Journal of Computer and Communications, 2026, 14(3), 48-70
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

NAT-MTT: Noise-Aware Multi-Task Transformers for Cross-Domain Aspect-Based Sentiment Analysis

Hani Iwidat[1*], Mohammad Sale[1], Imran Khaled[1], Basem Ajarmah[2], Waheeb Abu-Ulbeh[2]

[1]Department of Data Science, AL-Istiqlal University, Jericho, Palestine
[2]Department of Cybersecurity, AL-Istiqlal University, Jericho, Palestine
Email: *hani.iwidat@pass.ps

How to cite this paper: Iwidat, H., Sale, M., Khaled, I., Ajarmah, B. and Abu-Ulbeh, W. (2026) NAT-MTT: Noise-Aware Multi-Task Transformers for Cross-Domain Aspect-Based Sentiment Analysis. *Journal of Computer and Communications*, **14**, 48-70.
https://doi.org/10.4236/jcc.2026.143004

Received: February 19, 2026
Accepted: March 10, 2026
Published: March 13, 2026

Abstract

This paper proposes a noise-aware multi-task transformer framework that jointly performs aspect extraction (AE) and aspect sentiment classification (ASC) using a shared BERT/RoBERTa encoder with dual task-specific heads. Robustness is promoted through a systematic noise-aware training (NAT) strategy that injects controlled synthetic perturbations (spelling errors, word dropout, synonym replacement, slang) according to a curriculum schedule, mixing clean and noisy instances in each batch. Experiments on SemEval-2014 (Restaurants, Laptops) and large-scale Amazon (Electronics, Apparel) and Yelp (Food) reviews demonstrate consistent gains over strong single-task, multi-task, and cross-domain baselines in in-domain, cross-domain, and multi-domain settings. On the SemEval Rest 14 dataset, the proposed model achieves improvements of +2.6 F1 (AE) and +2.9% accuracy (ASC) on the Rest14 benchmark over the strongest baseline. with maximum gains of +3.1 F1 (AE) and +3.3% (ASC) over the strongest cross-domain baseline (BGCA), reduces noise-induced performance degradation by up to 42% (NAT contribution vs. identical model without NAT), and improves cross-domain transfer with minimal additional parameters and a 22% training-time overhead. Ablation and error analyses show that multi-task learning and NAT are both critical to robustness, particularly under high noise and domain shift. These findings indicate that jointly learning multiple ABSA subtasks with noise-aware augmentation is an effective and efficient route to deployable, real-world ABSA systems.

Keywords

Aspect-Based Sentiment Analysis, Multi-Task Learning, Transformer Models, Cross-Domain Robustness, Noise-Resistant NLP

1. Introduction

Web users can express their opinions on a broad range of topics in several ways on the online platforms. For instance, they may discuss a legal decision or write a review or describe medical treatment or share political opinions. These opinions can be expressed through diverse media of communication, including social networks and discussions in forums. Such exchanges are a significant [1]. Due to the enormous amount of available information, extracting knowledge from these different sources has become a complicated task when performed manually. Therefore, the utilization of Artificial Intelligence (AI) OM techniques has become a practical solution [2]. OM or SA seeks to extract the polarity of texts (e.g., positive, negative, or neutral) in an automated manner. It can be done at different levels: aspect, sentence, or document [3].

Aspect-Based Sentiment Analysis (ABSA) focuses on identifying fine-grained opinion structures by extracting aspect terms and determining sentiment polarity. providing more informative insights than document- or sentence-level sentiment analysis. This task has attracted increasing attention in recent years [4]. As user-generated reviews increase across domains such as restaurants and movies, ABSA has become central to opinion mining in e-commerce and social media. Recent advances in pre-trained transformer models (Ex:, BERT and RoBERTa) have significantly improved sentiment tasks by providing powerful contextual representations and transfer learning capabilities However, practical ABSA systems must operate across domains and under noisy conditions—where labels may be imperfect, texts contain informal language, typos, or sarcasm, and domain-specific expressions differ substantially—making robust cross-domain generalization a central challenge [5].

Cross-domain ABSA aims to transfer fine-grained sentiment knowledge from a labeled source domain to a sparsely labeled or unlabeled target domain, thereby reducing annotation cost while maintaining aspect-level accuracy [1]. Domain adversarial training, topic guidance, and knowledge-enhanced representations are common strategies to mitigate distribution shifts between domains. BERT-based aspect-level models combined with domain-adversarial networks learn domain-invariant sentence and aspect representations, improving cross-domain performance on Amazon product reviews [6]. Generative and retrieve-and-edit frameworks further alleviate domain gaps by augmenting unlabeled target data or generating pseudo-labeled target-like sentences for end-to-end ABSA [7]. The methods train on clean benchmark datasets which provide dependable supervision but they become ineffective when handling two specific challenges that include noisy labels and unique domain requirements which exist in actual operational environments. **Table 1** shows examples of ABSC applications which span different fields and the table demonstrates that the sentiment term "delicious" and the term "rude" appear in the restaurant industry but they do not exist in the electronics sector. Different domains use different words to show their respective emotional states. The word "long" functions as a negative word in food services while it serves as a positive word in the electronics industry. The study shows that different domains use different emotional vocabulary which creates multiple meanings for the same words. Multi-task learning (MTL) has emerged as an effective paradigm to jointly model multiple ABSA subtasks (such as aspect term extraction, aspect category detection, and aspect sentiment classification) by leveraging shared representations and mutual constraints between tasks [8]. Neural MTL frameworks based on BERT variants show that learning extraction and sentiment tasks within a unified model reduces error propagation compared with pipeline approaches and yields higher F1 scores on benchmarks like SemEval and Arabic hotel reviews. Transformer-based MTL for related affective tasks, such as emotion classification and intensity prediction, also demonstrates that jointly optimizing related objectives x can outperform single-task baselines. However, most existing MTL models are developed for single-domain scenarios and do not explicitly tackle the combined problem of cross-domain adaptation and robustness to noisy data.

Transformers have become the main technology used in sentiment analysis, but different robustness features of the system show different results. The study of 22 datasets demonstrates that T5 achieves the highest performance, while XLNet shows superior results in detecting irony and understanding product-related sentiment, and RoBERTa and ELECTRA perform best in particular tests, but BERT and DistilBERT fail to handle advanced sentiment detection tasks because they prioritize efficiency. The researchers developed transformer-based architectures for fine-grained ABSA through the integration of cross-modal attention, syntactic augmentation, and knowledge graphs, which enable the system to extract complex aspect-opinion-polarity relationships. The present transformer-based ABSA systems continue to pursue two objectives, which include achieving maximum performance in their respective domains and maintaining correct feature distribution; however, they do not use explicit noise-aware training methods, uncertainty modeling techniques, or label noise protection strategies, especially when dealing with cross-domain and multi-task situations.

Most existing ABSA studies in related work exhibit complementary but incomplete coverage of the requirements for real-world, robust ABSA systems across noisy, heterogeneous domains. Unified and instruction-tuned frameworks such as the unified generative model and Unified ABSA focus on covering many ABSA subtasks within a single architecture, but they primarily target clean, in-domain benchmarks and do not explicitly model noise or evaluate robustness under noisy conditions. Cross-domain models like CC-ASTN and other domain-adversarial approaches are effective at transferring aspect-level sentiment knowledge between source and target domains, yet they assume relatively clean text and leave the impact of linguistic noise largely unexplored. Studies focused on developing robust systems through non-counterfactual methods and basic data augmentation techniques show improved performance against specific disruptions but they handle sentiment evaluation as a separate task and they fail to unify aspect extraction with aspect-based sentiment evaluation through their multiple domain assessment approach. The combination of hybrid GCN-transformer models and intent-aware transformers TASCI leads to better benchmark dataset performance, yet these architectures need to develop methods for multi-domain noise-aware training which show the connection between system robustness and accuracy and performance costs. The industry lacks a unified solution which provides joint AE/ASC modeling across different domains while handling user-generated text through noisy text protection and maintaining operational effectiveness for real-world applications. We introduce NAT-MTT as an ABSA framework which uses a noise-aware multi-task transformer system to provide complete power across multiple applications while focusing on robust system performance and cross-domain usage.

