



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 15, Issue 2, February 2026**

# Comparing Generative, Traditional and Hybrid Data Augmentation Methods for Fine-Grained Emotion Classification

Akrshita Sharma<sup>1</sup>, Kushal Sharma<sup>1</sup>, Anamika Rangra<sup>2</sup>

P.G. Student, Department of Computer Science and Engineering, Himachal Pradesh Technical University, Hamirpur,  
Himachal Pradesh, India<sup>1</sup>

P.G. Student, Department of Computer Science and Engineering, Himachal Pradesh Technical University, Hamirpur,  
Himachal Pradesh, India<sup>1</sup>

Guest Faculty, Department of Computer Science and Engineering, Himachal Pradesh Technical University, Hamirpur,  
Himachal Pradesh, India<sup>2</sup>

**ABSTRACT:** The fine-grained emotion classification has turned out to be a crucial activity in natural language processing that allows learning more about the subtle human feelings than just broad categories of sentiment. Nevertheless, the lack of labeled data, as well as class imbalance, are still significant issues, especially with delicate or low-resource emotional classes. Another approach that has come up to curb such constraints is data augmentation. The conventional methods like backtranslation and paraphrasing have shown quantifiable increases in the performance of the affective classification by augmenting linguistic variety and maintaining semantics [6]. As the transformer-based architecture, such as DistilBERT [9], continues to develop, and more people switch to the use of large language models (LLM), methods of generative augmentation, such as persona-based and prompt-based text synthesis, become a matter of serious concern [3], [7]. Although the fine-tuning techniques enhance task-specific fine-tuning [2], [4], the issues of bias, robustness, and interpretability of the data generated by LLM are essential [5]. Hybrid frameworks: A significant number of papers use structured augmentation techniques alongside transformer representations and multimodal or self-supervised methods to improve generalization and fairness [1], [8].

The article systematically compares different data augmentation paradigms of generative, traditional, and hybrid data augmentation in the context of fine-grained emotion classification by comparing the efficacy of these data augmentation approaches, their computational efficiency, and their robustness, and ethical issues. Moreover, it points at the gap in research on evaluation consistency, domain adaptation, and bias mitigation and defines the future directions of the development of reliable, scalable, and cost-efficient emotion-aware NLP systems.

**KEYWORDS:** Fine-Grained Emotion Classification, Data Augmentation, Generative Models, Traditional Augmentation, Hybrid Methods, Large Language Models (LLMs).

## I. INTRODUCTION

Emotion classification is a core activity within the natural language processing (NLP) which tries to recognize and classify human emotions in text. Whereas traditional sentiment analysis looks at the overall category e.g. positive, negative, or even neutral, fine-grained emotion classification aims at detecting more particular emotional states e.g. joy, anger, fear, sadness, surprise and disgust. This more detailed classification is necessary in applications such as mental health applications, social media applications, customer feedback applications, conversational agents, and human-computer interaction systems.

Recent progress in Transformer architecture based systems, including DistilBERT [9], as well as the rapid growth of large language models (LLMs) have shown considerable benefits to text classification. Nevertheless, the classification of emotions on a fine-grained level is a difficult task because of the tendency of annotated datasets, the class distribution, the ambiguity of the context, and the domain variability. The research on fine-tuning the scientific and

domain-specific classification problem of LLM emphasizes the improvement in performance and the limitation in computing power [2], [4]. Also, the issue of bias and interpretability in the case of systems based on LLM has been highlighted in recent studies [5], especially when synthetic data generation is used. Data augmentation has become a key approach in order to solve these problems. Conventional methods like backtranslating and paraphrasing have been found to be useful in enhancing the performance of emotion classification by enhancing the linguistic diversity without altering the semantic meaning [6]. More recently, contextual rich and diverse training samples have been promising with generative augmentation with LLMs, such as persona-driven text generation, [3], [7]. Moreover, hybrid models that combine existing transformer based representation with artificial data creation approaches strive to trade off robustness, scalability, and fairness [1], [8]. Despite covering all of these augmentation paradigms separately, the comparison of traditional, generative, and hybrid data augmentation approaches regarding the use in the field of fine-grained emotion classification is not well done. Thus the review paper provides a systematic analysis of the current approaches, a critique of their limitations and strengths and helps identify gaps in the research to move towards the creation of reliable, bias-conscious as well as cost-effective emotion-aware NLP systems.

