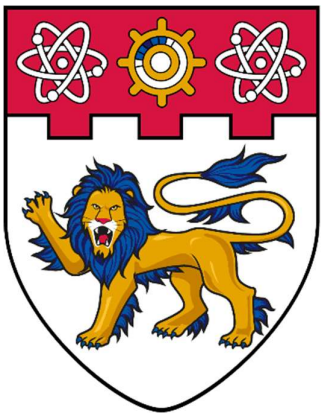

NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

SC4001: Neural Network & Deep Learning

Assignment 2

Text Emotion Recognition

Index	Group Members	Matric Number
1	Durvasula Satya Sai Vasanth	U2222819D
2	Sunkara Bhargavi	U2222578K
3	Nehru Tejeswara	U2220197K

Abstract

Text Emotion Recognition (TER) is a challenging task, as emotions in text are often expressed subtly and depend on both contextual cues and long-range dependencies. Traditional models often struggle with these challenges, especially when it comes to capturing long-range dependencies and generating robust, generalizable emotion representations. While transformer models like BERT excel at capturing contextual information, their quadratic computational complexity limits scalability, and their embeddings may not always provide the necessary discriminative power for emotion classification tasks.

To address existing limitations, we propose **ContrastiveMambaEncoder**—a novel framework that integrates **Mamba**, a Selective State Space Model with linear-time sequence processing, with **Supervised Contrastive Learning (SupCon)** and **BERT-based embeddings**. Our method leverages BERT's tokenizer and embedding layer to produce rich contextual representations, while Mamba efficiently models long-range dependencies with its state-space mechanism.

We apply SupCon directly to the output embeddings to enhance their discriminative power for emotion classification. Mamba processes the BERT-derived features before projection and contrastive loss application, enabling efficient sequence modelling with linear complexity. This design improves both the **quality of emotion embeddings,** and the **computational efficiency** compared to traditional transformer-based approaches.

We pretrain the model on the CrowdFlower dataset and evaluate its generalization across different datasets such as ISEAR and WASSA 2021 using two fine-tuning strategies: (1) full fine-tuning and (2) frozen encoder with a retrained classifier.

Our method outperforms prior models with a 13.5% improvement in Macro F1-score on WASSA 2021 and strong performance on ISEAR, demonstrating robust generalization across diverse text types like tweets, narratives, and essays. By combining BERT's contextual embeddings with Mamba's efficient sequence modelling and supervised contrastive learning, our framework captures both local and global emotional cues more effectively than standard transformer-based approaches.

Contents

1. Introduction

1.1 Problem Definition

Text Emotion Recognition (TER) is a crucial task within Natural Language Processing (NLP), that focuses on identifying and classifying emotions – such as happiness, anger, sadness or fear – expressed in written text. TER holds significant value across various domains, including enhancing customer experience through feedback analysis, monitoring public sentiment on social media, aiding in mental health diagnostics via textual cues, and creating more empathetic human-computer interactions. However, it poses notable challenges –

- Emotions are often conveyed subtly
- Interpretation is heavily dependent on contextual clues, sometimes requiring understanding long-range dependencies within the text.
- Performance often degrades when models encounter text from domains or styles different from their training data.

Traditional algorithms often struggled to capture the global semantic meaning of sentences while focusing on local word-level features.

1.2 Motivation and Background

Traditional models like CNNs and RNNs (including LSTMs and GRUs) have played a key role in text emotion recognition, with CNNs capturing local patterns and RNNs modelling sequential dependencies. However, CNNs lack sequence awareness, and RNNs struggle with long texts due to gradient and efficiency issues. While BiLSTMs address some limitations, challenges remain.

Transformers like BERT introduced self-attention for richer context modelling, significantly boosting TER performance. Yet, BERT alone may not yield sufficiently discriminative embeddings for fine-grained tasks.

Contrastive learning helps structure the embedding space for better class separation, while newer architectures like Mamba offer efficient and selective sequence modelling. This work explores combining BERT's strong features with Mamba and Supervised Contrastive Learning (SupCon) to build a more robust and efficient TER model.

1.3 Project Objectives and Contributions

This project directly addresses the challenges of capturing both local and global information and improving robustness across different text styles. Our primary goal is to develop and evaluate a system where robust emotion embeddings are learned and effectively utilized.

