Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.37226

Parameter-Efficient Adaptation of Modern Pretrained Encoders for Low-Resource Entity-Level Financial Sentiment Classification

Xiang Peng

School of AI and Advanced Computing, Xi'an Jiaotong-Liverpool University, Suzhou, China
Xiang.Peng23@student.xjtlu.edu.cn

Abstract. Financial news may express different sentiments towards the entities it mentions, while labelled examples for entity-level classification are often limited. This study compares the adaptation of three pretrained encoders to this task on FinEntity: FinBERT, a model based on Bidirectional Encoder Representations from Transformers (BERT) and further pretrained on financial text; DeBERTa-v3-base, the base-sized third version of Decoding-enhanced BERT with disentangled attention; and ModernBERT-base, a modern bidirectional Transformer encoder. Term frequency–inverse document frequency (TF-IDF) features with a linear support vector machine (SVM) serve as the traditional baseline. The pretrained encoders use explicit target-entity markers and a representation combining global and entity information. Three adaptation strategies are considered: full fine-tuning, a frozen encoder and low-rank adaptation (LoRA). Evaluation covers overall classification performance, trainable parameter counts, learning with reduced training data, variation across random seeds, entity representation and transfer to another dataset. In the overall comparison, FinBERT and DeBERTa-v3-base outperform the traditional baseline, whereas ModernBERT-base does not. DeBERTa-v3-base obtains the highest macro-averaged F1 score (Macro-F1) in this comparison. FinBERT performs better at the smallest training budget and produces more consistent results across the tested seeds, indicating that the preferred encoder depends on the evaluation setting. LoRA achieves competitive classification scores with only a small proportion of parameters updated, although this reduction does not shorten training time in the reported experiments. Combining global and target-entity representations improves performance for both encoders evaluated in the ablation study. Without further fine-tuning, the two LoRA models also achieve Macro-F1 scores above 0.70 on SEntFiN, but both score lower than on FinEntity. The results support parameter-efficient adaptation for this task while showing that performance under limited data, consistency across runs and transfer to another dataset require separate consideration.

Keywords: entity-level financial sentiment classification, parameter-efficient fine-tuning, LoRA, pretrained language models, low-resource learning

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.37226

1. Introduction

Financial news, corporate statements and market commentary make up part of the growing volume of text produced by digital finance and Internet services. Alongside facts about companies and market events, these texts convey sentiment relevant to investment decisions, corporate reputation monitoring and financial risk analysis. Extracting this information is the focus of financial sentiment analysis (FSA), a research area within natural language processing (NLP). Financial pretrained language models offer ways to represent domain-specific language, while benchmark datasets allow researchers to evaluate how well these models analyse financial texts [1-4].

Much of the existing research considers sentiment classification at the sentence or document level, assigning a single label to each text. Financial news, however, often mentions several companies or financial entities, and the sentiment towards them may differ. A single label cannot capture these differences, which limits its usefulness for corporate reputation monitoring, news filtering and financial risk management. Tang et al. introduced FinEntity for entity-level financial sentiment classification, while Sinha et al. developed SEntFiN 1.0 as another benchmark for this task [3, 4]. By associating sentiment with individual target entities, entity-level classification provides a more detailed account of the information expressed in financial news.

Although financial pretrained language models, modern pretrained encoders and parameter-efficient fine-tuning (PEFT) have advanced, their combined use for low-resource entity-level financial sentiment classification remains less explored. Existing studies often assess a particular model or adaptation strategy, leaving the interplay between pretrained encoders, entity-aware representations and fine-tuning methods less well understood. Entity-level financial datasets are also relatively small and costly to annotate [3]. These constraints make it necessary to examine how well models learn from limited data and how efficiently they can be adapted to the task [5, 6].

This paper studies low-resource entity-level financial sentiment classification on FinEntity. It compares FinBERT, DeBERTa-v3-base and ModernBERT-base using an entity-aware representation with explicit target markers, and examines full fine-tuning, a frozen encoder and LoRA as adaptation strategies. The experiments measure classification performance, learning with reduced training data and adaptation costs. They also assess variation across random seeds, the contribution of entity representation and transfer to an external dataset. These comparisons address practical choices: which encoder performs well when labels are scarce, how much of the model needs to be updated, and whether the observed performance is consistent across runs. Such choices are relevant when adapting sentiment classifiers for financial news filtering, corporate reputation monitoring and financial risk analysis.

2. Related work

2.1. Financial sentiment analysis

Financial texts use specialized terminology, and their sentiment often depends on implicit meanings and the market conditions being described. Research on market sentiment analysis examines how these cues can be extracted and interpreted from financial information [7]. A general-purpose text classifier may have difficulty with such expressions because their meaning depends on financial context. Early approaches relied mainly on sentiment lexicons, handcrafted features and classical machine learning methods. Their performance depended on the quality of the lexicons and features, which required domain knowledge to develop and could be difficult to transfer between financial settings.

DOI: 10.54254/2755-2721/2026.37226

With the emergence of pretrained language models, researchers began to further pretrain models on large-scale financial corpora to improve their understanding of financial language. Araci proposed FinBERT, a BERT model further pretrained on financial texts, and reported better performance than the original BERT on financial sentiment classification tasks [1]. Yang et al. also investigated financial domain pretraining, using financial communication data such as corporate reports and analyst conference calls. Their results showed the value of domain knowledge across multiple financial NLP tasks [2]. These studies support domain-specific pretraining as a way to improve financial text understanding. More recent studies have investigated further improvements to pretrained models and the use of large language models for financial sentiment analysis. Sun et al. incorporated financial dictionary knowledge and neutral features into pretrained language models to improve sentiment classification, particularly the recognition of neutral expressions [8]. Shen and Zhang studied large language models and FinBERT for sentiment analysis of financial news and reports, examining both domain-specific encoders and general-purpose large language models (LLMs) for financial text analysis [9].