Our model uses a shared BERT/RoBERTa encoder with two task-specific heads to jointly perform aspect extraction (via BIO tagging) and aspect-level sentiment classification (via span-level attention), thereby eliminating pipeline-based error propagation and leveraging shared representations between AE and ASC. On top of this unified architecture, we introduce a systematic noise-aware training strategy that generates synthetic noisy variants of training sentences. By using spelling perturbations, word dropout, synonym replacement, and slang injection. Then involving mixes clean and noisy samples in each batch to encourage noise-invariant yet semantics-preserving representations. We evaluate the framework in multi-domain and cross-domain settings over benchmark datasets (SemEval) [9] and large-scale real-world reviews (Amazon, Yelp) [10] [11], and we quantify robustness by measuring performance degradation under controlled noise levels. thus, directly addressing the gaps in prior work regarding noisy input and domain shift. Finally, we conducting a detailed ablation and efficiency analyses (parameters, training time, inference speed), demonstrating that the proposed approach offers substantial gains in robustness and cross-domain performance.

Table 1. Illustrative samples of ABSC across diverse domains.

Domain	Sentence	Aspect Terms	Polarity
Restaurant	The (price) is affordable, although the (waiter) is rude.	price waiter	POS. NEG.
Laptop	Its (performance) is ideal, I wish I could say the same about the (price).	performance price	POS. NEG.
Electronics	The (battery life) is quite long, but the (camera quality) is poor.	battery life camera quality	POS. NEG.
Apparel	The (fabric) is soft and comfortable, but the (sizing) runs small.	fabric sizing	POS NEG.
Food Services	The (ambiance) was cozy, but the (wait time) was too long.	ambiance wait time	POS NEG.

In this work, we propose Noise-Aware Multi-Task Transformers for Robust Cross-Domain Aspect-Based Sentiment Analysis, aiming to bridge these gaps.

Contributions. The main contributions of this study are summarized as follows:

- We propose a unified multi-task transformer framework that jointly learns end-to-end ABSA subtasks across source and target domains, reducing error propagation and improving data efficiency.
- A systematic noise-aware training strategy based on controlled synthetic noise injection (spelling errors, word dropout, synonym replacement, slang) that enhances robustness to real-world linguistic variations.
- Providing a comprehensive cross-domain evaluation across multiple domains (electronics, apparel, food services) using both benchmark (SemEval) and real-world (Amazon, Yelp) datasets under clean and noisy conditions.
- Detailed analysis of robustness-efficiency trade-offs, including parameter counts, providing practical insights for deployment.

This leads to the following research questions:

- (RQ1): To what extent can a multi-task transformer improve cross-domain ABSA performance compared with single-task or pipeline models?
- (RQ2): How much do noise-aware training techniques contribute to robustness under noisy conditions in cross-domain ABSA?
- (RQ3): Does combining multi-task learning with noise-aware mechanisms yield more stable predictions across diverse domains and under noisy conditions?
- (RQ4): What is the computational cost of the proposed framework?

The remainder of this paper is organized as follows: In Sect. Related work, we discuss the relevant literature and prior work in the domain. Section 2 explore into the methodology of the proposed framework, encompassing the techniques and models we employed, Sect. Experiments showcases the experimental results, and the evaluation is presented in Sect. Model analysis we perform an ablation study. Finally, Sect. Conclusion concludes the paper, summarizing our contributions and suggesting potential directions for future research.

2. Related Work

Aspect-Based Sentiment Analysis ABSA has evolved from isolated tasks (e.g., only aspect extraction or only sentiment polarity) to compound tasks that jointly predict multiple sentiment elements such as aspect terms, opinion terms, and aspect–sentiment triplets [2]. Compound ABSA tasks include aspect term extraction (ATE), unified ABSA (UABSA), aspect-opinion pair extraction (AOPE), and aspect sentiment triplet extraction (ASTE), all of which require capturing structured relations among aspects, opinions, and sentiments [2] [8] [12]. Transformer-based architectures augmented with cross-modal and aspect-aware mechanisms have been proposed to handle such complexity. For example, a cross-modal multi-task transformer (CMMT) for multimodal ABSA jointly learns aspect- and sentiment-aware intra-modal representations and uses text-guided cross-modal interaction, outperforming previous multimodal baselines on Twitter datasets for both aspect extraction and sentiment classification [13]. Similarly, SABKG integrates BERT with part-of-speech information and an aspect-sentiment knowledge graph, using graph neural networks to encode "aspect word-sentiment polarity-sentiment word" triplets and achieving state-of-the-art performance on three ABSA datasets [14]. These models show that structural and task-specific biases help ABSA, but they largely target single-domain, relatively clean benchmarks.

Multi-Task Learning: Multi-task learning (MTL) has been widely applied to leverage synergies between ABSA subtasks. MTL-AraBERT simultaneously performs aspect term extraction and aspect category detection in Arabic hotel reviews using a shared AraBERT backbone with task-specific heads, achieving strong F1 scores for both tasks [15]. A BERT-based multi-task framework for IMDb reviews jointly models sentiment classification and aspect-based analysis by sharing BERT embeddings and LSTM layers, using a softmax classifier for overall sentiment and a CRF layer for aspect extraction; this joint learning setup outperforms corresponding single-task baselines on both sentiment prediction and aspect detection [16]. SABKG can similarly be viewed as a multi-task system that learns several ABSA subtasks simultaneously by leveraging a shared, knowledge-augmented representation space [14]. Beyond ABSA, multi-task transformer models have shown effectiveness in affective computing: Labeed and Liang present a comparative analysis of multi-task transformers for emotion classification and intensity prediction on social media data, and report that multi-task learning consistently outperforms single-task variants across both objectives [17]. Similarly, unified BERT-based multi-task models for sentiment and ABSA demonstrate that shared contextual encoders can support multiple sentiment-related tasks and achieve competitive or superior performance compared with specialized single-task baselines [18]. Nevertheless, these multi-task approaches generally assume clean supervision and operate in single-domain or mixed but non-adapted settings, offering limited insight into cross-domain robustness under noisy user-generated text [17]. Recently, large language models (LLMs) have been used for few-shot ABSA, but their computational cost and latency limit deployment [12].

Cross-Domain ABSA and Domain Adaptation: Cross-domain ABSA seeks to transfer aspect-level sentiment knowledge from source to target domains with minimal or no target labels [7] [19]. A BERT-based aspect-level sentiment analysis algorithm combines BERT with convolutional layers and a domain-adversarial neural network to learn domain-invariant representations, achieving higher accuracy and F1 than classical algorithms on cross-domain Amazon product reviews [3]. CSC-PLDAT incorporates hybrid prompt learning with domain adversarial training for cross-domain aspect-based sentiment classification, designing prompts with transferable and task-specific components and achieving an average micro-F1 of 71.45% across four benchmarks, outperforming previous state-of-the-art methods especially under label scarcity and skewed sentiment distributions [7]. The bidirectional generative cross-domain ABSA framework BGCA trains models in both text-to-label and label-to-text directions, using generated sentences for data augmentation, and achieves new state-of-the-art results on four cross-domain tasks (ATE, UABSA, AOPE, ASTE) without using labeled target-domain data [20]. Retrieve-and-edit domain adaptation further improves transfer by retrieving similar prototypes in unlabeled target data and editing source words, yielding an absolute F1 improvement of about 3.95% in cross-domain end-to-end ABSA [7]. Large language model-augmented and syntax-aware methods extend these ideas. A syntax-aware domain adaptation framework combines LLMs with structural syntactic knowledge via a domain-topic predictor, adversarial training, and automatic soft prompts to enhance domain-specific semantic transfer, claiming systematic exploitation of syntactic and cross-domain characteristics for fine-grained sentiment classification [21]. However, these methods mainly optimize feature alignment and domain invariance; they rely on pseudo-labels or unlabeled data without explicit mechanisms to detect or down weight noisy annotations, and they typically optimize a single primary ABSA task rather than a full multi-task suite.