## II. LITERATURE SURVEY

[1] **Sourget et al. (2025)** examined the notions of fairness and robustness in the CLIP-based frameworks and examined the behavior of large multimodal architectures on the distributional changes. The significance of their study is that it emphasizes the need to consider robustness and bias in model adaptation frameworks, which proves to be extremely timely when synthetic augmentation is incorporated into classification pipelines. The conclusions highlight that there should be augmentation strategies that are robustness-conscious, to enable reliable performance [1].

[2] **Rostam and Kertész (2024)** made a comparison of scientific text classification of large language models through the application of fine-tuning. The authors have shown that fine-tuning of the models is quite effective to task-specific performance, however, the computational cost and data requirements are the significant drawbacks. Their study justifies the nature of adaptive fine-tuning in hybrid augmentation models [2].

[3] **Jeong et al. (2025)** suggested the persona-based method of text data augmentation based on the LLM. The study demonstrated statistically significant changes in the classification accuracy by creating emotionally varied samples by persona conditioning and this was especially true in the case of imbalanced classes. The work is a generative augmentation method that can be applied in the case of fine-grained emotion tasks [3].

[4] **Jahan et al. (2025)** compared the cost-efficient LLMs on biomedical benchmark tasks. They indicate that optimized or smaller LLM can compete when trained well, and thus, they become viable applications in augmentation enhanced classification systems [4].

[5] **Chandna et al. (2025)** discussed the bias in the LLMs through the lens of mechanistic interpretability. It was found that internal model representations can enhance the existing biases, particularly in cases where the synthetic generation of data is not controlled. The study sheds light on the necessity of bias-conscious augmentation of emotion classifications [5].

[6] **Radliński et al. (2025)** compared backtranslation and paraphrasing methods regarding the classification of emotions. Their empirical study revealed that conventional augmentation strategies enhance robustness and generalization especially with low-resource emotional classes. This paper serves as potent background information on traditional augmentation effectiveness [6].

[7] **Chai et al. (2025)** provided an in-depth overview of text data augmentation methods of large language models. The authors divided augmentation strategies into traditional, generative, and hybrid paradigms and presented their strengths and weaknesses, as well as open research issues. Their survey provides a theoretical basis of comparative assessment [7].

[8] **Imam et al. (2025)** investigated the idea of zero-shot classifier tuning with the help of big-sized pretrained models and self-supervised learning methods. They conclude that their hybrid frameworks using pretrained embeddings and unlabeled data could improve generalization, which can help to combine augmentation and representation learning [8].

[9] Sanh *et al.* (2020) proposed the DistilBERT, a smaller variant of BERT which does not lose its competitive performance but minimizes model size and inference cost. The article offers background information that lightweight transformer models can successfully hold augmented training strategies in practicable NLP systems [9].

### III. COMPARISION OF EXISTING METHODS

Table1: Comparison of Existing Methods

Ref. No.	Method Type	Core Approach	Strengths	Limitations	Relevance to Fine-Grained Emotion Classification
[1]	Robustness Evaluation	Fairness & robustness analysis in large models	Highlights distributional shift challenges; improves reliability assessment	Focused on multimodal domain	Useful for evaluating robustness of augmented emotion datasets
[2]	Fine-Tuning (Hybrid Support)	Task-specific LLM fine-tuning	High domain-specific performance	High computational cost; needs labeled data	Supports hybrid augmentation + fine-tuning framework
[3]	Generative Augmentation	Persona-driven LLM text generation	Generates diverse and context-rich emotional samples	Risk of hallucination and bias amplification	Effective for handling class imbalance in fine-grained emotions
[4]	Cost-Efficient LLMs	Optimized LLM evaluation strategies	Reduced computational requirements; scalable	Possible performance trade-offs	Suitable for practical emotion-aware NLP systems
[5]	Bias & Interpretability Analysis	Mechanistic bias evaluation in LLMs	Identifies fairness and ethical concerns	Does not propose direct augmentation technique	Important for bias-aware generative augmentation
[6]	Traditional Augmentation	Backtranslation & paraphrasing	Improves robustness; preserves semantic meaning	Limited linguistic diversity	Effective for low-resource emotion categories
[7]	Survey (All Types)	Taxonomy of augmentation methods	Provides theoretical foundation and research gaps	No task-specific experimental validation	Guides structured comparison of augmentation strategies
[8]	Hybrid Framework	Zero-shot tuning + pretrained models	Leverages unlabeled data; strong generalization	Dependent on pretrained model quality	Enhances hybrid augmentation pipelines
[9]	Transformer Baseline	DistilBERT (compressed transformer)	Lightweight; faster inference; cost-effective	Slight performance drop vs full-scale models	Strong baseline for augmented emotion classification