2.2. Entity-level financial sentiment classification

Compared with traditional sentence-level sentiment classification, entity-level financial sentiment classification focuses on sentiment expressed towards particular financial entities, such as companies and institutions. It can distinguish the sentiment associated with different entities within the same text, making it useful for applications such as corporate public opinion monitoring and financial risk analysis. To support research on this task, Tang et al. released FinEntity, which annotates target entities and their sentiment polarities in financial news, and provided benchmark models for entity-level financial sentiment classification [3]. Sinha et al. developed SEntFiN 1.0 by annotating sentiment towards multiple entities in financial news headlines. Their study showed that entities mentioned in the same financial text may have different sentiment polarities, illustrating why a single label for the whole text may be insufficient [4].

LLMs have also been used in recent years for entity-level financial sentiment analysis. Muhammad and Rospocher compared the performance of large language models and discriminative models for target-level sentiment analysis of financial news headlines, examining their performance under different learning settings [10]. While these studies have advanced entity-level financial sentiment analysis, much of the work discussed here focuses on improving prediction performance, constructing benchmark datasets or applying LLMs. The relationship between entity representation methods and model adaptation strategies under limited training data remains less explored and motivates the comparison in this study.

2.3. Pretrained encoders and parameter-efficient fine tuning

Recent advances in pretrained encoders and PEFT have provided new options for natural language processing in low-resource settings. DeBERTa-v3 improves language representation through changes to its pretraining objective and representation learning strategy, achieving strong results on a range of NLP tasks [5]. ModernBERT introduces improvements to model architecture and inference efficiency, with competitive performance on text classification, information retrieval and long-context understanding tasks [11]. These models were primarily developed for general-purpose NLP, however, and their performance on low-resource entity-level financial sentiment classification requires evaluation within this specific setting.

DOI: 10.54254/2755-2721/2026.37226

LoRA, introduced by Hu et al. and reviewed by Wang et al. [6], keeps pretrained model weights fixed while learning low-rank update matrices. This reduces the number of trainable parameters and provides a parameter-efficient way to adapt a pretrained model to a downstream task. Surveys of PEFT methods have discussed approaches such as LoRA, adapters and prompt tuning, as well as how their effectiveness varies with the amount of training data, backbone model and downstream task [12]. Sentence Transformer Fine-tuning (SetFit) also demonstrated that text classification models can perform well with few labelled examples [13]. These studies provide foundations for investigating low-resource learning and parameter-efficient fine-tuning, although they address different aspects of the problem.

The studies reviewed above address financial domain pretraining, entity-level sentiment classification, encoder design and parameter-efficient fine-tuning. They provide evidence for particular models and techniques, but leave open how encoder choice, entity representation and adaptation strategy interact when few labelled financial examples are available. This study compares these choices on the same entity-level classification task. It evaluates pretrained encoders and adaptation strategies under a shared experimental setup, and uses an ablation study to assess the contribution of entity representations. The comparison is intended to clarify which choices work well under the training-data constraints examined here.

3. Methodology

3.1. Problem definition

This work studies the entity-level financial sentiment classification task defined by the FinEntity dataset [3], especially in low-resource financial news settings. The goal is to determine the sentiment expressed towards a particular company, institution or financial entity and classify it as Positive, Neutral or Negative. Sentence-level classification assigns sentiment to a whole sentence, whereas entity-level classification predicts the sentiment associated with each target entity. Thus, different entities in the same financial news article may receive different sentiment labels.

Let the input financial news text be represented as $X = \{x_1, x_2, \ldots, x_n\}$, where e denotes the target financial entity contained in the text. The objective of the model is to use the input text and target entity to predict the corresponding sentiment label:

$$y \in \{\text{Positive}, \text{Neutral}, \text{Negative}\} \tag{1}$$

In Equation (1), y represents the sentiment label associated with the target entity. The input financial news text is denoted by X, and e denotes the target financial entity contained in the text. The function f represents the entity-level sentiment classification model, which learns to predict sentiment towards a specific target entity based on both the financial news context and the entity information. Since entity-level financial sentiment datasets are expensive to construct and high-quality annotated samples are relatively limited, practical applications often face low-resource constraints. This work therefore focuses on evaluating model performance with limited training data. A common experimental framework is used to compare financial-domain pretrained models, modern general-purpose pretrained encoders and adaptation strategies in terms of classification performance, training efficiency and data efficiency. Figure 1 was prepared for this study to show the implemented workflow, from target-entity marking and pretrained encoding to the extraction and fusion of global and entity representations for sentiment classification.

DOI: 10.54254/2755-2721/2026.37226

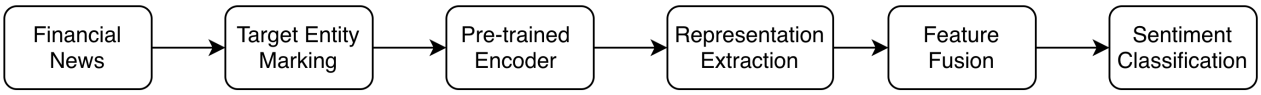

Figure 1. Overall framework of the proposed entity-level financial sentiment classification method (picture credit: original)

3.2. Input representation

Financial news articles can contain several companies, financial institutions or other economic entities that may have different sentiment polarities. When classification uses only the global semantic representation of the whole text, information associated with other entities may interfere with the prediction for the target entity. To address this problem, an explicit entity marker mechanism is used to mark the entity whose sentiment is to be predicted. This input construction is consistent with the entity-level task defined by FinEntity, where the annotated target entity is the prediction target [3].