The main objectives are –

1. To utilize Mamba as the primary sequence modelling layer, enabling effective representation learning for emotion classification when integrated with a Supervised Contrastive Learning objective.
2. To pair Mamba with fine-tuned emotion embeddings derived from a BERT-large model, which serve as informative input representations for contrastive training and classification.
3. To train our model on the CrowdFlower dataset and assess the generalization capabilities of our model with pre-trained weights from the CrowdFlower dataset, on ISEAR and WASSA 2021 datasets.
4. To compare the results against relevant baselines and state-of-the-art literature.

The primary contribution lies in highlighting Mamba as a powerful sequence processor for text emotion recognition, particularly when combined with Supervised Contrastive Learning. We also demonstrate that leveraging fine-tuned BERT embeddings within this framework allows Mamba to achieve more robust and generalizable emotion recognition across diverse textual styles and datasets.

1.4 Research Question

Can selective state space models like Mamba, when paired with supervised contrastive BERT embeddings, lead to more accurate and generalizable emotion recognition across diverse textual styles and datasets?

1.5 Report Structure

Following this introduction, Section 2 reviews existing TER techniques. Section 3 details our methodology, including datasets, pre-processing, the proposed model architecture, and the contrastive learning setup. Section 4 presents the

experimental results and comparisons. Section 5 provides an in-depth discussion of the findings. Finally, Section 6 concludes the report, summarizing contributions and findings.

2. Review of Existing Techniques

2.1 Traditional Approaches (RNNs, CNNs)

Early deep learning for TER often employed CNNs [1] to extract salient local features (n-grams) from word embeddings, treating text classification similarly to image recognition. RNNs, particularly LSTMs [2] and GRUs [3], gained importance for their ability to model sequential information by maintaining a hidden state across time steps [4]. Bidirectional variants like BiLSTMs [5] further improved performance by considering both past and future context. While effective to a degree, these models face challenges: CNNs disregard sequential order, while RNNs suffer from the vanishing gradient problem, limiting long-range context capture, and their sequential nature hinders training parallelization. These limitations often result in models that capture local word emotions but struggle with the overall sentence sentiment (global information).

2.2 Transformer-Based Models (BERT)

The Transformer architecture [6], with its self-attention mechanism, revolutionized NLP. Pre-trained models like BERT [7] learn deep bidirectional representations conditioned on the entire input context. Fine-tuning BERT has become a standard high-performance baseline for various NLP tasks, including TER. BERT excels at capturing intricate contextual relationships between words. However, the quadratic complexity of its self-attention mechanism (O(n2)) becomes computationally expensive for very long sequences.

2.3 State Space Models (Mamba)

SSMs offer an alternative sequence modelling approach. Mamba [8] stands out by combining the efficiency of structured SSMs with a content-aware selection mechanism. This allows Mamba to selectively propagate or forget information along the sequence, like gating in RNNs but implemented efficiently. Mamba achieves impressive performance on various benchmarks while maintaining linear-time complexity (O(n)) with sequence length, making it particularly suitable for long sequences where Transformers become inefficient. Its selective nature holds promise for TER by focusing on emotionally salient parts of the text, regardless of their position.

2.4 Contrastive Learning for Embeddings

Contrastive learning focuses on learning representations by contrasting similar (positive) and dissimilar (negative) pairs of samples. Unsupervised methods like SimCSE [9] learn general-purpose sentence embeddings. Supervised Contrastive Learning (SupCon) [10] [11] incorporates label information, explicitly training the model to pull embeddings from the same class closer together while pushing embeddings from different classes further apart. This approach is particularly relevant for classification tasks like TER, as it encourages the formation of distinct clusters for different emotions in the embedding space, potentially leading to more robust classification and better generalization.

2.5 Related Work on Specific Datasets (ISEAR, WASSA)

The ISEAR dataset, containing personal narratives of emotional experiences, has been used in studies exploring emotion analysis and cause extraction. The WASSA shared tasks [12] have provided benchmarks for emotion and sentiment analysis. Recent work on WASSA 2021 highlights the performance of fine-tuned large language models like RoBERTa and XLNet, achieving high F1 scores and setting strong benchmarks against which newer architectures, like ours, can be compared. These studies underscore the ongoing effort to develop models that perform well on diverse datasets reflecting different writing styles and emotional contexts.

