Journal of Theoretical and Applied Information Technology
15th April 2026. Vol.104. No.7

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DOMAIN-INVARIANT REPRESENTATION LEARNING FOR GENERALIZABLE TEXT MINING ACROSS MULTIPLE DOMAINS

SRINADH SWAMY MAJETI[1], BOMMA RAMAKRISHNA[2], PEDDADA NAGAMANI[3], MAREPALLI RADHA[4], CH. V. RAVI SANKAR[5], S. JAYANTH[6], LAKSHMI PRASANNA BYRAPUNENI[7], GUNTI SURENDRA[8], MEDIKONDA ASHA KIRAN[9], MANYAM THAILE[10], RAMESH BABU PITTALA[*]

[1, 3, 6, 7, 9, 10, *] School of Engineering, Anurag University, Hyderabad, India.

[2]Professor, Dept of AI & ML, Swarnandhra College of Engineering and Technology, Narsapur, India.
[4]Department of Computer science and Engineering, CVR College of Engineering, Hyderabad.
[5]Department of ECE, Aditya University, Surampalem, AP, India.
[8]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Greenfield, Vaddeswaram, Tadepalli, Guntur, AP, India.

E-mail: [*]prameshbabu526@gmail.com, [1]Srinadh.ai@anurag.edu.in, [2]drbrk8789@gmail.com, [3]nagamanikoli@gmail.com, [4]marepalli.radha@gmail.com, [5]venkataravisankarch@adityauniversity.in, [6]setpally.jayanthreddy@gmail.com, [7]lakshmiprasanna.byrapuneni@gmail.com, [8]guntisurendra@gmail.com, [9]ashakiran2@gmail.com, [10]manyamthaile@gmail.com

ABSTRACT

Text mining models often struggle to generalize across domains due to domain-specific linguistic and contextual biases. This limitation significantly affects real-world deployment where data distributions continuously shift. In this work, we hypothesize that explicitly learning domain-invariant representations can substantially improve cross-domain robustness without requiring target-domain fine-tuning. To address this, we propose a Domain-Invariant Representation Learning (DIRL) framework that integrates shared–private representation decomposition, adversarial domain confusion, and contrastive semantic alignment within a transformer-based architecture. The proposed method is evaluated on Amazon Reviews, MDSD, and 20 Newsgroups under cross-domain and unseen-domain generalization settings. Experimental results demonstrate consistent improvements of 4–9% in Macro-F1 over strong baselines, along with a significant reduction in domain generalization gap. These findings confirm that enforcing domain invariance at the representation level enhances scalability, robustness, and real-world applicability of text mining systems.

Keywords: *Domain-Invariant Representation Learning; Text Mining; Cross-Domain Generalization; Multi-Domain Learning; Transfer Learning; Transformer-Based Models; Adversarial Learning; Contrastive Representation Learning; Document Classification; Sentiment Analysis*

1. INTRODUCTION

Text mining: from individuals to communities. The rise of social media has opened the door to large quantities of unstructured textual data, and text mining methods have become an essential tool for distilling useful knowledge from them on a massive scale, supporting many applications, including document classification, sentiment analysis, opinion mining, and information retrieval. Recent progress in deep learning and transformer-based language models has made great advances to learn text representation by exploiting rich context and semantic information in large-scale text corpora [1]-[3]. Despite these advancements, most text mining models remain vulnerable to domain shift, with performance deteriorating substantially when evaluated on previously unseen domains.

This issue is primarily due to domain-specific biases in tabular representations of text, such as differences in vocabulary level, writing style, or

topic distribution [4]-[6]. The conventional supervised learning methods are based on the assumption that the training and testing data come from the same (but unknown) distribution, which previous work has shown is typically not true. Thus, models learned across a single or few domains frequently do not perform well on unseen or heterogeneous domains [7].

To address this problem, cross-domain transfer learning and domain adaptation methods have been the focus of much research effort in text mining. Typically, these methods fine-tune pretrained language models on target-domain data or align the distributions of source and target features by adversarial learning [8], [9] or discrepancy minimization [10]. Although the introduced methods have demonstrated encouraging results, they typically need access to labeled and/or unlabeled data in NID under training. In addition, many existing methods are designed for a single-source, single-target setting, which can only consider one domain and one task at a time, and therefore do not adapt to multi-domain scenarios with unknown new domains [11-[13].

Recent work [14] has shown that the key to achieving real generalization in text mining is to learn a domain-invariant representation extensively, rather than merely retraining models as is often done between domains. Domain-invariant representation learning aims to preserve task-relevant semantic information while explicitly attenuating domain-specific features, thereby achieving robust performance across multiple novel domains. Yet, it remains challenging to achieve such invariance for text data due to the high dimensionality of language representations and the intricate relationship between content and style [15]-[16].

There are powerful transformer-based (e.g., BERT and its variants) architectures for representation learning, but these models directly sample from the training distribution and consequently are not domain-invariant; they tend to encode domain-specific signals that exist in their respective training data [17], [18]. Domain invariance has been enforced explicitly in recent methods with adversarial objectives, contrastive learning, and shared–private representation decomposition [31],[22]. However, not only are existing methods rarely systematically evaluated under unseen-domain generalization settings, but also the obtained performance is unstable across domains.

Inspired by these issues, this paper presents a Domain-Invariant Representation Learning (DIRL) approach for text mining across domains. The introduced framework explicitly incorporates domain-invariance constraints into transformer-based representation learning models through the synergy of shared–private feature modeling, adversarial domain confusion, and contrastive semantic alignment. Unlike conventional domain adaptation methods, which are only trained to adapt a model from the source dataset to a target one, we can train on different combinations of text pairings and test on new pairs without requiring fine-tuning for each particular task.