In particular, the input text is augmented with predefined entity boundary markers, inserted before and after the target entity. Consider the following financial news text: "Apple acquired Beats for $3 billion." When predicting sentiment towards Apple, the input is represented as: "[ENT] Apple [/ENT] acquired Beats for $3 billion." When predicting sentiment towards Beats, the input becomes: "Apple acquired [ENT] Beats [/ENT] for $3 billion."

By specifying different target entities, multiple training instances can be created from the same news article. The text remains the same, while the marker positions identify the target entity for each prediction.

The marked text is then passed to a pretrained encoder to obtain both a global semantic representation of the whole news article and a local representation of the target entity. The global representation captures the overall semantic context of the news, while the entity representation captures contextual information associated with the target entity. These two representations provide complementary information and are concatenated to form the final feature representation:

$$h = [h_{cls}; h_{ent}] \tag{2}$$

In Equation (2) h_{cls} denotes the global representation produced by the encoder, h_{ent} denotes the representation of the target entity, and $[\,;\,]$ denotes concatenation. The resulting vector h combines the global and entity representations for classification.

This feature fusion is intended to strengthen the contribution of target-entity information and reduce interference from other entities in financial news. It retains the global context while giving the classifier a separate representation of the entity being evaluated. The method also keeps the classification component lightweight, without adding attention modules or complex network structures. The same input construction is used for all pretrained encoders to support a fair comparison between models.

DOI: 10.54254/2755-2721/2026.37226

3.3. Pretrained encoders

The encoder comparison includes FinBERT, DeBERTa-v3-base and ModernBERT-base. These models allow financial domain pretraining and general-purpose encoder designs to be examined on the same low-resource entity-level sentiment task. TF-IDF with Linear SVM serves as the traditional machine learning baseline.

FinBERT is further pretrained on financial corpora and is used here for its relevance to financial terminology and domain-specific expressions [1, 2]. DeBERTa-v3-base provides a general-purpose alternative, with strong reported results on natural language understanding tasks, including text classification [5]. ModernBERT-base introduces more recent changes to encoder architecture and inference efficiency [11]. Their performance is evaluated on FinEntity to establish how each model handles the financial context and the specified target entity.

The three pretrained encoders share the data preprocessing procedure, input representation and classification-head design. The experiments use a common training configuration, with the backbone encoder and adaptation strategy varied across comparisons. Holding these parts of the procedure constant limits differences arising from preprocessing or classifier design. The results therefore compare the encoders under the configuration used in this study.

TF-IDF with Linear SVM is included in the overall performance comparison to establish a reference score without pretrained neural representations. Its inclusion allows the experiments to test whether each pretrained encoder improves on a traditional classifier, rather than assuming that all three will do so.

3.4. Adaptation strategies

Three adaptation strategies are explored: Frozen Encoder, LoRA fine-tuning and Full Fine-Tuning. Their effectiveness is evaluated to understand how different parameter update strategies adapt pretrained models to entity-level financial sentiment classification with limited training data.

The Frozen Encoder strategy updates only the parameters of the classification head and keeps all pretrained encoder parameters fixed. This approach tests whether the pretrained representations can support the task without further encoder adaptation. It requires only a small number of trainable parameters, although classification performance depends on the suitability of the fixed representations. Full Fine-Tuning updates all model parameters, allowing both the encoder and classification head to adapt to the task. LoRA introduces trainable low-rank matrices into selected linear layers while leaving the original pretrained weights fixed [6]. This provides a way to adapt the encoder with a small number of additional parameters while aiming to maintain competitive classification performance.

All three adaptation strategies use the same data split, number of training epochs and optimizer, with a common hyperparameter configuration. The main distinction is how the model parameters are updated. Classification performance, training time and the number or percentage of trainable parameters are compared to assess the practical value of each strategy. Trainable parameter count and training time are considered separately, because a reduction in trainable parameters does not necessarily lead to a shorter training time.

3.5. Classification objective

The fused feature representation is passed to a fully connected classification head to predict the probability of sentiment towards the target entity belonging to one of three categories: *Positive*,

DOI: 10.54254/2755-2721/2026.37226

$Neutral$ or $Negative$. Let the fused representation be denoted by h.The output of the classifier is given by:

$$p = Softmax\left(Wh + b\right) \tag{3}$$

In Equation (3), W and b represent the weight matrix and bias vector of the classification head, respectively. The vector h is the fused feature representation, and Softmax converts the classifier output into the probability distribution p over the three sentiment categories.

The model is trained using cross-entropy loss with the AdamW optimizer. Considering potential class imbalance in entity-level financial sentiment classification, Macro-F1 is used as the main evaluation metric. Accuracy is also reported to describe overall classification performance. Training time and the number or percentage of trainable parameters are used to evaluate efficiency. Together, these metrics allow the models to be compared in terms of classification performance and adaptation cost under low-resource conditions.

4. Experiments

4.1. Experimental setup

FinEntity was used as the main benchmark, and SEntFiN 1.0 was used for external evaluation of entity-level financial sentiment classification [3, 4]. The models included TF-IDF + Linear SVM as the traditional machine learning baseline, FinBERT as the financial-domain pretrained model, and DeBERTa-v3-base and ModernBERT-base as general-purpose pretrained encoders.

Three adaptation strategies were assessed: Full Fine-Tuning, Frozen Encoder and LoRA fine-tuning. For LoRA, low-rank adaptation was applied to the Query, Key and Value projections in the attention layers . The experiments followed a common data split and training procedure, with matched hyperparameter settings for the comparisons. The best checkpoint was selected using Macro-F1 on the validation set.

