Semantic alignment in hyperbolic space for fine-grained emotion classification

Anonymous ACL submission

Abstract

Existing approaches to fine-grained emotion classification (FEC) often operate in Euclidean space, where the flat geometry limits the ability to distinguish semantically similar emotion labels (e.g., *annoyed* vs. *angry*). While prior research has explored hyperbolic geometry to capture fine-grained label distinctions, it typically relies on predefined hierarchies and overlooks semantically confusable negatives. In this work, we propose HyCoEM, a semantic alignment framework that leverages the Lorentz model of hyperbolic space. Our approach jointly embeds text and label representations into hyperbolic space via the exponential map, and employs a contrastive loss to bring text embeddings closer to their true labels while pushing them away from adaptively selected, semantically similar negatives. This enables the model to learn label embeddings without relying on a predefined hierarchy and better captures subtle distinctions by incorporating information from both positive and challenging negative labels. Experimental results on two benchmark FEC datasets demonstrate the effectiveness of our approach over baseline methods.[1]

1 Introduction

Fine-grained emotion classification (FEC) is a single-label task that assigns each text to a specific emotion from a set of closely related categories. Unlike coarse emotion recognition, which uses a small set of basic emotions (Ekman et al., 1999), FEC involves a larger and more nuanced label space. For instance, the two largest FEC datasets include up to 27 (Demszky et al., 2020) and 32 (Rashkin et al., 2019) emotion categories. Many of these labels exhibit subtle semantic differences, such as between *guilty* and *ashamed*, making FEC particularly challenging. Despite this complexity, recognizing fine-grained emotions is essential for capturing subtle human expressions and enabling more empathetic AI interactions.

Existing FEC approaches typically operate in Euclidean space, where the flat geometry makes it difficult to distinguish emotion labels with overlapping semantics (e.g., *fear* and *remorse*) (Yin and Shang, 2022; Suresh and Ong, 2021). In contrast, hyperbolic space, with its negative curvature and exponential growth of distances, is better suited to embed fine-grained emotions with subtle distinctions. The HypEmo (Chen et al., 2023) method utilizes hyperbolic space to learn label representations from a predefined emotion hierarchy (Parrott, 2001). However, this reliance on a fixed structure can be limiting, as emotion labels may not always conform to a strict parent–child organization. Moreover, its cross-entropy loss is weighted solely by the distance to the positive label, overlooking semantically similar negatives that may still mislead the model during prediction.

We propose HyCoEM, a semantic alignment framework that leverages the Lorentz model of hyperbolic space. The model uses RoBERTa (Liu et al., 2019) as the text encoder and treats label embeddings as learnable parameters. During training, both text and label embeddings are projected into hyperbolic space via the exponential map. To guide alignment, we apply a contrastive loss that pulls each text embedding closer to its correct label while pushing it away from semantically similar negative labels. These negatives are adaptively selected for each sample based on geodesic distance in hyperbolic space. The contrastive loss is then used to weight the cross-entropy loss, enabling the model to focus more on samples with weak text–label alignment. We adopt the Lorentz model for its numerical stability and reduced geometric distortion compared to other hyperbolic formulations (Nickel and Kiela, 2018; Chen et al., 2022). Our training setup follows a hybrid design simi-

[1] Code is available at: https://anonymous.4open.science/r/HyCoEM-8725/

lar to HypEmo: contrastive supervision is applied in hyperbolic space, while the cross-entropy loss is computed in Euclidean space. However, unlike HypEmo, our method does not rely on a predefined label hierarchy. Instead, it learns label embeddings directly from data, guided by contrastive alignment. Moreover, since the contrastive loss reflects how well a text aligns with its correct label relative to semantically similar negatives, it provides a more informative weighting signal than the isolated text–label distance used in HypEmo.