IV. PROBLEM DOMAIN

1. Data Scarcity in Fine-Grained Emotion Labels

Emotion classification Fine-grained emotion classification involves finer-grained datasets of annotated data of subtle emotion categories. But there exist only a limited collection of labeled datasets on specific emotions which are usually imbalanced.

2. Class Imbalance and Rare Emotion Categories

Some of the emotions like fear, disgust or surprise are less represented than the prevalent emotions like sadness or joy. This disproportion decreases model generalization and classification rates.

3. Contextual and Semantic Ambiguity

Emotional expressions are quite situational and might differ in various domains. Subtle or implicit emotional responses are usually very difficult to read properly by models.

4. Bias in Generative Augmentation Methods

Synthetic data generation with LLM can add or contribute to the underlying biases within the pretrained models. This has the potential to impact on impartiality and dependability in emotion conscious systems.

5. Computational Cost of Large Language Models

Large models tend to consume large amounts of computational resources when trying to fine-tune and generatively augment. This restricts scalability and applicability in real life applications.

6. Evaluation and Robustness Challenges

No uniform philosophy exists to undertake a consistent assessment of conventional, generative, and hybrid augmentation approaches. The need to be robust in domains and datasets is an unresolved issue.

V. COMPARATIVE EVALUATION OF EXISTING APPROACHES

Table 2: Comparative Evaluation of Existing Approaches

Ref. No.	Approach Category	Methodological Focus	Key Contributions	Challenges Identified	Implication for Fine-Grained Emotion Classification
[1]	Model Reliability Assessment	Evaluation of fairness and stability in large pretrained models	Establishes robustness benchmarks under distribution shifts	Limited direct focus on text emotion datasets	Provides evaluation perspective for augmentation robustness
[2]	Domain-Specific Adaptation	Supervised fine-tuning of LLMs for classification tasks	Demonstrates performance gains through adaptive learning	Requires substantial labeled data and resources	Supports integration with augmentation strategies
[3]	Synthetic Data Generation	Persona-conditioned emotional text generation using LLMs	Enhances diversity of minority emotion classes	Potential risk of semantic noise and bias	Improves representation of subtle emotional expressions
[4]	Efficient LLM Deployment	Performance analysis of lightweight and optimized LLMs	Balances efficiency and predictive capability	Trade-off between model size and accuracy	Encourages scalable emotion-aware applications
[5]	Bias Examination Framework	Internal representation and bias inspection in	Highlights ethical and fairness considerations	Does not directly enhance dataset size	Guides responsible generative augmentation

		LLMs			
[6]	Linguistic Transformation Techniques	Backtranslation and structured paraphrasing	Maintains semantic integrity while increasing data volume	Limited contextual creativity	Serves as controlled augmentation baseline
[7]	Augmentation Strategy Survey	Systematic categorization of augmentation methodologies	Identifies comparative strengths and open challenges	Lacks experimental focus on single task	Forms theoretical backbone for method comparison
[8]	Pretrained Model Enhancement	Zero-shot and representation-based tuning approaches	Utilizes unlabeled data effectively	Performance dependent on pretrained embeddings	Improves generalization in hybrid systems
[9]	Compact Transformer Architecture	Knowledge-distilled transformer model (DistilBERT)	Reduced computational overhead with competitive accuracy	Slight reduction in predictive performance	Acts as efficient baseline for augmented training