Accuracy and Macro-F1 were used to evaluate classification performance, with Macro-F1 as the primary metric. Training time and the percentage of trainable parameters were also reported to assess adaptation cost. The following sections examine overall performance, parameter-efficient fine-tuning, low-resource learning, robustness across random seeds, entity-aware representation and cross-dataset generalization.

4.2. Overall performance comparison

The traditional machine learning baseline and three pretrained encoders were evaluated on the FinEntity test set. Table 1 presents their Accuracy and Macro-F1 scores, providing an overall comparison before examining individual adaptation strategies and training budgets.

Table 1. Overall performance comparison on FinEntity test set

Model	Accuracy	Macro-F1
TF-IDF + Linear SVM	0.7500	0.7305
FinBERT	0.9031	0.9000
DeBERTa-v3-base	0.9235	0.9216

DOI: 10.54254/2755-2721/2026.37226

Table 1. (continued)

ModernBERT-base	0.6633	0.6009

DeBERTa-v3-base achieved the highest scores in Table 1, with a Macro-F1 of 0.9216 and an Accuracy of 0.9235. FinBERT followed with a Macro-F1 of 0.9000. Both models outperformed TF-IDF + Linear SVM. The Accuracy scores followed the same ranking: FinBERT reached 0.9031, the SVM baseline reached 0.7500, and ModernBERT-base reached 0.6633. The two metrics therefore gave a consistent overall comparison. The result for DeBERTa-v3-base is consistent with its strong language understanding performance reported in previous work [5]. FinBERT's performance also supports its usefulness for financial text classification, in line with earlier financial domain pretraining studies [1, 2].

TF-IDF + Linear SVM reached a Macro-F1 of 0.7305, below FinBERT and DeBERTa-v3-base but above ModernBERT-base. ModernBERT-base had the lowest Macro-F1 in this comparison, at 0.6009. Its general-purpose pretraining [11] may not provide the same financial-domain advantages as FinBERT. However, pretraining domain alone cannot explain the ranking, since DeBERTa-v3-base is also a general-purpose encoder. The result establishes weaker performance for ModernBERT-base under the tested configuration, rather than identifying a single cause.

Given the stronger results of FinBERT and DeBERTa-v3-base, the subsequent parameter-efficiency table, multi-seed analysis, representation ablation and external evaluation focus on these two models. ModernBERT-base remains included in the low-resource comparison.

4.3. Analysis of parameter-efficient fine-tuning

Full Fine-Tuning, Frozen Encoder and LoRA fine-tuning were compared to examine classification performance alongside the proportion of parameters updated. Table 2 reports the results for FinBERT and DeBERTa-v3-base. Figure 2 was plotted from the experiments conducted in this study to compare classification performance with the percentage of trainable parameters.

Table 2. Performance and parameter efficiency of different adaptation strategies

Model	Adaptation	Macro-F1	Trainable (%)	Training Time (s)
FinBERT	Full Fine-Tuning	0.9000	100.000	118.98
FinBERT	Frozen Encoder	0.6448	0.004	61.47
FinBERT	LoRA	0.9040	0.408	187.79
DeBERTa-v3-base	Full Fine-Tuning	0.9216	100.000	173.83
DeBERTa-v3-base	Frozen Encoder	0.4342	0.003	89.36
DeBERTa-v3-base	LoRA	0.9267	0.243	325.91

DOI: 10.54254/2755-2721/2026.37226

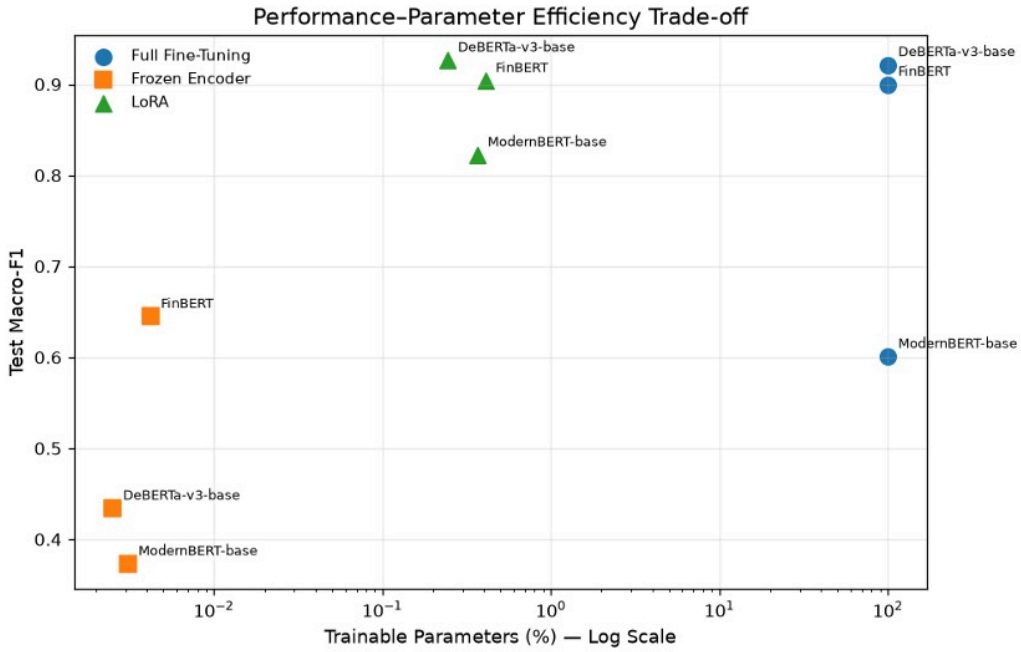

Figure 2. Performance–parameter efficiency trade-off of different adaptation strategies (picture credit: original)

Full Fine-Tuning achieved strong classification performance while updating all model parameters. The Frozen Encoder strategy updated only the classification head, but its Macro-F1 fell to 0.6448 for FinBERT and 0.4342 for DeBERTa-v3-base. The trainable proportions for the frozen models were just 0.004% for FinBERT and 0.003% for DeBERTa-v3-base. This was the most restrictive update strategy, but its parameter savings came with a clear loss in classification performance. These results indicate that the fixed pretrained representations were less effective for this task than representations adapted during training.