2 Related Work

Prior studies on FEC have largely focused on modeling within Euclidean space. (Khanpour and Caragea, 2018) use lexicon-derived features for emotion detection in health-related posts. (Yin et al., 2020) apply syntactic self-attention to better capture sentiment composition. (Mekala et al., 2021) use generative models with coarse emotion labels, while (Sosea and Caragea, 2021) use emotion-specific masking during pretraining. (Suresh and Ong, 2021) propose a label-aware contrastive loss that modulates sample influence based on model confidence. (Yin and Shang, 2022) enhance semantic separation via whitening transformation and nearest-neighbor retrieval. (Chen et al., 2023) adopts a hybrid approach by modeling label representations in hyperbolic space while encoding text inputs in Euclidean space. (Zhang et al., 2024) propose a GNN-based method that captures semantic and temporal patterns through anchor graphs built over token representations.

3 Hyperbolic geometry for Lorentz Model

Let $\mathbf{u} = (\mathbf{u}_s, u_t) \in \mathbb{R}^{n+1}$, where $\mathbf{u}_s \in \mathbb{R}^n$ is the *space*-like component and $u_t \in \mathbb{R}$ is the *time*-like component. The Lorentzian inner product is defined as: $\langle \mathbf{u}, \mathbf{v} \rangle_\mathcal{L} = \langle \mathbf{u}_s, \mathbf{v}_s \rangle - u_t v_t$, where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product. The Lorentzian norm is $\|\mathbf{u}\|_\mathcal{L} = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle_\mathcal{L}}$. The n-dimensional Lorentz model $\mathcal{H}^n$ with curvature $-k$ is represented as a submanifold of $\mathbb{R}^{n+1}$, defined as: $\mathcal{H}^n = \{\mathbf{u} \in \mathbb{R}^{n+1} : \langle \mathbf{u}, \mathbf{u} \rangle_\mathcal{L} = -1/k,\ u_t > 0\}$, where all vectors in $\mathcal{H}^n$ satisfy the constraint $u_t = \sqrt{1/k + \|\mathbf{u}_s\|^2}$. The **geodesic** distance denotes the shortest path between two points on $\mathcal{H}^n$ and is given by:

$$d(\mathbf{u}, \mathbf{v}) = \sqrt{1/k} \cosh^{-1}(-k \langle \mathbf{u}, \mathbf{v} \rangle_\mathcal{L}) \tag{1}$$

At any point $\mathbf{p} \in \mathcal{H}^n$, the **tangent space** $T_\mathbf{p}\mathcal{H}^n$ is a Euclidean vector space consisting of all vectors in $\mathbb{R}^{n+1}$ that are orthogonal to $\mathbf{p}$ as: $T_\mathbf{p}\mathcal{H}^n = \{\mathbf{q} \in \mathbb{R}^{n+1} : \langle \mathbf{p}, \mathbf{q} \rangle_\mathcal{L} = 0\}$. For $\mathbf{q} \in T_\mathbf{p}\mathcal{H}^n$, the **exponential map** projects the vector onto the hyperboloid $\mathcal{H}^n$ as:

$$\exp_\mathbf{p}(\mathbf{q}) = \cosh(\sqrt{k}\|\mathbf{q}\|_\mathcal{L})\mathbf{p} + \frac{\sinh(\sqrt{k}\|\mathbf{q}\|_\mathcal{L})}{\sqrt{k}\|\mathbf{q}\|_\mathcal{L}}\mathbf{q} \tag{2}$$

In this study, we fix $\mathbf{p}$ at the origin $\mathbf{O} = [\mathbf{0}, \sqrt{1/k}]$, where the *space* components are zero and the *time*-like component is $\sqrt{1/k}$.

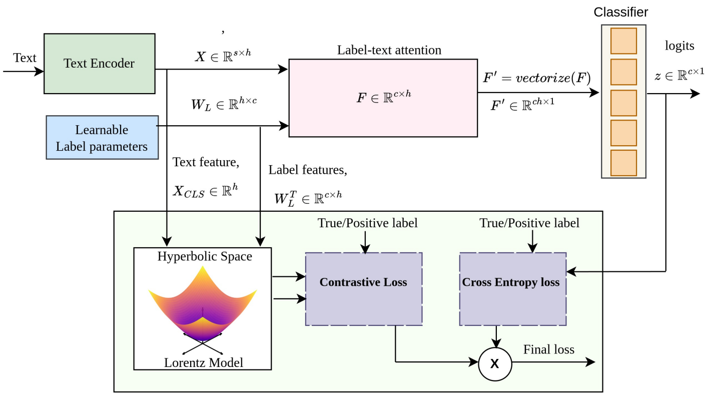


Figure 1: Architecture of HyCoEM. The forward pass generates label-aware features. During training, a contrastive loss is computed in hyperbolic space, which is used to weight the cross-entropy loss.

