Task Learning
joy		
f1-score	.82	.81
precision	.84	.88
recall	.81	.75
sadness		
f1-score	.64	.64
precision	.63	.73
recall	.65	.57
anger		
f1-score	.73	.70
precision	.76	.76
recall	.71	.65
fear		
f1-score	.67	.63
precision	.65	.52
recall	.71	.78
accuracy score	**.72**	**.81**

The single-task model shows higher and more consistent performance across individual emotion detection. The emotion joy has the highest performance despite being the second least represented emotion. This is likely because the datasets where emotion is labeled joy are more distinct and less noisy therefore making it easier for the model to identify compared to other emotions. Other emotions like sadness, anger, and fear also show a close range between precision and recall. This stability can be attributed to the single-task model's focused training, which avoids cross-task interference and noise introduced by shared parameters.

This is in contrast with the multi-task model, which jointly performs emotion classification and intensity regression simultaneously through shared parameter training. The emotion fear shows higher recall and lower precision because it's the most represented label, making the model overpredict its emotion and therefore causing more error and less precision in predicting it. In contrast, sadness and anger, which is one of the least represented labels, makes the model more careful when predicting the emotion. Therefore causing their recall to be lower and precision to be higher, because the model only predicts those emotions when it's fairly confident.

While the single-task model often achieves a higher f1-score due to task isolation, the highest overall accuracy score is obtained by the multi-task model. This is because the multi-task model tends to predict emotion dominated by the majority class therefore boosting the overall accuracy even if rare classes are quite poorly predicted. In addition, the shared representation between emotion labels and intensity score might contribute to the final accuracy of multi-task learning. Therefore, making the model able to recognize patterns in these emotions, hence the higher accuracy.

4.2.2. *Regression Model Evaluation*

Evaluation of the regression model uses Pearson's correlation coefficient with a score from 0 to 1.

Table 2. Comparison Pearson Correlation Scores

Method	a	f	j	s	avg
Single-Task Learning	.65	.66	.44	.65	.60
Multi-Task Learning	.73	.75	.72	.72	.73

Single-task learning managed to detect the intensity of fear and anger emotions quite well. This is because both of those emotions are more represented than joy and sadness. Joy shows the lowest pearson correlation score despite not being the least represented value. This is because joy might have a more complex or noisier intensity distribution. The variance value of fear and joy is quite large, which is 0.22. This shows that there is inconsistency in predicting emotion intensity in single-task learning.

This is in contrast to multi-task learning which has a lower variance of 0.03. The highest prediction is fear, followed by anger, joy, and sadness. The order is exactly the same as the order of represented emotion labels in the training sample. Therefore proves that the number of representations in the dataset affect the performance of the multi-task learning model. In addition, multi-task learning shows consistency in predicting emotion intensity, proving that joint learning helps stabilize performance across different emotional dimensions.

The increase in average pearson correlation score from single-task learning (0.60) to multi-task learning (0.73) shows that multi-task learning can not only detect emotion in text, but emotion intensity is also successfully detected quite well. Unlike the classification task benefits from parameter sharing, which allows the model to capture richer emotional representations and leads to more accurate and consistent intensity predictions.

5. Conclusion

In this work, we have proposed a single-task learning and multi-task learning model to do two tasks: emotion classification and intensity prediction of a text. We evaluated the classification using f1-score, precision, and recall. Meanwhile, the regression model was evaluated using Pearson's correlation coefficient. Experiment results show that the multi-task model performs better than single-task model. This is because multi-task learning models are trained with shared parameters which increases the knowledge of the models. While it might increase noise and task-interference, our experiment shows that the multi-task model still performs better than the single-task model. The difference is especially shown in the increase of the intensity regression score.

References

1. Devgan A. Contextual Emotion Recognition Using Transformer-Based Models. Authorea Prepr [Internet]. 2023 [cited 2025 Apr 19]; Available from: https://www.academia.edu/download/105292021/IRJET_V10I7150.pdf
2. Ekman P. An argument for basic emotions. Cogn Emot. 1992 May;6(3–4):169–200.
3. Devaram S. Empathic chatbot: Emotional intelligence for mental health well-being. ArXiv Abs201209130. 2020;
4. Poria S, Majumder N, Mihalcea R, Hovy E. Emotion Recognition in Conversation: Research Challenges, Datasets, and Recent Advances [Internet]. arXiv; 2019 [cited 2025 Apr 21]. Available from: http://arxiv.org/abs/1905.02947
5. Mohammad SM, Turney PD. CROWDSOURCING A WORD–EMOTION ASSOCIATION LEXICON. Comput Intell. 2013 Aug;29(3):436–65.