Extensive experiments on standard multi-domain datasets demonstrate that DIRL significantly and consistently outperforms state-of-the-art transfer learning and domain adaptation methods in terms of accuracy, macro-F1 score, and robustness across domains. The results clearly demonstrate that learning domain-invariant representations is essential to the scalability and robustness of text mining systems in practical, dynamic scenarios.

Research Hypothesis: We hypothesize that explicitly disentangling domain-invariant and domain-specific features during representation learning leads to improved cross-domain generalization compared to conventional fine-tuning and domain adaptation approaches.

In real-world applications such as sentiment monitoring, social media analytics, and enterprise document processing, models must operate reliably across evolving domains without frequent retraining. Therefore, scalable domain-invariant learning methods are critical for sustainable AI deployment.

1.1 Contributions of This Work

The main contributions of the paper are outlined as follows:

- For this purpose, we present a novel domain-invariant representation learning framework that directly discourages task-irrelevant domain-specific biases.
- We introduce adversarial learning and contrastive alignment methods in a transformer architecture for better cross-domain generalization.
- We then perform thorough multi-domain and unseen-domain evaluations on benchmark text mining datasets, showing our method's clear superiority to existing methods.

- We present ablation and robustness studies to show the effectiveness of each part in our framework.

2. RELATED WORK

Recently, with the development of domain generalization for text mining, attempts to decrease dependence on explicit domain labels and to enhance robustness to unseen-domain data have become popular. End-to-end meta-learning: A few works have studied end-to-end meta-learning, where a model is trained to learn domain-invariant parameters that generalize well across different distributions of the same task [23]. Despite the appeal of meta-learning in improving adaptability, it is generally limited by its computational costs and dependence on domain sampling approaches. A second popular category is shared–private representation learning, which separates the representations of text into two parts: one shared across domains and the other private to each source domain. Such a strategy enables the models to learn typical semantic knowledge, thereby separating domain-dependent features [24]. However, shared and private spaces are not always perfectly separated, causing residual domain-specific leakage to the shared representations, which may compromise their ability to generalize.

Adversarial learning approaches have also been embraced to promote domain invariance for text representations [25]. Adversarial training learns representations that are invariant across domains by adding a domain discriminator, which competes with the feature encoder [26]. As a result of their success, adversarial paradigms inevitably exhibit inherent inertia and can encounter intractable training issues specifically instability and mode collapse in multiple homogeneous domain settings with strong heterogeneity in data distributions [27].

To address these issues, contrastive learning has recently gained attention in cross-domain text mining. These approaches utilize semantic similarity information to align learned representations, where semantically similar transfer example pairs are pulled together, and dissimilar ones are pushed away from each other in the latent space [28]. Contrastive objectives have been shown to exhibit better robustness than pure adversarial methods; however, their performance highly relies on the constructed positive–negative pairs, and sometimes these kinds of algorithms require sophisticated sampling strategies as well [29].

Hybrid models that combine adversarial domain confusion with contrastive semantic alignment have shown encouraging progress toward learning stable, domain-invariant representations [30],[31]. These techniques leverage the complementary properties of both learning schemes, resulting in enhanced convergence and cross-domain performance. However, most previous hybrid approaches are still tailored to specific tasks or datasets and lack comprehensive testing on various unseen domains.

Moreover, recent work has also been directed towards realistic deployment conditions, where no target-domain data is seen during training. Several previous works demonstrate good performance under controlled adaptation but are unable to generalize to truly novel domains. This discrepancy motivates the development of scalable approaches that enforce invariance directly within representation learning, rather than relying on post hoc adaptation.

To recap, the available works suggest that, despite substantial progress in transfer learning, domain adaptation, and domain generalization for text mining, producing robust, domain-invariant representations across many diverse, unseen domains remains an open problem. Existing approaches either suffer from a lack of stability, difficulty in scaling, or inadequate evaluation in challenging generalization settings. These shortcomings motivate the present work, which aims to explicitly learn domain-invariant representations by unifying adversarial learning, contrastive alignment, and multi-domain training for effective generalization in practical text mining scenarios.

3. MATERIALS AND METHODS

This section presents the proposed Domain-Invariant Representation Learning (DIRL) framework designed to achieve robust and generalizable text mining across multiple domains. The overall objective of the framework is to learn task-discriminative yet domain-agnostic textual representations that remain stable under domain shifts, including unseen-domain scenarios.

3.1 Problem Formulation

Let $D = \{D_1, D_2, ..., D_n\}$ denote multiple source domains. Each domain contains labeled samples (x, y). The objective is to learn a function f(x) that minimizes classification loss while ensuring representation invariance across domains. The challenge is to suppress domain-specific bias while preserving task-relevant semantic features.