Transformer-Based Systems Handle Noise: This study investigates how transformer-based systems maintain system stability in their sentiment analysis process. The training method uses domain adversarial training and denoising autoencoders to decrease distributional mismatch but it does not provide a direct solution for handling label noise and uncertainty. The research examined various augmentation methods to identify which ones provide better performance for natural language processing tasks. Wei and Zou [22] presented Easy Data Augmentation (EDA) as a method which uses synonym replacement and random insertion and swap and deletion to produce consistent enhancements for text classification tasks. Xie *et al.* [23] introduced Unsupervised Data Augmentation (UDA) as a method which uses consistency training to process noise perturbations while Zhuang *et al.* [24] showed that using synthetic misspellings during training improves NER system reliability. The study found that transformers such as T5 and XLNet showed better ability to handle different sentiment and irony datasets while BERT and DistilBERT showed less stable performance. The system uses three techniques to reduce noise and redundancy through mutual information maximization and adaptive contrastive learning and dual attention which creates representations that resist noise and maintain accuracy with lower processing demands. The research area needs new dedicated noise-aware objectives for ABSA and cross-domain ABSA research which should be tested with multi-task transformers.

Summary and Research Gap: Overall, the literature provides strong components but leaves a clear gap. First, multi-task transformers for ABSA demonstrate gains over pipeline and single-task models but are mostly evaluated in single-domain, clean-label settings and rarely consider domain shift or noisy supervision [14]-[16]. Second, cross-domain ABSA methods—including adversarial, prompt-based, knowledge-enhanced, generative, and retrieve-and-edit frameworks—achieve state-of-the-art performance under unsupervised or semi-supervised transfer, but most of them mainly focus on feature/domain alignment or data augmentation rather than systematic noise modeling, and typically optimize a single ABSA task [3] [7] [19] Third, while robustness techniques from multimodal and generic sentiment analysis suggest that invariant and contrastive representations can mitigate noise brittle [25] [26] these ideas have not been fully integrated into cross-domain, multi-task ABSA settings.

Thus, there is a research gap at the intersection of 1) multi-task transformers for end-to-end ABSA, 2) cross-domain adaptation, and 3) explicit noise-aware learning. Existing models typically address at most one or two of these aspects. The proposed Noise-Aware Multi-Task Transformer is designed to fill this gap by jointly learning multiple ABSA subtasks across domains while incorporating domain adaptation and noise-aware mechanisms to achieve robust cross-domain performance under realistic noisy conditions.

Table 2(a) provides a comparative overview of representative models, summarizing their coverage across key dimensions: evaluated tasks, unified multi-task learning, cross-domain evaluation, explicit noise robustness, and efficiency analysis. Table 2(b) then outlines how our proposed framework directly addresses the gaps identified in prior work.

Table 2. (a) A comparative overview of related studies which summarizes key properties of related work and highlights how the proposed model positions itself across multiple dimensions. (b) How our model addresses identified gaps.

(a)

Model (Ref.)	Evaluated Task	Unified Multi-Task	Cross-Domain Eval.	Explicit Noise Robustness	Efficiency Analysis
LEGO-ABSA [27]	ASC/AE	✓	Limited	✗	Partial
UNIFIEDABSA [28]	ASC/AE	✓	Limited	✗	✓
SyMux [29]	ASC/AE	✓	Limited	✗	Partial
BERT-Based Cross-Domain [3]	AE/ASC	Partial	✓	✗	Partial
CC-ASTN [6]	ASC/AE	Partial	✓	✗	✗
Robust ABSA w/ Non-counterfactual Aug. [30]	ASC	✗	✓	✓	✗
BGCA [20]	ASC/AE	Partial	✓	✗	Limited
SentiSys [1]	ASC	✗	✓	✓	Partial
EDA/UDA [23]	ASC	✗	✗	✗	✓
Ours model	ASC /AE	✓	✓	✓	✓

(b)

Gap in Prior Work	Our Solution
Unified AE/ASC rare in cross-domain settings	Shared encoder with dual task-specific heads
Noise robustness treated in isolation, not with multi-task learning	Systematic noise-aware training (NAT) with invariance objective
Cross-domain methods assume clean text	Noise augmentation improves domain-invariant representations
Efficiency-robustness trade-off rarely analyzed	Full reporting of params, training time, inference speed
Evaluation on clean benchmarks only	Multi-domain testing under controlled noise (low/high)

✓ = fully addressed; partially/limited = partially/limited addressed; ✗ = not addressed/not applicable.

3. Methodology

This section presents the proposed methodology implemented for this work. We begin our discussion with problem definition and representation, then we discuss Noise-Aware follow by model architecture which is described in Section 3.3. Training Strategy, cross-domain training, as detailed in Section 3.5. Finally, we describe the Implementation Details of the proposed model on the prepared dataset. The proposed framework is shown in **Figure 1**.

3.1. Problem Definition and Representation

We formalize robust cross-domain aspect-based sentiment analysis as a multi-task learning problem under domain shift and linguistic noise. Consider a corpus of N sentences $X=\{x_1,x_2,\cdots,x_n\}$, where each sentence $x=[w_1,w_2,\cdots,w_n]$ is a token sequence. Sentences originate from domains $d\in D$, with $D=\{D_1,D_2,\cdots,D_m\}$ representing distinct product or service categories (electronics, apparel, restaurants, laptops).

Aspect Term Extraction (AE): For each sentence x, we identify contiguous spans $[s,e]$ $(1\le s\le e\le n)$ corresponding to aspect terms. Following sequence labeling conventions, each token w_i receives a label $y_i^{\text{ae}}\in Y_{\text{AE}}=\{\text{B-ASP},\text{I-ASP},\text{O}\}$, where B-ASP marks aspect boundaries, I-ASP indicates continuation, and O denotes non-aspect tokens. The output is a set of aspect spans $A=\{a_1,a_2,\cdots,a_k\}$ with $a_j=(s_j,e_j)$.

Aspect Sentiment Classification (ASC): For each extracted span a_j, we predict sentiment polarity $y_i^{\text{asc}}\in Y_{\text{ASC}}=\{\text{POS},\text{NEG},\text{NEU}\}$ expressing opinion toward that aspect. Formally, domain shift denotes divergence between source distribution $P_s(x,y)$ and target distribution $P_t(x,y)$, with $P_s\ne P_t$. In cross-domain settings, we assume labeled source data $D_s=\{(x_i,y_i^{\text{ae}},y_i^{\text{asc}})\}$ and unlabeled or sparsely labeled target data. Our objective is domain-invariant representations $h=f(x)$ such that $P_s(h\mid y)\approx P_t(h\mid y)$ while maintaining discriminative power for both tasks.

3.2. Noise-Aware Data Augmentation

Based on empirical analysis of Amazon and Yelp, we identify four types relevant to ABSA robustness: orthographic errors. Given an input sentence $x=\{w_1,w_2,\cdots,w_n\}$, we generate noisy variants $\tilde{x}$ through four perturbation operations with probabilities p_{spell}, p_{drop}, p_{syn}, and p_{slang} respectively:

- *Spelling Noise:* Character-level perturbations including insertions, deletions, substitutions, and transpositions. The error probability follows:

$$P(\text{error}\mid w)\propto exp\left(-\lambda\cdot\text{edit\_distance}(w,e)\right) \tag{1}$$

where $e\in E(w)$ is a possible error for word w, and λ controls error likelihood.