LoRA updated only 0.408% of the parameters for FinBERT and 0.243% for DeBERTa-v3-base, while reaching Macro-F1 scores of 0.9040 and 0.9267, respectively. Both scores were slightly above the corresponding Full Fine-Tuning results in Table 2. This supports the use of LoRA as a parameter-efficient alternative in the tested setting and is consistent with previous work showing that competitive downstream performance can be obtained by adapting a small fraction of model parameters [6, 12, 14].

Overall, LoRA maintained performance comparable to Full Fine-Tuning with substantially fewer trainable parameters. The small score differences in this comparison should not be taken as evidence that LoRA is consistently superior across runs.

Reducing the number of trainable parameters did not reduce training time in these experiments. LoRA took 187.79 seconds for FinBERT and 325.91 seconds for DeBERTa-v3-base, compared with 118.98 and 173.83 seconds for Full Fine-Tuning. Although LoRA optimized fewer parameters, the full encoder forward computation was still required, and the low-rank modules introduced additional operations. Trainable parameter count therefore did not directly correspond to wall-clock training time. This study uses the number and percentage of trainable parameters to assess parameter efficiency, with training time reported as a separate efficiency measure.

DOI: 10.54254/2755-2721/2026.37226

4.4. Low-resource learning performance

Experiments used 10% (171 samples), 25% (429 samples) and 50% (858 samples) of the original training set to assess learning with limited labelled data. Table 3 reports the Macro-F1 scores obtained in our low-resource experiments. These results were plotted in Figure 3 to show how each model and adaptation strategy performed at the three training budgets.

Table 3. Different low-resource training budgets and Macro-F1

Model	Strategy	10% (171)	25% (429)	50% (858)
FinBERT	Full Fine-Tuning	0.8041	0.8570	0.8578
FinBERT	LoRA	0.7895	0.8618	0.8665
DeBERTa-v3-base	Full Fine-Tuning	0.6594	0.8446	0.9287
DeBERTa-v3-base	LoRA	0.4148	0.6243	0.8793
ModernBERT-base	Full Fine-Tuning	0.3949	0.5068	0.5664
ModernBERT-base	LoRA	0.4502	0.5560	0.7498

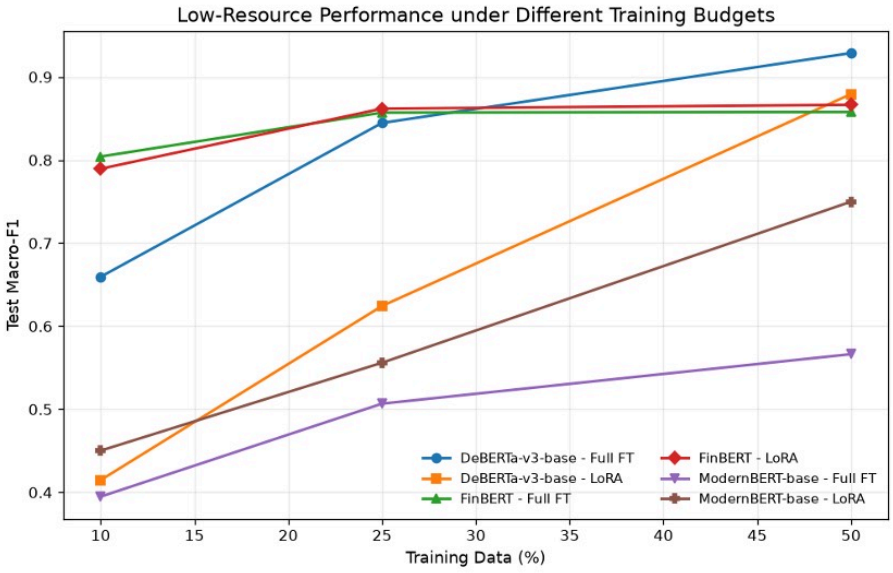

Figure 3. Macro-F1 under different low-resource training budgets (picture credit: original)

The models differed in their sensitivity to training data size. With only 10% of the training data, FinBERT performed best, reaching Macro-F1 scores of 0.8041 with Full Fine-Tuning and 0.7895 with LoRA. Both were higher than the corresponding DeBERTa-v3-base scores. This suggests an advantage for financial-domain pretraining at the smallest budget tested. As more training data became available, DeBERTa-v3-base improved more rapidly. Its Full Fine-Tuning model reached a Macro-F1 of 0.9287 with 50% of the training data, surpassing FinBERT.

Figure 3 shows that FinBERT was less sensitive to training budget than DeBERTa-v3-base. ModernBERT-base generally performed below the other two encoders, although its LoRA model outperformed DeBERTa-v3-base with LoRA at the 10% training budget. Its LoRA Macro-F1 improved from 0.4502 to 0.7498 as the training proportion increased from 10% to 50%. Additional data therefore benefited both general-purpose encoders, although their final performance differed.

DOI: 10.54254/2755-2721/2026.37226

These results suggest that the benefit of financial-domain pretraining is particularly useful with very limited labelled data, while some general-purpose encoders benefit more from larger training budgets. This pattern is relevant to previous research on financial pretraining and few-shot learning [1, 2, 13, 15], but it does not establish a universal ranking between domain-specific and general-purpose models.