4 Methodology

This section describes the components of our proposed framework. Fig. 1 illustrates the overall architecture.

4.1 Label-aware feature

We use RoBERTa to encode the input text. For a document D, the encoded token representations are given by: $X = f_{enc}(D)$, where $X \in \mathbb{R}^{s\times h}$, with s representing the token sequence length and h denoting the feature size. To compute the label-aware feature, we apply a label-text attention mechanism using a learnable parameter matrix $W_L \in \mathbb{R}^{h\times c}$, where c is the number of labels:

$$A = XW_L; \quad F = \text{softmax}(A^T)X \tag{3}$$

The resulting matrix $F \in \mathbb{R}^{c\times h}$ is then vectorized to obtain $F' \in \mathbb{R}^{ch\times 1}$ and fed into a classifier. The logit vector $\boldsymbol{z} \in \mathbb{R}^c$ is computed as:

$$F' = \text{vectorize}(F); \quad \boldsymbol{z} = W_c^T F' + \mathbf{b} \tag{4}$$

where $W_c \in \mathbb{R}^{ch\times c}$ and $\mathbf{b} \in \mathbb{R}^c$ represent the weights and bias of the classifier.

4.2 Projection onto the Lorentz Hyperboloid

Let $\mathbf{e}_{\text{enc}} \in \mathbb{R}^h$ be the encoded text/label vector. To project it onto the Lorentz hyperboloid $\mathcal{H}^h$ embedded in $\mathbb{R}^{h+1}$, we extend it as $\mathbf{e} = [\mathbf{e}_{\text{s}}, e_{\text{t}}] = [\mathbf{e}_{\text{enc}}, 0]$, where the *space* component is $\mathbf{e}_{\text{enc}}$ and the *time* component is zero. The vector $\mathbf{e}$ is orthogonal to the hyperboloid origin $\mathbf{O} = [\mathbf{0}, \sqrt{1/k}]$ under the Lorentzian inner product, and thus lies in the tangent space at $\mathbf{O}$. As $e_{\text{t}} = 0$, the exponential map can be used to parameterize only the space component $\mathbf{e}_{\text{s}}$, and the time-like component can be recomputed later to satisfy the constraint $e_t = \sqrt{1/k + \|\mathbf{e}_s\|^2}$. Thus, the exponential map derived from Eqn. 2 becomes:

$$\exp_{\mathbf{0}}(\mathbf{e}_{\mathbf{s}}) = \cosh(\sqrt{k}\|\mathbf{e}\|_{\mathcal{L}})\mathbf{0} + \frac{\sinh(\sqrt{k}\|\mathbf{e}\|_{\mathcal{L}})}{\sqrt{k}\|\mathbf{e}\|_{\mathcal{L}}}\mathbf{e}_{\mathbf{s}} \quad (5)$$

where the first term is zero. Furthermore, the Lorentzian norm simplifies to the Euclidean norm: $\|\mathbf{e}\|^2_{\mathcal{L}} = \langle \mathbf{e}, \mathbf{e} \rangle_{\mathcal{L}} = \langle \mathbf{e}_{\text{s}}, \mathbf{e}_{\text{s}} \rangle - 0 = \|\mathbf{e}_{\text{s}}\|^2$. The resulting expression after all substitutions is:

$$\phi(\mathbf{e}_{\text{s}}) = \exp_{\mathbf{0}}(\mathbf{e}_{\mathbf{s}}) = \frac{\sinh(\sqrt{k}\|\mathbf{e}_{\text{s}}\|)}{\sqrt{k}\|\mathbf{e}_{\text{s}}\|}\mathbf{e}_{\text{s}} \quad (6)$$