6. Alm CO, Roth D, Sproat R. Emotions from text: machine learning for text-based emotion prediction. In: Proceedings of human language technology conference and conference on empirical methods in natural language processing [Internet]. 2005 [cited 2025 Apr 19]. p. 579–86. Available from: https://aclanthology.org/H05-1073.pdf
7. Mohammad SM, Bravo-Marquez F. Emotion Intensities in Tweets [Internet]. arXiv; 2017 [cited 2025 Apr 21]. Available from: http://arxiv.org/abs/1708.03696
8. Poria S, Cambria E, Bajpai R, Hussain A. A review of affective computing: From unimodal analysis to multimodal fusion. Inf Fusion. 2017 Sep;37:98–125.
9. Wang Y, Song W, Tao W, Liotta A, Yang D, Li X, et al. A systematic review on affective computing: emotion models, databases, and recent advances. Inf Fusion. 2022 Jul 1;83–84:19–52.
10. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. Nat Lang Eng. 2017 Sep;23(5):649–85.
11. Buechel S, Hahn U. EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis [Internet]. arXiv; 2022 [cited 2025 Apr 21]. Available from: http://arxiv.org/abs/2205.01996
12. Mohammad S. A Practical Guide to Sentiment Annotation: Challenges and Solutions. In: Balahur A, van der Goot E, Vossen P, Montoyo A, editors. Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis [Internet]. San Diego, California: Association for Computational Linguistics; 2016 [cited 2025 Apr 21]. p. 174–9. Available from: https://aclanthology.org/W16-0429/
13. Mohammad SM, Bravo-Marquez F. WASSA-2017 Shared Task on Emotion Intensity [Internet]. arXiv; 2017 [cited 2025 Apr 19]. Available from: http://arxiv.org/abs/1708.03700
14. Köper M, Kim E, Klinger R. IMS at EmoInt-2017: Emotion Intensity Prediction with Affective Norms, Automatically Extended Resources and Deep Learning. In: Balahur A, Mohammad SM, van der Goot E, editors. Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis [Internet]. Copenhagen, Denmark: Association for Computational Linguistics; 2017 [cited 2025 Apr 20]. p. 50–7. Available from: https://aclanthology.org/W17-5206/
15. Kulshreshtha D, Goel P, Kumar Singh A. How emotional are you? Neural Architectures for Emotion Intensity Prediction in Microblogs. In: Bender EM, Derczynski L, Isabelle P, editors. Proceedings of the 27th International Conference on Computational Linguistics [Internet]. Santa Fe, New Mexico, USA: Association for Computational Linguistics; 2018 [cited 2025 Apr 20]. p. 2914–26. Available from: https://aclanthology.org/C18-1247/
16. Akhtar MS, Ghosal D, Ekbal A, Bhattacharyya P, Kurohashi S. A Multi-task Ensemble Framework for Emotion, Sentiment and Intensity Prediction [Internet]. arXiv; 2018 [cited 2025 Apr 20]. Available from: http://arxiv.org/abs/1808.01216
17. Hasan M, Rundensteiner E, Agu E. DeepEmotex: Classifying Emotion in Text Messages using Deep Transfer Learning [Internet]. arXiv; 2022 [cited 2025 Apr 20]. Available from: http://arxiv.org/abs/2206.06775
18. Mohammad F, Khan M, Nawaz Khan Marwat S, Jan N, Gohar N, Bilal M, et al. Text Augmentation-Based Model for Emotion Recognition Using Transformers. Comput Mater Contin. 2023;76(3):3523–47.
19. Alejo IN, Badia T, Barnes J. Cross-lingual Emotion Intensity Prediction [Internet]. arXiv; 2020 [cited 2025 Apr 20]. Available from: http://arxiv.org/abs/2004.04103
20. Sundari AS, Shenbagavali R. Dominant Lexicon Based Bi-LSTM for Emotion Prediction on a Text. Int J Innov Technol Explor Eng. 2019 Oct 11;8(11S):1272–7.
21. Teng Z, Vo DT, Zhang Y. Context-Sensitive Lexicon Features for Neural Sentiment Analysis. In: Su J, Duh K, Carreras X, editors. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing [Internet]. Austin, Texas: Association for Computational Linguistics; 2016 [cited 2025 Apr 20]. p. 1629–38. Available from: https://aclanthology.org/D16-1169/
22. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Burstein J, Doran C, Solorio T, editors. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) [Internet]. Minneapolis, Minnesota: Association for Computational Linguistics; 2019 [cited 2025 Apr 20]. p. 4171–86. Available from: https://aclanthology.org/N19-1423/
23. Mohammad S, Bravo-Marquez F, Salameh M, Kiritchenko S. SemEval-2018 Task 1: Affect in Tweets. In: Proceedings of The 12th International Workshop on Semantic Evaluation [Internet]. New Orleans, Louisiana: Association for Computational Linguistics; 2018 [cited 2025 Apr 20]. p. 1–17. Available from:

http://aclweb.org/anthology/S18-1001
24. Miller GA. WordNet: A Lexical Database for English. In: Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994 [Internet]. 1994 [cited 2025 May 24]. Available from: https://aclanthology.org/H94-1111/
25. Mohammad S. Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. In: Gurevych I, Miyao Y, editors. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) [Internet]. Melbourne, Australia: Association for Computational Linguistics; 2018 [cited 2025 May 24]. p. 174–84. Available from: https://aclanthology.org/P18-1017/
26. Mohammad SM, Kiritchenko S. Using Hashtags to Capture Fine Emotion Categories from Tweets.