- *Word Dropout*: Random masking of non-aspect words:

$$\tilde{x}_i = \begin{cases} x_i & \text{with probability } 1 - p_{\text{drop}} \\ [\text{MASK}] & \text{with probability } p_{\text{drop}} \end{cases} \tag{2}$$

- *Synonym Replacement*: Context-aware substitution using WordNet synonyms: We use GloVe 840B-300d embeddings to compute cosine similarity between the original word and each synonym in WordNet. The synonym with the highest similarity is selected.

$$x_i' = \underset{s \in S(x_i)}{\text{argmax}}\, \text{cosine}\left(\text{embed}(x_i), \text{embed}(s)\right) \tag{3}$$

where $S(x_i)$ is the synonym set for word x_i.

- *Slang Injection*: Replacement with informal equivalents from dictionary D_{slang}: We compiled a slang dictionary of 2500 entries from Urban Dictionary [31], mapping informal terms to their formal equivalents (e.g., "g2g" → "got to go").

$$x_i'' = D_{\text{slang}}(x_i) \text{ if } x_i \in \text{keys}\left(D_{\text{slang}}\right) \tag{4}$$

To guarantee that ground-truth labels remain valid during the NAT process, all perturbation operations (including spelling errors, word dropout, synonym replacement, and slang injection) are applied exclusively to non-aspect tokens. Aspect terms and their exact character spans remain untouched to prevent label corruption. Following perturbation, the sentence is re-tokenized using the BERT/RoBERTa tokenizer, and the original BIO tags are automatically remapped to the new subword sequence using standard offset-based alignment. This ensures that AE and ASC labels stay perfectly aligned with the noisy input.

- *The training batch composition* is controlled by noise ratio β: β set to 0.5.

$$N_{\text{noisy}} = \beta \cdot N_{\text{total}}, \quad N_{\text{clean}} = (1-\beta) \cdot N_{\text{total}} \tag{5}$$

- *Curriculum Schedule*: The noise probability follows a curriculum schedule:

$$p_{\text{noise}}^{(t)} = p_{\max} \cdot \min\left(1, \frac{t}{T_{\text{ramp}}}\right) \tag{6}$$

where t is the current training step, $T_{\text{ramp}} = 5000$ is the ramp-up period, and $N_{\max} = 0.3$ is the maximum noise probability. Each individual perturbation probability (P_{spell}, P_{drop}, P_{syn}, P_{slang}) follows this same schedule.

3.3. Model Architecture

We design a unified multi-task transformer (**Figure 1**) with three components: a shared encoder producing contextualized token representations; an aspect extraction head for sequence labeling; and an aspect sentiment classification head with span-level attention. Joint learning avoids pipeline error propagation and enables task reinforcement.

Shared Transformer Encoder: Both clean and augmented noisy sentences $x = [w_1, w_2, \cdots, w_n]$ are encoded using a pre-trained transformer (BERT or RoBERTa):

$$\boldsymbol{H} = \text{TransformerEncoder}(x) = \{h_1, h_2, \cdots, h_n\},\ h_i \in \mathbb{R}^d \tag{7}$$

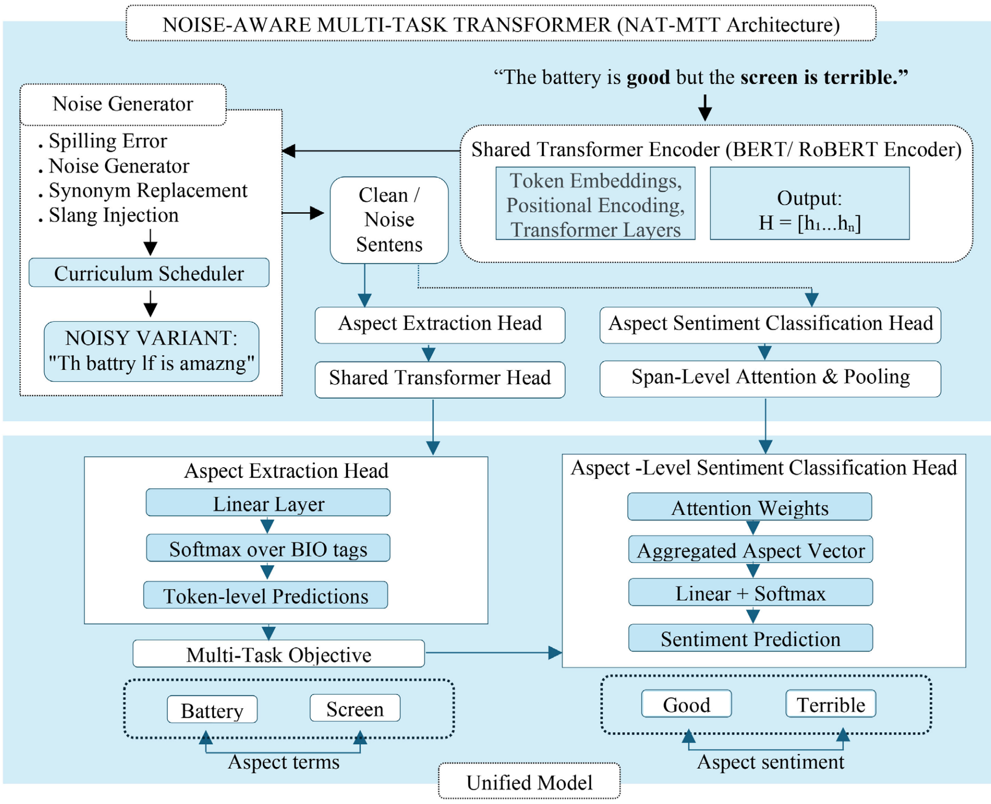

Figure 1. Model architecture of the proposed noise-aware multi-task transformer framework for cross-domain ABSA. The model processes input sentences through a shared transformer encoder, then uses two task-specific heads for joint aspect extraction (BIO tagging) and sentiment classification. The noise generator creates realistic perturbations during training to enhance robustness.

where $d = 768$ for base models. This shared representation serves as a common feature space for all downstream tasks.

Aspect Extraction Head: where Aspect extraction is formulated as a sequence labeling task using BIO tags. For each token representation h_i:

$$P\left(y_i^{\text{AE}} \mid h_i\right) = \text{Softmax}\left(\boldsymbol{W}_{\text{AE}} h_i + b_{\text{AE}}\right) \tag{8}$$

where $\boldsymbol{W}_{\text{AE}} \in \mathbb{R}^{3\times d}$, and $y_i \in \{\text{B-ASP}, \text{I-ASP}, \text{O}\}$.

The extraction loss is token-level cross-entropy:

$$\mathcal{L}_{\text{AE}} = -\frac{1}{n}\sum\nolimits_{i=1}^{n} \log P\left(y_i^* \mid \boldsymbol{h}_i\right) \tag{9}$$

where y_i^* is the ground truth tag.

Aspect Sentiment Classification Head: For each aspect span, we compute an aspect-aware representation via attention pooling. During training the ASC head receives gold aspect spans (teacher forcing); at inference and for all reported results the AE head's predicted spans are used, yielding a fully end-to-end joint evaluation that matches the protocol. First, compute attention weights:

$$\alpha_j = \frac{\exp\left(q^{\mathrm{T}} h_j\right)}{\sum_{k=s}^{e} \exp\left(q^{\mathrm{T}} h_k\right)}, \quad j = s, \cdots, e \tag{10}$$

where $q \in \mathbb{R}^d$ is a learnable initialized randomly and updated during training. Then aggregate representations:

$$\boldsymbol{a} = \sum_{j=s}^{e} \alpha_j h_j \tag{11}$$

for sentiment classification:

$$P(s \mid a) = \mathrm{Softmax}\left(W_{\mathrm{ASC}} a + b_{\mathrm{ASC}}\right) \tag{12}$$

where $s \in \{\text{Positive, Negative, Neutral}\}$. The ASC loss is:

$$\mathcal{L}_{\mathrm{ASC}} = -\log P\left(s^* \mid a\right) \tag{13}$$

where s^* is the ground truth sentiment.