4.3 Loss functions

4.3.1 Contrastive loss

We apply contrastive loss in hyperbolic space to align the text embedding with its correct label and separate it from negatives. For a sample $X_i \in \mathbb{R}^{s \times h}$, we use the first token ([CLS]), $x_i \in \mathbb{R}^h$, as the text feature. Label features are defined as the transpose $W_L^\top \in \mathbb{R}^{c \times h}$. Both are projected to hyperbolic space via the exponential map (Eqn. 6) as: $x_{\mathcal{H}_i} = \phi(\alpha_t x_i)$ and $L_{\mathcal{H}} = \phi(\alpha_l W_L^\top)$, where α_t and α_l are learnable scaling factors applied to ensure unit norm before projection. The set of hyperbolic label embeddings is: $L_{\mathcal{H}} = \{\ell_{\mathcal{H}_1}, \ell_{\mathcal{H}_2}, \ldots, \ell_{\mathcal{H}_c}\}$. For each sample-label pair (x_i, y_i), where $y_i \in \mathcal{M}$ (the set of emotion labels), we select the r labels closest to the text (excluding y_i) as negatives:

$$\mathcal{N}(i) = \underset{j \in \mathcal{M} \setminus \{y_i\}}{\text{argmin-r}}\, d(x_{\mathcal{H}_i}, \ell_{\mathcal{H}_j}) \quad (7)$$

where $d(.,.)$ represents the geodesic distance as defined in Eqn. 1 and $r \geq 1$ is a hyperparameter. This adaptive selection provides semantically similar, challenging negative labels, enabling contrastive loss to push the text away from these confusable negatives. Finally, the contrastive loss for sample i is formulated as:

$$CL_i = -\log\left(\frac{e^{(-d(x_{\mathcal{H}_i}, \ell_{\mathcal{H}_{y_i}})/\tau)}}{e^{(-d(x_{\mathcal{H}_i}, \ell_{\mathcal{H}_{y_i}})/\tau)} + \sum_{j \in \mathcal{N}(i)} e^{(-d(x_{\mathcal{H}_i}, \ell_{\mathcal{H}_j}))/\tau)}}\right) \quad (8)$$

where $\tau \in \mathbb{R}^+$ is the temperature hyperparameter.

4.3.2 Overall Loss

The overall loss is a weighted cross-entropy (WCE), where each sample is weighted by its contrastive loss CL_i. For a batch of m samples:

$$\text{Loss}_{\text{WCE}} = -\frac{1}{m}\sum_{i=1}^{m} CL_i \cdot \log \frac{e^{(z_i^{y_i})}}{\sum_{j=1}^{c} e^{(z_i^j)}} \quad (9)$$

where z_i^j is the logit score for class j. The contrastive weight CL_i is high when the text is either distant from its true label or close to confusable negatives, guiding the model to penalize such cases more strongly.

5 Experiments

5.1 Experiment Setup

5.1.1 Datasets and Evaluation metrics

We use two benchmark fine-grained emotion datasets: GoEmotions (GE) (Demszky et al., 2020) with 27 emotion labels, and Empathetic Dialogues (ED) (Rashkin et al., 2019) with 32 emotion labels. We follow the same preprocessing and evaluation setup as prior work (Suresh and Ong, 2021; Chen et al., 2023), including accuracy and weighted F1 as evaluation metrics. Further details on dataset statistics are provided in Appendix A.

5.1.2 Implementation Details

We use the pretrained `RoBERTa-base` [2] as the text encoder. Text and label features have dimension h, set to 768. The curvature k is a scalar initialized as 1, and the scalars α_t and α_l are initialized as $1/\sqrt{h}$. All scalars are learned in the logarithmic space as: $\log(k)$, $\log(\alpha_t)$, and $\log(\alpha_l)$. The negative label set size r is set to 6 for Go Emotions and 8 for Empathetic Dialogues, determined via grid search on the validation set with $r \in \{2, 3, \ldots, 10\}$. τ is fixed at 0.07 for both datasets. During training, the batch size is set to 64, and the Adam optimizer is used with a learning rate of 1e-5. We train the model end-to-end using PyTorch. Training stops if

[2] https://huggingface.co/FacebookAI/roberta-base

neither accuracy nor weighted F1 improves on the validation set over ten consecutive epochs.