The effect of LoRA also depended on the encoder. At 10%, LoRA performed below Full Fine-Tuning for FinBERT and DeBERTa-v3-base, whereas ModernBERT-base performed better with LoRA at all three budgets. For DeBERTa-v3-base, the gap between LoRA and Full Fine-Tuning narrowed as more data became available, although LoRA remained lower at 50%. FinBERT's LoRA model slightly exceeded Full Fine-Tuning at both 25% and 50%. The results therefore point to an interaction between training budget, encoder choice and adaptation strategy.

4.5. Multi-seed robustness analysis

To assess variation across runs, the LoRA versions of FinBERT and DeBERTa-v3-base were trained with three random seeds: 42, 123 and 2026. Table 4 summarizes the mean scores and standard deviations (SD) from these runs. Figure 4 was plotted from the individual Macro-F1 scores obtained with seeds 42, 123 and 2026, allowing the variation behind the summary statistics to be examined.

Table 4. Multi-seed robustness of LoRA models

Model	Accuracy (Mean ± SD)	Macro-F1 (Mean ± SD)
DeBERTa-v3-base	0.9031 ± 0.0442	0.8995 ± 0.0488
FinBERT	0.9048 ± 0.0078	0.9030 ± 0.0083

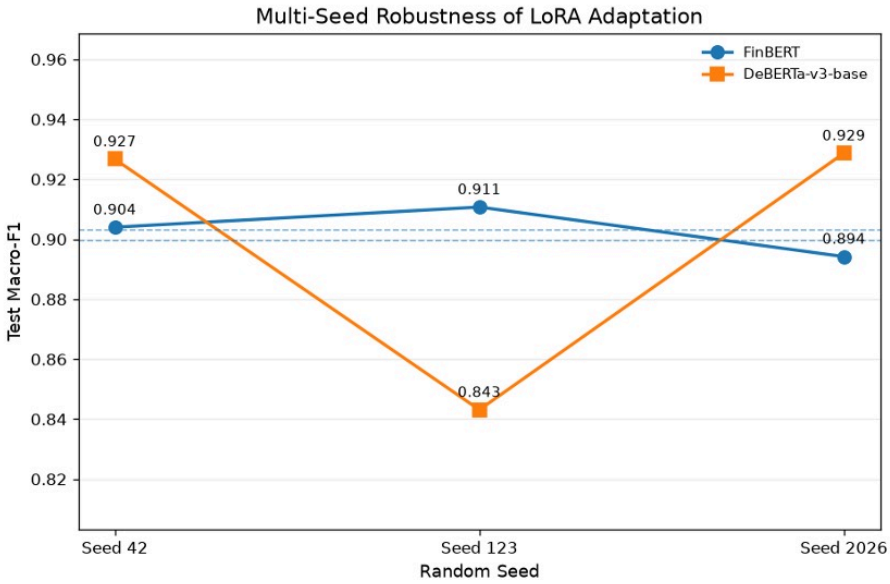

Figure 4. Macro-F1 across different random seeds for LoRA models (picture credit: original)

FinBERT achieved a mean Macro-F1 of 0.9030 with a standard deviation of 0.0083. DeBERTa-v3-base obtained a similar mean of 0.8995 but a larger standard deviation of 0.0488. The standard deviations describe the spread of the observed scores and should be considered alongside the means

DOI: 10.54254/2755-2721/2026.37226

when interpreting this comparison. The main distinction was therefore stability across seeds, rather than a substantial difference in average performance.

Figure 4 shows that DeBERTa-v3-base performed better under some seeds but varied more across repeated runs. FinBERT produced more consistent classification results across the three experiments. This helps qualify the single-run comparison: a higher score in one run did not translate into a higher mean across the seeds tested.

Taken together, Table 4 and Figure 4 suggest that FinBERT was more stable in these runs, while DeBERTa-v3-base achieved higher scores under some initializations. Repeated experiments provide a fuller view than a single score, which is relevant to the evaluation concerns discussed in low-resource learning research [13, 15]. FinBERT may therefore be preferable when consistency across runs matters, whereas DeBERTa-v3-base remains competitive when considering its best observed performance. These conclusions are limited to the three seeds evaluated.

4.6. Ablation study of entity-aware representation

An ablation experiment assessed the contribution of target-entity information by comparing two settings: Global-only, which uses the global text representation, and Global + Entity, which combines global and target-entity representations. Table 5 reports the results of this ablation experiment. Figure 5 was constructed from the same measurements to compare Macro-F1 with and without the additional target-entity representation.

Table 5. Effect of entity-aware representation on FinEntity

Model	Representation	Accuracy	Macro-F1	Δ Macro-F1
FinBERT	Global-only	0.7908	0.7791	–
FinBERT	Global + Entity	0.9031	0.9000	+0.1209
DeBERTa-v3-base	Global-only	0.8265	0.8102	–
DeBERTa-v3-base	Global + Entity	0.9235	0.9216	+0.1114