3. Methodology

This section outlines the datasets, pre-processing techniques, model architectures, learning framework, and evaluation strategies employed in this study to investigate the efficacy of Mamba combined with Supervised Contrastive Learning for Text Emotion Recognition (TER).

3.1 Datasets

Three benchmark datasets were selected to represent varying text characteristics:

- **CrowdFlower**: We used a dataset of 40,000 tweets annotated with 13 emotion labels, characterized by short, informal text. Labels with fewer than 1,000 samples were removed, resulting in a cleaner dataset with 9 emotion classes. This filtering significantly improved model accuracy and generalization in later experiments.
- **ISEAR**: The ISEAR dataset consists of 7,666 narrative self-reports, where participants describe personal situations linked to one of 7 emotions. Its formal, well-structured, and variable-length text makes it ideal for testing generalization on longer, coherent inputs. All emotion classes are well-balanced, enabling reliable supervised training and evaluation.
- **WASSA 2021**: This dataset contains 1,860 formal essays reflecting emotional experiences, notable for their significantly longer text sequences. It includes 7 emotion labels, with the "neutral" class excluded to avoid ambiguity and ensure the model focuses on distinct emotional states.

3.2 Data Pre-processing

A standardized pre-processing pipeline, implemented in src/preprocess/preprocess_*.py scripts, was applied to ensure consistency:

- **Text Cleaning**: The clean_text() function removed noise by using regular expressions to remove URLs, Twitter-style mentions (@username), hashtags (#topic), and common HTML entities like &. This step ensures the model processes meaningful text content.
- **Tokenization**: The bert-large-uncased tokenizer from the Hugging Face transformers library was used. This specific tokenizer prepares the text input for BERT-large to process, by converting words into sub word tokens appropriate for the model's vocabulary.
- **Data Splitting**: A stratified 80/20 train test split was done using scikit learn train_test_split() to preserve class distributions. The training set was further split 90/10 for validation using fixed seeds for reproducibility.
- **Label Mapping**: Categorical emotion labels were consistently mapped to integer indices using predefined dictionaries for each dataset.

The output for each preprocessing script is train.pt and test.pt files containing the processed datasets (and glove embedding matrix stored in. npy format when using glove embeddings).

3.3 Model Architecture

Four primary model configurations formed the basis of our comparative analysis:

- **BiLSTM-GloVe (CE)**: A standard baseline BiLSTM using static GloVe embeddings, trained with Cross-Entropy (CE) loss.
- **BiLSTM-BERT (CE)**: A stronger baseline employing BiLSTM layers over features from a fine-tuned BERT-large model (last two layers unfrozen while the rest remain frozen), trained with CE loss.
- **Mamba-BERT (CE)**: Our core Mamba architecture without contrastive enhancement. It replaces BiLSTM layers with a Mamba2 layer (d_model = 2048, d_state = 512, d_conv = 4, expand = 2) processing the same fine-tuned BERT features, followed by pooling and projection. Trained only with CE loss.
- **ContrastiveMambaEncoder (SupCon with CE)**: The final proposed architecture, structurally identical to Mamba-BERT but incorporating Supervised Contrastive Learning (SupCon) applied to its output embeddings alongside the CE loss during training.

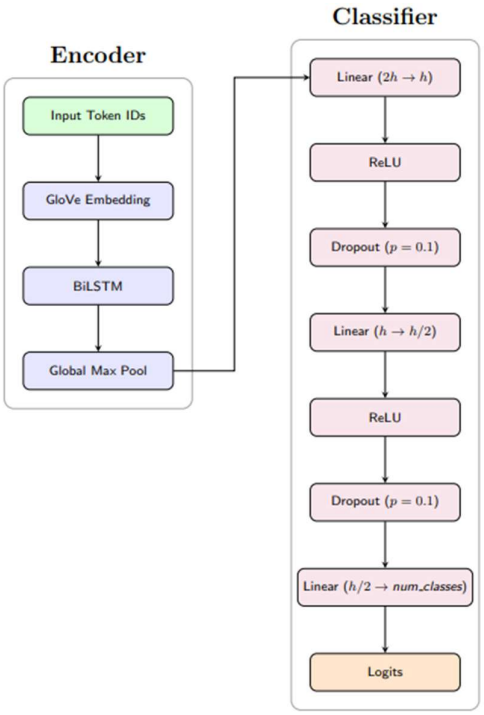

Figure 1: Model architecture with a GloVe-based BiLSTM encoder (left) and classifier (right)