5.2 Main results

Table 1 presents the results of our proposed approach alongside baseline comparisons (see details of baseline methods in Appendix B). The first part of the table shows a comparison with pretrained language models (BERT (Devlin et al., 2019), RoBERTa, ELECTRA (Clark et al., 2020)) fine-tuned for FEC, in both base and large variants. The second part of the table compares with HyperIM (Chen et al., 2020) and HIDDEN (Chatterjee et al., 2021), which leverage hyperbolic space for classification but were not originally trained for FEC. Our proposed approach, HyCoEM, significantly outperforms all methods across both these sections of the table.

In the third part of the table, we compare with existing FEC methods, namely KNNEC (Yin and Shang, 2022), LCL (Suresh and Ong, 2021), and HypEmo (Chen et al., 2023). For a fair comparison, KNNEC and LCL were trained using RoBERTa as the encoder, ensuring all FEC methods use the same text backbone. We also include a variant of our approach, EucCoEM, which performs contrastive learning in Euclidean space and does not use hyperbolic geometry. We did not compare with SEAN-GNN (Zhang et al., 2024) as the official implementation is not publicly available.

For our implemented methods (KNNEC, LCL, EucCoEM, and HyCoEM), we report the average performance across five runs with different seeds. Our approach outperforms the second-best method, HypEmo, with the same parameter count (125M), achieving an improvement of 1.3–1.9% in accuracy and 1–1.7% in weighted F1 across the two datasets. In contrast, the Euclidean variant, EucCoEM, underperforms, highlighting the importance of hyperbolic space for learning label embeddings and improving text-label alignment.

5.3 Additional Analysis

In Table 2, we compare our model with FEC baselines on challenging ED subsets identified by (Suresh and Ong, 2021), each comprising four confusable labels (details in Appendix C). Since each subset contains four confusable labels, we use the other three (excluding the positive) as negatives to help the model better distinguish between similar emotions. HyCoEM outperforms the second-best by 0.9–1.4% in weighted F1 across all subsets.

Model	**Go Emotions (GE)**		**Empathetic Dialogues (ED)**	
	Acc	Weighted F1	Acc	Weighted F1
$\text{BERT}^{*}_{\text{base}}$	60.9 ± 0.4	62.9 ± 0.5	50.4 ± 0.3	51.8 ± 0.1
$\text{RoBERTa}^{*}_{\text{base}}$	62.6 ± 0.6	64.0 ± 0.2	54.5 ± 0.7	56.0 ± 0.4
$\text{ELECTRA}^{*}_{\text{base}}$	59.5 ± 0.4	61.6 ± 0.6	47.7 ± 1.2	49.6 ± 1.0
$\text{BERT}^{*}_{\text{large}}$	64.5 ± 0.3	65.2 ± 0.4	53.8 ± 0.1	54.3 ± 0.1
$\text{RoBERTa}^{*}_{\text{large}}$	64.6 ± 0.3	65.2 ± 0.2	57.4 ± 0.5	58.2 ± 0.3
$\text{ELECTRA}^{*}_{\text{large}}$	63.5 ± 0.3	64.1 ± 0.4	56.7 ± 0.6	57.6 ± 0.6
HyperIM*	50.2 ± 0.9	49.7 ± 0.7	44.1 ± 1.2	43.6 ± 1.0
HIDDEN*	47.2 ± 1.1	49.3 ± 0.9	42.9 ± 1.4	44.3 ± 1.1
KNNEC	63.8 ± 0.3	64.7 ± 0.8	57.8 ± 0.8	58.7 ± 1.1
LCL	64.1 ± 0.2	64.8 ± 0.3	59.2 ± 0.4	59.3 ± 0.6
HypEmo	65.4 ± 0.2	66.3 ± 0.2	59.6 ± 0.3	61.0 ± 0.3
EucCoEM	64.2 ± 0.5	64.6 ± 0.6	58.9 ± 0.4	59.1 ± 0.3
HyCoEM	**66.7 ± 0.4**	**67.3 ± 0.5**	**61.5 ± 0.3**	**62.7 ± 0.4**
Δ	+1.3%	+1%	+1.9%	+1.7%

Table 1: Comparison of results. The results for methods marked with (*) are sourced from the HypEmo (Chen et al., 2023) study. Δ denotes the improvement compared to the underlined second-best method. ± denotes standard deviation.