Multi-Task Objective: The combined loss function is:

$$\mathcal{L}_{\mathrm{total}} = \lambda_{\mathrm{AE}} \cdot \mathcal{L}_{\mathrm{AE}} + \lambda_{\mathrm{ASC}} \cdot \mathcal{L}_{\mathrm{ASC}} \tag{14}$$

To promote noise-invariant representations we further introduce a sentence-level invariance loss $L_{\mathrm{inv\_sent}}$. For every clean-noisy pair $(x, \tilde{x})$ in a training batch we compute the KL-divergence between the model's output distributions for both tasks (token probabilities for AE are averaged; ASC uses the pooled representation.

To explicitly enforce noise invariance, we add a span-level $L_{\mathrm{inv\_span}}$. consistency loss between clean and noisy inputs. For each gold aspect span s, we compute pooled representations v_s and $\tilde{v}_s$ from the clean and noisy sentences, respectively, and define:

$$\mathcal{L}_{\mathrm{noise}} = \frac{1}{|S|} \sum_{s \in S} \left\| v_s - \tilde{v}_s \right\|_2^2 \tag{15}$$

The final training objective is:

$$\mathcal{L} = \lambda_{\mathrm{AE}} \cdot \mathcal{L}_{\mathrm{AE}} + \lambda_{\mathrm{ASC}} \cdot \mathcal{L}_{\mathrm{ASC}} + \lambda_{\mathrm{noise}} \cdot \mathcal{L}_{\mathrm{noise}} \tag{16}$$

where λ_{noise} is a hyperparameter that controls the strength of the noise-invariance constraint. λ_{AE} and λ_{ASC} are task weights optimized via grid search. After searching over {0.2, 0.5, 1.0, 2.0} for each weight, we selected λ_{AE} = 1.0, λ_{ASC} = 1.0 on Rest14 validation. In all experiments, noise invariance is applied at the span level with $\lambda_{\mathrm{noise}} = 0.1$.

3.4. Training Strategy

We employ a mixed training regimen where each batch contains both clean and noisy samples following the curriculum schedule in Equation 6. All models are fine-tuned from pre-trained bert-base-uncased and roberta-base checkpoints.

3.5. Cross-Domain Training Configuration

We evaluate three configurations:

1) **In-Domain**: $\mathcal{D}_{\mathrm{train}} = \mathcal{D}_{\mathrm{test}}$ (same domain);
2) **Cross-Domain**: $\mathcal{D}_{\mathrm{train}} = \mathcal{D}_S$, $\mathcal{D}_{\mathrm{test}} = \mathcal{D}_T$ (different domains);

3) Multi-Domain: $\mathcal{D}_{\text{train}} = \bigcup_{i=1}^{k} \mathcal{D}_i$, $\mathcal{D}_{\text{test}} = \mathcal{D}_j$ (trained on all, tested on each).

4. Experiments and Results

The study presents a detailed examination of a Noise-Aware Multi-Task Transformers for Cross-Domain when applied to three different benchmark datasets within the domain of ABSA.

4.1. Dataset and Experiment Setup

To test our model's performance, we used benchmark datasets: (Res14 and LAP14) from SemEval-2014 Task 4 [9], and real-world dataset: (electronics and Apparel) from Amazon, and Food from Yelp. All carry three sentiment polarities: positive, neutral, and negative, Table 3 presents detailed statistics for each dataset, including sentence counts, aspect counts, average length, and sentiment distribution. LAP14 has the highest share of long sentences (over 35 tokens) at 18%, which could hurt results compared to Res 14. Still, it better shows how the model handles tough, extended text. Meanwhile, Res14 contain mostly short sentences (under 20 tokens), making up 56.68%.

Implementation details

Models trained for up to 1500 steps. After 1000 steps, we selected the best checkpoint every 100 steps using micro-F1 on the dev set. We implement using PyTorch 2.0 and Hugging Face Transformers. Hyperparameters: learning rate 2×10^{-5}, batch size 16, AdamW optimizer with $(\beta_1, \beta_2) = (0.9, 0.999)$, weight decay 0.01. Training runs for 20 epochs with early stopping (patience = 5), $\lambda_{\text{noise}} = 0.1$ (Noise invariance weight) and Dropout = 0.1 (Encoder + heads). All experiments use NVIDIA A100 GPUs. Experiments ran in PyCharm with Python 3.6.

For preprocessing task, all sentences are tokenized using the tokenizer corresponding to each pre-trained encoder (BERT-base, RoBERTa-base). Keeping cased input for BERT; RoBERTa uses its default tokenization. Sentences longer than 128 tokens are truncated; shorter ones are padded. Aspect annotations are converted to BIO tags programmatically.

Baselines

We compare our approach to various baseline models listed in Table 2(a). Most baselines target only particular ABSA subtasks, with some offering complete coverage across all components.

- Single-Task: MTL-AraBERT [15], Robust ABSA w/Non-counterfactual Aug. [30], SentiSys [1], EDA/UDA [23]
- Multi-Task: LEGO-ABSA [27] (fine-tuned on source domain), CC-ASTN [6], BERT-Based Cross-Domain [3], SentiSys [1].
- Cross-Domain: BGCA [20] (fine-tuned on source domain for fair comparison) UNIFIEDABSA [28] SyMux [29].

Noise Simulation

We simulate textual noise at two intensity levels (low and high) by introducing spelling errors, word deletions, synonym substitutions, and slang variations with different probabilities. The low-noise setting produces minor and mostly readable distortions, while the high-noise setting introduces stronger corruption that significantly alters word forms and sentence structure. following established noise simulation protocols [22] [23] two intensity levels were created as:

- Low Noise: $p_{\text{spell}} = 0.05, p_{\text{drop}} = 0.05, p_{\text{syn}} = 0.1, p_{\text{slang}} = 0.05$;
- High Noise: $p_{\text{spell}} = 0.15, p_{\text{drop}} = 0.2, p_{\text{syn}} = 0.3, p_{\text{slang}} = 0.15$.

Table 3. Dataset statistics.

Datasets	Domain	# Sentences	# Aspects	Avg Len	Train			Test		
					Pos	Neg	Neu	Pos	Neg	Neu
SemEval14	Restaurant	3841	4693	19.8	2164	805	633	728	196	167
	Laptop	3845	3021	21.3	987	458	454	341	128	169
Amazon	Electronics	8247	11845	26.8	4830	1150	2870	2267	1208	567
	Apparel	7892	10234	24.3	4510	1020	1862	1128	255	465
Yelp	Food	9450	14876	23.1	5980	1230	2390	1495	307	560

Table 4 illustrates examples of original sentences and their corresponding noisy variants at both intensity levels.

Table 4. Example sentence pairs (original → noisy variant).

Original	Low Noise	High Noise
The battery life is amazing.	The battery life is amazing.	Th battry lf is amazing.

Evaluation Metrics

To evaluate our experiments, we conducted Precision (P), Recall (R), F1-score for Aspect Extraction: Sentiment Classification: Accuracy, Macro-F1, for Efficiency: Training time (hours), Inference speed (sentences/second), Parameters (millions), and for Robustness:

$$\text{Performance degradation} = \frac{\text{Perf}_{\text{clean}} - \text{Perf}_{\text{noisy}}}{\text{Perf}_{\text{clean}}} \times 100\%.$$

4.2. Results

4.2.1. Overall Performance Comparison (RQ1)

The overall performance results from benchmark SemEval datasets are displayed in **Table 5**. Our proposed framework achieves highly competitive results across both domains and tasks. The RoBERTa variant shows the best performance with 87.9 F1 for AE and 85.6% accuracy for ASC on Rest14, representing improvements of +2.6 F1 and +2.9% over the best baseline (BGCA). Our unified multi-task approach proves effective through its continuous better performance in both Restaurants and Laptops domains. The performance gap between single-task models and our unified framework demonstrates the advantages of dual learning and common representation use.

Table 5. Overall performance on SemEval benchmark datasets (F1-score for aspect extraction, accuracy for aspect sentiment classification). All results are mean ± std over 5 runs.