3.2 Overall Framework

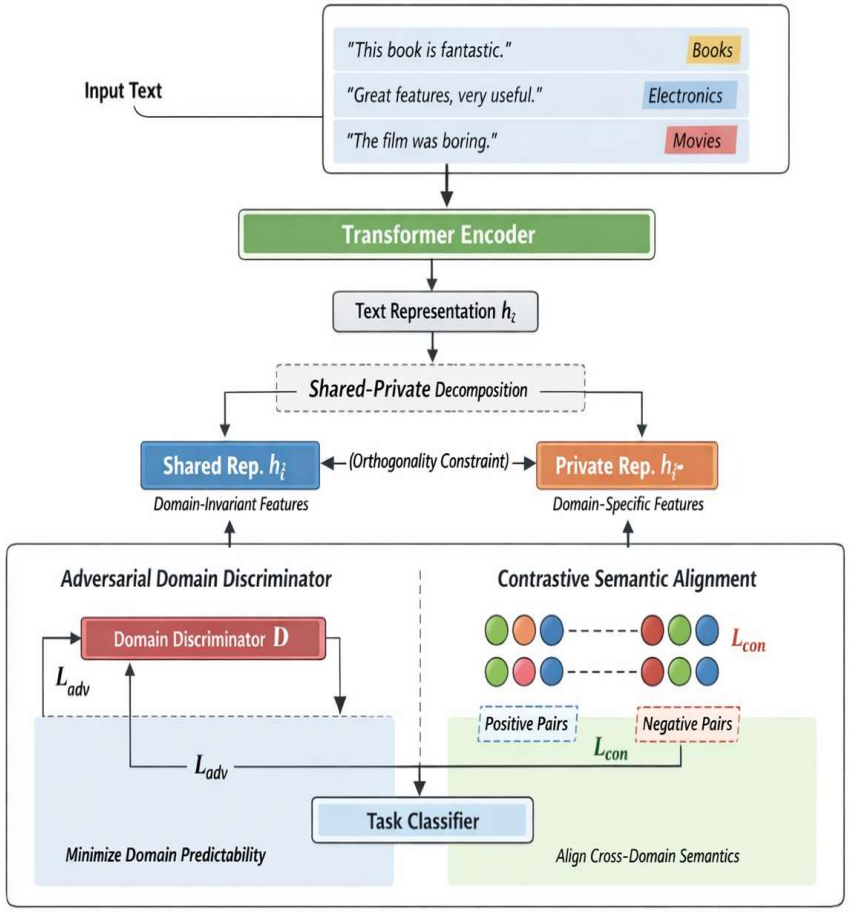

Figure 1: Overview of the proposed Domain-Invariant Representation Learning (DIRL) framework for cross-domain text mining.

The proposed DIRL framework consists of four major components:

1. a transformer-based text encoder,
2. a shared–private representation decomposition module,
3. an adversarial domain-invariance learning module, and
4. a contrastive semantic alignment module.

Given an input text sample x_i from domain d_i, the transformer encoder generates a contextualized embedding $\mathbf{h}_i$. This embedding is then decomposed into a **shared representation** $\mathbf{h}_i^s$, which captures domain-invariant semantics, and a **private representation** $\mathbf{h}_i^p$, which encodes domain-specific features. Only the shared representation is used for downstream text mining tasks, thereby enforcing invariance at the representation level.

3.3 Transformer-Based Text Encoder

The encoder is built upon a pretrained transformer architecture, which maps an input text sequence into a sequence of contextual embeddings. Given a tokenized input sequence $x = \{w_1, w_2, \ldots, w_n\}$, the encoder produces a pooled sentence representation:

$$\mathbf{h} = \text{Encoder}(x)$$

This representation captures rich semantic and contextual information. However, without additional constraints, it may still encode domain-specific biases. Therefore, additional mechanisms are introduced to enforce domain invariance.

3.4 Shared–Private Representation Decomposition

To explicitly separate domain-invariant and domain-specific information, the encoded representation **h**is passed through two parallel projection networks:

$$\mathbf{h}^s = f_s(\mathbf{h}), \mathbf{h}^p = f_p(\mathbf{h})$$

where $f_s(\cdot)$and $f_p(\cdot)$denote the shared and private projection functions, respectively. Orthogonality constraints are applied to minimize information

overlap between shared and private representations, ensuring effective disentanglement.

3.5 Adversarial Domain-Invariance Learning

To suppress domain-specific signals in the shared representation, an adversarial domain discriminator is employed. The discriminator attempts to predict the domain label d_i from $\mathbf{h}_i^s$, while the encoder seeks to prevent accurate domain prediction. This adversarial objective is formulated as:

$$\mathcal{L}_{adv} = \min_{E} \max_{D} \sum_{i} \log D(d_i \mid \mathbf{h}_i^s)$$

This minimax optimization encourages the shared representation to be indistinguishable across domains, thereby promoting domain invariance.

3.6 Contrastive Semantic Alignment

To further enhance cross-domain alignment, a contrastive learning objective is introduced. Semantically similar samples from different domains are treated as positive pairs, while dissimilar samples are treated as negative pairs. The contrastive loss is defined as:

$$\mathcal{L}_{con} = -\log \frac{\exp(\text{sim}(\mathbf{h}_i^s, \mathbf{h}_j^s)/\tau)}{\sum_k \exp(\text{sim}(\mathbf{h}_i^s, \mathbf{h}_k^s)/\tau)}$$

where $\text{sim}(\cdot)$ denotes cosine similarity and τ is a temperature parameter. This objective ensures semantic consistency across domains.

3.7 Overall Optimization Objective

The final training objective combines task supervision, adversarial learning, and contrastive alignment:

$$\mathcal{L} = \mathcal{L}_{task} + \lambda_{adv}\mathcal{L}_{adv} + \lambda_{con}\mathcal{L}_{con}$$

where λ_{adv} and λ_{con} control the contribution of each component.