Model	subset_a	subset_b	subset_c	subset_d
$\text{RoBERTa}_{\text{base}}$	56.9	64.6	55.6	79.1
LCL	58.8	66.1	57.1	80.3
HypEmo	63.1	69.3	59.9	81.0
HyCoEm	**64.0**	**70.4**	**61.3**	**82.2**
Δ	+0.9%	+1.1%	+1.4%	+1.2%

Table 2: Weighted F1 scores on the most challenging subsets of the ED dataset, as proposed by (Suresh and Ong, 2021). Δ denotes the improvement over the second-best method.

We demonstrate the encoder-agnostic nature of our approach in Appendix D. A t-SNE visualization of the learned text representations, highlighting improved separation of confusable emotion labels in HyCoEM compared to other methods, is presented in Appendix E. We further ablate key design choices of HyCoEM in Appendix F, including the role of contrastive loss, the impact of label initialization, the effect of adaptive negative label selection, and the choice of hyperbolic geometry variant.

6 Conclusion

We propose HyCoEM for FEC, leveraging contrastive learning in hyperbolic space to align a text with its emotion label while separating it from confusable negatives. The contrastive loss helps learn label embeddings without a predefined hierarchy and serves as a weighting signal for cross-entropy loss, penalizing weak text-label alignments. Comparisons with baselines show that HyCoEM improves performance on benchmark datasets.

References

Soumya Chatterjee, Ayush Maheshwari, Ganesh Ramakrishnan, and Saketha Nath Jagarlapudi. 2021. Joint learning of hyperbolic label embeddings for hierarchical multi-label classification. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2829–2841, Online. Association for Computational Linguistics.

Boli Chen, Xin Huang, Lin Xiao, Zixin Cai, and Liping Jing. 2020. Hyperbolic interaction model for hierarchical multi-label classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):7496–7503.

Chih Yao Chen, Tun Min Hung, Yi-Li Hsu, and Lun-Wei Ku. 2023. Label-aware hyperbolic embeddings for fine-grained emotion classification. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10947–10958, Toronto, Canada. Association for Computational Linguistics.

Weize Chen, Xu Han, Yankai Lin, Hexu Zhao, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2022. Fully hyperbolic neural networks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5672–5686, Dublin, Ireland. Association for Computational Linguistics.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators. In *International Conference on Learning Representations*.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A Dataset of Fine-Grained Emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *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)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Paul Ekman, Tim Dalgleish, and Michael Power. 1999. Basic emotions. *Handbook of cognition and emotion*, 98(45-60):16.

Hamed Khanpour and Cornelia Caragea. 2018. Fine-grained emotion detection in health-related online posts. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1160–1166, Brussels, Belgium. Association for Computational Linguistics.

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*.

Dheeraj Mekala, Varun Gangal, and Jingbo Shang. 2021. Coarse2Fine: Fine-grained Text Classification on Coarsely-grained Annotated Data. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 583–594, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Maximillian Nickel and Douwe Kiela. 2018. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *International conference on machine learning*, pages 3779–3788. PMLR.

W Gerrod Parrott. 2001. *Emotions in social psychology: Essential readings*. psychology press.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381, Florence, Italy. Association for Computational Linguistics.

Tiberiu Sosea and Cornelia Caragea. 2021. eMLM: A New Pre-training Objective for Emotion Related Tasks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 286–293, Online. Association for Computational Linguistics.

Varsha Suresh and Desmond Ong. 2021. Not all negatives are equal: Label-aware contrastive loss for fine-grained text classification. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4381–4394, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Da Yin, Tao Meng, and Kai-Wei Chang. 2020. SentiBERT: A transferable transformer-based architecture for compositional sentiment semantics. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3695–3706, Online. Association for Computational Linguistics.

Wenbiao Yin and Lin Shang. 2022. Efficient Nearest Neighbor Emotion Classification with BERT-whitening. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4738–4745, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Pinyi Zhang, Jingyang Chen, Junchen Shen, Zijie Zhai, Ping Li, Jie Zhang, and Kai Zhang. 2024. Message passing on semantic-anchor-graphs for fine-grained emotion representation learning and classification.