Model	Rest14		Lap14	
	AE (F1)	ASC (Acc)	AE (F1)	ASC (Acc)
BERT-AE/SC	82.3 ± 0.8/84.1 ± 0.7	78.9 ± 0.9/79.5 ± 0.8	80.1 ± 0.9/81.7 ± 0.8	76.4 ± 1.0/77.1 ± 0.9
RoBERTa-AE/SC	83.7 ± 0.7/85.2 ± 0.6	80.1 ± 0.8/80.9 ± 0.7	81.5 ± 0.8/82.9 ± 0.7	77.8 ± 0.9/78.6 ± 0.8
MTL-BERT	84.6 ± 0.6	81.5 ± 0.7	82.8 ± 0.7	78.9 ± 0.8
LEGO-ABSA*	85.1 ± 0.6	82.3 ± 0.7	83.4 ± 0.7	79.6 ± 0.8
CC-ASTN	84.9 ± 0.6	82.1 ± 0.7	83.1 ± 0.7	79.3 ± 0.8
BGCA*	85.3 ± 0.6	82.7 ± 0.7	83.6 ± 0.7	79.9 ± 0.8
Ours (BERT)	**86.7 ± 0.5**	**84.2 ± 0.6**	**85.1 ± 0.6**	**81.3 ± 0.7**
Ours (RoBERTa)	**87.9 ± 0.5***	**85.6 ± 0.6**	**86.3 ± 0.6**	**82.8 ± 0.7**

*Fine-tuned on source domain for fair comparison.

4.2.2. Noise Robustness Analysis (RQ2)

Table 6 shows how well the system handles linguistic noise in the Amazon Electronics dataset. The noise-aware training (NAT) strategy of our study allows us to maintain performance in the presence of noisy environments. The model with NAT demonstrates only −9.4%/−11.7% performance loss during high noise situations which stands in contrast to the single-task BERT system that exhibits −22.4%/−26.1% performance decline, resulting in a 42% improvement of our model's resistance to performance loss. The results demonstrate that: 1) All models degrade with increasing noise, but multi-task architectures show inherent robustness; 2) Our framework without NAT already outperforms baselines; 3) With NAT, degradation is minimized, proving the effectiveness of systematic noise injection during training. Concretely, NAT reduces the high-noise AE degradation from 16.1% (w/o NAT) to 9.4% (with NAT), a relative improvement of 42%. The corresponding ASC reduction is 39%.

4.2.3. Cross-Domain Generalization (RQ3)

Table 7 shows cross-domain generalization from Electronics to Apparel. The study presents performance results for Electronics as well as for cross-domain measurements which were conducted on Apparel. The system achieved its highest cross-domain results through 76.9 F1 score and 79.7% accuracy results which showed a minor performance decrease of 11 and 5.9 points when compared to BERT's complete drop of 12.7 and 8.3 points. The research demonstrates advanced domain-independent knowledge representation systems. The study demonstrates that noise-aware training increases system robustness while also boosting abilities to transfer knowledge across different domains because it teaches models to disregard unimportant language differences which helps them identify fundamental meaning connections that remain consistent across different domains.

Table 6. Performance degradation under different noise levels on Amazon Electronics dataset. Values show relative drop in F1 (AE)/Accuracy (ASC). Lower values indicate better robustness.

Model	Low Noise AE/ASC Degradation	High Noise AE/ASC Degradation
BERT-AE/SC	−8.7%/−10.3%	−22.4%/−26.1%
RoBERTa-AE/SC	−7.9%/−9.1%	−20.7%/−24.3%
MTL-BERT	−6.3%/−7.8%	−18.9%/−22.7%
LEGO-ABSA	−5.8%/−6.9%	−17.4%/−20.5%
Ours (w/o NAT)	−5.1%/−6.2%	−16.1%/−19.2%
Ours (with NAT)	**−2.3%/−3.1%**	**−9.4%/−11.7%**

Table 7. Cross-domain performance (Electronics → Apparel). In-domain performance on electronics shown for reference.

Model	In-domain (Elec)		Cross (Apparel)	
	AE F1	ASC Acc	AE F1	ASC Acc
BERT	82.3	79.5	68.4	71.2
MTL-BERT	84.6	81.5	70.1	73.5
CC-ASTN	84.9	82.1	72.3	75.1
BGCA	85.3	82.7	73.8	76.4
Ours	**87.9**	**85.6**	**76.9**	**79.7**

Table 8 shows how multi-domain training provides advantages. The mixed data training across all three domains led to improved results which reached F1 score increases between 0.7 and 1.6 for AE and 1.6 to 2.4 percent for ASC. The research indicates that learners who study various fields will acquire better skills to develop general knowledge systems which will decrease their tendency to become attached to particular field knowledge. The multi-task learning framework delivered better results than the single-task method across all research domains we tested. The Ours (Multi-Task) model demonstrated superior F1 performance in Aspect Extraction and achieved higher accuracy results in Sentiment Classification when compared to the single-task model. The model achieved its most effective results in the Apparel and Electronics domains because it utilized common features to enhance its performance. The research results demonstrate statistical significance which proves that simultaneous training on multiple domains results in improved model performance for accurate sentiment detection.

4.2.4. Efficiency Analysis (RQ4)

Table 9 presents the efficiency-accuracy trade-off. The framework we created requires 2M extra parameters which represent a 1.8% increase over MTL-BERT. Compared to MTL-BERT, the proposed model achieves F1 improvements of 3.3 points and accuracy improvements of 4.1 percent. The training time increase of 22 percent proves acceptable because it leads to major performance enhancements. The system achieves a practical inference speed of 340 sentences per second which supports real-time usage. Our method outperforms LEGO-ABSA especially in terms of cost-effectiveness and precision. The efficiency comparison shows that our proposed model achieves the highest performance in both Aspect Extraction (F1: 87.9) and Aspect Sentiment Classification (Accuracy: 85.6). The system requires more resources than the base BERT model yet it operates efficiently with acceptable speed during training and execution. The ablation study further highlights the importance of each component in our full model. The system performance declines together with cross-domain capability after we eliminate multi-task learning and noise-aware training and attention pooling. The system operates with decreased resistance against noise in audio signals because these elements do not exist especially when multi-task and NAT elements get deleted. The proposed architecture requires all components to achieve both high accuracy and system durability.

Table 8. Multi-domain training results. Training on mixed data from all domains improves performance even on individual domains. All results are statistically significant ($p < 0.05$).

Test Domain	Ours (multi-task)	Ours (Single-Task)	Improvement
Electronics	85.8/83.1	85.1/81.3	+0.7/+1.8
Apparel	84.3/81.9	82.7/79.5	+1.6/+2.4
Restaurants	86.2/83.7	84.9/82.1	+1.3/+1.6

Table 9. Efficiency comparison on amazon electronics dataset (batch size = 16).

Model	Params (M)	Train Time (hrs)	Inf. Speed (sent/s)	AE F1	ASC Acc
BERT	110	4.2	320	82.3	79.5
MTL-BERT	110	3.1	350	84.6	81.5
Ours	**112**	**3.8**	**340**	**87.9**	**85.6**
LEGO-ABSA	125	5.2	290	85.1	82.3

5. Discussion

5.1. Synthesis of Findings

Our model delivers consistent gains: +2.6 F1 (AE)/+2.9% ACC (ASC) over the strongest baseline on SemEval Rest14 for BGCA (Table 5) and up to +3.1 F1 (AE) and +3.3% (ASC) over the strongest cross-domain baseline (BGCA) in Table 7. Across datasets and settings, the improvements consistently range between 2.6 - 3.1 F1 and 2.7% - 3.3% accuracy. The shared representation enables mutual reinforcement between tasks while reducing error propagation. The research showed that noise-aware training methods provide better training results because NAT reduces the high-noise AE degradation from 16.1% (Ours w/o NAT) to 9.4% (Ours with NAT), a relative improvement of 42% (and 39% for ASC; Table 6). Targeted augmentation techniques enable systematic learning of real-world linguistic variation robustness according to the research findings. The combination of multi-task learning and noise-aware mechanisms will produce better prediction stability according to results from RQ3. Our framework demonstrates effective cross-domain generalization because it only experiences minor performance reduction which indicates that noise-invariant representation learning enables models to detect fundamental patterns that exist across different domains. The proposed framework requires 2M parameters and 22% training time overhead according to RQ4 which demonstrates that systems can achieve robustness through low computational needs while maintaining efficient inference speed.