## VI. CONCLUSION

The classification of emotions in a fine-grained fashion has been a key factor in the creation of emotion conscious natural language processing systems. Nevertheless, issues in the form of little labeled data, class imbalance, contextual ambiguity, computational limits, and bias of generative models still obstruct sound performance. This was a systematic comparison of traditional, generative and hybrid methods of data augmentation through the latest studies. The more traditional approaches like backtranslation and paraphrasing have stable and cost effective enhancements but little contextual variety. Large language model based generative methods are shown to have good capacity to generate varied and semantically strong samples of emotions, especially in minority groups, however, they posed a threat to bias, hallucination and computation costs. Hybrid models, which are based on augmentation techniques but also include transformer-based refinement and pretrained representations, seem to be the most reasonable compromise in robustness, scalability, and performance. In general, although augmentation can remarkably improve the fine-grained emotion classification, the selection of methods, bias regulation, and consistency of evaluations are critical to achieving the creation of efficient and ethically responsible systems.

## VII. FUTURE WORK

- Creation of prejudice-conscious generative augmentation models to reduce ethical hazards.
- Research on lightweight and cost-effective LLMs to be used in scalable emotion classification.
- Planning of standardized benchmarks of evaluation of traditional, generative and hybrid augmentation.
- Combination of zero-shot and few-shot prompting methods on low resource categories of emotions.
- Research on domain-adaptive augmentation of cross-domain emotion classification.
- Integration of multimodal information (text + audio + visual stimuli) in order to enhance the accuracy of fine-grained emotion detection.
- Creation of an integrated framework of benchmarking that directly compares generative, traditional and hybrid augmentation strategies in the same fine-grained emotion datasets, and in the same evaluation metrics.
- Architecture of dynamically-configured adaptive hybrid augmentation pipelines that choose between traditional and generative techniques depending on the complexity of the emotion category and the properties of the dataset.

**REFERENCES**

- [1] T. Sourget, D. Restrepo, C. Hudelot, E. Ferrante, S. Christodoulidis and M. Vakalopoulou, “Fairness and Robustness of CLIP-Based Models for Chest X-Rays,” in *arXiv preprint*, arXiv:2507.21291, 2025.
- [2] Z. R. K. Rostam and G. Kertész, “Fine-Tuning Large Language Models for Scientific Text Classification: A Comparative Study,” in *arXiv preprint*, arXiv:2412.00098, 2024.
- [3] H. S. Jeong, H. K. Ko, S. Y. Park and T. Kim, “LLM-Based Persona-Driven Text Data Augmentation,” in *IEEE Access*, vol. 13, pp. 167560–167574, 2025.
- [4] I. Jahan, M. T. R. Laskar, C. Peng and J. X. Huang, “Evaluating the Effectiveness of Cost-Efficient Large Language Models in Benchmark Biomedical Tasks,” in *arXiv preprint*, arXiv:2507.14045, 2025.
- [5] B. Chandna, Z. Bashir and P. Sen, “Dissecting Bias in LLMs: A Mechanistic Interpretability Perspective,” in *arXiv preprint*, arXiv:2506.05166, 2025.
- [6] Ł. Radliński, M. Guściora and J. Kocoń, “Backtranslation and Paraphrasing in the LLM Era? Comparing Data Augmentation Methods for Emotion Classification,” in *Proc. International Conference on Computational Science*, Springer, Cham, pp. 3–17, 2025.
- [7] Y. Chai, H. Xie and J. S. Qin, “Text Data Augmentation for Large Language Models: A Comprehensive Survey of Methods, Challenges, and Opportunities,” in *arXiv preprint*, arXiv:2501.18845, 2025.
- [8] M. F. Imam, R. F. Marew, J. Hassan, M. Fiaz, A. F. Aji and H. Cholakkal, “CLIP Meets DINO for Tuning Zero-Shot Classifier Using Unlabeled Image Collections,” in *arXiv preprint*, arXiv:2411.19346, 2025.
- [9] V. Sanh, L. Debut, J. Chaumond and T. Wolf, “DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter,” in *arXiv preprint*, arXiv:1910.01108, 2020.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details