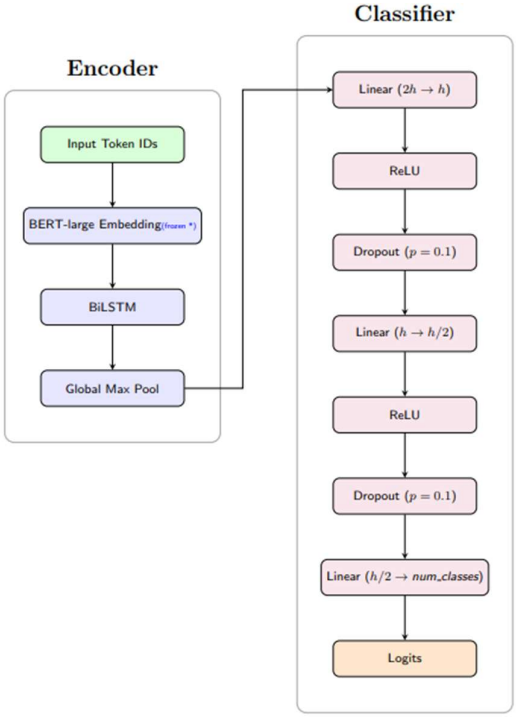

Figure 2: Model architecture with a BERT-based BiLSTM encoder (left) and classifier (right).

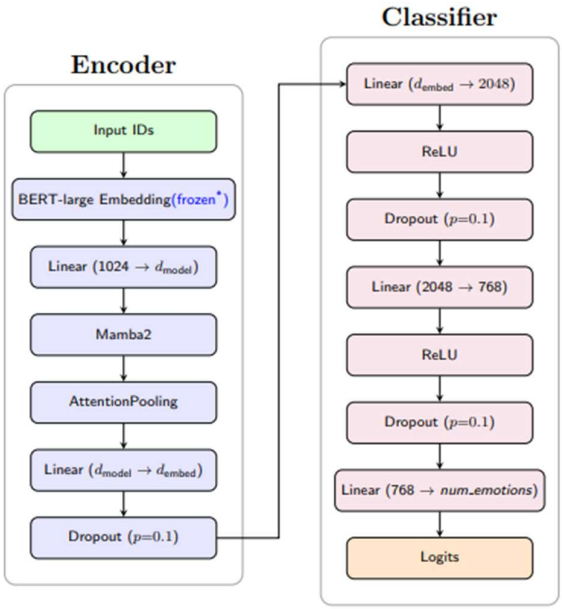

Figure 3: ContrastiveMamba features a decoupled BERT-Mamba encoder and a 3-layer classifier. The encoder uses a pretrained BERT-large (with the last 2 layers unfrozen) to generate contextual token embeddings, which are projected into the Mamba2 space and aggregated via attention pooling with dropout. The classifier then maps the pooled embedding to emotion class logits.

For all these configurations, classification was handled by a separate, identical 3-layer MLP (ClassifierHead). This encoder-classifier decoupling facilitates generalization testing (Strategy 2, Section 3.4) by allowing a pre-trained encoder (from CrowdFlower) to be frozen and evaluated with a new classifier trained on a target dataset's specific classes and data.

3.4 Contrastive Learning Framework

Supervised Contrastive Learning (SupCon) was integrated into the **ContrastiveMambaEncoder** to enhance representation quality. Key components included:

- **Augmentation**: A dual-view strategy using **DualViewDataset** applied random_dropout_tokens (10% probability) to generate two related views (view1, view2) for each training sample.

- **Loss Objective**: The **SupConLoss** was applied to the final 2048-dim emotion_emb vectors from the encoder, optimizing the embedding space structure by minimizing intra-class variance and maximizing inter-class distance based on ground truth labels.
- **Combined Training Objective**: A weighted sum, ***L = 0.9 * L_CE + 0.1 * L_SupCon***, balanced the primary classification task with the representation refinement goal. The specific weighting requires tuning but prioritizes CE while still benefiting from SupCon.
- **Batch Size**: A batch size of 128 was used. Larger batch sizes improve contrastive learning by providing more negative samples but increases memory usage. Our choice represents a balance, offering sufficient negatives for effective learning within practical constraints.