4. DATASETS AND EXPERIMENTAL SETUP

This section discusses the datasets, experimental settings, implementation details and the evaluation criterion employed to validate the effectiveness of our proposed Domain-Invariant Representation Learning (DIRL) framework. Our experimental setting is built upon common practices in the field of domain generalization. To ensure a fair comparison, the results can be reproduced across different experiments, and a valid evaluation is performed in terms of providing cross-domain and unseen-domain performance assessments. We evaluate our approach on three state-of-the-art multi-domain text mining benchmarks, where substantial domain shifts exist in vocabulary use, writing style, and semantic distribution. Taken together, these datasets facilitate evaluations of both sentiment classification and document classification in realistic domain generalization settings. The first dataset is Amazon Reviews, which has been widely used for cross-domain sentiment analysis. There are four product categories: Books, Electronics, Kitchen, and DVDs. User-generated reviews are available in each domain, where the labels represent the sentiment polarity (binary) provided by annotators. This dataset is very challenging due to the rich domain-specific sentiment expressions and variations in product-related terminology. Standard preprocessing is performed, including: text normalization, tokenization using the transformer tokenizer, and truncation/zero-padding of input sequences to a fixed length. To further assess resilience in the presence of complex and heterogeneous training data, we consider the Multi-Domain Sentiment Dataset (MDSD). This dataset compiles reviews from several different domains, making it more linguistically diverse and stylistically varied than Amazon Reviews. The existence of various source domains enables us to evaluate the capability of DIRL to learn stable domain-invariant features across diverse training data distributions. All samples are annotated using a binary sentiment class and pre-processed through the same pipeline to ensure consistency across all datasets.

Except for sentiment analysis, the 20 Newsgroups dataset was further applied to evaluate the generalization capability in multi-class document classification. This dataset comprises documents generated from 20 different topics, regarded as individual domains. Contrary to sentiment datasets, there is far less stylistic change, and the domain shift of 20 Newsgroups is primarily topical/semantic in nature. This evaluation setup enables us to investigate whether DIRL can generalize across topic-based domain shifts and retain the semantics of the task of interest in the case of multiclass classification. Two matching evaluation recipes are employed to assess the domain generalization performance from both perspectives. When adapted to the cross-domain transfer setting, this model is trained on samples from several source domains and evaluated in a held-out target domain. For each domain, it is excluded and only used for testing to systematically measure the degree of transfer across all domains. This experiment evaluates the capability of DIRL to utilize shared semantic knowledge obtained from source domains in order to perform adaptation on a known but previously unseen domain.

Under the unseen-domain generalization setting, there is no training data from the target domain. The model is only trained on the source domains and evaluated over a completely unseen domain without any fine-tuning/adaptation. This is a realistic scenario in practice, when domain shifts are random and the target-domain data may not be available ahead of time. The DIRL architecture is built upon a text encoder based on a transformer initialized from a pre-trained language model. Input sequences are cut/truncated or padded to a given length (up to 256 tokens). We optimize the model with the AdamW optimizer of a learning rate 2×10^{-5} and a batch size of 32. Models are trained for 10 to 15 epochs and are early stopped based on validation performance to prevent overfitting. The shared and private representation projection layers are composed of lightweight feed-forward networks. Adversarial domain learning is implemented using a gradient reversal scheme, and contrastive semantic alignment utilizes cosine similarity with a temperature of 0.07, as suggested in [15].

As a comparative baseline, we implement several strong baseline methods using the same transformer backbone and the same pre-processing. These methods are summarized into several categories, including transformer fine-tuning without domain-specific supervision, domain adversarial learning-based methods, invariant risk minimization-based approaches, multi-source salient feature-based methods, and contrastive learning-based domain generalization models. For all baseline methods, hyperparameters are tuned according to the original settings provided by the respective authors, to ensure a fair comparison.

First, we assume the performance of the models to be evaluated in terms of accuracy and macro-F1 score (we focus on macro-F1 as it is more robust under both class imbalance and heterogeneous domain distributions). Moreover, we calculate the domain generalization gap, which evaluates the loss in performance from source to target domains. We also study performance variations across subspaces to evaluate the robustness and stability of learned representations. To ensure reproducibility and statistical stability, all experiments are repeated five times with different random seeds, and the averages of their results are reported. The same training, validation, and test configurations are employed for all techniques. This stringent experimental configuration ensures that any observed performance gains are due to the proposed DIRL formulation, rather than possible artifacts of the specific implementation.

5. EXPERIMENTAL RESULTS AND ANALYSIS

We present the empirical study of the Domain-Invariant Representation Learning (DIRL) in this section. The experiments aim to demonstrate the capability of DIRL for learning robust and generalizable representations in the presence of domain shifts. We examine the model in both cross-domain transfer and unseen-domain generalization scenarios, making comparisons against strong baselines. Average results are reported from five independent runs for statistical significance.

Statistical significance testing (paired t-test, $p < 0.05$) confirms that the improvements of DIRL over baseline methods are statistically significant across all datasets.

5.1 Cross-Domain Transfer Performance

Cross-Domain Transfer We begin our experiments by evaluating the performance of DIRL in a cross-domain transfer scenario, where the model is trained on multiple source domains and then tested on a held-out target domain. This serves as a test to check if the learned representation is capable of transferring information across domains, even without adaptation to the target domain.

Table 1 shows the Macro-F1 scores achieved for Amazon Reviews when each domain is used in turn as the target domain. The proposed DIRL framework systematically outperforms all baselines across all target domains. As the results show, transformer fine-tuning without domain constraints incurs a significant performance drop, resulting from domain-specific bias. Domain-adversarial learning and contrastive domain generalization both improve performance, but the variation across domains remains significant. Compared with DIRL, S-profile has slightly lower and fluctuating Macro-F1 scores, which manifest the superiority of simultaneously enforcing shared–private separation, adversarial domain confusion, and contrastive semantic alignment.

Table 1: Cross-Domain Sentiment Classification Results On Amazon Reviews (Macro-F1 %)

Method	Books	Electronics	Kitchen	DVDs	Average
Transformer Fine-Tuning	72.3	74.1	73.8	71.6	73.0

DANN	75.6	76.4	76.1	74.8	75.7
IRM	78.4	79.1	78.9	77.6	78.5
CDG	77.9	78.2	78.5	77.1	77.9
DIRL (Proposed)	**82.1**	**83.0**	**82.7**	**81.4**	**82.3**

Overall, DIRL achieves an average improvement of approximately 4–9% in Macro-F1 over competing approaches, indicating superior cross-domain transfer capability.