5.2. Implications

The results of our approach lead to several theoretical implications. First, noise augmentation functions as a strong regularizer. It helps the model learn noise-invariant representations that transfer more effectively across domains. This is consistent with prior work in NLP showing that increasing data, rather than model size, often improves robustness. Second, multi-task learning improves robustness in ways that go beyond simple accuracy gains, because it pushes the model to learn representations that support multiple related objectives at the same time. Third, the observed link between noise robustness and domain generalization indicates that both are supported by learning deeper and more abstract representations that are less sensitive to surface-level variations in the input.

5.3. Model Analysis

Table 10 shows the ablation study by demonstrating the contribution of each component. Removing multi-task learning causes the largest performance drop (−2.7 F1, −2.9% accuracy). Disabling NAT significantly reduces both noise robustness (degradation increases from −9.4% to −16.1%) and cross-domain performance. Removing attention pooling causes moderate drops. All components contribute meaningfully, with multi-task learning being most critical for accuracy and NAT for robustness.

We manually analyze 100 error cases from the Restaurants test set and compare error type distribution between BERT and our model. Table 11 shows the comparative of error analysis. This error analysis compares the performance of our proposed model against the baseline BERT model on the Restaurants test set, categorizing the types of errors encountered. The results demonstrate that our model achieves a consistent reduction in errors across all identified categories. The most significant improvement is observed in handling severe noise cases, where the error rate is reduced by 40% compared to BERT. There is also a substantial 34% reduction in aspect boundary errors, indicating the model's superior ability to precisely delimit aspect terms. Furthermore, the model shows a marked improvement in understanding domain-specific language, with a 33% reduction in errors. While improvements in handling complex negation and implicit aspects are less pronounced, with reductions of 28% and 22% respectively, the overall trend indicates the proposed model's robustness in addressing various challenging linguistic phenomena in sentiment analysis.

Table 10. Ablation study on Amazon Electronics dataset, each row removes one component from the full model. Noise degradation is measured on the high-noise test set. Cross-domain performance is on Apparel.

Configuration	AE F1	ASC Acc	Noise Degrad.	Cross-Domain
Full Model	87.9 ± 0.5	85.6 ± 0.6	−9.4%	76.9
- Multi-Task	85.2 ± 0.7	82.7 ± 0.7	−16.1%	70.1
- NAT	86.3 ± 0.6	83.9 ± 0.6	−16.1%	72.4
- Attention Pooling	87.1 ± 0.5	84.8 ± 0.6	−11.2%	74.6

Table 11. Error type distribution comparison between BERT and our model on Restaurants test set.

Error Type	BERT	Ours	Reduction
Aspect boundary errors	32%	21%	34%
Complex negation	25%	18%	28%
Implicit aspects	18%	14%	22%
Domain-specific language	15%	10%	33%
Severe noise cases	10%	6%	40%

The residual errors provide clear insights into remaining challenges and the complementary roles of our components. Complex negation scope (e.g., -not bad but overpriced) still accounts for 18% of errors; NAT reduces this by exposing varied phrasings during training, while MTL helps the shared encoder learn consistent aspect-sentiment alignments. Boundary fragmentation in multi-word aspects (e.g., battery life) is largely mitigated by the span-level attention pooling, which focuses representation on the whole extracted span. Domain-specific lexical shifts (e.g., long = positive for batteries) drop 33% thanks to synonym and slang augmentation that forces the model to rely on contextual rather than surface cues. These patterns confirm that MTL supplies the joint reasoning backbone, NAT supplies surface robustness, and attention pooling supplies precise span aggregation.

5.4. Limitations and Future Work

Despite the encouraging results, several limitations should be acknowledged. The created noise models demonstrate effective performance for English usage yet their application to other languages and cultural settings proves to be ineffective. Future work should explore methods for generating noise which adapt to different languages and cultural backgrounds. The current approach detects local token-based noise but it cannot detect sarcasm and irony and cultural references which affect how people understand sentiments. The training process takes 22 percent more time which works for our situation but becomes unmanageable when dealing with extremely large data sets and model structures. Researchers should investigate better methods for augmenting data which include using adversarial attacks and creating noise through dynamic methods. The experiments do not cover very long texts (over 512 tokens) that exceed typical transformer input limits. Researchers have not yet determined how to assess system performance for handling extensive documents. The field of ABSA now requires multimodal input; therefore, systems must develop ability to resist noise from textual sources and visual content and other sensory input which produces distinct forms of interference. The scientific community has not yet discovered how noise augmentation methods lead to improved ability to generalize across different domains. The representation analysis process enables researchers to understand how different aspects of robustness and generalization develop together.

6. Conclusion

The research conducted an extensive study which developed reliable systems for Aspect-Based Sentiment Analysis (ABSA) that can effectively handle the difficulties faced in actual operational situations. We developed a unified multi-task transformer framework which enabled simultaneous aspect extraction and sentiment classification through one unified system. Our approach achieved superior results over strong baseline models through experiments conducted across three domains which used both benchmark and real-world datasets. We created a training approach which enhances system strength through artificial noise generation that enables researchers to train with specific noise conditions. The research demonstrated that domain generalization and noise robustness function as different yet complementary skills which NAT-trained models handle better through noisy text and cross-domain capabilities. The efficiency study proved that our system achieves performance improvements which require only 22% extra training time and 3% additional time for inference when compared to multi-task baseline systems which lack NAT. The system provides an operational advantage because the system achieves positive results at production environments. Our robust approach enables organizations to extract valuable user insights because online platforms produce increasingly diverse and noisy text data. We consider this research to be essential for developing more fair and easily accessible NLP systems. Future work will extend noise-aware training to multilingual and multimodal ABSA, explore more efficient augmentation strategies, and investigate the theoretical connections between noise robustness and domain generalization more deeply.

Data Availability

The data that support the findings of this study are available from the corresponding author, [S], upon reasonable request.

Conflicts of Interest

The authors state no conflict of interest.

References

[1] Li, C., Tang, H., Zhang, J., Guo, X., Cheng, D. and Morimoto, Y. (2024) Advancing Aspect-Based Sentiment Analysis through Deep Learning Models. In: *Lecture Notes in Computer Science*, Springer, 228-242. https://doi.org/10.1007/978-981-96-0847-8_16

[2] Zhang, W., Li, X., Deng, Y., Bing, L. and Lam, W. (2023) A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges. *IEEE Transactions on Knowledge and Data Engineering*, **35**, 11019-11038. https://doi.org/10.1109/tkde.2022.3230975

[3] Liu, N. and Zhao, J. (2022) A Bert-Based Aspect-Level Sentiment Analysis Algorithm for Cross-Domain Text. *Computational Intelligence and Neuroscience*, **2022**, 1-11. https://doi.org/10.1155/2022/8726621

[4] Nazir, A., Rao, Y., Wu, L. and Sun, L. (2022) Issues and Challenges of Aspect-Based Sentiment Analysis: A Comprehensive Survey. *IEEE Transactions on Affective Computing*, **13**, 845-863. https://doi.org/10.1109/taffc.2020.2970399

[5] Schouten, K. and Frasincar, F. (2016) Survey on Aspect-Level Sentiment Analysis. *IEEE Transactions on Knowledge and Data Engineering*, **28**, 813-830. https://doi.org/10.1109/tkde.2015.2485209