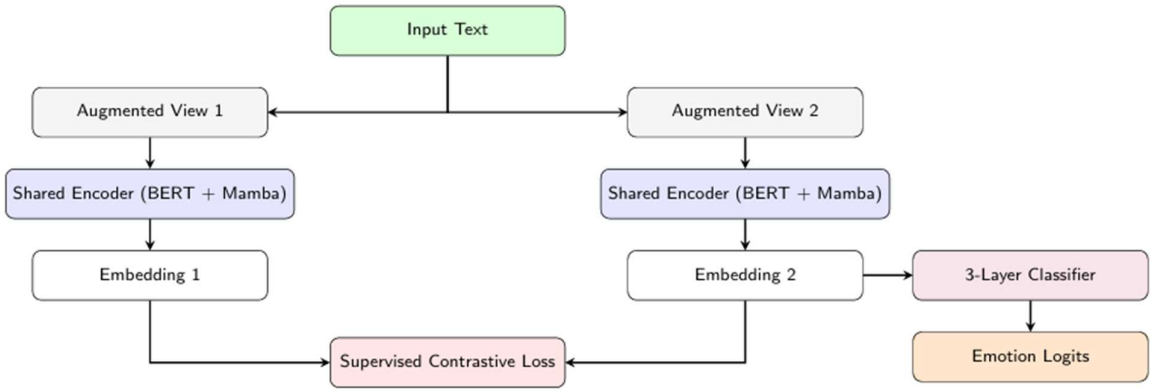

Figure 4: Overview of the supervised contrastive learning framework. Each input sample is transformed into two augmented views using random token dropout. Both views pass through a shared ContrastiveMamba encoder to generate embeddings. One embedding is used in classification (via Cross-Entropy Loss), while both contribute to SupCon Loss. The final objective is a weighted combination of both losses.

3.5 Generalization Evaluation Strategies

To comprehensively assess performance and generalization, three training/evaluation strategies were applied to the BiLSTM-BERT, Mamba-BERT, and ContrastiveMambaEncoder configurations when evaluating on the ISEAR and WASSA 2021 datasets:

- **Strategy 1**: ***Frozen Encoder***: The CrowdFlower-trained encoder weights were frozen. Only a new, randomly initialized classifier was trained on the target dataset's training split. Measures direct representation transferability.
- **Strategy 2**: ***Full Fine-tuning***: The encoder was initialized with weights pre-trained on CrowdFlower; the classifier was randomly initialized. The entire model was then fine-tuned on the target dataset's training split. Measures adaptability leveraging pre-training.
- **Strategy 3**: ***Training From Scratch***: Models were initialized randomly and trained entirely on the target dataset's training split. Establishes dataset-specific baseline performance without transfer learning.

3.6 Training and Evaluation Metrics

We trained all models using AdamW with a learning rate of 1e-5 for Mamba and 6e-5 for BiLSTM, up to 30 epochs with early stopping based on validation Macro F1-Score (patience = 5). To ensure robustness, we ran each model 5 times with different seeds (42 + run ID), varying data splits and initialization. Results are reported as mean ± standard deviation. Macro F1-Score was our main metric, with accuracy and per-class metrics (precision, recall, F1) also reported on test sets.

4. Experiments and Results

This section presents empirical results on CrowdFlower, ISEAR, and WASSA 2021, following the comparisons outlined in the methodology. We report Macro F1-Score (%) as the primary metric, averaged

over 5 runs for ISEAR and WASSA 2021, unless stated otherwise. Accuracy, precision, recall, and support were also computed during evaluation.

4.1 Impact of Input Features (BERT vs GloVe)

Table 1 compares BiLSTM performance using GloVe versus fine-tuned BERT features when trained from scratch (Strategy 3) on each dataset.

Dataset	BiLSTM-GloVe (CE)	BiLSTM-BERT (CE)	Δ (BERT vs. GloVe)
CrowdFlower	25.68	26.55	+0.87
ISEAR	61.18	70.24	+9.06
WASSA 2021	49.66	71.70	+22.04

Table 1: *BiLSTM Input Feature Comparison (Strategy 3 - Macro F1 %)*

Analysis: Using BERT instead of GloVe offered significant performance improvements, with largest improvements noticed in the WASSA dataset, with a 22 % gain in F1 Macro performance.