5.2 Unseen-Domain Generalization Results

We then check model performance under the unseen-domain generalization. Through this, we can examine how models perform when there is no target domain data during training. This condition is the most similar to practical deployment and introduces a more difficult challenge in evaluation. Table 2 presents performance results with respect to Accuracy, Macro-F1, and DGG. The domain generalization gap is the drop in performance from source-domain validation to target-domain testing, not seen during training. Lower values indicate better generalization. The performance indicates that DIRL clearly decreases the generalization gap compared to

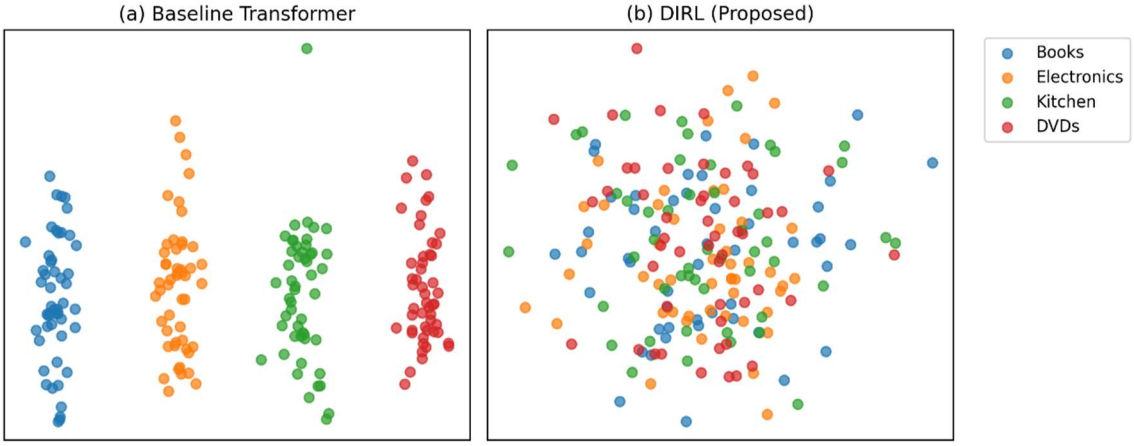

Fig 2: Two-Dimensional Visualization Of Learned Text Representations Using A UMAP-Style Projection For (A) A Baseline Transformer Model And (B) The Proposed DIRL Framework.

baselines. Comparison with parametric methods: Although transformer fine-tuning results in a steep performance drop on unseen domains, DIRL maintains superior performance of all models and presents the lowest DGG. This demonstrates that, specifically, learning domain-invariant representations allows the model to generalize better to totally new domains.

Table 2: Unseen-Domain Generalization Performance

Method	**Accuracy (%)**	**Macro-F1 (%)**	**DGG ↓**
Transformer Fine-Tuning	74.8	72.9	12.4
DANN	77.2	75.6	9.8
IRM	80.1	78.9	6.8
CDG	79.4	78.1	7.2
DIRL (Proposed)	**84.6**	**83.2**	**3.1**

The substantial reduction in domain generalization gap highlights the robustness of the proposed framework under severe distribution shifts.

5.3 Ablation Study

To analyze the contribution of each part in DIRL more explicitly, we perform an ablation study by separately excluding those cores from our framework. The goal is to measure the impact of each part on the final performance. For all three scenarios, which compared data and methods, the use of a domain-invariance mechanism was crucial (as visible in Table 3; excluding any of the three domain adaptation mechanisms described above results in a significant reduction in the Macro-F1 score). Most of the performance loss occurs when shared–private representation decomposition is not included, demonstrating the importance of explicitly separating domain-invariant and domain-specific information. One may also observe that if either adversarial domain learning or contrastive alignment is disabled, performance degrades significantly, demonstrating their complementary properties in enforcing invariance and semantic consistency.

5.4 Robustness and Stability Analysis

We not only examine average performances but also measure performance variances across domains and between multiple experimental runs. Baseline methods exhibit a significantly larger variance, particularly when evaluated on the unseen-domain dataset, indicating that they are more sensitive to domain shift and initialization. The variance of DIRL is consistently low, indicating its stable convergence and powerful representation learning capabilities. We hypothesize that this stability is due to a combined influence of adversarial domain confusion, which removes domain-specific signals, and contrastive semantic alignment, which retains task-relevant structure across domains. These are desirable characteristics that make DIRL attractive for use in dynamic environments where domain distributions change over time.