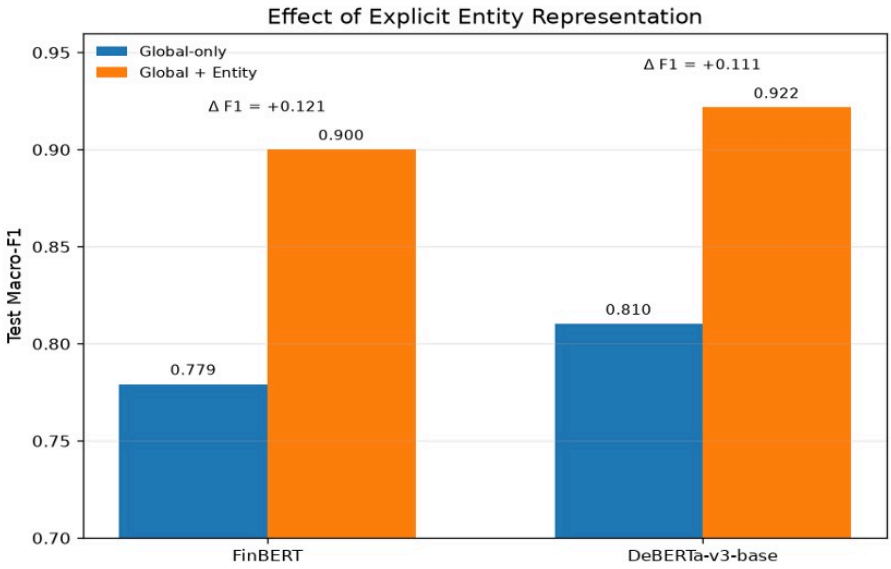

Figure 5. Effect of entity-aware representation on Macro-F1 (picture credit: original)

DOI: 10.54254/2755-2721/2026.37226

Adding the target-entity representation improved both models. For FinBERT, Macro-F1 increased from 0.7791 to 0.9000, an improvement of 0.1209. DeBERTa-v3-base improved from 0.8102 to 0.9216, a gain of 0.1114. The direction of the improvement was the same for both encoders, although the magnitude differed. The gain was also reflected in Accuracy, rather than being confined to Macro-F1. Accuracy increased by approximately 11.23 and 9.70 percentage points, respectively.

As shown in Figure 5, combining global and entity representations was more effective than using the global representation alone for both encoders. The result supports supplying the classifier with target-specific information alongside the overall news context. It is consistent with the broader motivation for entity-aware modelling in FinEntity [3]. The comparison evaluates the combined representation, however, and does not isolate the contribution of entity markers alone.

4.7. Cross-dataset generalisation

Cross-dataset generalization was assessed by directly evaluating the LoRA models trained on FinEntity on SEntFiN 1.0, without additional fine-tuning on the target dataset. Table 6 presents the results of our external evaluation. Figure 6 combines these scores with the FinEntity LoRA results in Table 2 to show the change in Macro-F1 when the trained models were evaluated on SEntFiN without further fine-tuning.

Table 6. Cross-dataset external evaluation results on SEntFiN 1.0

Model	Accuracy	Macro-F1	Negative F1	Neutral F1	Positive F1
FinBERT	0.7065	0.7062	0.7304	0.7145	0.6736
DeBERTa-v3-base	0.7188	0.7222	0.7641	0.7175	0.6850

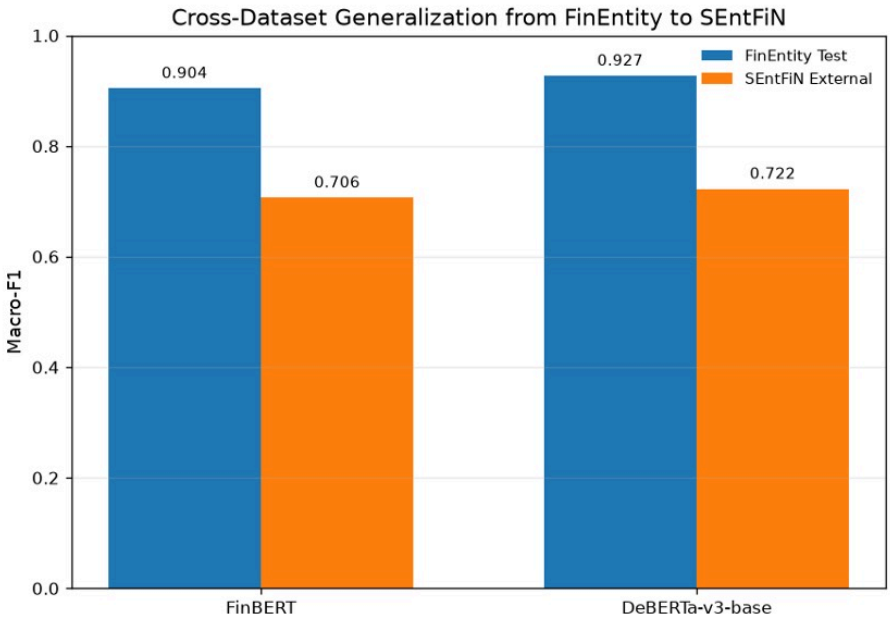

Figure 6. Cross-dataset Macro-F1 comparison between FinEntity and SEntFiN (picture credit: original)

On SEntFiN, DeBERTa-v3-base achieved an Accuracy of 0.7188 and a Macro-F1 of 0.7222. FinBERT obtained 0.7065 and 0.7062, respectively. Both models had their highest class-level F1 on

DOI: 10.54254/2755-2721/2026.37226

Negative and their lowest on Positive. DeBERTa-v3-base outperformed FinBERT in all three sentiment categories in this evaluation.

Compared with FinEntity, FinBERT's Macro-F1 decreased from 0.9040 to 0.7062. This is an absolute decrease of 0.1978, or a relative decrease of approximately 21.88%. DeBERTa-v3-base fell from 0.9267 to 0.7222, an absolute decrease of 0.2045 and a relative decrease of approximately 22.07%. Both models therefore lost performance when transferred to the external dataset. Nevertheless, their Macro-F1 scores remained above 0.70 without target-dataset fine-tuning, suggesting that some of the learned entity-level sentiment information transferred across datasets.

The decline also shows that performance on FinEntity did not carry over unchanged to SEntFiN. Differences between the datasets may contribute to this loss, although the present comparison does not isolate the sources of the difference. Future work could examine domain adaptation or other parameter-efficient fine-tuning methods to improve cross-dataset performance.