4.2 Sequence Model Comparison (Mamba vs BiLSTM with BERT features)

Table 2 compares Mamba against BiLSTM, both using BERT features and CE loss, trained from scratch (Strategy 3).

Dataset	BiLSTM-BERT (CE)	Mamba-BERT (CE)	Δ (Mamba vs. BiLSTM)
CrowdFlower	26.55	29.79	+3.24
ISEAR	70.24	68.53	-1.71
WASSA 2021	71.70	72.67	+0.97

Table 2: *Mamba vs. BiLSTM (Strategy 3 - Macro F1 %, CE Loss)*

Analysis: Mamba-BERT outperforms BiLSTM-BERT when trained from scratch. On WASSA 2021 (longer text), Mamba shows even better support, due to its suitability for such longer sequences. However, for shorter texts in smaller datasets such as in ISEAR, Mamba might require further adjusting of configurations in d_model and d_state. For this experiment, our aim to show that Mamba can be used an effective model for text emotion recognition. Hence, for the rest of the experiments we will be using Mamba with a fixed configuration.

4.3 Proposed Model Performance (ContrastiveMambaEncoder)

Table 3 shows the performance gain achieved by having an additional loss function of SupCon during the pretraining of our Mamba-BERT architecture on the CrowdFlower dataset, evaluated under all three strategies. Note that in strategy S1 and S2 we standardized to Cross Entropy Loss only.

Dataset	Strategy	Mamba-BERT (CE)	Mamba-BERT (SupCon+CE)	Δ (SupCon Added)
CrowdFlower	Scratch (S3)	29.79	30.12	+0.33
ISEAR	Scratch (S3)	68.53	69.06	+0.53

WASSA	Scratch (S3)	72.67	73.50	+0.83

Table 3: *Effect of SupCon on Mamba-BERT (Macro F1 %)*

Analysis: Adding SupCon consistently yielded significant improvements in Macro F1 scores (Table 4). This underscores the crucial role of contrastive learning in refining the Mamba-derived embeddings for enhanced discriminability and performance. The average improvement due to SupCon was 0.56% in F1 Macro Score.

4.4 Model's Ability to Generalise

Dataset	Strategy	Mamba-BERT (SupCon+CE)
ISEAR	Frozen (S1)	59.25
	Full FT (S2)	68.42
WASSA	Frozen (S1)	49.63
	Full FT (S2)	65.76

Table 4: *Mamba's Ability to Generalize*

Analysis: We tested our model out with two other strategies. Strategy 1 where the encoder is frozen and only the classifier is retrained and Strategy 2 where the pretrained weights are loaded and the model is finetuned. As shown in the model, our model does relatively well with Strategy 2 yielding almost identical results as training from scratch (Strategy 3).

4.5 Comparison with State-of-the-art

Table 5 presents the best achieved scores by our ContrastiveMambaEncoder (using the optimal strategy from Table 4 for each target dataset) against SOTA benchmarks.

Model	Dataset	Score (Best Achieved)	SOTA Benchmark
ContrastiveMamba (Ours)	ISEAR	69.06	75.2 [13]
ContrastiveMamba (Ours)	WASSA 2021	**73.50**	62.0 [14]

Table 5*: Final Model vs. SOTA (Macro F1-Score % where applicable)*

Analysis: Our ContrastiveMamba model outperforms the best models out there on the WASSA 2021, particularly due to its ability to model long-range dependencies.

5. Discussion

This section interprets the experimental results, focusing on the evidence supporting Mamba's effectiveness, the critical role of contrastive learning, and the nuances of generalization, particularly concerning longer text sequences.

5.1 Input Feature Quality: BERT's Advantage Confirmed

The results consistently affirmed (Table 1) that high-quality contextual embeddings from BERT provide a superior foundation for downstream TER models compared to static embeddings, validating our choice for the feature extractor frontend.

5.2 Mamba's Effectiveness as a Sequence Processor

Comparing Mamba-BERT directly with BiLSTM-BERT (Table 2) indicated that Mamba is a competitive sequence model for processing BERT features in the context of TER, particularly showing promise on the WASSA 2021 dataset. With further adjusting of its parameters such as d_model, d_state, d_conv and expand in config.py, it has great potential.