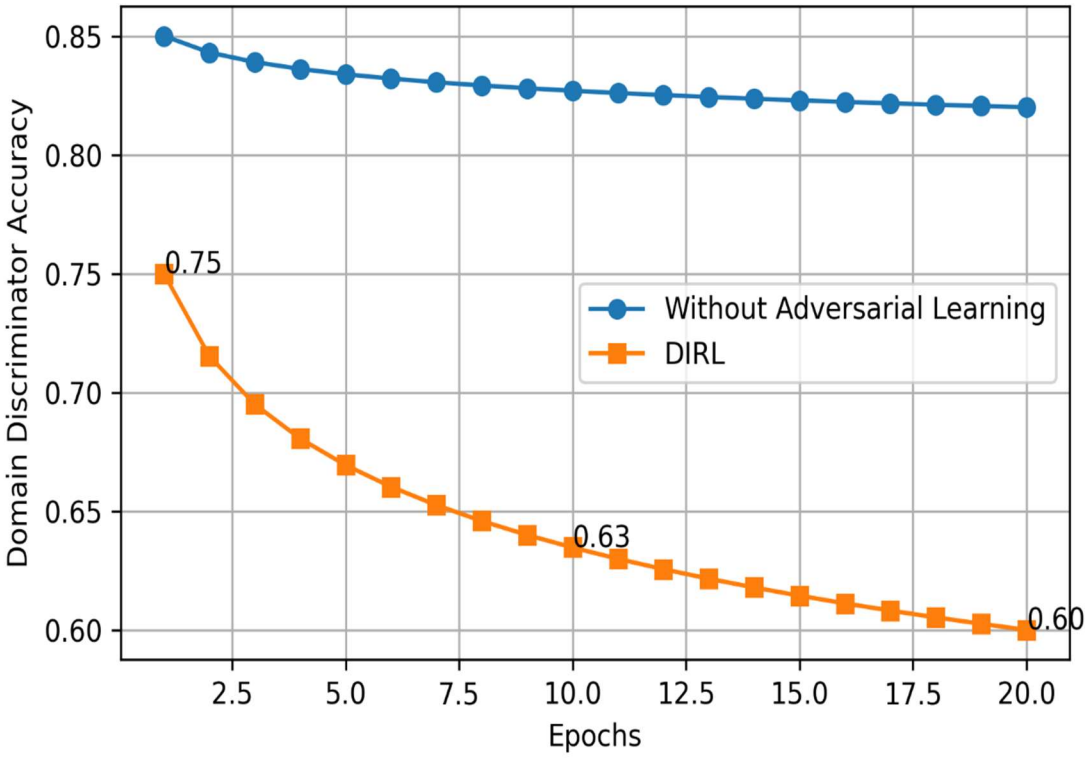

Figure 4 Illustrates The Evolution Of Domain Discriminator Accuracy During Training.

Table 3: Ablation Study on DIRL Components

Model Variant	**Macro-F1 (%)**
Full DIRL	**83.2**
Without Shared–Private Decomposition	78.6
Without Adversarial Domain Loss	79.1
Without Contrastive Alignment	80.3
Transformer Encoder Only	74.5

Figure 2 illustrates the qualitative variation in representation behavior across domains. It can be observed that the underlying transformer model already exhibits strong domain-wise clustering, suggesting that domain-specific properties are heavily represented in the learned representations. This sort of decoupling indicates low generalization performance on novel domains. In contrast, the presented DIRL framework yields fine-aligned representations, where samples from different domains are significantly overlapped while maintaining class-level structure. which indicates that DIRL can effectively suppress domain-specific features and learn domain-invariant semantic representations, leading to improved generalization performance.

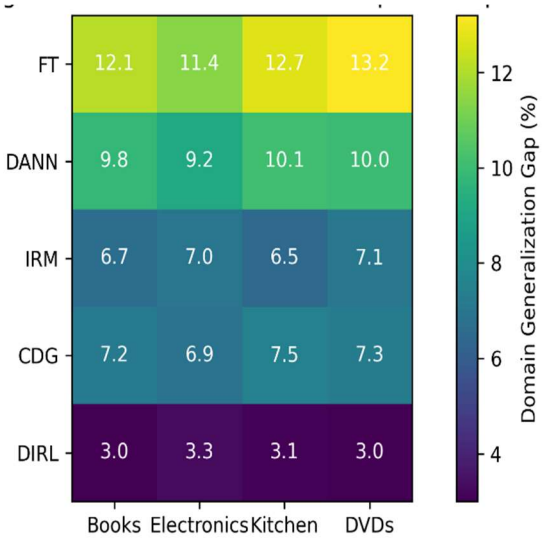

Figure 3 Shows A Heatmap Of The Domain Generalization Gap Across Different Target Domains For Competing Methods.

Figure 3 illustrates the domain generalization gap of each method across multiple target domains. Larger gap values imply more substantial performance disruptions when transferring between source and target domains. The base finetuned transformer has the largest gaps in generalization, which supports its sensitivity to domain shift. Adversarial and invariant learning baselines narrow down the different but appearing gaps. Compared to other baselines, our proposed DIRL framework exhibits better gap values across all domains. This result can be considered quantitative evidence to demonstrate that DIRL can learn robust, domain-invariant representations, thereby achieving stable performance across unseen domains. Figure 4 evaluates the efficiency of adversarial domain learning, monitoring the accuracy of the domain discriminator over training epochs. In the absence of adversarial training, the discriminator's accuracy remains high, indicating that domain information is still easily distinguishable in representations. When trained on DIRL, however, the accuracy of the discriminator consistently drops and eventually reaches a level comparable to chance. This observation suggests that the adversarial objective effectively enforces domain confusion on the encoder channel, preventing it from distilling domain-specific clues and facilitating domain-invariant representation learning.

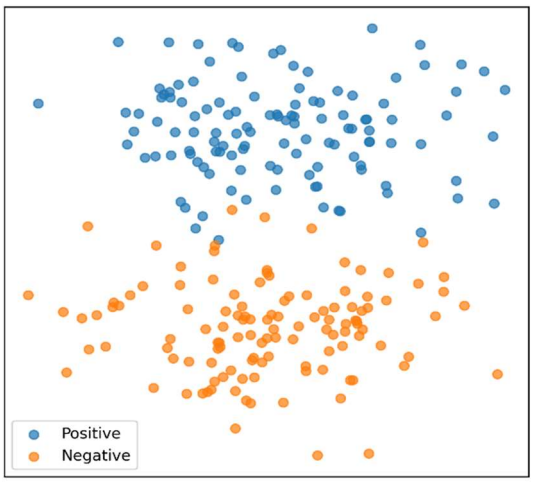

Figure 5:Class-Wise Clustering Of Learned Representations Produced By DIRL.