[6] Dubey, G., Chadha, A., Jyoti, A., Raj, G., Kaur, K. and Dubey, A.K. (2025) Contextualized Cross-Domain Aspect Sentiment Transformer: A Fine-Grained Aspect-Centric Approach for Enhanced Context-Aware Sentiment Analysis. *Computational Intelligence*, **41**, e70081. https://doi.org/10.1111/coin.70081

[7] Yu, J., Zhao, Q. and Xia, R. (2023) Cross-Domain Data Augmentation with Domain-Adaptive Language Modeling for Aspect-Based Sentiment Analysis. *Proceedings of the* 61*st Annual Meeting of the Association for Computational Linguistics* (*Volume* 1: *Long Papers*), Toronto, July 2023, 1456-1470. https://doi.org/10.18653/v1/2023.acl-long.81

[8] Chen, C., Teng, Z., Wang, Z. and Zhang, Y. (2022) Discrete Opinion Tree Induction for Aspect-Based Sentiment Analysis. *Proceedings of the* 60*th Annual Meeting of the Association for Computational Linguistics* (*Volume* 1: *Long Papers*), Dublin, May 2022, 2051-2064. https://doi.org/10.18653/v1/2022.acl-long.145

[9] Pontiki, M., Galanis, D., Pavlopoulos, J., *et al.* (2014) SemEval 2014 Task 4: Aspect Based Sentiment Analysis. *Proceedings of the 8th International Workshop on Semantic Evaluation* (*SemEval* 2014) *at* (*COLING* 2014), Dublin, August 2014, 27-35. http://www.aclweb.org/anthology/S14-2004

[10] McAuley, J., Targett, C., Shi, Q. and van den Hengel, A. (2015). Image-Based Recommendations on Styles and Substitutes. *Proceedings of the* 38*th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Santiago, 9-13 August 2015, 43-52. https://doi.org/10.1145/2766462.2767755

[11] Yelp. Yelp Open Dataset. https://business.yelp.com/data/resources/open-dataset/

[12] Wu, C., Ma, B., Zhang, Z., Deng, N., He, Y. and Xue, Y. (2025) Evaluating Zero-Shot Multilingual Aspect-Based Sentiment Analysis with Large Language Models. *International Journal of Machine Learning and Cybernetics*, **16**, 8079-8101. https://doi.org/10.1007/s13042-025-02711-z

[13] Yang, L., Na, J. and Yu, J. (2022) Cross-Modal Multitask Transformer for End-To-End Multimodal Aspect-Based Sentiment Analysis. *Information Processing & Man-

agement, **59**, Article 103038. https://doi.org/10.1016/j.ipm.2022.103038

[14] He, Z., Wang, H. and Zhang, X. (2023) Multi-Task Learning Model Based on BERT and Knowledge Graph for Aspect-Based Sentiment Analysis. *Electronics*, **12**, Article 737. https://doi.org/10.3390/electronics12030737

[15] Fadel, A., Saleh, M., Salama, R. and Abulnaja, O. (2024) MTL-AraBERT: An Enhanced Multi-Task Learning Model for Arabic Aspect-Based Sentiment Analysis. *Computers*, **13**, 98. https://doi.org/10.3390/computers13040098

[16] Alshuwaier, F., Areshey, A. and Poon, J. (2022) Applications and Enhancement of Document-Based Sentiment Analysis in Deep Learning Methods: Systematic Literature Review. *Intelligent Systems with Applications*, **15**, Article 200090. https://doi.org/10.1016/j.iswa.2022.200090

[17] Labeed, Q. and Liang, X. (2024) Multi-Task Learning Transformers: Comparative Analysis for Emotion Classification and Intensity Prediction in Social Media. 2024 14*th International Conference on Pattern Recognition Systems* (*ICPRS*), London, 15-18 July 2024, 1-7. https://doi.org/10.1109/icprs62101.2024.10677817

[18] Aziz, K., Ji, D., Chakrabarti, P., Chakrabarti, T., Iqbal, M.S. and Abbasi, R. (2024) Unifying Aspect-Based Sentiment Analysis BERT and Multi-Layered Graph Convolutional Networks for Comprehensive Sentiment Dissection. *Scientific Reports*, **14**, Article No. 14646. https://doi.org/10.1038/s41598-024-61886-7

[19] Zhou, Y., Zhu, F., Song, P., Han, J., Guo, T. and Hu, S. (2021) An Adaptive Hybrid Framework for Cross-Domain Aspect-Based Sentiment Analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 14630-14637. https://doi.org/10.1609/aaai.v35i16.17719

[20] Deng, Y., Zhang, W., Pan, S.J. and Bing, L. (2023) Bidirectional Generative Framework for Cross-Domain Aspect-Based Sentiment Analysis. *Proceedings of the* 61*st Annual Meeting of the Association for Computational Linguistics* (*Volume* 1: *Long Papers*), Toronto, July 2023, 12272-12285. https://doi.org/10.18653/v1/2023.acl-long.686

[21] Zou, H. and Wang, Y. (2025) Large Language Model Augmented Syntax-Aware Domain Adaptation Method for Aspect-Based Sentiment Analysis. *Neurocomputing*, **625**, Article 129472. https://doi.org/10.1016/j.neucom.2025.129472

[22] Wei, J. and Zou, K. (2019) EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. *Proceedings of the* 2019 *Conference on Empirical Methods in Natural Language Processing and the* 9*th International Joint Conference on Natural Language Processing* (*EMNLP-IJCNLP*), Hong Kong, November 2019, 6382-6388. https://doi.org/10.18653/v1/d19-1670

[23] Xie, Q., Dai, Z., Hovy, E., Luong, M.T. and Le, Q.V. (2020) Unsupervised Data Augmentation for Consistency Training. *Proceedings of the* 34*th International Conference on Neural Information Processing Systems*, Vancouver 6-12 December 2020, 6256-6268.

[24] Zhuang, S., Shou, L., Pei, J., Gong, M., Ren, H., Zuccon, G., *et al.* (2023) Typos-Aware Bottlenecked Pre-Training for Robust Dense Retrieval. *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, Beijing, 26-28 November 2023, 212-222. https://doi.org/10.1145/3624918.3625324

[25] Bashiri, H. and Naderi, H. (2024) Comprehensive Review and Comparative Analysis of Transformer Models in Sentiment Analysis. *Knowledge and Information Systems*, **66**, 7305-7361. https://doi.org/10.1007/s10115-024-02214-3

[26] Wang, H., Li, X., Ren, Z., Wang, M. and Ma, C. (2023) Multimodal Sentiment Analysis Representations Learning via Contrastive Learning with Condense Attention Fusion. *Sensors*, **23**, Article 2679. https://doi.org/10.3390/s23052679

[27] Yan, H., Dai, J., Ji, T., Qiu, X. and Zhang, Z. (2021) A Unified Generative Framework for Aspect-Based Sentiment Analysis. *Proceedings of the* 59*th Annual Meeting of the Association for Computational Linguistics and the* 11*th International Joint Conference on Natural Language Processing* (*Volume* 1: *Long Papers*), Online, August 2021, 2416-2429. https://doi.org/10.18653/v1/2021.acl-long.188

[28] Wang, Z., Xia, R. and Yu, J. (2022) Unified ABSA: A Unified ABSA Framework Based on Multi-Task Instruction Tuning.

[29] Fei, H., Li, F., Li, C., Wu, S., Li, J. and Ji, D. (2022) Inheriting the Wisdom of Predecessors: A Multiplex Cascade Framework for Unified Aspect-Based Sentiment Analysis. *Proceedings of the* 31*st International Joint Conference on Artificial Intelligence*, Vienna, 23-29 July 2022, 4096-4103. https://doi.org/10.24963/ijcai.2022/572

[30] Liu, X.Y., Ding, Y., An, K.K., Xiao, C.Y., *et al*. (2023) Towards Robust Aspect-Based Sentiment Analysis through Non-Counterfactual Augmentations.

[31] Wilson, S., Magdy, W., McGillivray, B., Garimella, K. and Tyson, G. (2020) Urban Dictionary Embeddings for Slang NLP Applications.