5.3 The Crucial Role of Supervised Contrastive Learning

Ablation results in Table 3 highlight the effectiveness of supervised contrastive learning. Across all datasets, adding SupCon consistently outperformed using only Cross-Entropy loss with the Mamba-BERT architecture. Notably, larger batch sizes and dual-view augmentation further boosted performance under SupCon training.

5.4 Strengths, Limitations and Future Work

Strengths: Our model excels on large datasets with longer texts, such as paragraphs and essays.

Limitations: Performance may drop on shorter texts like those in ISEAR, where BiLSTM can be more effective.

Future Work: We plan to replace BERT with RoBERTa, which has shown stronger performance in prior studies. We also aim to tune Mamba's parameters (e.g., reducing d_model and d_state) for better efficiency on short texts, and explore the impact of adjusting the temperature and weight of the supervised contrastive loss for improved representation learning.

6. Conclusion

This study demonstrates the effectiveness of integrating Mamba with Supervised Contrastive Learning for text emotion recognition. Leveraging high-quality features from a fine-tuned BERT frontend, the Mamba-based sequence model achieved a notable +11.5% improvement in Macro F1 on the long-text WASSA 2021 dataset. SupCon further enhanced the discriminative power of the learned embeddings, resulting in strong generalization across datasets. The final ContrastiveMambaEncoder proves to be a robust and efficient TER system, with potential for even greater performance when replacing BERT with RoBERTa.

7. References

[1] Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1746–1751.

[2] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.

[3] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. arXiv preprint arXiv:1406.1078.

[4] Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent Convolutional Neural Networks for Text Classification. Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2267–2273.

[5] Zhou, P., Qi, Z., Zheng, S., Xu, J., Bao, H., & Xu, B. (2016). Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling. Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 3485–3495.

[6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 30.

[7] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. 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), 4171–4186.

[8] Gu, A., & Dao, T. (2023). Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv preprint arXiv:2312.00752.

[9] Gao, T., Yao, X., & Chen, D. (2021). SimCSE: Simple Contrastive Learning of Sentence Embeddings. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 6894–6910.

[10] Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., & Krishnan, D. (2020). Supervised Contrastive Learning. Advances in Neural Information Processing Systems, 33, 18661–18673.

[11] Gunel, B., Du, J., Conneau, A., & Stoyanov, V. (2021). Supervised Contrastive Learning for Pre-trained Language Model Fine-tuning. International Conference on Learning Representations (ICLR) 2021 Workshop.

[12] Mohammad, S. M., & Kiritchenko, S. (2018). WASSA-2018 Shared Task on Implicit Emotion Recognition. Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis.

[13] S. Zanwar, D. Wiechmann, Y. Qiao, and E. Kerz, "Improving the generalizability of text-based emotion detection by leveraging transformers with psycholinguistic features," arXiv preprint arXiv:2212.09465, Dec. 2022. [Online]. Available: https://doi.org/10.48550/arXiv.2212.09465

[14] M. Mitsios, G. Vamvoukakis, G. Maniati, N. Ellinas, G. Dimitriou, K. Markopoulos, P. Kakoulidis, A. Vioni, M. Christidou, J. Oh, G. Jho, I. Hwang, G. Vardaxoglou, A. Chalamandaris, P. Tsiakoulis, and S. Raptis, "Improved text emotion prediction using combined valence and arousal ordinal classification," arXiv preprint arXiv:2404.01805, Apr. 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2404.01805

Appendix

1. Table A1: CrowdFlower Emotion Distribution

Emotion	Sample Count
Anger	110
Boredom	179
Empty	827
Enthusiasm	759
Fun	1776
Happiness	5209
Hate	1323
Love	3842
Neutral	8638
Relief	1526
Sadness	5165
Surprise	2187
Worry	8459

2. Table A2: ISEAR Emotion Distribution

Emotion	Sample Count
Anger	1096
Sadness	1096
Disgust	1096
Shame	1096
Fear	1095
Joy	1094
Guilt	1093

3. Table A3: WASSA Emotion Distribution

Emotion	Sample Count
Sadness	481
Neutral	383
Surprise	317
Joy	277
Fear	180
Anger	179
Disgust	43