In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 2771–2783, Miami, Florida, USA. Association for Computational Linguistics.

A Details on Datasets

GoEmotions (GE) (Demszky et al., 2020) and Empathetic Dialogues (Rashkin et al., 2019) (ED) are two widely recognized benchmark datasets commonly used for fine-grained emotion classification. These datasets are considered challenging, as they contain a large number of labels with overlapping semantics.

GoEmotions consists of 54,000 Reddit comments, each annotated with one or more of 27 emotion categories, along with a neutral class. Similar to prior studies (Suresh and Ong, 2021; Chen et al., 2023), we include only the single-labeled examples and remove the instances with the neutral label. After this selection, the dataset contains 23,485 / 2,956 / 2,984 examples for the train, validation, and test splits, respectively.

The Empathetic Dialogues dataset features multi-turn conversations between a speaker and a listener, with each conversation labeled with a single emotion. These conversations can extend up to six turns. Similar to prior studies (Suresh and Ong, 2021; Chen et al., 2023), we use only the first turn of each conversation. The dataset contains 24,850 samples labeled with 32 emotions, split into 19,533 / 2,770 / 2,547 examples for the training, validation, and test sets, respectively.

B Details on baseline methods

We compare our approach with three different categories of baseline methods.

Pretrained language models (PLMS). We fine-tuned base and large variants of BERT (Devlin et al., 2019), RoBERTa (Liu et al., 2019), and ELECTRA (Clark et al., 2020) for FEC.

Hyperbolic classification methods. These include approaches that leverage hyperbolic space but were not originally trained for FEC. HyperIM (Chen et al., 2020) jointly embeds text and labels in hyperbolic space, whereas HIDDEN (Chatterjee et al., 2021) learns label embeddings based on label co-occurrence information without assuming a predefined label hierarchy. Both methods utilize the Poincaré ball model of hyperbolic space.

FEC methods. KNNEC (Yin and Shang, 2022) incorporates a whitening transformation along with nearest-neighbor retrieval to improve sentence semantics. LCL (Suresh and Ong, 2021) uses a label-aware contrastive loss to modulate sample influence based on model confidence. HypEmo (Chen et al., 2023) uses hyperbolic text-label distance to weight the cross-entropy loss. We also include EucCoEM, a variant that operates in Euclidean space, with the rest of the components identical to HyCoEM.

C Details on Hard Subsets of ED

The hard subsets of Empathetic Dialogues (ED), selected by (Suresh and Ong, 2021), represent the most challenging and confusable emotion groups. These were identified by evaluating all possible four-label combinations to find sets with high semantic overlap. The selected subsets are: (a) {Anxious, Apprehensive, Afraid, Terrified}, (b) {Devastated, Nostalgic, Sad, Sentimental}, (c) {Angry, Ashamed, Furious, Guilty}, and (d) {Anticipating, Excited, Hopeful, Guilty}.

D Encoder agnostic performance

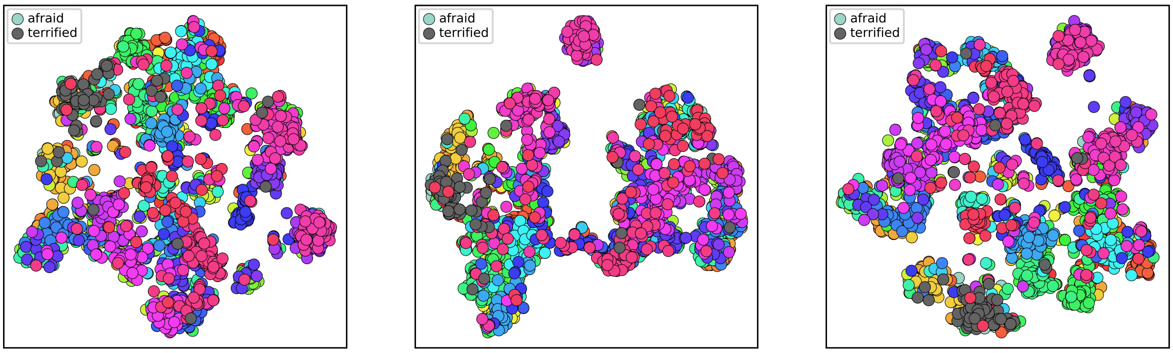
We propose HyCoEM as an encoder-agnostic approach that can improve FEC performance regardless of the text encoder used. Table 3 compares the weighted F1 scores with and without HyCoEM across different pretrained language models used as text encoders. The results demonstrate that incorporating HyCoEM improves performance across all encoders, highlighting the encoder-agnostic nature of our approach.