5. Conclusion

This study examined low-resource entity-level financial sentiment classification, using FinEntity as the main benchmark. The classification framework combined explicit entity markers with global and target-entity representations and was evaluated with different model adaptation strategies. A traditional machine learning baseline, financial-domain pretraining and modern general-purpose encoders were compared under the same experimental setup. The experiments covered overall classification performance, parameter-efficient fine-tuning, reduced training budgets, variation across random seeds, the contribution of entity representation and transfer to an external dataset.

In the overall comparison, FinBERT and DeBERTa-v3-base outperformed TF-IDF + Linear SVM, while ModernBERT-base performed less well. DeBERTa-v3-base obtained the highest score in that comparison. At the smallest training budget, however, FinBERT performed better and also produced more consistent results across the tested random seeds. The choice of encoder therefore depends on the available labelled data and the importance of consistency across runs. LoRA achieved performance comparable to Full Fine-Tuning while updating only a small proportion of model parameters. Training took longer in the reported experiments despite this reduction, so fewer trainable parameters should not be taken to imply faster training. For both encoders included in the ablation study, combining global and target-entity representations improved classification performance. Both LoRA models also achieved Macro-F1 scores above 0.70 on SEntFiN without further fine-tuning, although their scores were lower than on FinEntity.

The evaluation was mainly based on FinEntity, with SEntFiN providing the only external dataset. Testing on additional financial datasets would help establish whether the performance differences and transfer patterns observed here also occur in other settings. The comparison of adaptation strategies was limited to LoRA, Frozen Encoder and Full Fine-Tuning. Including adapters and prompt tuning would allow a wider assessment of classification performance and parameter efficiency when training data are scarce. Entity information was represented through explicit markers and feature concatenation. Future work could examine more detailed representations of the relationships between target entities and financial events. Knowledge-enhanced approaches could also be investigated for financial terminology and expressions whose interpretation depends on complex financial contexts. These extensions would broaden the evaluation and clarify the conditions under which parameter-efficient adaptation is useful for entity-level financial sentiment classification.

DOI: 10.54254/2755-2721/2026.37226

References

[1] Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*. https://arxiv.org/abs/1908.10063

[2] Yang, Y., Uy, M. C. S., & Huang, A. (2020). FinBERT: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*. https://arxiv.org/abs/2006.08097

[3] Tang, Y., Yang, Y., Huang, A., Tam, A., & Tang, J. (2023). FinEntity: Entity-level sentiment classification for financial texts. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 15465–15471. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.956

[4] Sinha, A., Kedas, S., Kumar, R., & Malo, P. (2022). SEntFiN 1.0: Entity-aware sentiment analysis for financial news. *Journal of the Association for Information Science and Technology, 73*(9), 1314–1335. https://doi.org/10.1002/asi.24634

[5] He, P., Gao, J., & Chen, W. (2023). DeBERTaV3: Improving DeBERTa using ELECTRA-style pre-training with gradient-disentangled embedding sharing. In *The Eleventh International Conference on Learning Representations (ICLR 2023)*. https://openreview.net/forum?id=sE7-XhLxHA

[6] Wang, L., Chen, S., Jiang, L., Pan, S., Cai, R., Yang, S., & Yang, F. (2025). Parameter-efficient fine-tuning in large language models: A survey of methodologies. *Artificial Intelligence Review, 58*, Article 227. https://doi.org/10.1007/s10462-025-11236-4

[7] Heydarian, P., Bifet, A., & Corbet, S. (2025). Understanding market sentiment analysis: A survey. *Journal of Economic Surveys, 39*(3), 1125–1147. https://doi.org/10.1111/joes.12645

[8] Sun, Y., Yuan, H., & Xu, F. (2025). Financial sentiment analysis for pre-trained language models incorporating dictionary knowledge and neutral features. *Natural Language Processing Journal, 11*, 100148. https://doi.org/10.1016/j.nlp.2025.100148

[9] Shen, Y., & Zhang, P. K. (2024). Financial sentiment analysis on news and reports using large language models and FinBERT. *arXiv preprint arXiv:2410.01987*. https://arxiv.org/abs/2410.01987

[10] Muhammad, I., & Rospocher, M. (2025). On assessing the performance of LLMs for target-level sentiment analysis in financial news headlines. *Algorithms, 18*(1), 46. https://doi.org/10.3390/a18010046

[11] Warner, B., Chaffin, A., Clavié, B., Weller, O., Hallström, O., Taghadouini, S., Gallagher, A., Biswas, R., Ladhak, F., Aarsen, T., Adams, G. T., Howard, J., & Poli, I. (2025). Smarter, better, faster, longer: A modern bidirectional encoder for fast, memory efficient, and long context finetuning and inference. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2526–2547. Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.acl-long.127

[12] Han, Z., Gao, C., Liu, J., Zhang, J., & Zhang, S. Q. (2024). Parameter-efficient fine-tuning for large models: A comprehensive survey. *Transactions on Machine Learning Research*. https://openreview.net/forum?id=lIsCS8b6zj

[13] Tunstall, L., Reimers, N., Jo, U. E. S., Bates, L., Korat, D., Wasserblat, M., & Pereg, O. (2022). Efficient few-shot learning without prompts. arXiv preprint arXiv:2209.11055. https://arxiv.org/abs/2209.11055

[14] Balne, C. C. S., Bhaduri, S., Roy, T., Jain, V., & Chadha, A. (2024). Parameter efficient fine tuning: A comprehensive analysis across applications. arXiv preprint arXiv:2404.13506. https://arxiv.org/abs/2404.13506

[15] Wang, Y., Yao, Q., Kwok, J. T., & Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys, 53*(3), 63:1–63:34. https://doi.org/10.1145/3386252