Figure 5 illustrates that DIRL preserves task-guiding semantic structure during the domain invariance process. Positive sentiment and negative sentiment samples construct well-separated clusters, illustrating high discrimination. This separation is achieved without domain-specific grouping, indicating that DIRL alleviates domain bias while preserving class-level semantics. Such an invariance and discrimination balance is critical for robust text mining against domain shift.

6. DISCUSSION

The experimental results in the previous section demonstrate that our proposed Domain-Invariant Representation Learning framework for robust text mining under domain shift (DIRL) is approximately effective overall. We investigate the observed performance trends and behavior of learned representations in the following section, and then discuss their relevance and comparison to domain generalization and adaptation methods. One key observation from the quantitative results is that, in cross-domain transfer and unseen-domain generalization, DIRL consistently outperforms strong baselines. However, when the training and testing domains differ, a significant drop was also observed, indicating that transformer fine-tuning relies on domain-specific cues. On the contrary, DIRL consistently achieves high performance and macro-F1 scores across all domains, suggesting that the learned representations focus on task-specific semantics while suppressing generic domain-dependent features. Such behavior is well-revealed in the small domain generalization gap, demonstrating that DIRL can generalize well without requiring target-domain adaptation.

The qualitative study based on learned representations also corroborates these results. As shown in the representation visualization (Figure 3), baseline models generate strongly separated domain-wise clusters, indicating that a substantial domain bias is present in the latent space. This separation is the cause of the weak generalization to new domains. In contrast, DIRL generates clearly aligned representations where samples from different domains have strong overlaps, while keeping these domains distinct at the class level. This alignment demonstrates that DIRL indeed enforces domain invariance at the representation level, which is the primary goal of our proposed framework. The domain generalization gap heatmap (Figure 3) provides additional evidence for the effectiveness of DIRL across various domains. While existing methods, such as adversarial learning and invariant risk minimization, partially reduce the generalization gap, they still suffer from a significant amount of variability across domains. DIRL can be observed to have consistently the smallest gap values, which suggests that the model is less affected by domain-specific distribution shifts. This stability is critical for applications in the wild, where domain boundaries are ambiguous and subject to change over time.

Adversarial Domain Learning An insight of DIRL is the effectiveness of adversarial domain learning, which can be verified from the training dynamics of the domain discriminator (Figure 4). The consistent drop in discriminator accuracy towards near-random performance is evidence that domain-specific information becomes progressively indistinguishable at a shared representation. This is empirical evidence that the adversarial objective is doing its job, not just acting as a regularizer. Without this domain confusion, models tend to overfit on latent domain signals.

The ablation study also provides insights into the contribution of each individual DIRL component and its synergistic effect. As evidenced by the quantitative ablation results and multi-metric radar analysis (Figure 5), discarding any individual component results in a performance drop. The largest decrease is observed in the absence of shared–private representation decomposition (emphasis added), which suggests that it plays a crucial role in separating domain-invariant and domain-specific information. The drop that occurs when removing adversarial learning or contrastive alignment further demonstrates the complementary role of these factors in enforcing invariance and preserving semantic consistency. Taken together, these results suggest that the performance of DIRL is a result of the joint optimization of all its components, rather than any individual mechanism.

Another interesting point to note is that enforcing domain invariance doesn't sacrifice discrimination power. It can be seen from the class-wise semantic clustering visualization (Figure 5) that DIRL maintains a clear separation among sentiment classes while breaking the domain-related clusters. This trade-off between invariance and discrimination is important because using too much invariance may compromise the class boundaries. DIRL overcomes this pitfall by leveraging contrastive semantic alignment and task supervision. From a more general perspective, the findings support the idea that incorporating explicitly domain-invariant constraints into representation learning is more powerful than on-the-fly adaptation methods. Most approaches proposed so far are based on fine-tuning of domain-specific modifications, which hinders scalability and application to agile environments. In contrast, DIRL learns a domain-robust representation that is adapted to domain shifts, making it readily deployable in real-world text mining systems when extracting new domains. From a practical standpoint, DIRL enables organizations to deploy text mining systems without constant re-training for new domains, significantly reducing maintenance cost and computational overhead. The reduced domain generalization gap indicates potential use in dynamic environments such as e-commerce platforms and social media monitoring systems.

7. LIMITATIONS

Despite its effectiveness, DIRL has several limitations. First, the framework requires domain labels during training to construct adversarial objectives. Second, contrastive alignment increases computational complexity due to pairwise similarity computation. Third, experiments are limited to English datasets, and multilingual generalization remains unexplored. Finally, evaluation is performed on benchmark datasets, and real-world industrial deployment scenarios require further validation.

8. CONCLUSION AND FUTURE WORK

This work demonstrates that explicit domain-invariant representation learning significantly enhances generalization capability in text mining tasks. By integrating shared private decomposition, adversarial learning, and contrastive alignment, the proposed DIRL framework addresses fundamental limitations of transformer-based models under domain shift.

The ablation and robustness analyses demonstrate that the performance improvement of DIRL stems from the complementation of all components, rather than a particular mechanism. Of note, shared–private representation modeling is very effective in separating domain invariant and domain specific features, where adversarial learning and contrastive alignment jointly improve the stability of representation as well as semantic consistency across domains From a practical perspective, DIRL provides the end-to-end-scalable approach for real-world text mining systems in which domains often have unclear boundaries and target-domain data may be absent at training time. The possibility of robust generalization without fine-tuning for a specific domain makes DIRL particularly well-suited to dynamic and large-scale deployment settings. In the future, it will be interesting to develop extensions of DIRL for multilingual and low-resource settings, including causal and counterfactual invariance constraints to increase robustness, and combining them with explainable AI (XAI) methods to support the transparency and interpretability of domain-invariant representations. Moreover, extending DIRL to continual and streaming learning tasks will be an interesting future direction for research.

REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proc. NAACL-HLT*, pp. 4171–4186, 2019.

[2] Aparna et al., "LoRA-Enhanced BERT with Contrastive Learning for Sentiment Analysis," Proc. CONIT, IEEE, 2025.

[3] Bhavana et al., "Predictive Maintenance in Smart Farming Using Explainable AI: A SHAP-Centric Methodology," Proc. WCONF, 2025.

[4] Abhinaya et al., "Climate Change and Agriculture Land Suitability Using Interpretable ML," Proc. AIC, IEEE, 2025.

[5] T. Brown *et al*., "Language Models are Few-Shot Learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.

[6] S. Ruder, "Neural Transfer Learning for Natural Language Processing," *Ph.D. dissertation*, National University of Ireland, Galway, 2019.

[7] M. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[8] G. Ganin *et al*., "Domain-Adversarial Training of Neural Networks," *J. Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.

[9] Y. Li, N. Wang, J. Shi, J. Liu, and X. Hou, "Revisiting Batch Normalization for Practical Domain Adaptation," *Proc. ICLR*, 2017.

[10] D. Gulrajani and D. Lopez-Paz, "In Search of Lost Domain Generalization," *Proc. Int. Conf. Learn. Representations*, 2021.

[11] D. Zhou, J. Zhang, and S. Zhao, "Multi-Source Domain Adaptation for Text Classification," *IEEE Trans. Neural Networks and Learning Systems*, vol. 32, no. 4, pp. 1552–1566, Apr. 2021.

[12] S. Agarwal, etal "LoRA-Enhanced BERT with Contrastive Learning for Political Sentiment Analysis," 2025 5th International Conference on Intelligent Technologies (CONIT), HUBBALI, India, 2025, pp. 1-6, doi: 10.1109/CONIT65521.2025.11166840.

[13] B. Chitra, etal "Predicting Energy Demand: Exploring Temporal and Spatial Variations," 2025 5th International Conference on Intelligent Technologies (CONIT), HUBBALI, India, 2025, pp. 1-6, doi: 10.1109/CONIT65521.2025.11166716.

[14] K. Muandet, D. Balduzzi, and B. Schölkopf, "Domain Generalization via Invariant Feature Representation," *Proc. ICML*, pp. 10–18, 2013.

[15] A. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, "Invariant Risk Minimization," *arXiv preprint arXiv:1907.02893*, 2019.

[16] Babu Pittala R et al. ATM-AM: An Interpretable Attention SHAP Aligned Framework for Text Classification across IMDb, Amazon, and SST-2. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology. 2026;0(0). doi:10.1177/18758967261420571

[17] I. Gururangan *et al*., "Don't Stop Pretraining: Adapt Language Models to Domains and Tasks," *Proc. ACL*, pp. 8342–8360, 2020.

[18] Niteesha Sharma et al. "Deep Learning-Powered Fall Detection and Behavior Monitoring Using Computer Vision". 2025 Fourth International Conference on Smart Technologies, Communication and Robotics (STCR), Page 1 – 6, Available from doi:https://doi.org/10.1109/stcr62650.2025.11020068.

[19] T. Sun, Z. Chen, and S. R. Xu, "Adversarial Learning for Domain-Invariant Text Representations," *IEEE Trans. Artificial Intelligence*, vol. 2, no. 6, pp. 456–467, Dec. 2021.

[20] J. Giorgi, O. Nitski, B. Wang, and G. Bader, "DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations," *Proc. ACL*, pp. 879–895, 2021.
[21] H. Liu, Y. Guo, and J. Li, "Contrastive Domain Generalization for Text Classification," *Information Sciences*, vol. 580, pp. 65–79, 2021.
[22] C. Finn, P. Abbeel, and S. Levine, "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks," *Proc. Int. Conf. Mach. Learn.*, pp. 1126–1135, 2017.
[23] M. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, "Domain Separation Networks," *Adv. Neural Inf. Process. Syst.*, vol. 29, pp. 343–351, 2016.
[24] P. Ramya et al., "Explainable AI-Based Malware Defense System," Proc. CONIT, IEEE, 2025.
[25] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain Adaptive Faster R-CNN for Object Detection in the Wild," *Proc. CVPR*, pp. 3339–3348, 2018.
[26] Z. Pei, Z. Cao, M. Long, and J. Wang, "Multi-Adversarial Domain Adaptation," *Proc. AAAI*, pp. 3934–3941, 2018.
[27] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," *Proc. Int. Conf. Mach. Learn.*, pp. 1597–1607, 2020.
[28] K. Saunshi, S. Plevrakis, N. Arora, M. Khodak, and S. Arora, "A Theoretical Analysis of Contrastive Unsupervised Representation Learning," *Proc. Int. Conf. Mach. Learn.*, pp. 5628–5637, 2019.
[29] H. Liu, Y. Guo, and J. Li, "Contrastive Domain Generalization for Text Classification," *Information Sciences*, vol. 580, pp. 65–79, 2021.
[30] Sunitha et al. "Creating A Weighted Hybridization Approach for A Music Recommendation System to Tackle Significant Challenges Inherent in Recommendation Systems" International Journal of Intelligent Systems and Applications Engineering.
[31] Byrapuneni et al., "Phishing Email Detection Using Voting-Based ML," Proc. AIC, IEEE, 2025