Dataset	Encoder	w/o HyCoEM	with HyCoEM
GE	BERT$_{\text{base}}$	62.9	**66.1**
GE	RoBERTa$_{\text{base}}$	64.0	**67.3**
GE	ELECTRA$_{\text{base}}$	61.6	**64.5**
ED	BERT$_{\text{base}}$	51.8	**58.6**
ED	RoBERTa$_{\text{base}}$	56.0	**62.7**
ED	ELECTRA$_{\text{base}}$	49.6	**58.9**

Table 3: Weighted F1 score when HyCoEM is used with different text encoders

E Visualization of Representations

Figure 2 shows t-SNE visualizations of the learned text representations on the ED test set. For fair comparison, t-SNE is applied with default settings across all methods. We compare with a standard cross-entropy variant that shares the same architecture as HyCoEM but is trained without contrastive



(a) Standard cross entropy loss

(b) HypEmo

(c) HyCoEM

Figure 2: Qualitative comparison of learned representations on the ED dataset. For the confusable emotion label pair *afraid* and *terrified*, HyCoEM shows increased separation compared to the other methods.

supervision(Fig. 2(a)), as well as with HypEmo (Fig. 2(b)). The analysis focuses on the confusable label pair *afraid* and *terrified*. In the standard cross-entropy setting, the representations of these labels are heavily entangled. In HypEmo, there is some improvement, but significant overlap still remains. HyCoEM (Fig. 2(c)) shows clearer separation between *afraid* and *terrified* compared to the other two, with reduced entanglement. Thus, HyCoEM helps in learning representations that better distinguish semantically similar and confusable emotions.

F Ablation study

We ablate the key components of our model, with results summarized in Table 4. First, removing contrastive loss supervision (*w/o CL*) and training the model using only cross-entropy leads to a substantial performance drop, highlighting the role of contrastive supervision in enhancing semantic alignment. Next, we initialized label embeddings using the average of RoBERTa token embeddings for each label name (*Label name init*). The observed decline suggests that random initialization is more effective than name-based initialization for this task. We also replaced the selection of top r negatives based on geodesic distance with random sampling (*Random negatives*). The underperformance of this variant underscores the value of adaptive negative selection.

We further replaced the label-text attention mechanism with simple elementwise multiplication between the text feature $x_i \in \mathbb{R}^h$ and the label features $W_L^\top \in \mathbb{R}^{c\times h}$, resulting in $F_i \in \mathbb{R}^{c\times h}$ (*w/o Label-text att.*). The lower performance of this variant confirms the importance of label-text attention, which computes label-specific features via weighted token aggregation.

Model	**Go Emotions (GE)**		**Empathetic Dialogues (ED)**	
	Acc	Weighted F1	Acc	Weighted F1
w/o CL	63.2 ± 0.6	64.1 ± 0.2	54.9 ± 0.7	56.6 ± 0.4
Label name init	64.9 ± 0.5	65.1 ± 0.4	58.7 ± 0.6	59.3 ± 0.2
Random negatives	64.1 ± 0.3	64.9 ± 0.4	55.9 ± 0.6	57.8 ± 0.5
w/o Label-text att.	63.9 ± 0.3	64.4 ± 0.5	55.2 ± 0.7	57.5 ± 0.7
PoincaréCoEM	65.3 ± 0.5	65.8 ± 0.6	59.3 ± 0.5	59.7 ± 0.6
HyCoEM	**66.7 ± 0.4**	**67.3 ± 0.5**	**61.5 ± 0.3**	**62.7 ± 0.4**

Table 4: Ablation study results for HyCoEM

Finally, we substituted the Lorentz model with the Poincaré ball for hyperbolic projection (*PoincaréCoEM)*. The resulting performance degradation demonstrates the superiority of the Lorentz model in our setup.